



A Cluster of ABA-Regulated Genes on *Arabidopsis thaliana* BAC T07M07

Ming Li Wang, Stephen Belmonte, Ulandt Kim, et al.

Genome Res. 1999 9: 325-333

Access the most recent version at doi:[10.1101/gr.9.4.325](https://doi.org/10.1101/gr.9.4.325)

References This article cites 40 articles, 22 of which can be accessed free at:
<http://genome.cshlp.org/content/9/4/325.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a woman wearing a red and white superhero cape and mask, and the logo for "CELLECTA" which consists of a green molecular structure.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

A Cluster of ABA-Regulated Genes on *Arabidopsis thaliana* BAC T07M07

Ming Li Wang,^{1,2,3} Stephen Belmonte,² Ulandt Kim,^{2,4} Maureen Dolan,^{2,5}
John W. Morris,^{2,6} and Howard M. Goodman^{1,2,7}

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115 USA; ²Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114 USA

Arabidopsis thaliana BAC T07M07 encoding the abscisic acid-insensitive 4 (*ABI4*) locus has been sequenced completely. It contains a 95,713-bp insert and 24 predicted genes. Most putative genes were confirmed by gel-based RNA profiling and a cluster of ABA-regulated genes was identified. One of the 24 genes, designated *PP2C5*, encodes a putative protein phosphatase 2C. The encoded protein was expressed in *Escherichia coli*, and its enzyme activity in vitro was confirmed.

[The sequence data described in this paper have been submitted to GenBank under accession no. AF085279.]

Sequencing the entire genomes of model organisms is fundamentally shifting the way we study gene expression and function. Traditionally, the search for gene function started with a phenotypic mutant and proceeded to gene cloning and functional analysis—"from phenotype to gene". Now, as genome sequencing is revealing the whole genotype of an organism, in principle, a reverse approach can be taken—"from gene to phenotype" (reverse genetics). Rather than studying a single gene's function and expression, with the use of sequence arrays one can view the genome as a whole to study multiple genes' function and expression (functional genomics) (DeRisi et al. 1997; Rowen et al. 1997).

Arabidopsis thaliana is an excellent plant organism for both genome and biological studies because of its small genome size, small physical size, and short life cycle (Meyerowitz 1994; Goodman et al. 1995). In addition, an international effort has been established for sequencing the *A. thaliana* genome (Kaiser 1996). We have been concentrating our efforts on *Arabidopsis* chromosome 2. We constructed a yeast artificial chromosome (YAC) physical map for chromosome 2 (Zachgo et al. 1996) and in a 2-Mb region in the vicinity of 80 cM, a higher resolution physical map composed primarily of bacterial artificial chromosome (BAC) clones (Wang et al. 1997). This 2-Mb region is being sequenced in collaboration with The Institute for Genome Research (TIGR). As part of this effort, we have sequenced several BACs on chromosome 2, and have begun to examine the encoded genes. Different software and database search tools (Xu et al. 1994;

Burge and Karlin 1997) were used for gene prediction. Prior to functional studies, a gel-based RNA profiling method was used to confirm whether these putative genes were indeed expressed. In general, the model that emerged by comparison of all the predictions was in good agreement with the experimental data. We report here the sequence of BAC T07M07, prediction of putative genes, confirmation by gel-based RNA profiling, and identification of a cluster of ABA-regulated genes on this BAC.

RESULTS

Subcloning and Sequencing of BAC DNA

Random fragmentation of high molecular weight (HMW) BAC DNA and size selection were critical for obtaining a complete set of sequencing clones. An initial T07M07 library created from a nebulized DNA sample with a broad range of size fragments (0.5–4.0 kb) did not appear to be random, as sequence assembly from 1269 reads formed deep, but not extended contigs. A second successful library used less DNA (4 vs. 6 µg) and nebulization produced a more narrow range of fragment sizes (primarily 0.5–1.6 kb). From this second library, three contigs were assembled from 2211 reads prior to gap closure by directed sequencing. Bacterial contamination as judged by blast matches was ~1% to 2%, a number lower than that reported for the yeast genomic DNA contamination (~9%) when sequencing libraries were constructed from *Caenorhabditis elegans* YAC DNA (Vaudin et al. 1995).

Structural Features of the Sequenced BAC T07M07

BAC T07M07 contained a 95,713-bp insert (GenBank accession no. AF085279). The generally high A+T content of *A. thaliana* was reflected in this BAC that had a level of 64.9%. Overall, this BAC was devoid of repeti-

Present addresses: ³Cereon Genomics, LLC., Cambridge, Massachusetts 02139 USA; ⁴Marine Biological Laboratory, Woods Hole, Massachusetts 02543 USA; ⁵DuPont Agricultural Products, Newark, Delaware 19711 USA; ⁶Millennium Pharmaceuticals Inc., Cambridge, Massachusetts 02139 USA.
Corresponding author.
E-MAIL Howard.Goodman@molbio.mgh.harvard.edu; **FAX** (617) 726-3535.

tive sequences and did not contain any CpG islands (CG-rich regions) commonly found in human sequences. This was consistent with our fingerprint and Southern hybridization results (M.L. Wang, unpubl.). Working primarily from the Grail predictions, genes were spaced relatively evenly every ~4.0 kb (average 3988 bp). In one exceptional region, between basepair positions 2300 and 11,400, there were no predicted genes and no significant blast sequence similarities. Most genes (17 of 24 predicted genes) contained introns, with an average intron size of 141 nucleotides (range 24–723). Of the genes with introns, the average number of introns in a gene was 4 (range 1–9). The average exon size for the 7 predicted genes without introns was 1091 bp (range 528–1794), whereas for the 17 predicted genes with introns, it was 217 bp (range 35–1352). For genes with introns, the average size of the gene was 1759 bp (range 376–3505), and the average predicted amino acid coding region was 1149 bp (range 209–2397). The genes were distributed evenly on the upper and lower strands.

Identifying Gene Function by Sequence Similarity

A summary of the results of blast analysis of the BAC sequence is shown in Table 1. These putative genes can

be classified into five groups: (1) identical to known *Arabidopsis* genes [2 genes, gene 5 identical to *AtEm6* and gene 10 identical to *ABI4* (abscisic acid insensitive 4)]; (2) similar to an *Arabidopsis* EST or an *Arabidopsis* predicted protein (11 genes, 3, 4, 9, 12, 13, 14, 16, 19, 22, 23, and 24); (3) similar to other plant sequences (6 genes, 2, 6, 8, 11, 15, and 18); (4) similar to nonplant sequences but with no significant match to plant sequences (4 genes, 1, 7, 17, and 21); (5) not similar to any database sequence (1 gene, 20).

