Research

Characterizing the stress/defense transcriptome of *Arabidopsis*

Ramamurthy Mahalingam*, AnaMaria Gomez-Buitrago†, Nancy Eckardt‡, Nigam Shah§, Angel Guevara-Garcia¶, Philip Day§, Ramesh Raina† and Nina V Fedoroff*

Addresses: *Life Sciences Consortium, †Department of Biology, and §Integrative Biosciences Graduate Degree Program, Pennsylvania State University, State College, PA 16802, USA. ‡American Society of Plant Biologists, Rockville, MD 20855, USA. ¶Instituto de Biotecnología-UNAM, Av. Universidad No. 2001Col. Chamilpa, Cuernavaca, Morelos CP 62271, Mexico.

Correspondence: Nina V Fedoroff. E-mail: nvf1@psu.edu. Ramesh Raina. E-mail: rxr21@psu.edu

Abstract

Background: To understand the gene networks that underlie plant stress and defense responses, it is necessary to identify and characterize the genes that respond both initially and as the physiological response to the stress or pathogen develops. We used PCR-based suppression subtractive hybridization to identify *Arabidopsis* genes that are differentially expressed in response to ozone, bacterial and oomycete pathogens and the signaling molecules salicylic acid (SA) and jasmonic acid.

Results: We identified a total of 1,058 differentially expressed genes from eight stress cDNA libraries. Digital northern analysis revealed that 55% of the stress-inducible genes are rarely transcribed in unstressed plants and 17% of them were not previously represented in *Arabidopsis* expressed sequence tag databases. More than two-thirds of the genes in the stress cDNA collection have not been identified in previous studies as stress/defense response genes. Several stress-responsive *cis*-elements showed a statistically significant over-representation in the promoters of the genes in the stress cDNA collection. These include W- and G-boxes, the SA-inducible element, the abscisic acid response element and the TGA motif.

Conclusions: The stress cDNA collection comprises a broad repertoire of stress-responsive genes encoding proteins that are involved in both the initial and subsequent stages of the physiological response to abiotic stress and pathogens. This set of stress-, pathogen- and hormone-modulated genes is an important resource for understanding the genetic interactions underlying stress signaling and responses and may contribute to the characterization of the stress transcriptome through the construction of standardized specialized arrays.

Background

Plants respond to invasion by pathogens with an array of biochemical and genetic changes, including the production of reactive oxygen species, antimicrobial compounds, antioxidants and signaling molecules such as salicylic acid (SA) and jasmonic acid (JA). They also respond by the localized activation of a cell-death program, designated the hypersensitive response (HR), and by the systemic activation of cellular and molecular defenses, termed systemic acquired resistance (SAR) [1-7]. Second messengers that contribute to

the development of the systemic response include reactive oxygen species (ROS), SA, JA and ethylene [8-10]. There is evidence for commonalities between plant responses to pathogens (referred to as defense responses) and environmental stresses (referred to as stress responses) [11,12]. Despite similarities, however, a plant's response to each environmental challenge is unique and tailored to increasing the plant's ability to survive the inciting stress [13-16].

A comprehensive understanding of the networks of genes, proteins and small molecules that underlie plant stress and defense responses requires identification and characterization of the molecular components, including the genes that respond both initially and as the physiological response to the stress or pathogen develops. Several groups [17,18] have used existing expressed sequence tag (EST) collections to carry out microarray experiments in initial efforts to identify genes whose expression levels change in response to pathogens and various abiotic stresses. However, existing EST collections are not complete and most were derived from cDNA libraries made from plants grown under normal environmental conditions; hence, ESTs representing stress- and pathogen-induced transcripts are likely to be under-represented in them.

