

# The Complement of Protein Phosphatase Catalytic Subunits Encoded in the Genome of Arabidopsis<sup>1</sup>

David Kerk\*, Joshua Bulgrien, Douglas W. Smith, Brooke Barsam, Stella Veretnik, and Michael Gribskov

Department of Biology, Point Loma Nazarene University, 3900 Lomaland Drive, San Diego, California 92106 (D.K., J.B., B.B.); Division of Biology, 0116, University of California San Diego, La Jolla, California 92093–0116 (D.W.S.); and San Diego Supercomputer Center, 0505, University of California San Diego, La Jolla, California 92093–0505 (S.V., M.G.)

Reversible protein phosphorylation is critically important in the modulation of a wide variety of cellular functions. Several families of protein phosphatases remove phosphate groups placed on key cellular proteins by protein kinases. The complete genomic sequence of the model plant Arabidopsis permits a comprehensive survey of the phosphatases encoded by this organism. Several errors in the sequencing project gene models were found via analysis of predicted phosphatase coding sequences. Structural sequence probes from aligned and unaligned sequence models, and all-against-all BLAST searches, were used to identify 112 phosphatase catalytic subunit sequences, distributed among the serine (Ser)/threonine (Thr) phosphatases (STs) of the protein phosphatase P (PPP) family, STs of the protein phosphatase M (PPM) family (protein phosphatases 2C [PP2Cs] subfamily), protein tyrosine (Tyr) phosphatases (PTPs), low- $M_r$  protein Tyr phosphatases, and dual-specificity (Tyr and Ser/Thr) phosphatases (DSPs). The Arabidopsis genome contains an abundance of PP2Cs (69) and a dearth of PTPs (one). Eight sequences were identified as new protein phosphatase candidates: five dual-specificity phosphatases and three PP2Cs. We used phylogenetic analyses to infer clustering patterns reflecting sequence similarity and evolutionary ancestry. These clusters, particularly for the largely unexplored PP2C set, will be a rich source of material for plant biologists, allowing the systematic sampling of protein function by genetic and biochemical means.

Reversible protein phosphorylation modulates many cellular functions including cell cycle events, growth factor response, hormone and other environmental stimuli, metabolic control, and developmental processes (Andreeva and Kutuzov, 1999; Chernoff, 1999; den Hertog, 1999; Iten et al., 1999; Schillace and Scott, 1999; Luan, 2000). As a paradigm, protein kinases and phosphatases add or remove phosphate groups on critical enzymes and regulatory proteins. Kinases differ in their phosphoryl amino acid substrate specificity: some act at Ser/Thr residues, some act at Tyr residues, and some can act at both ("dual-specificity" kinases). Despite these differences, all protein kinases share characteristic structural motifs and similar folded structures (Hanks and Hunter, 1995).

Phosphatases can be similarly grouped by substrate specificity into Ser/Thr, Tyr, and dual-specificity classes. However, in contrast to kinases, phosphatases represent a more structurally and evolutionarily diverse group. The STs originally were subdivided into the protein phosphatase 1 (PP1) and protein phosphatase 2 (PP2) groups based upon dif-

ferential sensitivity to small molecule inhibitors. PP2 proteins are further distinguished by metal ion requirements: PP2Cs require  $Mg^{+2}$  and protein phosphatases 2B (PP2Bs) require  $Ca^{+2}$ , whereas protein phosphatases 2A (PP2As) have no ion requirement (Cohen, 1989). The proteins of the PP1, PP2A, and PP2B groups share sequence similarity and now comprise the PPP sequence family. PP2C sequences, however, lack sequence similarity to the protein phosphatase P (PPP) family, and together with pyruvate dehydrogenase phosphatase and other  $Mg^{+2}$ -dependent Ser/Thr phosphatases (STs), comprise the protein phosphatase M (PPM) sequence family (Barford, 1996; Cohen, 1997). Despite their lack of sequence similarity, members of the PPP and PPM families share a similar structural fold (Das et al., 1996), suggesting a common mechanism of catalysis. Several conserved acidic residues complex metal ions, which are essential to activity (Egloff et al., 1995; Goldberg et al., 1995; Griffith et al., 1995). A metal-bound water molecule acts as a nucleophile to directly displace phosphate from the substrate amino acid (Lohse et al., 1995) in an acid base catalytic mechanism.

Tyr phosphatases have a distinct evolutionary origin and catalytic mechanism from the STs. The "conventional" Tyr phosphatases are those specific for phosphorylated Tyr residues (PTPs), whereas dual-specificity phosphatases (DSPs) act at both Tyr and Ser/Thr residues. Both phosphatase types comprise a common evolutionary family. They contain a cata-

<sup>1</sup> This work was supported by the National Science Foundation (grant nos. NSF ROA DBI-9975808/PTLOMA and NSF DBI: 9975808).

\* Corresponding author; e-mail dkerk@ptloma.edu; fax 619-849-2598.

Article, publication date, and citation information can be found at [www.plantphysiol.org/cgi/doi/10.1104/pp.004002](http://www.plantphysiol.org/cgi/doi/10.1104/pp.004002).

lytic core motif with a conserved Cys residue, which acts as a nucleophile, displacing the phosphate group from the substrate and forming a phosphoryl-cysteinyl intermediate. A positionally conserved Asp participates in the removal of the phosphate group (Fauman and Saper, 1996). The low- $M_r$  protein Tyr phosphatases (LMW-PTPs) constitute an evolutionarily distinct group, which have converged on a similar catalytic mechanism (Ramponi and Stefani, 1997).

Most knowledge of the structure, mechanism, and function of protein phosphatases has been derived from work in animal and fungal systems. Knowledge of plant proteins has begun to emerge in the last few years, but is still incomplete. The structure and expression patterns of the Ser/Thr phosphatases have been intensively investigated, but relatively little is known about their function. In contrast, functional information is available for a few well-studied PP2Cs, but relatively few members of this large family have been thoroughly investigated. The recent completion of the genomic sequence of Arabidopsis now permits the analysis of a complete set of plant protein phosphatases and their evolutionary relationships, which will facilitate their functional study in plant development and physiology.

## RESULTS

### Assessment of Gene Prediction Quality

Sequence annotations were subjected to the quality assessment procedure detailed in "Materials and Methods." Gene predictions (122) were examined, with the following breakdown: PP2C (76), DSP and "DSP like" (17), ST (27), PTP (1), and LMW-PTP (1). Twelve sequences were ultimately rejected at the final alignment stage (sequences indicated in figure legends), leaving a total of 110 initial candidate Arabidopsis phosphatase sequences. Two additional sequences (At5g04540 and At3g10550) were accepted into the data set during revision without an examination of their annotation quality. Table I summarizes our assessment of the quality of the gene structures for phosphatases predicted by the Arabidopsis genome sequencing project, using the error detection procedures detailed in "Materials and Methods." Of the 110 predictions for confirmed protein phosphatases, we found five instances of extra exons in the annotated sequence, four instances of missing exons, and 15 instances of duplicated sequences. For sequences with exon errors, corrected versions have been deposited with the PlantsP database (<http://plantsp.sdsc.edu>). For the duplicated sequences, we determined if the duplicates were likely to be biologic in origin (i.e. genuine genomic gene duplications) or artifacts of data recording during the genome project. We examined the chromosome of origin of the duplicates, the gene order of neighbors of the duplicates, and searched the expressed sequence tag (EST) database for hits supporting alter-

**Table I.** Annotation errors in protein phosphatase sequences

Annotation Error	Arabidopsis Sequence No.	National Center for Biotechnology Information (NCBI) Protein gi Nos.
Extra exon	At2g20630	4512655
	At4g28400	7488280
	At4g31750	7486331
	At2g35680	4263787
	At2g42810	4512673
Missing exon	At4g28400	7488280
	At1g68410	6714350
	At3g52180	4678947
	At5g53140	9757995
Duplication	At4g28400	2842482, 7269694
	At4g31860	7270089, 4584525
	At4g26080	4538937, 7269459
	At4g27800	7269633, 4972111
	At3g15260	6997172, 9294253
	At4g33920	727034, 3297816
	At4g38520	4467139, 7270835
	At4g31750	3281853, 7270078
	At4g32950	3688176, 7270242
	At1g43900	8778653, 7523669
	At3g12620	9294356, 12321955
	At1g71860	7239513, 12324541 <sup>a</sup>
	At4g11240	5596470, 7267823
	At4g26720	4455195, 7269524
	At4g03080	3924596, 7270178

<sup>a</sup> Fragmentary duplication—sequence truncated by end of BAC.

native splice variants. By these criteria, all duplicates appear to be project artifacts.

### Inventory of Phosphatase Sequences

Of the 112 Arabidopsis phosphatase candidates, there is a preponderance (69) of PP2Cs, only one PTP, 23 STs, 18 DSPs, and one LMW-PTP. Some sequences, in several phosphatase classes, achieved high scores in database searches (either by profile analysis or by repetitive family BLAST analyses), but lacked critical catalytic residues, and therefore were rejected as functional phosphatases. However, these sequences, indicated in figure legends, may be evolutionarily related to the functional sequences. Two sequences (At1g05000 and At2g32960) are "DSP like." These sequences have a Glu residue (E) at a critical catalytic position usually occupied by an Asp (D). Though this is usually considered a conservative substitution, its uniqueness to these sequences renders their catalytic activity as problematic. Eight sequences are classified here as new candidate phosphatases (i.e. not previously annotated as phosphatases): five DSPs and three PP2Cs (see figure legends).