Of the 24 predicted genes, 15 genes had matches with a level of significance ($P < 1.0 \times 10^{-20}$) that suggested that they might have a function related to their matched sequences. We would not normally consider a BLASTP P -value in the range of 0.001–0.0001 as significant. However, it is worth noting that the best two matches to gene 21 (mouse cation-dependent mannose-6-phosphate receptor and a *Schizosaccharomyces pombe* hypothetical protein) were to the same region of the gene, and many of the residues that were conserved between sequences were conserved in all three proteins. The match region included part of the mouse receptor's transmembrane domain and cytoplasmic tail which (in the bovine homolog) is important to proper trafficking of the protein (Rohrer et al. 1995). It

Table 1. Predicted Genes Encoded by BAC T07M07

Predicted gene	BLASTP match	Organism	GenBank accession no.	P value exponent	EST match	P value exponent
1	Hypothetical Ser-Thr protein kinase	<i>S. pombe</i>	Q096990	–57	H36565	–55
2	ATP-dependent protease ATP-binding subunit CLPL	<i>Lactococcus lactis</i>	Q06716	–9		
3	Hypothetical Cys3His zinc finger protein, BAC6D20	<i>A. thaliana</i>	1871192	–123	T43596	–134
4	Hypothetical protein, BAC F7F1	<i>A. thaliana</i>	3201617	–26	H37615	–115
5	EM6	<i>A. thaliana</i>	Q02973	identity	N37884	–115
6	Protein phosphatase 2 C, MSMP2C	<i>Medicago sativa</i>	Y11607	–100		
7	Hypothetical transmembrane protein	<i>Caenorhabditis</i>	P53993	–83	T04743	–134
8	Maize Intensifier 1	<i>Zea mays</i>	1420924	–8		
9	Hypothetical protein, BAC F21E10	<i>A. thaliana</i>	3047075	–17		
10	ABI4, AP2 domain family transcription factor hom.	<i>A. thaliana</i>	3282693	identity		
11	Hypersensitivity related gene	<i>Nicotiana tabacum</i>	1171577	–33	T14908	–120
12	Hypothetical salt-inducible protein	<i>A. thaliana</i>	2827663	–18		
13	Hypothetical APG protein, BAC T9D9	<i>A. thaliana</i>	2347208	–91		
14	Hypothetical cytoskeletal protein, BAC F14M14	<i>A. thaliana</i>	3335377	–25		
15	Ser-Thr protein kinase NPK15	<i>N. tabacum</i>	S52578	–23	T45416	–54
16	Hypothetical Ankyrin-like protein, BAC F13P17	<i>A. thaliana</i>	3337361	–203	T22001	–113
17	EIF-2-Alpha	<i>S. pombe</i>	S56286	–77	T45955	–129
18	Ferritin subunit cowpea2 precursor	<i>Vigna unguiculata</i>	2970654	–93	T20461	–105
19	Hypothetical polygalacturonase, BAC T9J22	<i>A. thaliana</i>	2739388	–212		
20	No Homology					
21	Cation-Dep Mannose-6-P receptor	<i>Mus musculus</i>	P24668	–3		
22	Hypothetical protein, BAC F7F1	<i>A. thaliana</i>	3201617	–38		
23	Hypothetical protein, BAC T6K21	<i>A. thaliana</i>	2894596	–41	Z34206	–101
24	ABI4, AP2 domain family transcription factor hom.	<i>A. thaliana</i>	3282693	–16	H77041	–32

is interesting that two small gene families seem to be present on the BAC. The first, genes, 4 and 22, produced significant matches to the same set of hypothetical proteins. The second, represented by *ABI4* and gene 24, both contain the AP2 DNA-binding motif. Finally, one predicted gene had no significant match to any known gene. Some putative genes (3, 5, and 15) are probably highly expressed, because several ESTs were identified (data not shown).

ABI4 and *AtEm6* are both regulated by ABA. *AtEm6* (gene 5) is ABA-inducible and highly expressed in immature and dry seeds (Finkelstein 1993; Gaubier et al. 1993). Interestingly, database searching identified another potential ABA-related gene, 6, that has homology to *ABI1* and *ABI2* and is tentatively designated as *PP2C5*. Both *ABI1* and *ABI2* are protein phosphatase 2C (PP2C) proteins and have a high degree of sequence similarity to one another. Mutants in *ABI1* and *ABI2* have identical amino acid substitutions (Gly → Asp) in the same relative positions, residues 168 and 180, respectively (Leung et al. 1997). The amino acids in the *ABI1* and *ABI2* mutated region, as well as in the other metal and the phosphate-binding sites, are also conserved in *PP2C5* (Fig. 5, below). The closest BLAST match (GenBank release 107) to the predicted *PP2C5* amino acid sequence was to *MSMP2C*, a PP2C from alfalfa shown to be active in the stress-activated MAPK pathway (Meskiene et al. 1998). The other significant *Arabidopsis* matches to *PP2C5* identified through database searching included the recently described AtP2C-HA (Rodriguez 1998; GenBank accession no. 3242077) and *AthPP2Ca*, a PP2C gene identified through yeast complementation studies (Kuromori and Yamamoto 1994; GenBank accession no. D38109).

Confirmation of Putative Genes by Reverse Northern Analysis

The positions of the putative genes relative to a minimal set of plasmid sequencing subclones of BAC T07M07 are schematically shown in Figure 1. Most of the putative genes spanned two to four plasmid clones, and individual plasmid clones represented only two putative genes (5 and 20). DNA fragments amplified from the plasmid clones by PCR using primers designed from the pBluescript II KS+ vector sequences

were checked on a 1% agarose gel (Fig. 2). As shown in Figure 2, PCR was highly specific. A single major band was amplified from most plasmid clones (a minor second band was observed for only three clones that may have come from amplification of contaminating DNA templates). After gel electrophoresis, the amplified DNA was transferred to a nylon membrane and hybridized with a single-stranded cDNA probe synthesized from whole *Arabidopsis* seedling poly(A) RNA. The expression signals seen in the reverse Northern analysis (gel-based RNA profiling, Fig. 3) of nine putative genes (1, 5, 6, 11, 17, 18, 20, 21, and 23) were as predicted. For 15 of the predicted genes, the expression signals were in the expected region but either extended into slightly different plasmid clones or showed uneven signals across the plasmid clones. These differences may be explained, at least in part, by the presence in the probe of the 5'- and 3'-untranslated regions of the corresponding mRNA that could hybridize and extend the region of hybridization beyond that predicated by the computer programs and/or the existence of nonhybridizing nonpredicted introns within the genomic fragments. However, in addition to the expected regions of expression, there were expression signals in some regions of the BAC (N1–N5) that were not predicted to encode genes on the basis of the gene finding software. For example, there was no putative gene prediction in the region between 2300–11,400 bp, but there were clear hybridization signals (N1) in this region.