To characterize the stress/defense transcriptome of *Arabidopsis* more thoroughly, we sought to identify genes whose expression levels change in response to abiotic stress, known second messengers and bacterial and fungal pathogens. We used the bacterial pathogen *Pseudomonas syringae* and the oomycete pathogen *Peronospora parasitica* to evoke the pathogen defense response. The availability of both avirulent and virulent strains of these pathogens facilitates identification of genes that are important for resistance, as well as those expressed during disease development [19-21]. Moreover, both of these pathogens cause severe damage to crops of economic importance, such as tomato and crucifers [19]. We also treated plants with SA and methyl jasmonate (MJ) because both are important signaling molecules implicated in plant responses to pathogens, herbivory and wounding and they induce different aspects of the SAR [22].

We used ozone as an abiotic stressor because the plant's response at the biochemical and molecular level shows extensive overlap with the pathogen defense response and includes the production of ROS, as well as induction of HR and SAR [23,24]. At the phenotypic level, acute ozone exposure (high dose for a short interval) causes necrotic lesions similar to those caused by avirulent pathogen infections, whereas chronic ozone exposure (low dose for an extended period of time) accelerates foliar senescence, producing similar symptoms to those caused by virulent pathogen infections [25,26]. There is evidence that elicitor-evoked ROS production, called the oxidative burst, is an essential signaling component of the defense response [27]. Moreover ozone is a component of photochemical smog and itself

represents an oxidative stress to living organisms, damaging crops and forests [28].

There are several strategies for identifying differentially expressed transcripts, including differential display (DD), representational difference analysis (RDA), serial analysis of gene expression (SAGE), enzymatic degradation subtraction and subtractive hybridization [29-33]. We chose the PCR-based suppression subtractive hybridization (SSH) procedure for several reasons: it includes a normalization step, it enriches for differentially expressed transcripts, and it yields cDNA fragments that can be used directly for the construction of DNA microarrays. We viewed the normalization step as particularly important because a few stress-activated genes, such as those encoding the pathogenesis-related (PR) proteins, are abundantly induced by a variety of stresses, potentially obscuring important stress-specific transcripts expressed at much lower levels. The SSH procedure developed by Diatchenko *et al.* [33] has the additional advantage that it exploits the suppression PCR effect, eliminating the need for physical separation of single- and double-stranded cDNAs [34]. We have cloned and sequenced cDNA fragments representing 1,058 stress-induced genes from eight different SSH cDNA libraries. We describe and discuss the stress/defense-induced genes we have identified, many of which have either not previously been associated with stress responses or are not represented in existing cDNA libraries.

Results
SSH cDNA library construction

Plants of the *Arabidopsis* ecotype Col-0 show disease symptoms when infiltrated with virulent *Pseudomonas syringae* pv. *tomato* DC3000 or when sprayed with the oomycete *Peronospora parasitica* strain Ahco. The avirulent bacterial strain of *P. syringae* expressing the *avrRpm1* gene elicits a marked HR in Col-0 plants, whereas the *P. parasitica* strain Emwa elicits a microscopic HR. All treatments, including pathogen infection and treatment with ozone, SA and MJ, were carried out on 3-4-week-old plants. All treatments affect foliar tissue, which was the material used for library construction.

To capture a wide spectrum of differentially expressed genes, leaf tissue was collected at different intervals after the treatment and pooled before RNA extraction (Table 1). Leaves were harvested at 1, 3, 6, 9 and 12 hours after acute ozone treatment and at 1, 8 and 24 hours after bacterial infection and after MJ and SA treatments. Because fungal spores take almost a day to germinate and penetrate the host cell, samples were collected at 12 hours, 2 and 3 days in the incompatible oomycete interaction. In the case of the compatible interaction the host fails to recognize the pathogen early and mounts a response only when the pathogen has proliferated extensively (3-5 days). Samples were therefore collected up to day 5 after inoculation. For the chronic ozone treatment, samples were harvested 2, 4 and 6 days after exposure.