In the PP2C class, a set of conserved motifs has been defined (Bork et al., 1996), and a crystal structure obtained (Das et al., 1996). Four conserved Asp (D) residues coordinate a divalent metal ion ( $Mg^{2+}$  or  $Mn^{2+}$ ) essential for catalytic activity. In addition to

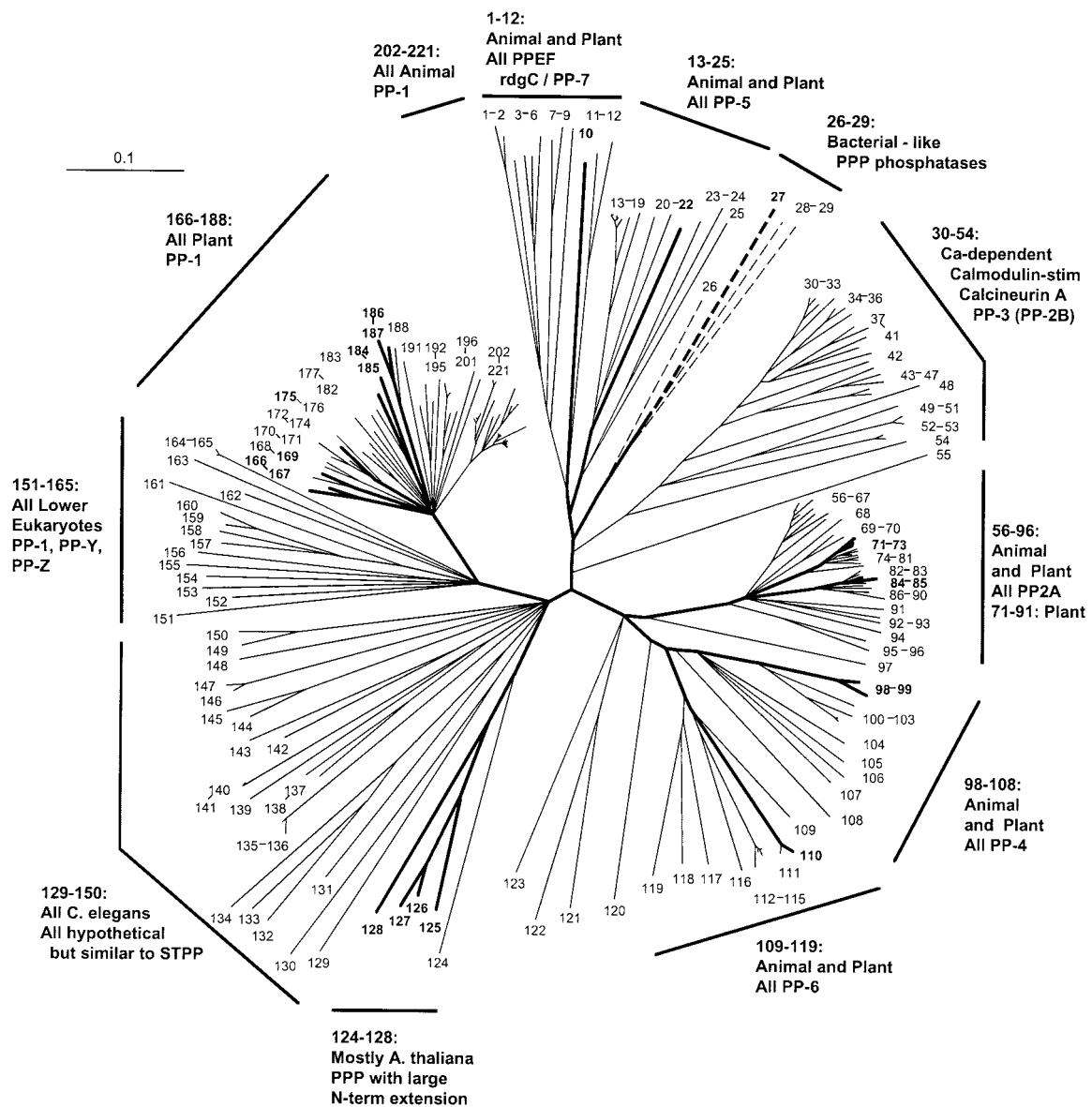
these acidic residues, there are other conserved motifs in the catalytic domain. In the fourth motif, a Thr (T) was found to be invariant in the first PP2C to be characterized, including the sequence whose three-dimensional structure has been solved. More recently, however, some plant phosphatases have been characterized that contain a Cys (C) in this position. PP2C5 of *Arabidopsis* (sequence At2g40180) and MP2C of *Medicago sativa* (gi:7488754) have been expressed in bacteria and shown to have phosphatase activity in vitro (Meskiene et al., 1998; Wang et al., 1999). Our search retrieved a number of other *Arabidopsis* sequences having a Cys at this same position: At1g67820, At3g17090, At4g38520, At2g28890, At5g-

02400, At2g30020, At3g12620, At3g55050, At1g07160, At5g02760, At3g09400, At4g33920, At1g07630, At3g16560, At5g66080, At5g06750, and At3g51370. We include these as valid PP2C candidates based upon this experimental work.

### Structural, Evolutionary, and Functional Relationships among Phosphatase Sequences

#### ST Phosphatases

An alignment of 221 ST catalytic domain sequences produced the pattern of relationships depicted in Figures 1, 2, and 3. These relationships are presented



**Figure 1.** Radial phylogenetic tree from ST protein phosphatase sequence comparisons. Correspondence between taxon or sequence number and NCBI protein gi number is given in Figures 2 and 3. Branch lengths are in arbitrary units. *Arabidopsis* sequence numbers are in bold and branches leading to these taxa are broad. Branches shown as dashed lines are presented as one-half of their true length.

as a radial phylogenetic tree with neighbor joining (NJ) (Saitou and Nei, 1987) branch lengths (Fig. 1) and as a topographic cladogram with representative bootstrap values (Figs. 2 and 3). The sequences are from a broad array of organisms: animals, plants, fungi, protists, and bacteria.

PP7 (At5g63870) is part of a cluster (sequences 1–12) containing animal EF hand-containing protein phosphatases (PPEFs) sequences and the *Drosophila melanogaster* protein RdgC. A PP5 cluster (sequences 13–25) contains one Arabidopsis sequence. Sequences 26 through 29 form a cluster consisting of three bacterial sequences plus an Arabidopsis sequence. The latter contains a chloroplast motif, and together with this clustering pattern, this suggests a chloroplast (and ultimately bacterial) origin. A large and well-supported cluster (sequences 30–54) contains the PP2B (calcineurin) subclass, all of which are animal or fungal sequences. This confirms previous failures to find PP2B catalytic subunit sequences in plants. A PP2A cluster is formed by sequences 56 through 96, which contains five Arabidopsis sequences. A PP4 cluster (sequences 98–108) contains two Arabidopsis sequences, whereas a PP6 cluster (sequences 109–119) has one Arabidopsis protein. Sequences 124 through 128 form a phosphatase cluster that contains four Arabidopsis members; these proteins all have a large N-terminal extension, but their function is unknown. Sequences 166 through 188 comprise a divergent plant PP1 cluster, containing eight Arabidopsis sequences (summarized in Table II).

The majority of ST phosphatase sequences belong to the same protein homology group (Group 1), with the exceptions being sequence 10 (Group 4), sequence 22 (Group 3) and sequence 27 (Group 2; Figs. 2 and 3). There is a single genomic DNA homology subgroup represented within each cluster defined by protein phosphatase domain similarity (homology group letters, Figs. 2 and 3). Most of the genes have one or more homologs (83%; Table III). The dispersion pattern of these genes shows a minority of tandem duplicates (42%) and a majority of more widely distributed copies (58%).

#### Dual-Specificity Phosphatases (DSPs)

An alignment of 169 DSP catalytic domain sequences produced the pattern of relationships depicted in the radial tree of Figure 4 and the topographic cladograms shown in Figures 5 and 6. Most of the sequences are derived from non-plant species—animal being the most numerous, with fungal, bacterial, and viral sequences also represented. The single largest functional group represented is formed by the mitogen-activated protein kinase (MAPK) phosphatases, which together comprise several clusters: sequences 42 through 56, 57 through 65, 70 through 78, and 91 through 93. There are no plant sequences in these clusters.

Sequences 113 through 122 form a cluster that receives high bootstrap support in both NJ (95%) and parsimony (85%). There are eight sequences from animal species, and two from Arabidopsis. The animal sequences are phosphatase and tensin homologs (PTENs). Sequence At3g50110 (gi: 11358703) is annotated as a putative protein phosphatase. Sequence At5g39400 is annotated as a putative protein phosphatase under gi number 10177687, but is noted as “PTEN like” under gi number 15241737 (a clearly synonymous no. with a few small exon differences from 10177687).

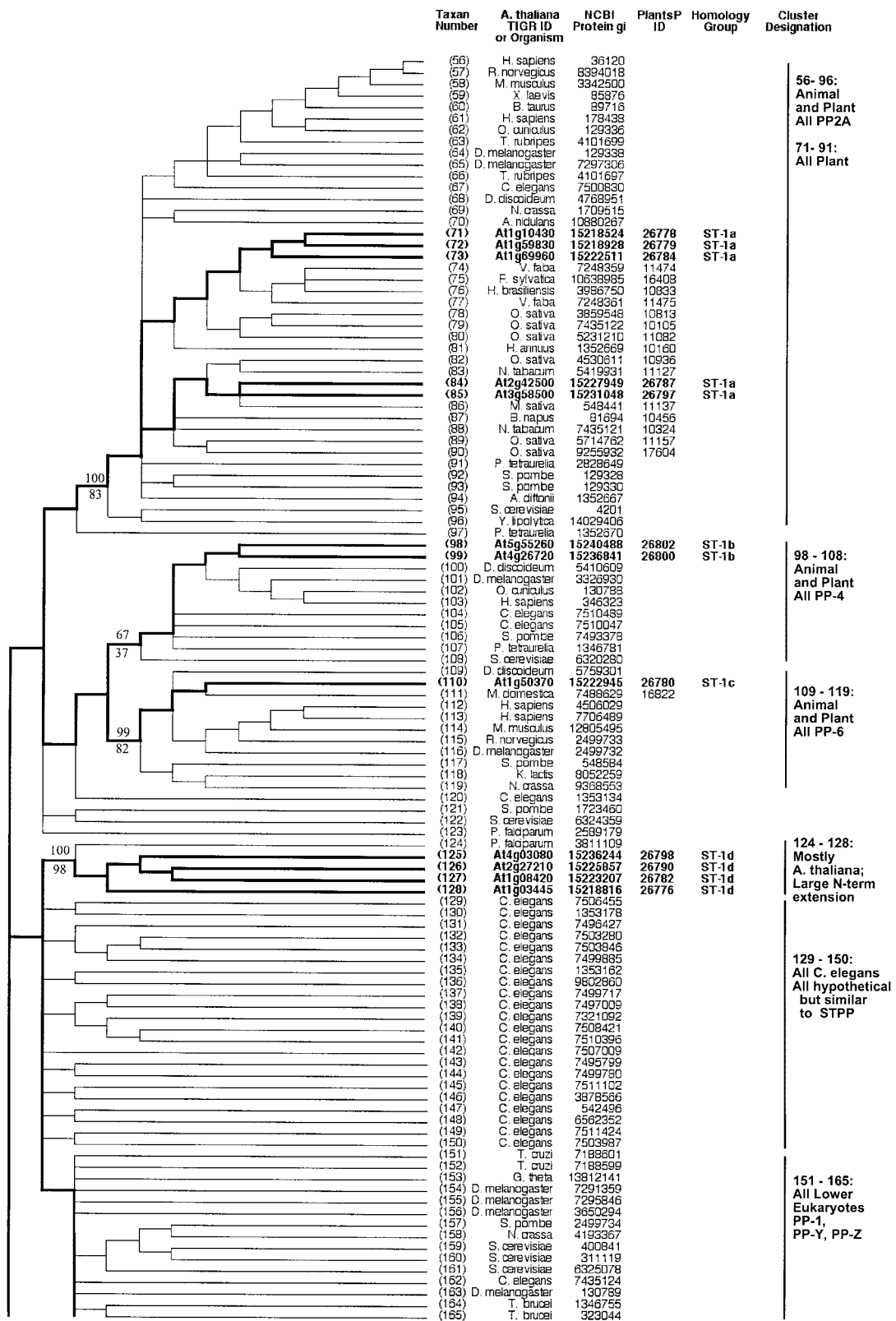
Sequences 123 through 126 are Arabidopsis proteins annotated as unknown function. At2g32960 and At1g05000 are “DSP like” as previously noted. Sequences 137 through 142 comprise a cluster with high bootstrap support in both NJ (100%) and parsimony (82%). This cluster contains animal “myotubularin” proteins. The two Arabidopsis sequences are annotated as “myotubularin like.” Sequences 143 through 150 form a cluster with moderate to strong bootstrap support (96% NJ and 70% parsimony), and consist of a mixture of animal and plant sequences, including sequences from Arabidopsis. Sequences 151 through 154 are from animals, and form a tight cluster. These are “Laforins,” from a particular form of epilepsy called Lafora disease. Sequences 155 through 157 are three Arabidopsis proteins that are associated with the Laforin cluster with moderate support in NJ (79% bootstrap support) and somewhat weaker support in parsimony (41% bootstrap support).