Because there were two characterized ABA-regulated genes identified on this BAC, we examined the entire BAC for genes whose expression was regulated by ABA. Using probes developed from RNA isolated from ABA treated or water (control) treated plants, we found by reverse Northern analysis that gene 3, a zinc finger protein, was also ABA responsive (data not shown, but see Northern analysis below). We have thus identified a small cluster of three ABA-regulated genes, located within 25 kb on BAC T07M07.

Confirmation of Putative Genes by Northern Analysis

To verify the results of reverse Northern analysis, three ABA-regulated genes (3, 5, and 10) and one unregu-

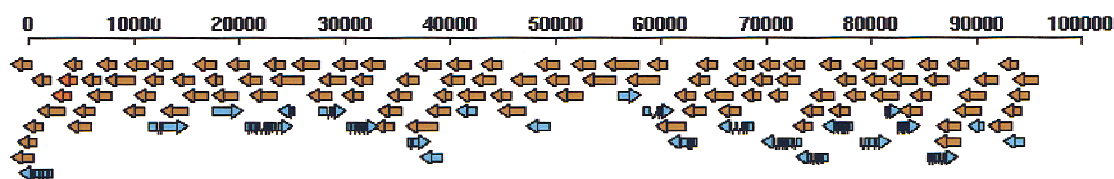


Figure 1 Relative positions of putative genes and plasmid clones sequenced on BAC T07M07. (Brown arrows) Plasmid clones; (blue arrows) predicted genes with arrowheads indicating the direction of transcription. Bent lines between the blue exons indicate predicted introns. The scale is in base pairs.

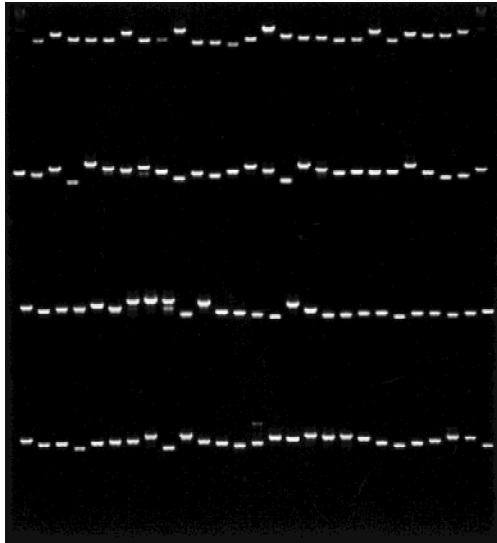


Figure 2 Amplified DNA fragments and putative genes on BAC T07M07. A minimal set of 104 sequencing clone inserts were amplified with primers designed from the pBluescript II KS+ vector sequences. The DNA samples were electrophoresed for 30 min at 60 V in a 1% agarose gel and stained with ethidium bromide. A 1-kb ladder is in each corner lane of the first row. The DNA samples for each set were loaded on the gel in the order corresponding to their positions on the BAC (Fig. 1).

lated gene (*15*, as a control) were chosen for standard Northern analysis (Fig. 4). Transcripts (2.0, 0.7, 1.8, and 2.5 kb) were detected for genes *3*, *5*, *10*, and *15*, respectively, confirming the results of the reverse Northern analysis. Relative to the control water treatment, gene *3* (zinc finger protein) expression was repressed (approximately threefold) by ABA at the seedling stage (Fig. 4, lane 2) but not at the flowering or silique stages (Fig. 4, lanes 4 and 6, respectively). Gene *5* (*AtEm6*, late embryogenesis abundant protein) expression was induced (~3- to 10-fold) by ABA at the seedling and flowering stages; however, at the silique stage, it is highly expressed in both control and ABA-treated plants. Gene *10* (*ABI4*, regulatory protein) expression was induced (approximately two- to threefold) by ABA at the flowering and silique stages but not at the seedling stage, at which there was a high level of expression in both the control and treated plants. Gene *15* was not responsive to ABA treatment (only seedling stage data shown).

Functional Analysis of *PP2C5* in Vitro

The coding region of *PP2C5* (Fig. 5) was amplified by PCR from a cDNA library with primers designed from the predicted gene sequence. The amplified DNA fragment was cloned and sequenced. Comparison of the cDNA to the genomic sequence exactly confirmed the two predicted splicing sites (data not shown). Recombinant protein encoded by *PP2C5* was expressed in *E. coli* and purified from cell culture by affinity chroma-

tography. *PP2C5* encodes an ~35-kD protein (Fig. 6A). Because the enzymatic activity of a PP2C protein is phosphate cleavage, the recombinant protein was assayed by measuring its ability to release ^{32}P into solution from labeled casein. Column-purified labeled casein used as the substrate contained only 1% unincorporated ^{32}P in solution (data not shown). Within 140 min of reaction initiation, a plateau was reached and the enzyme had dephosphorylated close to 30% of the labeled casein when 10 mM Mg^{2+} and 10 mM Mn^{2+} were provided in the reaction (Fig. 6B). The enzyme activity was abolished when 20 mM EDTA was added to the solution (Fig. 6B). These results were reproducible in two different experiments and consistent with the activity of other known PP2C enzymes (Stone et al. 1994; Bertauche et al. 1996; Leube et al. 1998; McGowan et al. 1988).