Table 1

Treatments used for generating stress cDNA libraries

Age (weeks)	Treatment	Control	Time points	Library/clone designation
4	350 ppb of O_3 for 6 h	Clean air	1, 3, 6, 9, 12, 24 h	Aoz
3	150 ppb of O_3 for 6 h/day for 6 days	Clean air	2, 4 and 6 days	Coz
4	1 mM SA in water	Water	1, 8 and 24 h	SA
4	50 μM of MJ in 0.001% ethanol	0.001% ethanol	1, 8, 24 and 48 h	MJ
4	*Pst* DC3000 5×10^7 CFU/ml	10 mM $MgCl_2$	1, 8 and 24 h	DC
4	*Pst* DC3000 (avrRpm1) 5×10^7 cfu/ml	10 mM $MgCl_2$	1, 8, and 24 h	RPM
3	*P. parasitica* Ahco 2×10^4 spores/ml	Water	1, 3 and 5 days	VPP
3	*P. parasitica* Emwa 2×10^4 spores/ml	Water	12 h, 1, 2 and 3 days	APP

Aoz, acute ozone; APP, avirulent *P. parasitica* infection; Coz, chronic ozone; DC, virulent *P. syringae* infection; MJ, methyl jasmonate; Rpm, avirulent *P. syringae* infection; SA, salicylic acid; VPP, virulent *P. parasitica* infection.

The efficiency of subtraction was evaluated by PCR amplification of a housekeeping gene, that for gyceraldehyde-3-phosphate dehydrogenase (*G3PDH*), and one of several differentially expressed genes. If subtraction is efficient, transcripts of housekeeping genes should be reduced, while those of differentially expressed genes should be substantially enriched in the population of cDNA fragments used for library construction. Figure 1 shows that the *G3PDH* fragment is barely detectable even after 30 cycles of amplification in the subtracted sample, while it is clearly detectable in the unsubtracted sample after 20 cycles. To test enrichment for differentially expressed genes, we amplified the *PR1* gene for the biotic stressors and SA treatment, the plant defensin gene *PDF1.2* for MJ treatment, and the amino-cyclopropane synthase gene (*ACS1*) for the ozone treatments [9,19,35]. The genes tested showed strong amplification in the subtracted samples after 15 cycles of PCR, whereas in the unsubtracted samples the PCR product was seen only after 10 additional cycles (Figure 1). On the basis of the number of PCR cycles required for equal amplification of the corresponding PCR products from the subtracted and unsubtracted cDNA samples, we estimated that the subtracted libraries were 32-64-fold enriched for differentially expressed genes.

One of the main advantages of SSH is that it normalizes the cDNA abundance so that cDNAs encoded by genes that are expressed infrequently, but nonetheless differentially, can be identified readily [33]. The efficiency of normalization is illustrated in Figure 2. The more uniform distribution of hybridization intensities obtained using the subtracted cDNA probe (Figure 2c) reflects the equalization in the concentrations of individual species present at markedly different concentrations in the initial unsubtracted cDNA populations (Figure 2b).