There are six protein homology groups represented within the DSP phosphatase class (Figs. 5 and 6). For the most part, each cluster defined by phosphatase domain similarity comprises a single protein homology group. The exceptions are sequences 38 and 86, which are dispersed members of protein homology Group 1. A minority of the genes have homologs (44%; Table III). All duplicate genes are dispersed on different chromosomes.

#### PP2C Phosphatases

An alignment of 169 PP2C phosphatase catalytic domain sequences produced the pattern of relationships depicted in the radial tree of Figure 7 and the topographic cladograms of Figures 8 and 9. A majority of the sequences (90) are from plants, with the remainder representing a diversity of organisms: animals, protists, bacteria, and viruses.

There are eight distinct clusters that are designated as “all plant,” containing the following groups of sequences: 2 through 14, 15 through 20, 21 through 33, 38 through 39, 52 through 60, 61 through 72, 106 through 109, and 165 through 167. Two sequence clusters consist of Arabidopsis sequences only: 73 through 77 and 90 through 98. The sequences in “Plants #6” and “Arabidopsis #1” have substantial similarity to animal (“78–84:Animal #1”) and fungal



**Figure 2.** Topographic cladogram with additional information for ST protein phosphatase sequences 56 through 165. Branch lengths for the cladogram are unit length. Representative bootstrap values are shown; the value above the line is the ClustalW/NJ value, and the value below the line is the maximum parsimony value (see “Materials and Methods”). Taxon number is as shown in Figure 1, and an appropriate NCBI gi number is provided for each taxon. Information for Arabidopsis sequences is in bold and branches leading to these taxa are broad. The PlantsP plant phosphorylation database (*Legend continues on facing page.*)

("85–87:Fungal #1") sequences, as evidenced by the moderate bootstrap support for a common node connecting these clusters (75% NJ bootstrap support and 68% parsimony bootstrap support).

Very few of the Arabidopsis sequences have been experimentally characterized. The cluster "Plants #1" (2–14) contains sequences At5g57050 ("ABI2") and At4g26080 ("ABI1"), two proteins involved in the abscisic acid (ABA) signaling pathway. Several sequences in this cluster are annotated as being "ABA1 like." Sequence 45 is kinase-associated protein phosphatase (KAPP), which has been shown to be a modulator of a receptor-like kinase signaling pathway. It forms a small cluster with two other plant sequences from *Oryza sativa* and *Zea mays* ("45–47:KAPP Cluster"), but is not closely related to any other Arabidopsis protein.

Most PP2C phosphatase sequences belong to a single protein homology group (Group 1; Figs. 8 and 9). The exceptions are sequences 74 through 77, which are in protein homology Group 2, and sequence 73, which represents a fusion of components from both of these protein homology groups. Most clusters defined by phosphatase domain similarity contain more than one DNA homology group. A majority of the genes (74%) have homologs (Table III). The dispersion pattern of gene copies shows about one-half as many tandem duplicates (31%) as more widely distributed copies (69%).

## DISCUSSION

### ST Phosphatases

The structures of ST phosphatases of the PPP family have been intensively investigated. Our results confirm, with a larger data set and more comprehensive analysis methods, patterns of relationships previously published by others. Cohen (1997) produced a distance tree showing that the PP4s and PP6s are most closely related to the PP2As, and that these are all in turn more closely related to PP1s than they are to PP2Bs. An examination of our radial ST tree shows that the same relationships hold true when two tree inference methods are used. Andreeva and Kutuzov (1999) showed that Arabidopsis PP7 clustered with animal PPEF and *D. melanogaster* rdgC, and that this cluster is most closely related to the PP5s. This pattern is confirmed in our data set. Finally, the relative topological relationships of all major ST protein phosphatase groups (PP1, PP2A, PP4,

PP5, PP6, PPEF/rdgC/PP7, and PP2B) displayed in these two previous distance trees are confirmed in our analysis using both neighbor joining and maximum parsimony. As has been noted in previous investigations, we failed to find any recognizable PP2B catalytic subunit sequences encoded by the genome of Arabidopsis.

Before the completion of the Arabidopsis genomic sequence, Lin et al. (1998, 1999) performed a survey of known PP1 phosphatase sequences. They produced trees by both distance and parsimony methods with consistent topologies, showing distinct clusters containing animal and plant sequences. Our analysis, performed with a larger data set, confirms their findings. In particular, our data reveals a pattern of well-supported subclusters among the plant sequences, which is very similar to that observed by Lin et al. (1999). Table II presents these subclusters, with their bootstrap support, and includes common sequence nomenclature to facilitate comparison with this earlier work.

Finally, our analysis of the ST phosphatases of the PPP family should assist future experimental work by allowing classification of some Arabidopsis sequences into established groups, and focusing attention on a novel group of unknown function. Our trees show that sequence At2g42810 belongs to the PP5 class, and that sequence At1g50370 is a PP6. Sequences 124 through 128 form a well-supported cluster of unknown function, whose first sequence is from the protist *Plasmodium falciparum*, and contains a long N-terminal extension preceding the phosphatase catalytic domain. All of the Arabidopsis sequences except At2g27210 contain a putative transmembrane region.

### DSPs

Proteins of the DSP class dephosphorylate substrate proteins at both Tyr and Ser/Thr residues. These proteins have received a great deal of attention in animal and fungal systems as regulators of signaling cascades involving multiple levels of protein kinase activity. Classic examples of these systems are the MAPK pathway in mammalian cells involving RAS-associated factor, MAP and ERK kinase, and EGF-regulated kinase (Lewis et al., 1998); the environmental stress-activated mammalian pathway involving p38 and c-Jun kinase (Lewis et al., 1998); and the mating pheromone MAPK cascade in budding

**Figure 2.** (Legend continued from facing page.)

(Gribskov et al., 2001) identification number is shown for all plant sequences. The cluster designations shown correspond to those shown in Figure 1. The Institute for Genomic Research ID numbers are shown for the Arabidopsis taxa. For all other taxa, the organism encoding the protein is shown. Standard nomenclature as taken from the NCBI taxonomy database (Wheeler et al., 2000) is used for all free-living organism names; virus abbreviations are shown in "Materials and Methods." Homology group, based on protein and DNA sequence alignments, is as defined in "Materials and Methods." The following sequences were rejected at the final alignment stage: At3g19980, At1g48120, At1g20320, and At5g10900.