DISCUSSION

Although gene prediction programs continue to improve, it will always be necessary to confirm the existence of the putative genes prior to gene functional studies. For BAC T07M07, the software prediction programs had a stringent bias, all of the predicted genes were confirmed (no false predictions), but several genes were missed. In a strict sense, expression has only been confirmed from the cloned regions, but not the exact exon sequences of the predicted genes. Where genes span an entire clone, there is of course, no concern that the hybridization signal might arise from a neighboring gene. However, we cannot tell from the reverse Northern blots which region of the amplified clone produced the positive hybridization signal. A traditional method to confirm the existence of the computer-predicted genes is to amplify them by PCR, clone, and sequence the clones to compare with genomic sequences. If the putative genes have introns, after cloning and sequencing, the splicing sites are accurately determined (*PP2C5* as an example); if the putative genes have no introns, they need to be amplified from total mRNA by RT-PCR (e.g., 3' rapid amplification of cDNA ends) and then cloned. After sequencing, the presence of a poly(A) tail on the cDNA clone confirms gene expression and eliminates the possibility of genomic DNA contamination as a possible template source. This approach can only confirm gene existence one by one, with a low efficiency, but with high fidelity. The development of DNA-expression chip technology provides a powerful new tool for confirmation of the existence of putative genes. The method we describe here, gel-based RNA profiling (reverse Northern blot) is an alternative to DNA microarray technology. The only difference is spotting expressed sequence tags (ESTs) on glass slides versus transferring amplified genomic DNA fragments from plasmid clones onto ny-

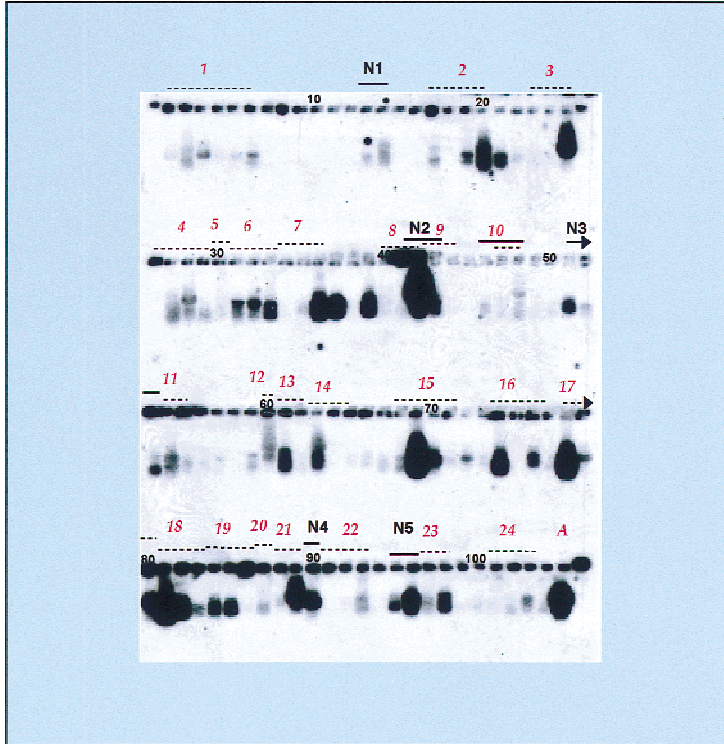


Figure 3 Gene expression profile of BAC T07M07 as determined by reverse Northern (gel-based RNA profiling). Following electrophoresis, DNA samples (see Fig. 2) were transferred onto a nylon membrane and the blot was hybridized to a single-stranded cDNA probe synthesized from whole *Arabidopsis* seedling poly(A) RNA. Hybridization signals indicate the presence of a transcript from the cloned region or gene region (with the caveat that transcripts from genes of similar sequence may also cross-hybridize). Dashed lines indicate the positions of predicted genes numbered above the line. Arrows indicate that the predicted gene extends to the next row. Small numbers mark the DNA sample in every tenth lane. New (N) indicates expression from regions not predicted to encode a gene. A Actin gene (A) used as an internal control for gene expression.

lon membranes. One limitation of the technique relative to DNA microarray technology is lower spotting density, resulting in low rates of data accumulation. Another more fundamental limitation with this technique, as well as microarray technology, is that gene identification relies on mRNA isolated from a limited set of treatment time points and physiological growth stages. mRNAs that are expressed in a small number of cells, for short time periods or at low levels, may be missed in the assay. It is therefore possible that the five genes that were identified on the hybridization but missed by the gene prediction software actually underestimate the number of missed genes. This is a tempting speculation given the large gap between the predicted genes 1 and 2. Despite the limitations, our results demonstrate that gel-based RNA profiling is an efficient method for gene confirmation and is especially useful and practical for conducting experiments on a small scale in laboratories without DNA microarray facilities.

We have made profitable use of the technique to examine changes in mRNA levels following Abscisic acid (ABA) treatment. ABA, a plant phytohormone, regulates many aspects of plant growth and development (such as seed development, dormancy, and germination) and adaptation to environmental stimuli (such as cold, wounding, and water or salt stress). Regulation of ABA-related gene expression in crop plants has the potential to be of agricultural importance. Interestingly, there is a cluster of ABA-regulated genes on this BAC. *AtEm6* (gene 5) encodes one member of the late embryogenesis-abundant protein (*lea*) family that is ABA inducible and highly expressed in immature and dry seeds (for review, see Skriver and Mundy 1990). Its expression level is also regulated by some *ABI* genes (especially *ABI3*, *ABI4*, and *ABI5*) (Finkelstein 1993, 1994; Parcy et al. 1994). While positional cloning *ABI4* (Finkelstein et al. 1998), we found that both *ABI4* and *AtEm6* were located on the same BAC, 17 kb apart from each other. *ABI4* (gene 10), a new member of the AP2 domain family, encodes a regulatory protein. Intriguingly, we identified another ABA-related gene in the vicinity of the locus *AtEm6*. Gene 3 (Cys-3-His zinc finger protein) is 500 bp downstream from *AtEm6*. Its expression is repressed by ABA at the seedling stage. Gene 6, *PP2C5*, which has high sequence similarity to *ABI1* and *ABI2* in the metal and phosphate-binding domains, is also located within the ABA-related cluster (2 kb upstream of *AtEm6*). However, unlike *ABI1* and *ABI2* that are up-regulated by ABA treatment, we did not detect a significant *PP2C5* response to ABA in the Northern analysis experiments (data not shown). Plant tissue samples for mRNA isolation were collected 24 hr after treatment with ABA; therefore, the gene expression changes identified are probably an indirect response (secondary message) to ABA application. Because the reverse Northern profiling relied on mRNA samples from whole plants, and examined only a limited portion of the life cycle, it is an open question as to whether there are other ABA-regulated genes around the *AtEm6* locus. It is an interesting, but as yet unanswered question, whether the spatial proximity of these genes and other genes reflects a functional interaction. One possible approach to answer this question would be to introduce a genomic fragment encompassing all of these genes into transgenic plants. Phenotypic changes might be observed, and changes in other gene expression patterns might be observed on a cDNA microarray (Schena et al. 1995; Lashkari et al. 1997). Such an analysis might help discern the transgene