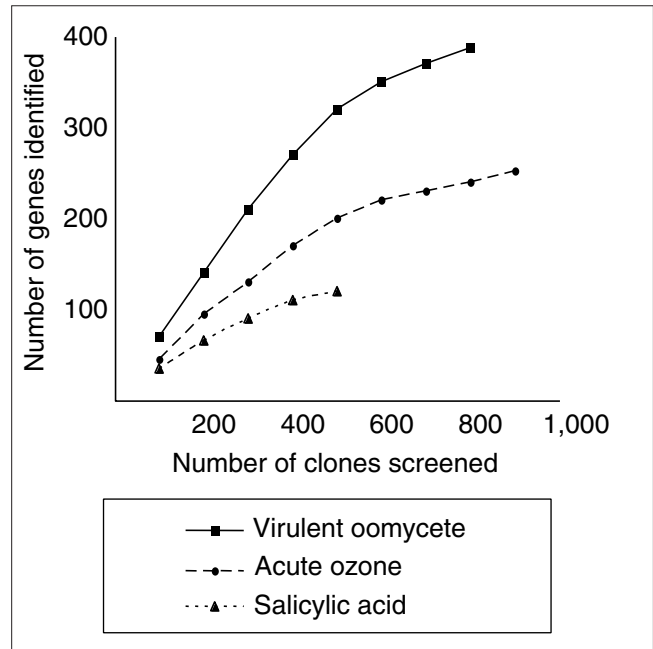### Differential expression of genes identified by SSH
cDNA clones for differentially expressed genes were identified by successively screening new clones first with the unsubtracted driver cDNA (Figure 2a) and tester (Figure 2b) cDNA pools as the probes, then the forward-subtracted (Figure 2c) and reverse-subtracted (Figure 2d) cDNA pools as probes. The first screen identifies cDNAs corresponding to only the most abundant differentially expressed genes, while the second screen identifies genes that are expressed less abundantly, but still differentially. The results from the four hybridizations were recorded for each clone, and those showing the most marked differential expression were selected for sequencing. Although selecting clones that showed strong hybridization only with the forward subtracted pool was straightforward, it was more difficult to select differentially expressed genes when signals were detected in both the forward-subtracted and unsubtracted pools. We endeavored to select those clones that showed 4-5-fold differential hybridization. However, as this was done by visual inspection, such genes constitute a potential source of false positives.
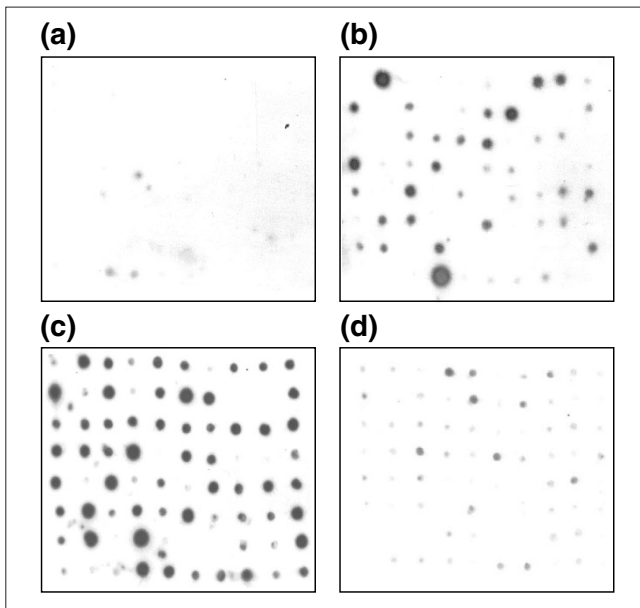
After screening the first three libraries (acute ozone, SA and virulent oomycete infection), we incorporated an additional procedure that permitted us to identify clones already represented in our collection. This was necessary simply because there is significant overlap at the molecular level in plant responses to different stresses and cDNA fragments were being identified and sequenced redundantly. We therefore pooled and labeled aliquots of DNA from sequenced clones, using them to probe each new set of clones. This procedure not only permitted identification of clones representing genes in previous libraries, but also allowed us to monitor the completeness of screening of each library. Figure 3 shows that the yield of new clones decreased as the number of clones examined increased, indicating that by the time several hundred clones had been examined, few new genes remain to be identified in a given library. As the libraries were both normalized and enriched for differentially expressed sequences, it is likely that the recovery of cDNAs
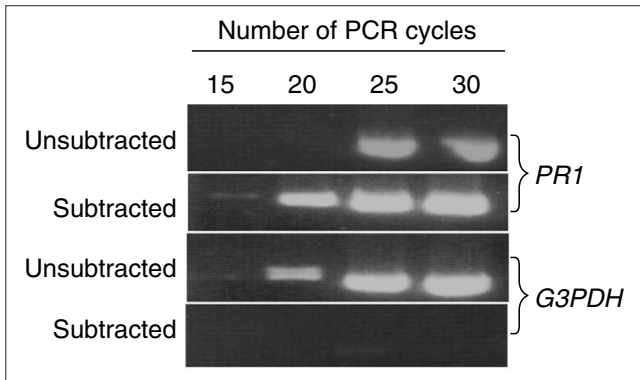
Figure 1
Analysis of subtraction efficiency using PCR. Tester cDNA was prepared from the poly(A)$^+$ RNA of plants sprayed with the virulent oomycete *P. parasitica* and the driver cDNA was from water-treated control plants. The unsubtracted and subtracted pools of cDNA were amplified using primers for the pathogen-inducible *PR1* gene or the constitutively expressed *G3PDH* gene. Aliquots of the samples were taken after 15, 20, 25 and 30 cycles of PCR amplification and the products were analyzed on a 2% agarose gel.