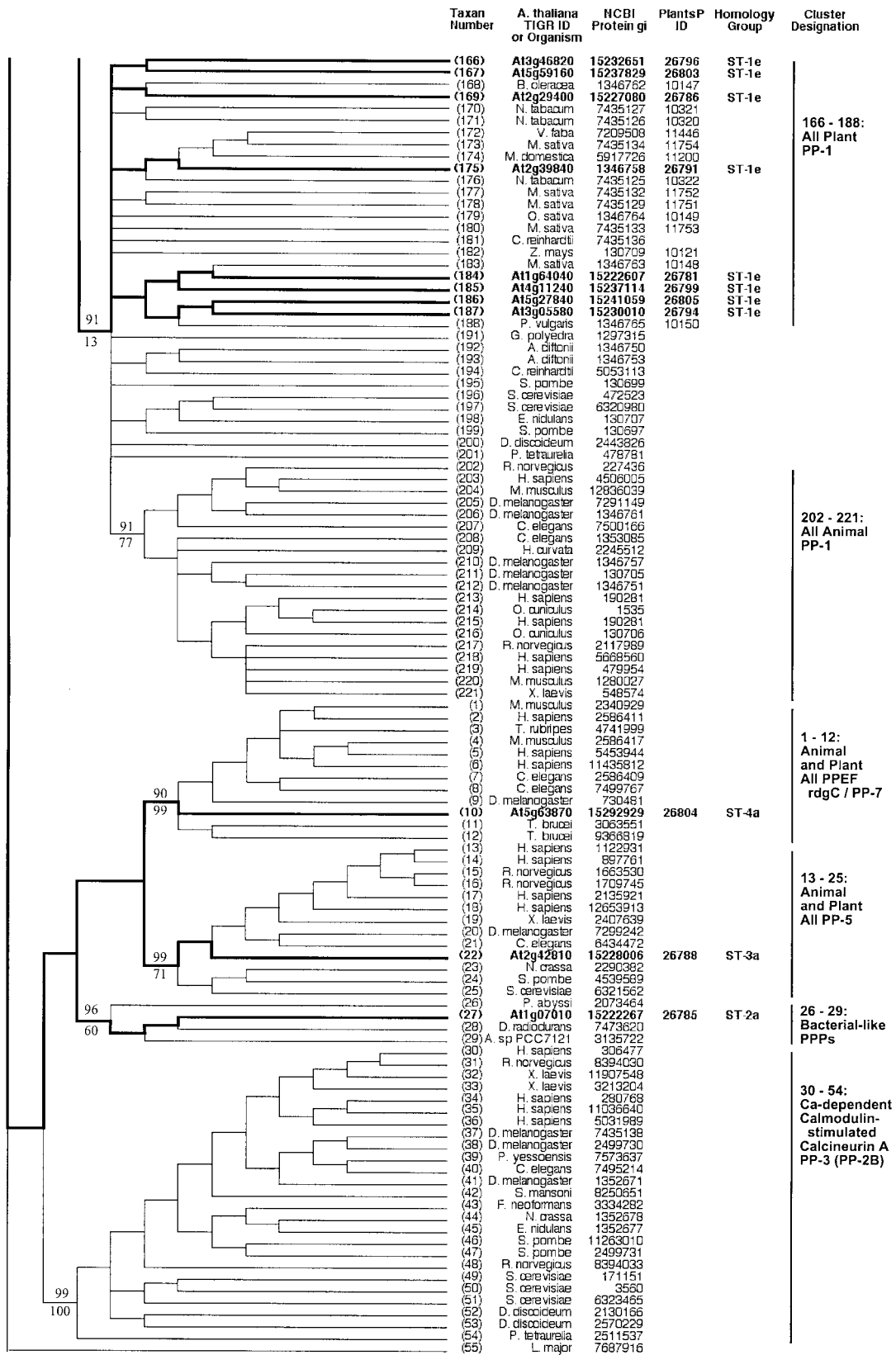


Figure 3. Topographic cladogram with additional information for ST protein phosphatase sequences 166 through 55. Continuation of Figure 2; see legend to Figure 2.

**Table II.** *Arabidopsis* sequences in plant PP1 subclusters of the PPP phosphatase family

ST Class	Percent Bootstrap Support for Cluster <sup>a</sup>	Arabidopsis Member(s)	Tree Taxon No.	Comments
Subcluster 1	99% (NJ)	At3g46820	166	Annotated, confirmed by tree ("TOPP5")
	100% (Pars)	At5g59160	167	
Subcluster 2	100% (NJ)	At2g29400	169	("TOPP1")
	99% (Pars)			
Subcluster 4	81% (NJ)	At2g39840	175	("TOPP4")
	53% (Pars)			
Subcluster 5	96% (NJ)	At4g11240	185	("TOPP7")
	89% (Pars)	At1g64040	184	
Subcluster 6	100% (NJ)	At5g27840	186	Annotated, confirmed by tree ("TOPP8")
	100% (Pars)	At3g05580	187	

<sup>a</sup> 1,000 total neighbor joining (NJ) bootstrap replicates, 200 total maximum parsimony (Pars) bootstrap replicates.

yeast (*Saccharomyces cerevisiae*; Wittenberg and Reed, 1996). Recently, plant systems have been intensively investigated for the presence of components of MAPK signaling pathways (for summary, see Ichimura et al., 2000; Jouannic et al., 2000). There is now evidence for the use of several such cascades in plants; for example, the ethylene hormone response pathway, responses to various biotic and abiotic stresses (e.g. pathogen infection, touch and wounding, dehydration, and low temperature), and cell cycle regulation.

The functional significance of MAPK phosphatases (MKPs) in animals is clearly reflected in their prominence in the DSP tree. There are four clusters, together comprising 30 sequences. Interestingly, there are no plant sequences in any of these clusters. The only Arabidopsis sequences in our DSP set with demonstrated MKP activity are At3g23610 ("AtDSPTP1"), which has been shown to inactivate in vitro an Arabidopsis MAPK ("AtMKP4"; Gupta et al., 1998), and At3g55270 ("AtMKP1"), which was identified as a mutation that increased Arabidopsis sensitivity to genotoxic stress treatments. The wild-type protein was shown to dephosphorylate MAPK proteins (Ulm et al., 2001). The former sequence clusters with sequence At3g06110 in our tree, suggesting

that this might be a logical target for testing for MKP activity. Aside from this one association, there are no other clustering patterns in our tree that might suggest possible Arabidopsis MKP candidates.

We observed the clustering of two Arabidopsis sequences (one of them previously unrecognized) with animal PTENs (sequences 113–122). Human PTEN was originally noted as a tumor suppressor protein whose loss through mutation is involved in a number of human cancers (Maehama and Dixon, 1999; Vazquez and Sellers, 2000). Arabidopsis proteins have been previously reported which share sequence similarity and cluster with two other groups of animal proteins that contain members implicated in human disease (Ganesh et al., 2001; Laporte et al., 2001). These relationships were confirmed in our DSP tree: myotubularins (sequences 137–142) and Laforins (sequences 151–154). The prototype myotubularin (MTM1) is mutated in myotubular myopathy, a disorder of skeletal muscle development. Mutations in the human Laforin gene (EPM2A [Epilepsy progressive myoclonus 2A]) produce a form of inherited epilepsy with neurological degeneration known as Lafora disease (Minassian et al., 2000). For each of these three groups, it has been demonstrated that the animal protein has a multidomain structure, and that regions other than the phosphatase domain used in this study for our alignments is functionally significant, and perhaps involved in disease. Assessment of the significance of the corresponding Arabidopsis proteins will have to await reports detailing structural analysis of these non-phosphatase protein regions, and in vivo functional studies.

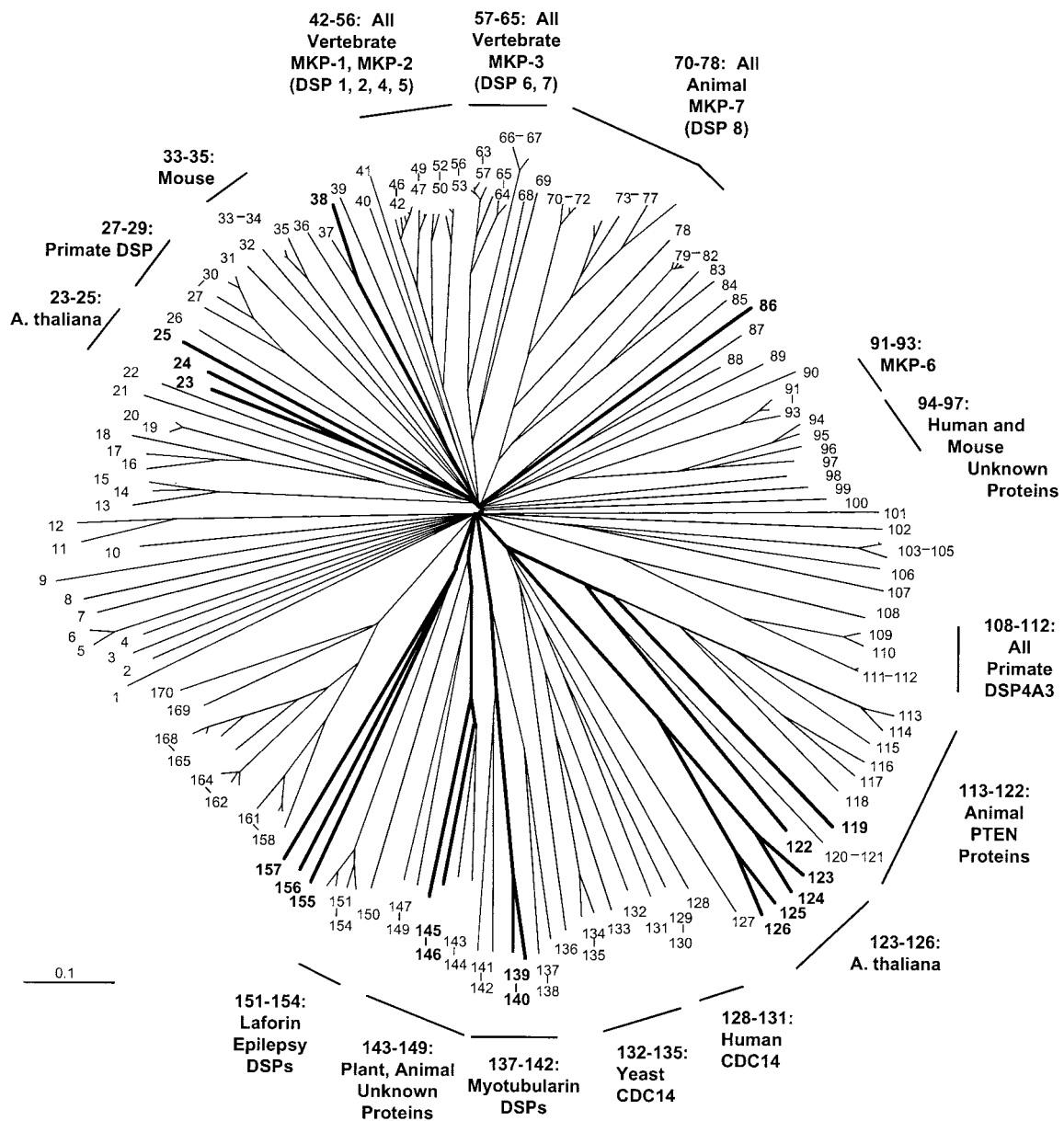
### PP2C Phosphatases

Evidence from a variety of organisms implicates protein phosphatases of the PP2C class as negative modulators of protein kinase pathways activated by various types of environmental stress. In mammalian cells, PP2C activity modulates stress signaling mediated by AMP-activated protein kinase in response to

**Table III.** *Gene homology data*

	Phosphatase Class		
	ST	PP2C	DSP
No. of genes	23	69	18
Proportion of genes with homolog [% (num)]	82.6 (19)	73.9 (51)	44.4 (8)
No. of homologous groups	8	34	13
Tandem duplications [% (num)]	42.1 (8)	31.4 (16)	0
Duplications in same chromosome [% (num)]	15.8 (3)	7.8 (4)	0
Duplications in different chromosomes [% (num)]	42.1 (8)	60.8 (31)	100 (8)