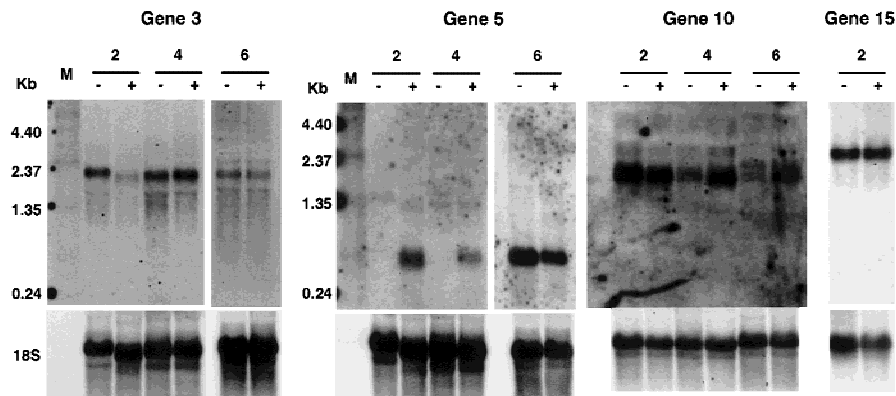


Figure 4 Level of RNA regulated by ABA treatment in selected putative genes. Total RNAs were isolated from tissues of 2-, 4-, and 6-week-old plants (seedling, flowering, and silique stages, respectively) and probed with gene-specific DNAs. (Lane M) RNA marker; (-) water-treated plants; (+) ABA-treated plants. The main hybridization bands corresponded to the transcripts predicted for each putative gene. The 18S rDNA hybridization band serves as a sample loading control.

function(s) in vivo. The binary bacterial artificial chromosome (BIBAC) system that has been developed recently for the delivery of HMW DNA into plants (Hamilton et al. 1996) should be useful in this regard.

The reversible phosphorylation of proteins, catalyzed by protein kinases and phosphatases, is a major mechanism for the regulation of many cellular functions from metabolism to signal transduction. PP2C is a Mg^{2+} -dependent enzyme (Ingebritsen and Cohen 1983). There are at least five genes for protein phosphatase 2Cs experimentally characterized in *Arabidopsis* and several more putative PP2C proteins identified on sequenced *Arabidopsis* BACs. Although these five gene paralogs for protein phosphatases 2C are all expressed in *Arabidopsis*, they may not serve a single function, spacial and temporal expression patterns have yet to be completely determined and may be expected to vary between the genes. As shown in Figure 5, the regions of similarity among characterized PP2C proteins is primarily in the carboxy-terminal end and this also appears to be true of the putative PP2C proteins from sequenced BACs (F17I5, F28M20, F21O9, T01B08, T20B5, T29F13; data not shown). It is inviting to speculate that amino-terminal domains will be shown to determine substrate specificity. *ABI1*, *ABI2*, *ATP2CHA*, and *ATHPP2C* are actually more closely related to each other than alfalfa *MsMP2c* and *PP2C5*, suggesting a divergence of functional partners. The three PP2C proteins *ABI1*, *ABI2*, and *AthPP2Ca* have been studied in vivo. In our case, the enzyme activity of *PP2C5* was only tested in vitro. Recently, it has been demonstrated that in plants, *ABI1*, *ATHPP2C*, and *MsMP2c* function as negative regulators in ABA or stress-activated MAPK pathways (Meskiene et al. 1998; Sheen 1998). A single cell system (Sheen 1996) is a rapid way to characterize in vivo activity of *PP2C5*

genes. Expression of *PP2C5*, perhaps in combination with known kinases or phosphatases in vivo, may provide additional insights into the ABA signal-transduction pathway.

METHODS

DNA Nebulization and Library Construction

BAC T07M07 from the TAMU *Arabidopsis* library (Choi et al. 1995) was isolated from a 500-ml culture by a modified alkaline lysis method. BAC DNA (4–6 μ g) was resuspended in 2 ml of 50 mM Tris (pH 8.0), 15 mM $MgCl_2$, 25% glycerol and transferred into a prepared nebulizer (IPI Medical Products). DNA was nebulized for 150 sec at 30 psi at the tank outlet.

The fragmented DNA sample (1800 μ l) was collected and concentrated to 400 μ l by five to six butanol extractions, precipitated with 2.5 volumes of ethanol (2 hr at $-20^\circ C$) and pelleted by microcentrifugation at $4^\circ C$ for 15 min. The DNA pellet was washed with 70% cold ethanol, dried at room temperature and dissolved in ~ 100 μ l of $1 \times TE$. Ends of the fragmented DNA (~ 1.5 μ g) were filled in by *Pfu* DNA polymerase (Stratagene). The reaction solution was loaded onto a 1%, 8-cm-long low melting point (LMP) agarose (Kodak) gel in $1 \times TBE$ and subjected to electrophoresis. Gel slices containing DNA fragments ranging from 500 bp to 3.0 kb were excised. The gel was melted at $65^\circ C$ for 10 min and the agarose digested with β -agarase I (New England Biolabs) at $40^\circ C$ for 2 hr. The digested agarose solution was chilled on ice for 15 min and microcentrifuged at $4^\circ C$ for 15 min to pellet any undigested agarose. The supernatant was extracted with an equal volume of phenol/chloroform, then chloroform, and the DNA was precipitated with two volumes of isopropanol. Following centrifugation, the DNA pellet was washed with 70% cold ethanol, dried at room temperature for 15 min, dissolved in 20 μ l of $1 \times TE$, and stored at $4^\circ C$ for later ligation. Bluescript II KS+ vector (Stratagene) DNA was digested with *EcoRV* at $37^\circ C$ for 3 hr and dephosphorylated with calf intestinal alkaline phosphatase (CIP) (NEB). The dephosphorylated vector DNA (~ 100 ng) was mixed with the size-selected BAC DNA (~ 100 ng) and ligated with T4 DNA ligase (NEB) at $16^\circ C$ overnight. After ligation, the ligase was heat inhibited at $65^\circ C$ for 10 min. DNA in the ligation solution was electrotransformed into *E. coli* DH5 α -competent cells by use of a Bio-Rad gene pulser. Transformants were recovered in 1 ml of SOC (2% Bacto-tryptone, 0.5% Bacto-yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 10 mM glucose) at $37^\circ C$ with shaking at 300 rpm for 1 hr. Transformants in the recovery solution were mixed with an equal volume of 30% glycerol, divided into aliquots (100 μ l), and stored at $-80^\circ C$ for later use.