Figure 2
Differential screening of clones from the stress libraries generated using SSH. Subtracted cDNA fragments obtained by the SSH procedure were cloned (see Materials and methods for details) and maintained as bacterial cultures in 96-well plates for each library. Quadruplicate colony dot blots were prepared and the membranes hybridized with labeled unsubtracted cDNA probes derived from **(a)** the driver, **(b)** the unsubtracted cDNA probes from the tester, **(c)** the forward subtracted cDNAs or **(d)** the reverse-subtracted cDNA.

Figure 3
Recovery of differentially expressed genes as a function of the number of clones screened. To reduce the redundant sequencing of clones, we pooled DNA from previously sequenced clones and used it as probes on new filters prepared from the stress libraries. As more clones were screened within a library, the fraction of genes that had not yet been recovered decreased.

for differentially expressed genes is reasonably complete. It should be noted, however, that because we screened each new library for clones we had already identified in previous libraries, the libraries are not independent of each other. This procedure might also lead us to miss highly homologous members of multigene families.

We cloned and sequenced the reverse-subtracted cDNAs for just two libraries, those prepared from ozone-treated plants and those infected with a virulent strain of the oomycete pathogen. Although we screened almost 600 clones in these two libraries, we identified cDNAs for only 48 differentially expressed genes, most of which encode proteins involved in photosynthesis. Thus it appears that a relatively small number of genes is downregulated by stress and suggests that different kinds of stress downregulate the same genes.