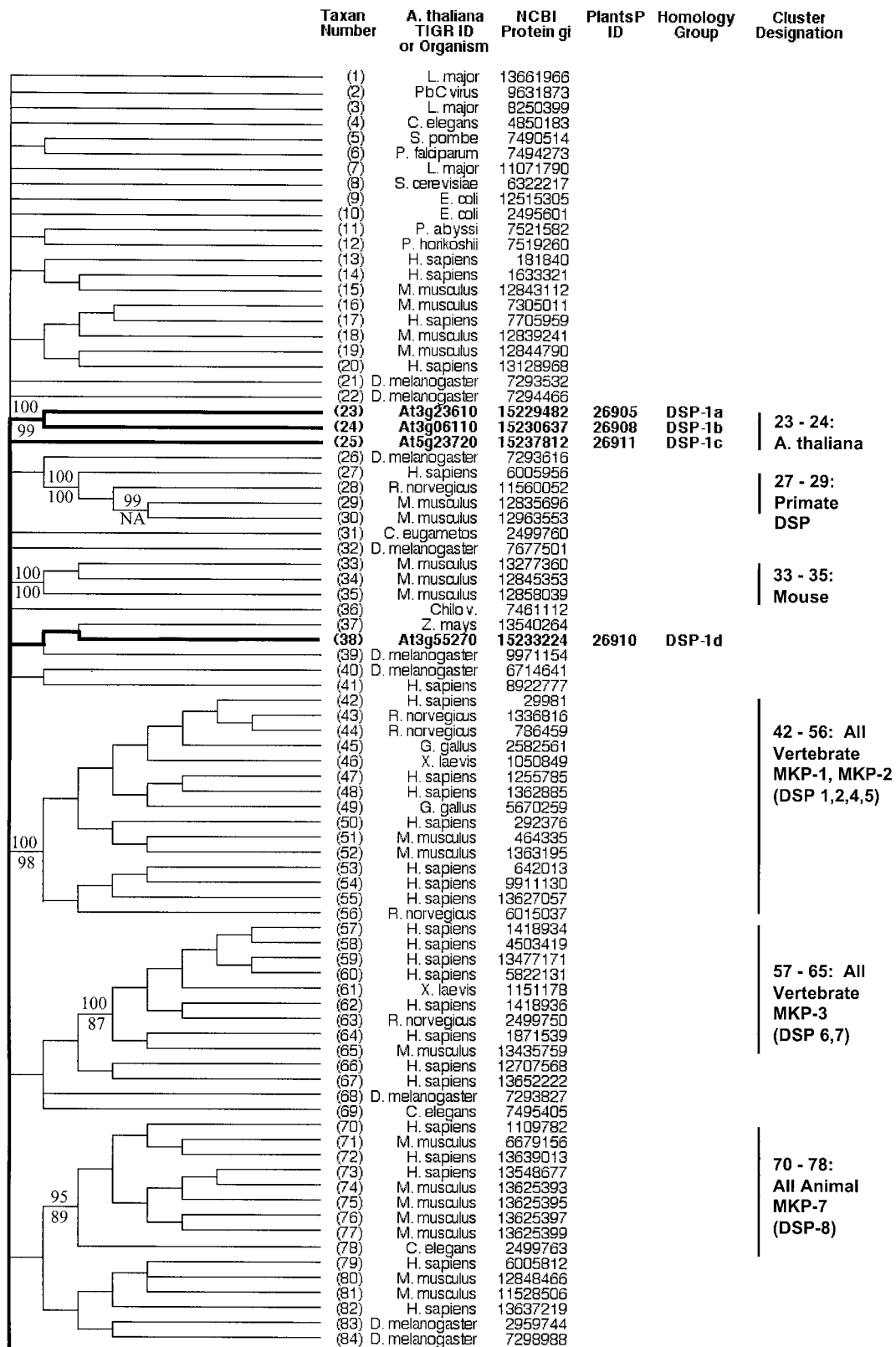
**Figure 4.** Radial phylogenetic tree from DSP sequence comparisons. Correspondence between taxon or sequence number and NCBI protein gi number is given in Figures 5 and 6. Figure characteristics are as described in legend to Figure 1.

energy depletion (Moore et al., 1991; Corton et al., 1994; Rodriguez, 1998) and the p38 and c-Jun kinase MAPK pathways responding to environmental factors (Hanada et al., 1998; Takekawa et al., 1998; Luan, 2000). In budding yeast, PP2C proteins appear to antagonize the hyperosmotic stress response mediated by the pathway ending in the MAPK HOG1 (Maeda et al., 1994; Rodriguez, 1998). PP2Cs are also involved in the negative modulation of stress response signaling in fission yeast (*Schizosaccharomyces pombe*; Shiozaki and Russell, 1995; Gaits et al., 1997; Rodriguez, 1998).

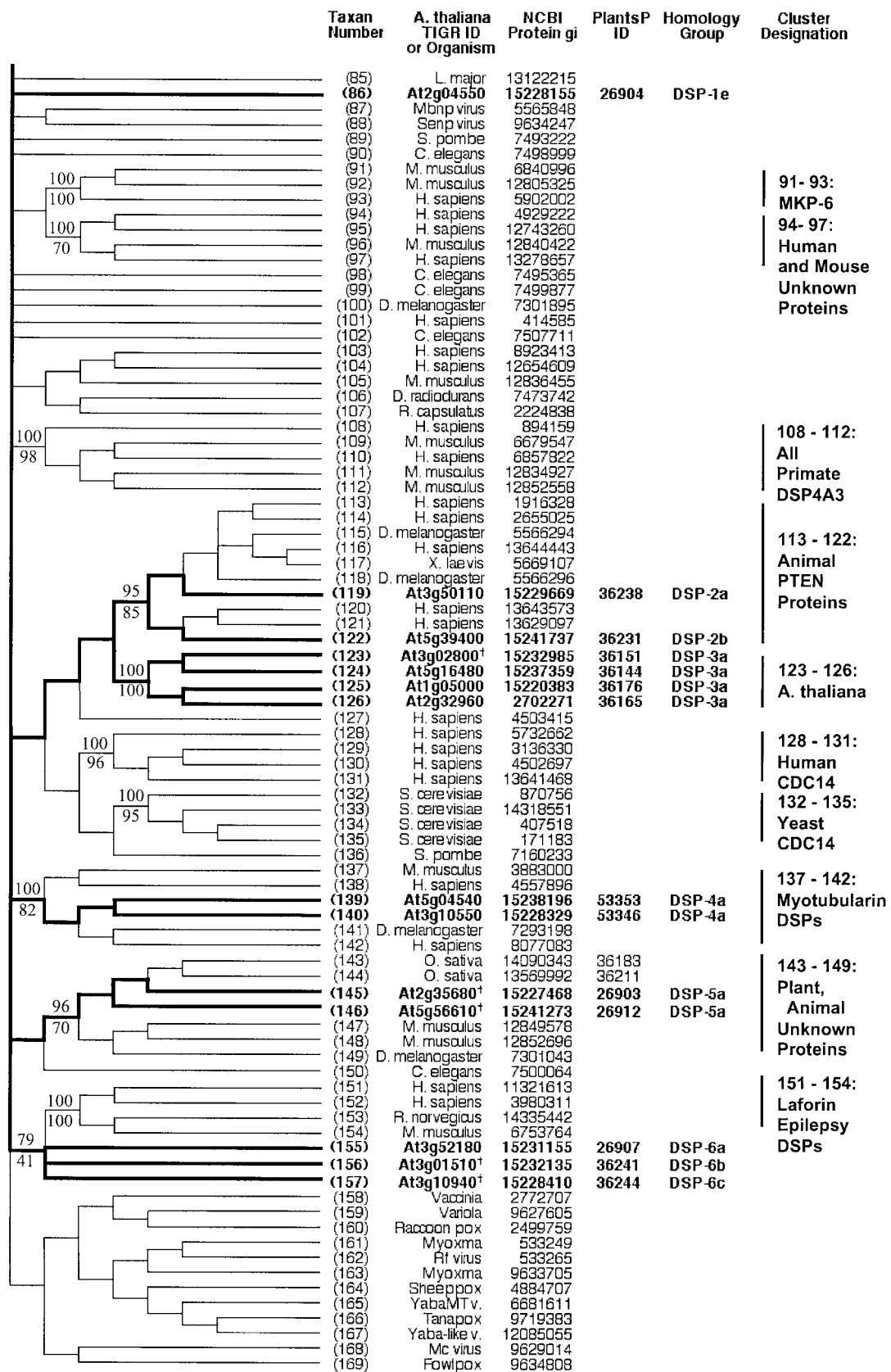
Before the completion of the Arabidopsis genome sequencing project, relatively few PP2C proteins

were known in plants. ABI1 and ABI2 were originally identified as a result of mutations, which conferred a phenotype of insensitivity to the actions of the hormone ABA (Leung et al., 1994, 1997; Meyer et al., 1994). KAPP was originally identified as an interacting protein with a receptor-like kinase (RLK5; Stone et al., 1994). Biochemical and genetic studies suggest that KAPP acts as a negative modulator of the CLV1 signaling pathway (Williams et al., 1997; Stone et al., 1998; Luan, 2000).