DNA Sequencing

The frozen transformant stock was thawed briefly and diluted 50 times with LB liquid medium and then plated onto selective LB agar plates containing 50 μ g/ml ampicillin. Recombi-



Figure 5 Peptide sequence comparison of PP2C genes. Multiple sequence alignment of PP2C5 to *Medicago sativa* MSMP2C (Y11607) and *A. thaliana* PP2C proteins; ABI1 (GenBank accession no. X77116), ABI2 (Y08965), AtPP2C-HA (ATAJ3119), and ATHPP2Ca (D38109). Note the high level of conservation at the putative phosphate-binding site (○), putative metal binding sites (●) and the position of the ABI1 and ABI2 mutation (□). Sites were determined from homology to previously characterized PP2C proteins (Leung et al. 1994, 1997; Meyer et al. 1994; Das et al. 1996). Shading, on the basis of the PP2C5 sequence, shows conserved residues with a black background and similar residues with a gray background.

nants were identified with a blue-white (X-gal and IPTG indicator) screening system. White colonies were inoculated into 96-well titer plates (each well containing 1.3 ml of liquid LB medium) and cultured at 37°C with shaking at 300 rpm for 22 hr. Plasmid DNA was isolated by rapid extraction alkaline plasmid kits (QIAGEN) and dissolved in 50 µl of double-distilled water for later use as DNA-sequencing templates. DNA was sequenced on an ABIprism 377 DNA sequencer by dye terminator chemistry as described by the manufacturer (Perkin Elmer). The Phred, Phrap, and Consed sequence assembly and viewing programs (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998) were used to remove BAC vector sequences and to assemble contigs. After large sequence contigs were assembled and only two gaps remained, the gaps were closed by amplifying the intervening DNA by PCR from the original BAC DNA with primers on the basis of adjacent sequences.

Gene Prediction

Putative genes were identified by a combination of three gene-prediction programs. Grail (Xu et al. 1994) was run with client server software and selecting for *Arabidopsis* as the DNA source. Gene assembly by the Grail program requires the investigator to define a region of the DNA sequence to assemble. This was done somewhat subjectively on the basis of the clustering of the predicted exons. Genscan (Burge and Karlin 1997) was run with the *Arabidopsis* organism option and an exon cutoff of 1.0. Genefinder (P. Green, pers. comm.) was

run with the *Arabidopsis* tables developed by ESSA (European Scientists Sequencing *Arabidopsis*). As partial confirmation of the predicted genes, the BAC sequence was compared with sequences deposited in GenBank by use of blastn and blastx, and translations of the predicted genes were compared by use of blastp (Altschul et al. 1990; Gish et al. 1993).

Amplification from the BAC Sequences; Northern and Reverse Northern Analysis

Minimum-overlapping plasmid clones were selected and insert DNAs were amplified by use of primers designed from pBluescript II KS+ vector sequences. The DNA concentration in PCR was quantified, 1 µg of DNA was electrophoresed on a 1% agarose gel for 30 min, and DNA fragments were transferred onto a nylon membrane (Biodyne, Pal Biosupport, East Hills, NY; 0.45 µm) by alkaline transfer. These blots were used for reverse Northern analysis. Tissue was collected from 2-, 4-, and 6-week-old glass-house-grown whole plants 24 hr following spraying with water (control) or ABA (100 µM), immediately frozen in liquid nitrogen, and stored at -80°C. Stage-specific, total or poly(A) RNA were isolated with Trizol (GIBCO-BRL) and the Oligotex mRNA mini kit (Qiagen). Total RNA for each sample (10 µg) was separated on a 1% agarose gel and transferred onto a nylon membrane. For reverse Northern analysis, a labeled single-stranded cDNA probe was synthesized from whole *Arabidopsis* seedling poly(A) RNA and used as probes on Southern blots containing DNA fragments covering the whole BAC T07M07. After hybridization, the

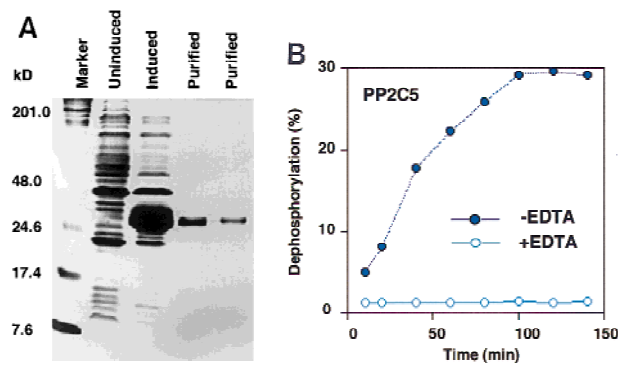


Figure 6 Purification of recombinant protein and enzyme activity assay. (A) Gene expression and protein purification of PP2C5. *E. coli* containing the recombinant expression plasmid was grown for 3 hr either under inducing or noninducing conditions. Following protein isolation, samples were analyzed on a 12% SDS-polyacrylamide gel. From left to right, the samples were as follows: protein marker (prestained SDS-PAGE standards, Bio-Rad, 25 μ g in 20 μ l); PP2C5-noninduced total protein (40 μ g in 40 μ l); PP2C5-induced total protein (40 μ g in 40 μ l); purified PP2C5 protein (5 μ g in 5 μ l); and purified PP2C5 protein (2.5 μ g in 2.5 μ l). (B) Enzymatic activity assay for PP2C. One microgram of purified PP2C5 was used in each reaction. Each experiment was repeated twice. Each value represents the average of the two different experiments. Labeled protein was dephosphorylated at 30°C in a 10- μ l solution [containing 50 mM Tris-HCl at pH 7.0, 0.1% (vol/vol) 2-mercaptoethanol, $\sim 3 \times 10^5$ cpm 32 P-labeled casein, ~ 1 μ g of recombinant protein, and either 10 mM Mg^{2+} and 10 mM Mn^{2+} or the same concentration of cations plus 10 mM EDTA]. At each time point (min), 100 μ l of 20% trichloroacetic acid (TCA) was added, incubated on ice for 5 min, and microcentrifuged at room temperature for 5 min to precipitate the protein. Radioactivity in 90 μ l of supernatant and 20 μ l of pellet was measured by scintillation counting. Dephosphorylation was expressed as a percentage of 32 P released into the supernatant. (●) -EDTA; (○) +EDTA.

blots were washed four times at 65°C for 15 min with 40 mM Na_2HPO_4 (pH 7.2), 1% SDS, and 1 mM EDTA (pH 8.0). For Northern analysis, DNA fragments encompassing putative genes were amplified from BAC T07M07, labeled by [α - 32 P]dATP and used as probes for hybridization. Primers were designed on the basis of the sequences of putative genes from start codons to stop codons and synthesized in the Department of Molecular Biology core facility. The blots were washed with 1 \times SSC, 0.1% SDS, at 65°C for 15 min twice; 0.1 \times SSC, 0.1% SDS, at 65°C for 15 min twice, and then exposed to X-ray film.