Although differential expression of every gene has not yet been tested under every stress condition, of the more than 700 genes tested to date, roughly 90% have exhibited differential expression by either northern blotting or DNA microarray analysis (see Additional data files), each of which has different sensitivity limitations. Some representative northern blots are shown in Figure 4. Because these tests have been carried out under a more limited set of conditions than those represented by the treatments used for library construction, it appears likely that most of the cDNA clones described here represent differentially expressed genes.
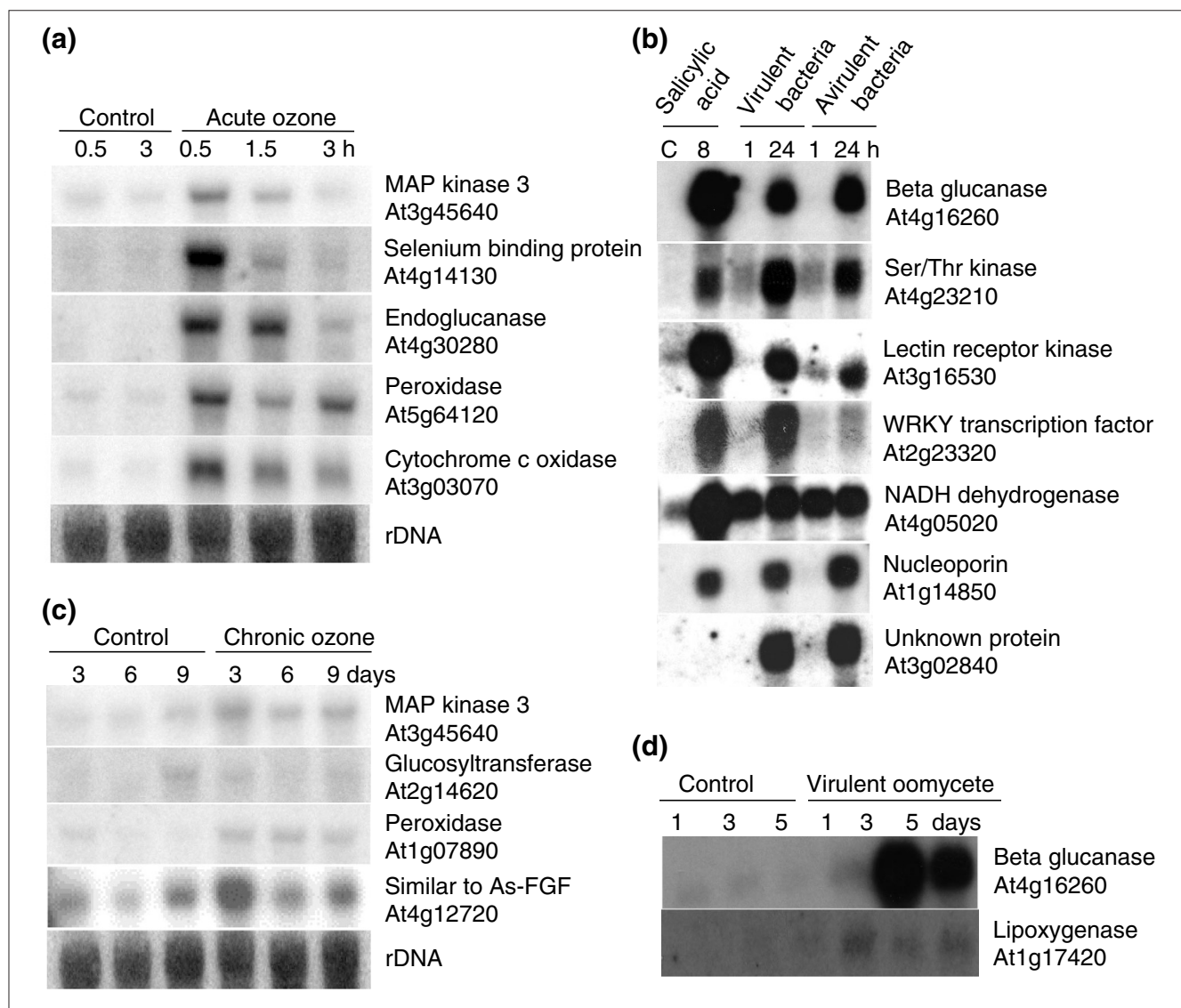
Figure 4
Representative northern blot analyses of stress-modulated genes using cloned cDNA fragments from the SSH libraries. **(a)** Acute ozone treatment; **(b)** comparison of treatment with salicylic acid, virulent bacteria or avirulent bacteria; **(c)** chronic ozone treatment; **(d)** treatment with virulent oomycete. Control leaves (C) were infiltrated with 10 mM MgCl$_2$. Tissue for RNA isolations was harvested at the indicated time points post-treatment.

Sequence analysis of stress ESTs

We sequenced 1,461 clones selected for differential expression as described above from among more than 6,000 clones in the eight stress cDNA libraries. On average, 12% of the sequences in each library were redundant (Table 2). The frequency with which identical cDNA fragments were isolated and sequenced was low (approximately 2%), in large part because of the prescreening to eliminate such redundancy. Most of the redundancy within libraries (10%), as well as between libraries (25%), resulted from the use of *Rsa*I-restricted cDNA fragments in the SSH procedure. Because of the restriction step, two or more different cloned cDNA fragments can represent a single transcript. This type of

redundancy was rapidly identified using coincidence of the Munich Information Center for Protein Sequences (MIPS) identifiers. Thus the 1,461 different fragments sequenced identified a total of 1,058 different genes (Table 2).

With the information gathered from the MIPS *Arabidopsis* database [36] and InterPro protein domain searches [37] we were able to identify or assign putative functions to about three-quarters (764) of the genes in the stress cDNA collection. More than 290 SSH clones encoded proteins with insufficient similarity to proteins of known function to assign a function with confidence, and we therefore classified them as being of unknown function. The genes of known function were