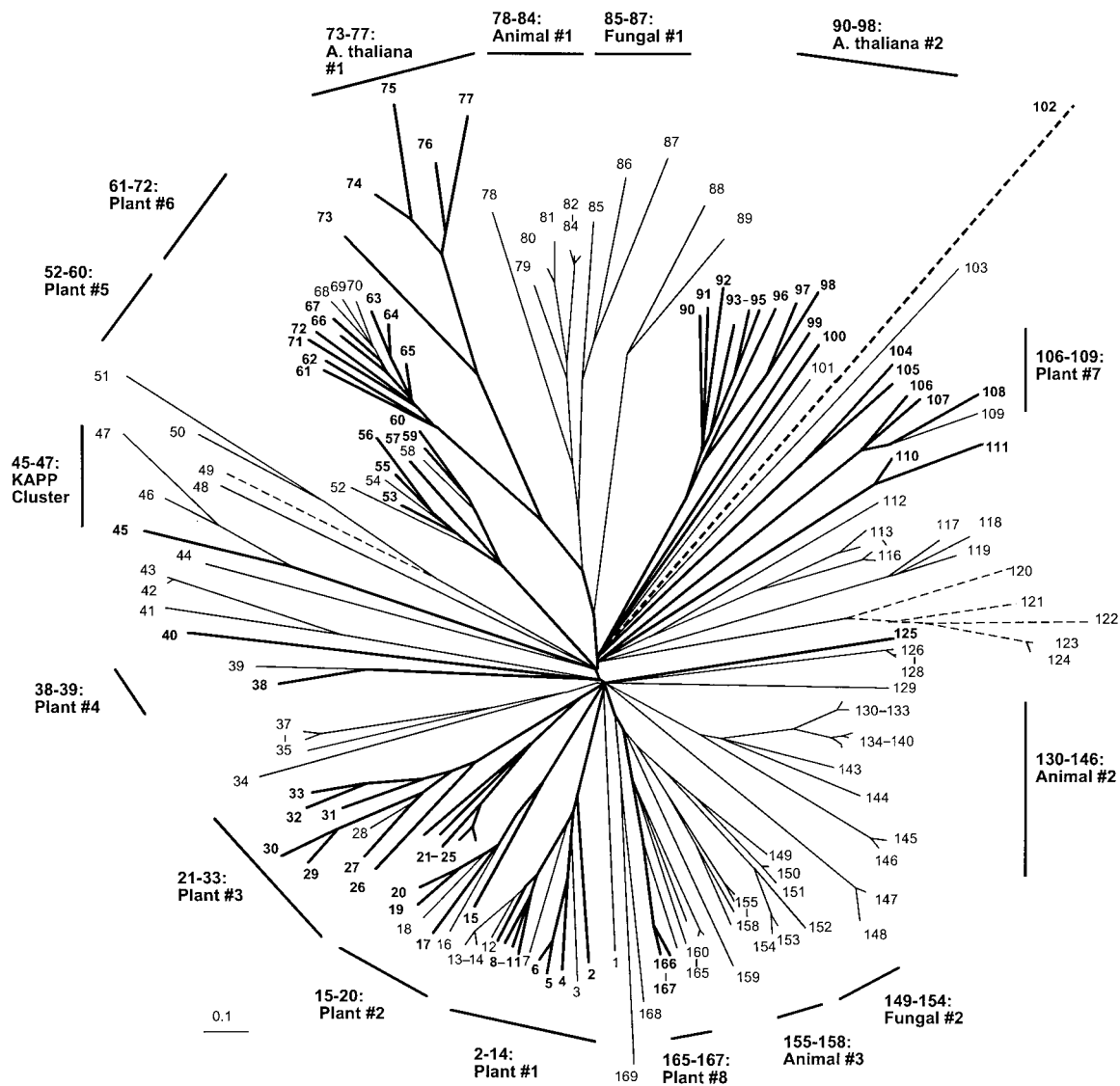
In this study, we have analyzed 69 PP2C proteins encoded by the Arabidopsis genome, including three that were not previously recognized as protein phosphatases. Several sequences in the "Plants #1" cluster



**Figure 5.** Topographic cladogram with additional information for DSP protein phosphatase sequences 1 through 84. Taxon number is as shown in Figure 4. Figure characteristics are as described in legend to Figure 2. The following sequences are newly identified protein phosphatase candidates (previously annotated as putative or unknown): At3g10940, At2g35680, At3g01510, At5g56610, and At3g02800 (designated in the figure with the symbol †). Sequences At1g05000 and At2g32960 are designated as “DSP like” (see text and “Methods and Materials”). Sequence At3g19420 was rejected at the final alignment stage.



**Figure 6.** Topographic cladogram with additional information for DSP protein phosphatase sequences 85 through 169. Continuation of Figure 5; see legend to Figure 5.



**Figure 7.** Radial phylogenetic tree from PP2C protein phosphatase sequence comparisons. Correspondence between taxon or sequence number and NCBI protein gi number is given in Figures 8 and 9. Figure characteristics are as described in legend to Figure 1.

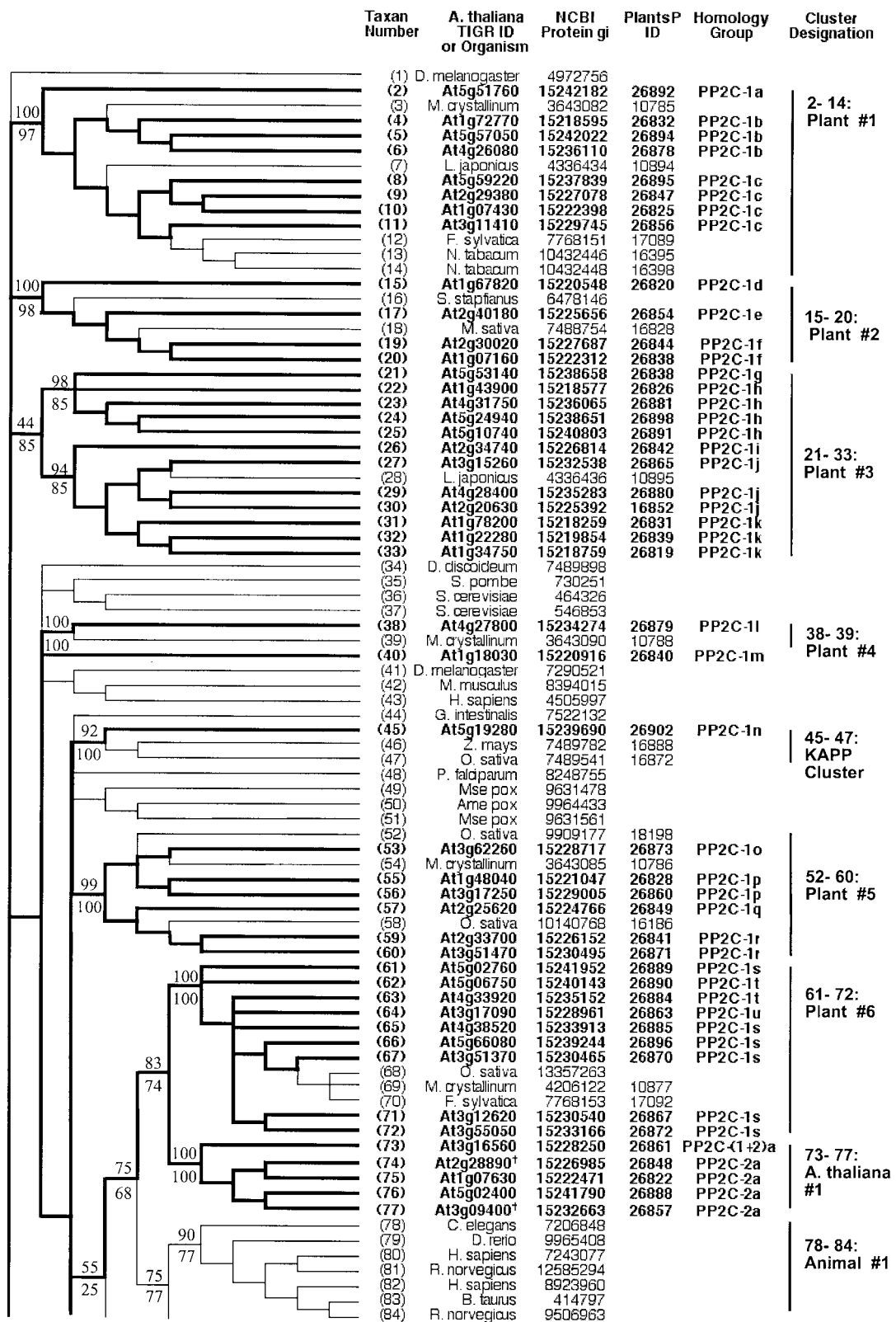
have marked similarity to ABI1, and might be logical candidates for testing in ABA signaling. The “KAPP cluster” is small, and contains only one sequence from Arabidopsis. It has been suggested that this might imply a promiscuous function of the KAPP protein, binding to a number of substrate molecules (Rodriguez, 1998). There is little information to guide hypotheses concerning the functions of the other sequences located in the various PP2C clusters.

Therefore, the bulk of these PP2C sequences hold out both challenge and promise for future understanding of plant biology. The clusters with high bootstrap support documented in this study represent groups of sequences with highly similar structures, which might be expected to serve similar functions. Thus, a systematic use of these clusters to guide the isolation of knockout lines and the design

of biochemical experiments should allow the plant research community to most rapidly and efficiently canvas the diversity of functional capacities the collection of clusters represents.

### Evolution of Phosphatase Genes in Arabidopsis

Gene homology can be inferred by reference to conservation of exon/intron architecture, by conservation of protein structure, and by the degree of similarity of the coding DNA and encoded amino acid sequence. We have used these properties together to assess the pattern of duplication of genes encoding the various classes of protein phosphatases in Arabidopsis. We have adapted the criteria of Riechmann et al. (2000), in a comparative genomic study of transcription factors in Arabidopsis and other organ-



**Figure 8.** Topographic cladogram with additional information for PP2C protein phosphatase sequences 1 through 84. Taxon number is as shown in Figure 7. Figure characteristics are as described in legend to Figure 2. The following sequences are newly identified protein phosphatase candidates (previously annotated as putative or unknown): At2g28890, At3g09400, and At1g75010 (designated in the figure with the symbol †). The following sequences were rejected at the final alignment stage: At2g46920, At2g35350, At3g23360, At3g27140, At4g08260, At4g11040, and At1g17550.