Amplification and Cloning of PCR Fragments from a cDNA Library

DNAs prepared from a cDNA library (Minet et al. 1992) were used as DNA templates for PCR amplification of putative genes. High-fidelity enzyme mix (Boehringer Mannheim) and *Taq* extender (Stratagene) were used in the PCR. PCR-amplified fragments corresponding to the ABA regulated genes on T07M07 were extracted with phenol/chloroform/isoamylalcohol (25:24:1), then chloroform/isoamylalcohol (24:1), precipitated with ethanol, dried at room temperature, dissolved in 1 \times TE and cloned into the pET-19b expression vector (Novagen) that encodes 10 histidine residues at the amino terminus of the expressed protein. PCR fragments (~ 1000 ng) and pET-19b vector (500 ng) DNA were digested

with *Xho*I and dephosphorylated with CIP. The digested PCR fragment DNA was electrophoresed on a 1% LMP agarose gel. The gel slice containing the DNA band was excised under long UV light and melted at 65°C. Digested vector DNA and PCR-amplified DNA were ligated at 16°C for 16 hr and then electroporated into *E. coli* DH5 α -competent cells. Recombinants were identified by outgrowth on LB agar plates containing ampicillin (50 μ g/ml). Four clones from each putative gene were sequenced and aligned with genomic sequences. Plasmid DNA from a sense clone without mutations and pET-19b vector DNA (as a control) were transformed into *E. coli* BL21pLysS-competent cells (Novagen) (expression host) by heat-shock for later gene expression.

Expression and Purification of Recombinant Protein

A single colony was picked from LB agar plates containing 50 μ g/ml ampicillin and 34 μ g/ml chloramphenicol and inoculated into 5 ml of LB solution with the same antibiotics and cultured at 37°C with shaking at 300 rpm for 16 hr. Five milliliters of cultured cell solution was inoculated into 500 ml of LB solution with the same antibiotics. When the OD₆₀₀ reached 0.6–0.7, isopropyl β -D-thiogalactoside (IPTG, 1 mM) was added to the culture for induction of expression from the T7 promoter. The solution was cultured for another 3 hr to allow accumulation of recombinant protein. Cells were harvested by centrifugation and suspended in 5 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl (pH 7.9), and lysed by sonication at the following settings (time, hold; duty cycle, 80; output control, 9) on a Sonifier 250 (Branson) at 4°C for 1 min. Protein was purified by nickel affinity chromatography Novagen kits (His-bind resin; His-bind buffer) and dialyzed at 4°C overnight. Protein was concentrated by use of centricon concentrators (Amicon) and its concentration determined by a Bio-Rad protein assay.

Casein Labeling and Protein Enzyme Activity Assay

J.M. Stone (pers. comm.) provided the casein-labeling method. Before labeling, 5 \times incubation buffer (250 mM Tris at pH 7.4, 0.5 mM EGTA, 50% glycerol, 0.5% β -mercaptoethanol) and stop buffer (100 mM EDTA, 500 mM NaF) were prepared, and 250 units of the catalytic subunit of protein kinase from bovine heart (Sigma) were dissolved in 25 μ l of 40 mM dithiothreitol (DTT). Partially dephosphorylated casein from bovine milk (Sigma) was labeled by [γ - 32 P]ATP (NEN/DuPont) in a total volume of 500 μ l (containing 1 \times incubation buffer, 10 mM Mg-acetate, 1 U/ μ l protein kinase, 0.4 μ Ci/ μ l [32 P]ATP, 11.2 μ g/ μ l casein) at room temperature for 16 hr. The reaction was terminated by adding 55 μ l of stop buffer and left at room temperature for 10 min. Denatured protein was removed by brief centrifugation. To remove unincorporated [32 P]ATP, the reaction solution was chromatographed on a 10-ml size Sephadex G-50 column pre-equilibrated with 1 \times incubation buffer. The labeled protein solution was loaded onto the column, 32 fractions (500 μ l each) were eluted by addition of 1 \times incubation buffer, collected into 1.5-ml Eppendorf tubes and stored at 4°C. Incorporated label in each fraction was determined by scintillation counting and the initial highly radioactive fraction (4×10^5 cpm/ μ l) was used as a substrate for the protein assay. Phosphatase activity was determined as described (Stone et al. 1994).