Table 2

Sequence analysis of stress cDNA libraries

Library	Aoz (%)	Coz (%)	APP (%)	VPP (%)	MJ (%)	Rpm (%)	DC (%)	SA (%)	**Total**
Sequenced cDNAs	387	80	137	252	91	235	159	120	1461
Genes identified	257 (66)	69 (86)	111 (81)	218 (87)	57 (63)	165 (68)	113 (71)	68 (57)	1058 (72)
No EST match	46 (18)	11 (16)	19 (17)	29 (13)	7 (12)	32 (19)	21 (19)	13 (19)	178 (17)
Known function	182 (71)	53 (77)	83 (75)	155 (71)	42 (74)	120 (73)	78 (69)	51 (75)	764 (72)
Unknown function	75 (29)	16 (23)	28 (25)	63 (29)	15 (26)	45 (27)	35 (31)	17 (25)	294 (28)

Abbreviations as in Table 1.

sorted into the 12 primary functional categories [38]. The distribution of the genes with known or predicted functions is represented as a pie chart in Figure 5. The largest set of genes (15%) was assigned to the metabolism category, while genes involved in cell growth/division constituted the smallest group, comprising less than 2% of the genes. Genes involved in signal transduction and protein destination/storage formed the second (13%) and third largest groups (12%), respectively. Genes implicated in stress/defense response and genes involved in transcription together constituted 20% of the stress cDNA collection (Figure 5).

Digital northern analysis

EST datasets have been used recently to extract information on gene-expression levels [39-41]. The underlying assumption of such a 'digital northern' analysis is that the number of EST clones is proportional to the abundance of the mRNA used for constructing the library [42]. Most *Arabidopsis* EST collections are derived from non-normalized cDNA libraries, making them useful for this purpose [43]. We recorded the number of EST hits from the MIPS summary report link or by BLAST searches with the *Arabidopsis* EST database for each clone in our stress collection. Using 1/20,000 as the definition of a rare message [44-46] and knowing that there are more than 100,000 *Arabidopsis* ESTs in GenBank, we consider genes with fewer than six ESTs to represent rarely transcribed genes.

By the foregoing criteria, 577 (55%) of the stress-inducible genes are in the rarely transcribed category and 178 of these (17%) are not represented by an EST in the *Arabidopsis* EST database. Only 2 (0.0018%) of the genes in the stress cDNA collection are represented by more than 200 ESTs and can be classified as abundantly transcribed. These are the Rubisco activase (AozUF12, At2g39730) and a polyubiquitin gene (APP-FD09, At5g05320). The remaining 479 (45%) cDNAs correspond to genes transcribed at moderate rate (7-200 ESTs) in unstressed plants. This analysis does not, of course, reflect the stress-altered levels of these transcripts. Nonetheless, it reveals that a significant fraction of stress-modulated genes is not represented in existing EST collections and that many are infrequently transcribed in normal plants, underlining the need to construct specialized libraries.

Promoter analysis

The number, order, and type of protein-binding sequences present in promoters are major determinants of the differences in expression patterns of genes. Because the transcript levels of the genes represented in the present collection change in response to stress, protein-binding motifs associated with stress-inducibility should be over-represented in the promoters of the stress collection when compared with the total complement of *Arabidopsis* promoters. We compared the frequency of occurrence of 16
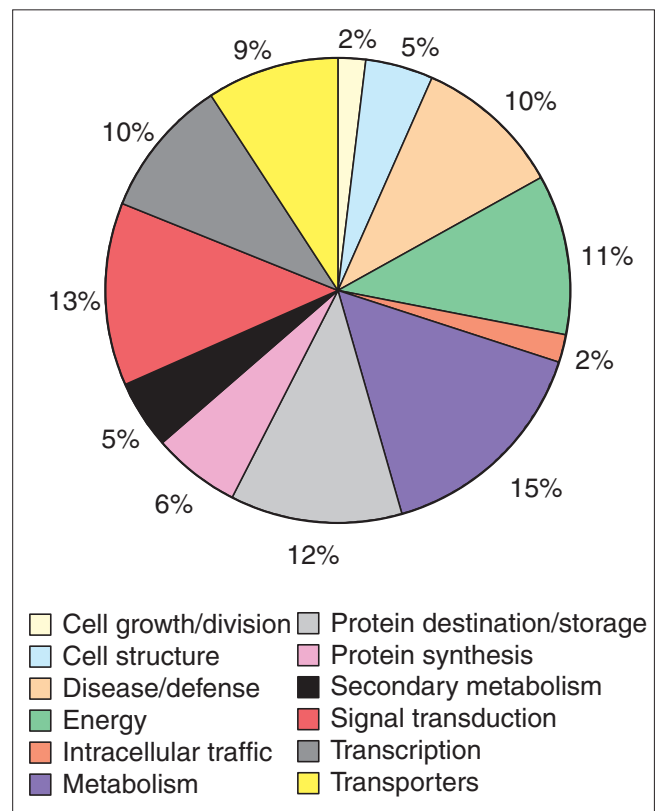
Figure 5
A pie chart showing the fraction of stress-modulated genes in each of the functional categories described in Bevan *et al.* [38].