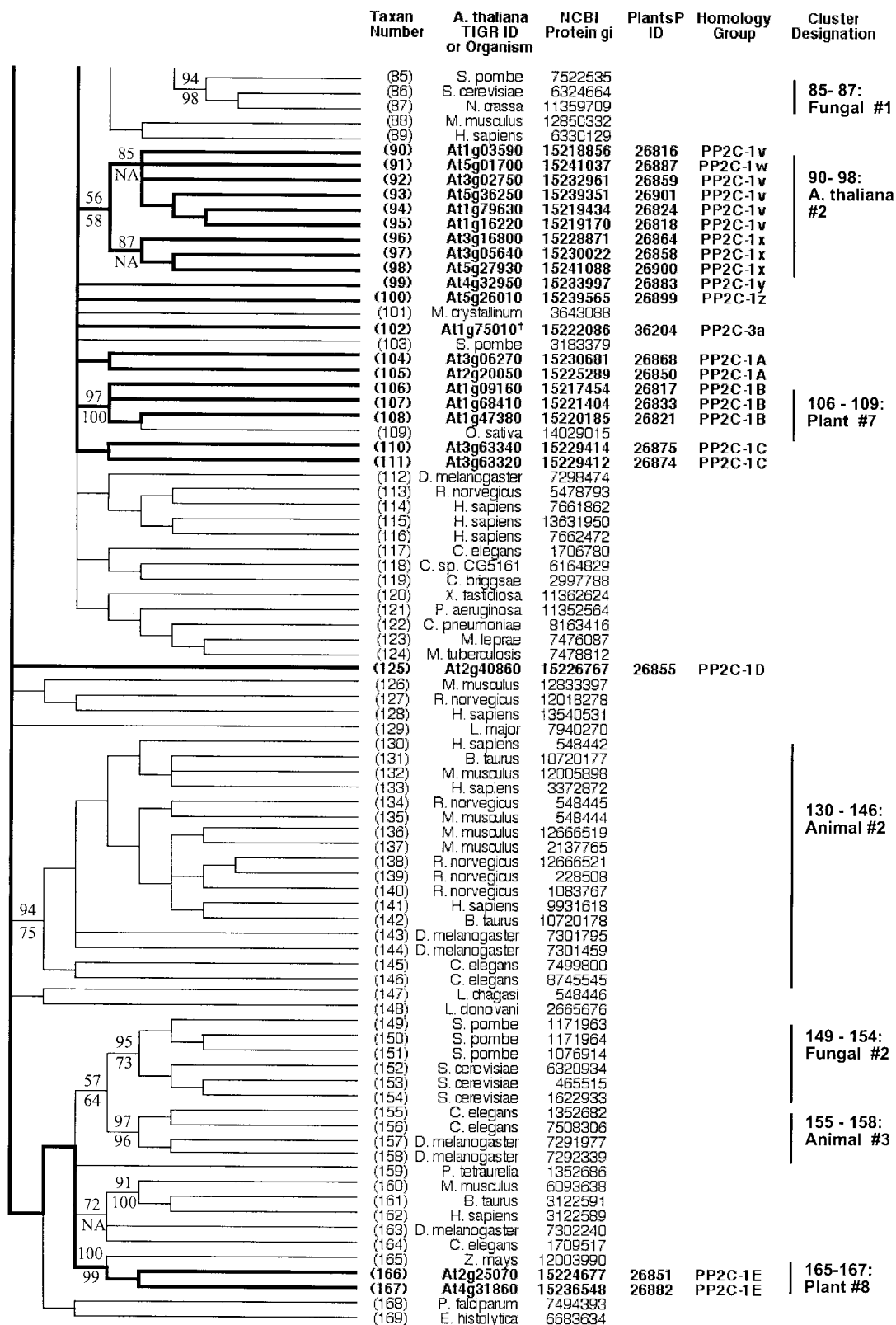


Figure 9. Topographic cladogram with additional information for PP2C protein phosphatase sequences 85 through 169. Continuation of Figure 8; see legend to Figure 8.

isms, for assessing the presence and chromosomal distribution of homologs. The result, summarized in Table III, is an intriguing data set where there are distinct differences between the various protein phosphatase structural classes in the proportion of genes with homologs and their chromosomal distribution, and the correspondence between clusters as revealed by phosphatase domain similarity and homology groups derived from whole protein and genomic DNA sequence similarity analysis. These data should prove useful for future detailed analyses of protein phosphatase gene evolution.

## MATERIALS AND METHODS

### Assessment of the Quality of Genome Project Gene Predictions

The following procedure was developed to assess the quality of predicted genomic sequence. A keyword search of the NCBI nonredundant protein database was performed in ENTREZ to retrieve Arabidopsis sequences annotated as putative protein phosphatases. The nucleotide sequence corresponding to each protein was submitted to analysis by two gene prediction programs: GENSCAN (Burge and Karlin, 1997) and GENMARK (Lukashin and Borodovsky, 1998). The amino acid sequence predicted from these gene-finding programs was used as a query in a BLASTP search of the nonredundant database. The annotated sequence appeared as a high-scoring hit, and the alignment of the two sequences was compared. Discrepancies (e.g. amino acid residues present in one sequence but missing in the other) were investigated further. This was done by using as an experimental reference the nucleotide sequence present in the NCBI EST database from Arabidopsis. Because an EST is derived from a single-pass read of a transcribed mRNA, it provides an anchor for a gene prediction. Our predicted nucleic acid coding sequence was used as a query in a BLASTN search of the EST database. Strong hits to Arabidopsis sequences were examined further. The EST sequence was then translated, and its amino acid sequence aligned with both the annotated amino acid sequence and our predicted amino acid sequence. The amino acid sequence, which agreed with the EST translation, was deemed to be correct.

### Construction of Sequence-Specific Probes for Protein Phosphatase Structural Subclasses

We constructed sequence probes specific for the various subclasses of protein phosphatase: PTP, ST, DSP, and PP2C. The procedure was as follows.

First, a starting alignment was obtained, either from the published literature, or from the "seed" alignment in the protein families (Pfam) database (Bateman et al., 2000) of Hidden Markov models. Sequences in each subset were aligned with the multiple sequence alignment program CLUSTALW (Higgins et al., 1996), using default settings for gap penalties. These alignments were then examined and edited by hand as required. Finished alignments were

then used to construct sequence "profiles" according to the refined method of Gribskov and Veretnik (1996; <http://motifweb.sdsc.edu/>). In brief, this involves weighting the various sequences to reflect a more accurate random distribution, and construction of a position-specific scoring matrix that summarizes the probability of occurrence of each possible amino acid residue at each sequence position. This profile was then used to search databases (NCBI nonredundant database, NCBI all-plants database, or The Institute for Genomic Research Arabidopsis database) for high-scoring sequences. In certain instances, an alternative procedure was used to acquire candidate sequences for further analysis. Sets of unaligned sequences were searched for conserved motifs by application of the "expectation maximization" approach, using the MEME program (Bailey and Elkan, 1995; <http://meme.sdsc.edu/meme/website/meme.html>). These motifs were then used to search the NCBI nonredundant protein database for high-scoring Arabidopsis sequences using the MAST program (Bailey and Gribskov, 1998; <http://meme.sdsc.edu/meme/website/mast.html>).

Resulting candidate sequences were then placed into the multiple sequence alignment using the "Profile/Structure" alignment feature of CLUSTALW. Resulting alignments were then examined and edited by hand. In particular, we sought to confirm the presence of amino acid residues known to be highly conserved and catalytically active in the phosphatase domain of the various protein phosphatase structural subclasses. A very useful compilation of these conserved motifs and residues for various phosphatase subclasses is presented and referenced in the study of Shi et al. (1998). A rigorous standard was applied, and only those sequences containing all necessary residues were retained in the alignment. The modified alignments were then used to construct new sequence profiles, and search the databases again, in an iterative fashion, until no further sequences were retained in the alignments. Finally, the alignment was purged of duplicate sequences.

### Repetitive Search of Databases for Candidate Phosphatase Sequences

The BLAST algorithm (Altschul et al., 1997) was used repetitively to search databases for new phosphatase candidate sequences (Gribskov et al., 2001). The procedure used was similar to that previously published under the moniker "family pair-wise search" (Grundy, 1998). In brief, starting groups containing known phosphatase sequences were used in sequential queries in BLASTP database searches. Those database sequences that obtained high scores from a number of these query sequences (i.e. were common "hits" for this query sequence group) were then retained for further analysis. These candidate protein phosphatase sequences were then entered into the multiple sequence alignments described above, and screened for the presence of critical conserved and catalytic residues. Those sequences possessing these residues were retained as new candidate protein phosphatases.

## Phylogenetic Tree Inference

Multiple sequence alignments constructed as described above were subjected to "bootstrap resampling." In brief, this entails randomly removing columns of data in the multiple sequence alignment and replacing them with replicated columns from elsewhere in the alignment, so that the alignment size is not altered. These bootstrap replicate alignments were then utilized to construct phylogenetic trees by the neighbor joining method (Saitou and Nei, 1987) and by maximum parsimony using appropriate PHYLIP (Felsenstein, 1996) programs. "Consensus" trees summarizing the topologies found among the bootstrap replicate trees are presented. In the figures, clusters are usually displayed, labeled, and discussed where the topology of neighbor joining and maximum parsimony trees agreed, and where the bootstrap support obtained from each method exceeded 50%. Exceptions are made in a few instances for nodes with lower support, where the relationships described seemed especially noteworthy.

In the course of our database searches, we collected a large number of sequences from a variety of organisms: animals, plants, fungi, protists, bacteria, archaea, and viruses. Because the sequences are presented in our various phosphatase class phylogenetic trees below, their species of origin is abbreviated by a standard genus initial (capitalized) followed by the species name in lowercase. Details about species referenced can be obtained by going to the NCBI taxonomy web site (<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>). Abbreviations for viruses encoding protein phosphatases analyzed here are as follows: Ame pox, *Amsacta moorei* entomopoxvirus; Chilo v., Chilo iridescent virus; Fowlpox, Fowlpox virus; Mbnp virus, *Mamestra brassicae* nucleopolyhedrovirus; Mc virus, *Molluscum contagiosum* virus subtype 1; Mse pox, *Melanoplus sanguinipes* entomopoxvirus; Myxoma, *Myxoma* virus; PbC virus, *Paramecium bursaria* Chlorella virus 1; R<sub>F</sub> virus, Rabbit fibroma virus; Raccoon pox, Raccoon pox virus; Senp virus, *Spodoptera exigua* nucleopolyhedrovirus; Sheeppox, Sheeppox virus; Tanapox, Tanapox virus; Vaccinia, *Vaccinia* virus; Variola, *Variola* virus; Yaba-like v., Yaba-like disease virus; and YabaMT v., Yaba monkey tumor virus.

## Gene Evolution via Homology Analysis

The main quantitative criteria used to define homology was sequence similarity based on expect values from two-sequence BLAST comparisons:  $E < e^{-10}$  for protein comparisons and  $E < e^{-6}$  for genomic DNA comparisons. Homologs must satisfy both criteria. NCBI protein gi numbers were used for protein sequence, and the NCBI RefSeq NM sequences for the genomic DNA sequence of the gene. Protein homology group is indicated by number, and genomic DNA homology group within each protein group is indicated by first lowercase letter and then uppercase letter, in the Homology Group column of Figures 2, 3, 5, 6, 8, and 9. Homology groups are assumed to arise from genetic recombination events that have dispersed gene copies to different locations within the Arabidopsis genome. The distribution of this dispersion is summarized in

Table III. Utilizing the criteria of Riechmann et al. (2000), we classify homologs as "tandem duplications" if they lie within 50 kb on the same chromosome. Homologs lying on the same chromosome with a separation greater than 50 kb are designated as "duplications in same chromosome."

Received February 7, 2002; returned for revision March 5, 2002; accepted May 4, 2002.