ACKNOWLEDGMENTS

We thank Drs. Jen Sheen and Gabor Lazar for suggestions and

stimulating discussions regarding the experiments, Dr. Julie M. Stone for instructions for the casein labeling and the recombinant protein assay, Dr. Hawk-Bin Kwon for suggestions in RNA isolation, and Dr. Julie Nardone for critical reading of the manuscript. We acknowledge Dr. Michèle Minet for providing the *Arabidopsis* pFL61 cDNA library. This work was supported by a grant from Hoechst to H.M.G.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Altschul, S.F., W. Gish., W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Bertauche, N., J. Leung, and J. Giraudat. 1996. Protein phosphatase activity of abscisic acid insensitive 1 (ABI1) protein from *Arabidopsis thaliana*. *Eur. J. Biochem.* **241**: 193–200.
- Burge, C. and S. Karlin. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**: 78–94.
- Choi, S., R.A. Creelman, J.E. Mullet, and R.A. Wing. 1995. Construction and characterization of a bacterial artificial chromosome library of *Arabidopsis thaliana*. *Plant Mol. Biol. Rep.* **13**: 124–128.
- Das, A.K., N.R. Helps, P.T.W. Cohen, and D. Barford. 1996. Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* **15**: 6798–6809.
- DeRisi, J.L., V.R. Iyer, and P.O. Brown. 1997. Exploring the metabolic and genetic control of gene expression on a genome scale. *Science* **278**: 680–686.
- Ewing, B. and P. Green. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8**: 186–194.
- Ewing, B., L. Hillier, M.C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185.
- Finkelstein, R. 1993. Abscisic acid-insensitive mutations provide evidence for stage-specific signal pathways regulating expression of an *Arabidopsis* late embryogenesis-abundant (*lea*) gene. *Mol. & Gen. Genet.* **238**: 401–408.
- . 1994. Mutations at two new *Arabidopsis* ABA response loci are similar to the *abi3* mutations. *Plant J.* **5**: 765–771.
- Finkelstein, R., M.L. Wang, T.J. Lynch, S. Rao, and H.M. Goodman. 1998. The *Arabidopsis* abscisic acid response locus *ABI4* encodes an APETALA2 domain protein. *Plant Cell* **10**: 1043–1054.
- Gaubier, P., M. Raynal, G. Hull, G.M. Huestis, F. Grellet, C. Arenas, M. Pagès, and M. Delseny. 1993. Two different *Em*-like genes are expressed in *Arabidopsis thaliana* seeds during mutation. *Mol. & Gen. Genet.* **238**: 409–418.
- Gish, W. and D.J. States. 1993. Identification of protein coding regions by database similarity search. *Nat. Genet.* **3**: 266–272.
- Goodman, H.M., J.R. Ecker, and C. Dean. 1995. The genome of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **92**: 10831–10835.
- Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. *Genome Res.* **8**: 195–202.
- Hamilton, C.M., A. Frary, C. Lewis, and S.D. Tanksley. 1996. Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc. Natl. Acad. Sci.* **93**: 9975–9979.
- Ingebritsen, T.S. and P. Cohen. 1983. Protein phosphatases: Properties and role in cellular regulation. *Science* **221**: 331–338.
- Kaiser, J. 1996. First global sequencing effort begins. *Science* **274**: 30.
- Kuromori, T. and M. Yamamoto. 1994. Cloning of cDNAs from *Arabidopsis thaliana* that encode putative protein phosphatase 2C and a human Dr1-like protein by transformation of a fission yeast mutant. *Nucleic Acids Res.* **22**: 5296–5301.
- Lashkari, D.A., J.H. McCusker, and R.W. Davis. 1997. Whole genome analysis: Experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. *Proc. Natl. Acad. Sci.* **94**: 8945–8947.
- Leube, M.P., E. Grill, and N. Amrhein. 1998. ABI1 of *Arabidopsis* is a protein serine/threonine phosphatase highly regulated by the proton and magnesium ion concentration. *FEBS Lett.* **424**: 100–104.
- Leung, J., M. Bouvier-Durand, P. Morris, D. Guerrier, F. Chefdor, and J. Giraudat. 1994. *Arabidopsis* ABA response gene *ABI1*: Features of a calcium-modulated protein phosphatase. *Science* **264**: 1448–1452.
- Leung, J., S. Merlot, and J. Giraudat. 1997. The *Arabidopsis* *ABSCISIC ACID-INSENSITIVE2 (ABI2)* and *ABI1* gene encode homologous protein phosphatases 2C involved in abscisic acid signal transduction. *Plant Cell* **9**: 759–771.
- McGowan, C.H. and P. Cohen. 1988. Protein phosphatase-2C from rabbit skeletal muscle and liver: An Mg²⁺-dependent enzyme. *Methods Enzymol.* **159**: 416–427.
- Meskiene, I., L. Bogre, W. Glaser, J. Balog, M. Brandstotter, K. Zwerger, G. Ammerer, and H. Hirt. 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.* **95**: 1938–1943.
- Meyer, K., M.P. Leube, and E. Grill. 1994. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* **264**: 1452–1455.
- Meyerowitz, E.M. 1994. *Arabidopsis: Structure and organization of the Arabidopsis thaliana nuclear genome* (ed. E.M. Meyerowitz and C.R. Somerville), pp. 21–36. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Minet, M., M.E. Dufour, and F. Lacroute. 1992. Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. *Plant J.* **2**: 417–422.
- Parcy, F., C. Valon, M. Raynal, P. Gaubier-Comella, M. Delseny, and J. Giraudat. 1994. Regulation of gene expression programs during *Arabidopsis* seed development: Roles of the *ABI3* locus and of endogenous abscisic acid. *Plant Cell* **6**: 1567–1582.
- Rodriguez, P.L., M.P. Leube, and E. Grill. 1998. Molecular cloning in *Arabidopsis thaliana* of a new protein phosphatase 2C (PP2C) with homology to *ABI1* and *ABI2*. *Plant Mol. Biol.* **38**: 879–883.
- Rohrer, J., A. Schweizer, K.F. Johnson, and S. Kornfeld. 1995. A determinant in the cytoplasmic tail of the cation-dependent mannose 6-phosphate receptor prevents trafficking to lysosomes. *J. Cell. Biol.* **130**: 1297–1306.
- Rowen, L., G. Mahairas, and L. Hood. 1997. Sequencing the human genome. *Science* **278**: 605–607.
- Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Sheen, J. 1996. Ca²⁺-dependent protein kinases and stress signal transduction in plant. *Science* **274**: 1900–1902.
- . 1998. Mutational analysis of protein phosphatase 2C involved in abscisic acid signal transduction in higher plants. *Proc. Natl. Acad. Sci.* **95**: 975–980.
- Skriver, K. and J. Mundy. 1990. Gene expression in response to abscisic acid and osmotic stress. *Plant Cell* **2**: 503–512.
- Stone, J.M., M.A. Collinge, R.D. Smith, M.A. Horn, and J.C. Walker. 1994. Interaction of a protein phosphatase with an *Arabidopsis* serine-threonine receptor kinase. *Science* **266**: 793–795.
- Vaudin, M., A. Roopra, L. Hillier, R. Brinkman, J. Sulston, R.K. Wilson, and R.H. Waterson. 1995. The construction and analysis of M13 libraries prepared from YAC DNA. *Nucleic Acids Res.* **23**: 670–674.
- Wang, M.L., L. Huang, D.K. Bongard-Pierce, S. Belmonte, E.A. Zachgo, J.W. Morris, M. Dolan, and H.M. Goodman. 1997. Construction of an ~2 Mb contig in the region around 80 cM of *Arabidopsis thaliana* chromosome 2. *Plant J.* **12**: 711–730.
- Xu, Y., J.R. Einstein, R.J. Mural, M. Shah, and E.C. Uberbacher. 1994. An improved system for exon recognition and gene modeling in human DNA sequences. *Intell. Systems Mol. Biol.* **2**: 376–384.
- Zachgo, E.A., M.L. Wang, J. Dewdney, D. Bouchez, C. Camilleri, S. Belmonte, L. Huang, M. Dolan, and H.M. Goodman. 1995. A physical map of chromosome 2 of *Arabidopsis thaliana*. *Genome Res.* **6**: 19–25.

Received November 5, 1998; accepted in revised form January 29, 1999.