## LITERATURE CITED

- Altschul SF, Madden TL, Schäffer 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
- Andreeva AV, Kutuzov MA (1999) RdgC/PP5-related phosphatases: novel components in signal transduction. *Cell Signal* **11**: 555–562
- Bailey TL, Elkan C (1995) Unsupervised learning of multiple motifs in biopolymers using EM. *Machine Learning* **21**: 51–80
- Bailey TL, Gribskov M (1998) Combining evidence using *p*-values: application to sequence homology searches. *Bioinformatics* **14**: 48–54
- Barford D (1996) Molecular mechanisms of the protein serine-threonine phosphatases. *Trends Biochem Sci* **21**: 407–412
- Bateman A, Birney E, Durbin R, Eddy SR, Howe KI, Sonnhammer ELL (2000) The Pfam protein families database. *Nucleic Acids Res* **28**: 263–266
- Bork P, Brown NP, Hegyi H, Schultz J (1996) The protein phosphatase 2C (PP2C) superfamily: detection of bacterial homologues. *Protein Sci* **5**: 1421–1425
- Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* **268**: 78–94
- Chernoff J (1999) Protein tyrosine phosphatases as negative regulators of mitogenic signaling. *J Cell Physiol* **180**: 173–181
- Cohen P (1989) Structure and regulation of protein phosphatases. *Ann Rev Biochem* **58**: 453–508
- Cohen PTW (1997) Novel protein serine-threonine phosphatases: variety is the spice of life. *Trends Biochem Sci* **22**: 245–251
- Corton JM, Gillespie JG, Hardie DG (1994) Role of the AMP-activated protein kinase in the cellular stress response. *Curr Biol* **4**: 315–324
- Das AK, Helps NR, Cohen PTW, Barford D (1996) Crystal structure of the protein serine-threonine phosphatase 2C at 2.0 Å resolution. *EMBO J* **15**: 6798–6809
- den Hertog J (1999) Protein-tyrosine phosphatases in development. *Mech Dev* **85**: 3–14
- Egloff MP, Cohen PTW, Reinemer P, Barford D (1995) Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *J Mol Biol* **254**: 942–959
- Fauman EB, Saper MA (1996) Structure and function of the protein tyrosine phosphatases. *Trends Biochem Sci* **21**: 413–417



- Felsenstein J** (1996) Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol* **266**: 418–426
- Gaits F, Shiozaki K, Russell P** (1997) Protein phosphatase 2C acts independently of the stress-activated kinase cascade to regulate the stress response in fission yeast. *J Biol Chem* **272**: 17873–17879
- Ganesh S, Agarwala KL, Amano K, Suzuki T, Delgado-Escueta AV, Yamakawa K** (2001) Regional and developmental expression of Epm2a gene and its evolutionary conservation. *Biochem Biophys Res Commun* **283**: 1046–1053
- Goldberg J, Huang HB, Kwon YG, Greengard P, Nairn AC, Kuriyan J** (1995) Three-dimensional structure of the catalytic subunit of protein serine-threonine-phosphatase-1. *Nature* **376**: 745–753
- Gribskov M, Fana F, Harper J, Hope DA, Harmon AC, Smith DW, Tax FE, Zhang G** (2001) PlantsP: a functional genomics database for plant phosphorylation. *Nucleic Acids Res* **29**: 111–113
- Gribskov M, Veretnik S** (1996) Identification of sequence patterns with profile analysis. *Methods Enzymol* **266**: 198–211
- Griffith JP, Kim JL, Kim EE, Sintchak MD, Thomson JA, Fitzgibbon MJ, Fleming MA, Caron PR, Hsiao K, Navia MA** (1995) X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. *Cell* **82**: 507–522
- Grundy WN** (1998) Homology detection via family pairwise search. *J Comput Biol* **5**: 479–492
- Gupta R, Huang Y, Kieber JJ, Luan S** (1998) Identification of a dual-specificity protein phosphatase that inactivates MAP kinase from *Arabidopsis*. *Plant J* **16**: 581–590
- Hanada M, Kobayashi T, Ohnishi M, Ikeda S, Wang H, Katsura K, Yanagawa Y, Hiraga A, Kanamaru R, Tamura S** (1998) Selective suppression of stress-activated protein kinase pathway by protein phosphatase 2C in mammalian cells. *FEBS Lett* **437**: 172–176
- Hanks SK, Hunter T** (1995) The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* **9**: 576–596
- Higgins DG, Thompson JD, Gibson TJ** (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* **266**: 383–401
- Ichimura K, Mizoguchi T, Yoshida R, Yuasa T, Shinozaki K** (2000) Protein phosphorylation and dephosphorylation in environmental stress responses in plants. *Adv Bot Res* **32**: 355–377
- Iten M, Hoffmann T, Grill E** (1999) Receptors and signaling components of plant hormones. *J Recept Signal Transduct Res* **19**: 41–58
- Jouannic S, Leprince A-S, Hamal A, Picaud A, Kreis M, Henry Y** (2000) Plant mitogen-activated protein kinase signalling pathways in the limelight. *Adv Bot Res* **32**: 299–354
- Laporte J, Blondeau F, Buj-Bello A, Mandel J-L** (2001) The myotubularin family: from genetic disease to phosphoinositide metabolism. *Trends Genet* **17**: 221–228
- Leung J, Bouvier-Durand M, Morris PC, Guerrier D, Chedford F, Giraudat J** (1994) *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phosphatase. *Science* **264**: 1448–1452
- Leung J, Merlot S, Giraudat J** (1997) The *Arabidopsis* *ABSCISIC ACID-INSENSITIVE2* (*ABI2*) and *ABI1* genes encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. *Plant Cell* **9**: 759–771
- Lewis TS, Shapiro PS, Ahn NC** (1998) Signal transduction through MAP kinase cascades. *Adv Cancer Res* **74**: 49–139
- Lin Q, Buckler IVES, Muse SV, Walker JC** (1999) Molecular evolution of type 1 serine/threonine protein phosphatases. *Mol Phylogenet Evol* **12**: 57–66
- Lin Q, Li J, Smith RD, Walker JC** (1998) Molecular cloning and chromosomal mapping of type one serine/threonine protein phosphatases in *Arabidopsis thaliana*. *Plant Mol Biol* **37**: 471–481
- Lohse DL, Denu JM, Dixon JE** (1995) Insights derived from the structures of the Ser-Thr phosphatases calcineurin and protein phosphatase 1. *Structure* **3**: 987–990
- Luan S** (2000) Protein phosphatases: structure, regulation, and function. *Adv Bot Res* **32**: 67–107
- Lukashin AV, Borodovsky M** (1998) GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* **26**: 1107–1115
- Maeda T, Wurgler-Murphy SM, Saito H** (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**: 242–245
- Maehama T, Dixon JE** (1999) PTEN: a tumour suppressor that functions as a phospholipid phosphatase. *Trends Cell Biol* **9**: 125–128
- Meskiene I, Bogre L, Glaser W, Balog J, Brandstotter M, Zwerger K, Ammerer G, Hirt H** (1998) MP2C, a plant protein phosphatase 2C, functions as a negative regulator of mitogen-activated protein kinase pathways in yeast and plants. *Proc Natl Acad Sci USA* **95**: 1938–1943
- Meyer K, Leube MP, Grill E** (1994) A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* **264**: 1452–1455
- Minassian BA, Ianzano L, Meloche M, Andermann E, Rouleau GA, Delgado-Escueta AV, Scherer SW** (2000) Mutation spectrum and predicted function of laforin in Lafora's progressive myoclonus epilepsy. *Neurology* **55**: 341–346
- Moore F, Weekes J, Hardie DG** (1991) Evidence that AMP triggers phosphorylation as well as direct allosteric activation of rat liver AMP-associated protein kinase. *Eur J Biochem* **199**: 691–697
- Ramponi G, Stefani M** (1997) Structural, catalytic, and functional properties of low Mr phosphotyrosine protein phosphatases. Evidence for a long evolutionary history. *Int J Biochem Cell Biol* **29**: 279–292
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR et al.** (2000) *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science* **290**: 2105–2110
- Rodriguez PL** (1998) Protein phosphatase 2C (PP2C) function in higher plants. *Plant Mol Biol* **38**: 919–927

- Saitou N, Nei M** (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425
- Schillace RV, Scott JD** (1999) Organization of kinases, phosphatases, and receptor signaling complexes. *J Clin Invest* **103**: 761–765
- Shi L, Potts M, Kennelly PJ** (1998) The serine, threonine, and/or tyrosine-specific protein kinases and protein phosphatases of prokaryotic organisms: a family portrait. *FEMS Microbiol Rev* **22**: 229–253
- Shiozaki K, Russell P** (1995) Counteractive roles of protein phosphatase 2C (PP2C) and a MAP kinase kinase homolog in the osmoregulation of fission yeast. *EMBO J* **14**: 492–502
- Stone JM, Collinge MA, Smith RD, Horn MA, Walker JC** (1994) Interaction of a protein phosphatase with an *Arabidopsis* serine-threonine receptor kinase. *Science* **266**: 793–795
- Stone JM, Trotochaud AE, Walker JC, Clark SE** (1998) Control of meristem development by CLAVATA1 receptor kinase and kinase-associated protein phosphatase interactions. *Plant Physiol* **117**: 1217–1225
- Takekawa M, Maeda T, Saito H** (1998) Protein phosphatase 2C- $\alpha$  inhibits the human stress-responsive p38 and JNK MAPK pathways. *EMBO J* **17**: 4744–4752
- Ulm R, Revenkova E, di Sansebastiano G-P, Bechtold N, Paszkowski J** (2001) Mitogen-activated protein kinase phosphatase is required for genotoxic stress relief in *Arabidopsis*. *Genes Dev* **15**: 699–709
- Vazquez F, Sellers WR** (2000) The PTEN tumor suppressor protein: an antagonist of phosphoinositide 3-kinase signaling. *Biochim Biophys Acta* **1470**: M21–M35
- Wang ML, Belmonte S, Kim U, Dolan M, Morris JW, Goodman HM** (1999) A cluster of ABA-regulated genes on *Arabidopsis thaliana* BAC T07M07. *Genome Res* **9**: 325–333
- Wheeler DL, Chappay C, Lash AE, Leipe DD, Madden TL, Schuler GD, Tatusova TA, Rapp BA** (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **28**: 10–14
- Williams RW, Wilson JM, Meyerowitz EM** (1997) A possible role for kinase-associated protein phosphatase in the *Arabidopsis* CLAVATA1 signaling pathway. *Proc Natl Acad Sci USA* **94**: 10467–10472
- Wittenberg C, Reed SI** (1996) Plugging it in: signaling circuits and the yeast cell cycle. *Curr Opin Cell Biol* **8**: 223–230