different stress-related motifs in the total collection, and in each library individually, with their frequencies in the total population of *Arabidopsis* promoters (see Materials and methods).

We found that six of the 16 stress-related motifs analyzed were over-represented in the promoter sequences of the stress collection as a whole (Table 3a), and four additional motifs were over-represented in the promoters of at least one of the libraries (Table 3b). The ABRE-like motif, W-box motif, W-box-like motif, SA-inducible motif, the G-box and the TGA motifs were significantly over-represented in the promoters of the stress collection as a whole, compared with the total population of *Arabidopsis* promoters (Table 3a). The H-box factor (HBF) motif was over-represented in the promoters of genes represented in the virulent bacterial library, heat-shock element (HSE) and Myb4 motifs were over-represented in the avirulent bacterial library, and the Myc motif was over-represented in the SA library (Table 3b). We also looked for combinations of motifs that had been reported to function together [47]. The G-box and H-box motifs occur together in 31 promoters in our collection (*p* = 0.00587). Over all, more than half of the tested stress motifs are over-represented in the promoters of the genes in the stress cDNA collection, reflecting its enrichment for stress-responsive genes. The two ethylene-related motifs (GCC-box and EIN3), and the drought-response element (DRE), AtMyb1, AtMyb2 and AtMyb3 motifs were not statistically significantly over-represented in the promoters of genes identified by the stress cDNA collection.

Discussion

Analysis of the total SSH library

To assess the contribution of the present stress cDNA collection to characterizing the *Arabiopsis* stress transcriptome, we examined five recent reports of *Arabidopsis* stress/defense-modulated genes that used either commercial or investigator-constructed cDNA microarrays [7,17,18,48,49]. Using the accession numbers provided in each publication, we retrieved each sequence from the GenBank database. We then used BLAST at the TAIR site to search the AGI genome database and identified the MIPS code for the differentially expressed genes reported in each of these studies. Once the MIPS code had been identified, we were able to estimate the number of genes differentially expressed in each of these studies and determine the overlap between each group and the stress cDNA collection (Table 4).

The extent of overlap between the stress cDNA collection and the subset of stress- or defense-modulated genes ranged from a low of 16% with the 308 genes identified by Maleck *et al.* [7]) to a high of 32% with the 507 genes reported by Schenk [17]. Thus almost 70% of the genes in the stress cDNA collection have not previously been identified as stress/defense modulated. Conversely, a significant number of genes reported to be induced or repressed in the stress/defense response is not represented in the stress cDNA collection. This may be attributable in part to the stringency of the differential expression criterion (4-5-fold difference between experimental and control expression levels) we used in selecting clones to sequence. As the criterion for differential expression in the microarray studies analyzed was 1.5-2.5-fold over control

Table 3

Motif representation

(a) Motifs over-represented in the stress cDNA collection

Motif name	Motif sequence	*p*-value	Promoters	Libraries
ABRE-like [95]	BACGTGKM	3.83E-05	271	Aoz,VPP, MJ
GBF[74]	CACGTG	3.50E-04	197	MJ
WRKY [70]	TTGACY	0.0051	732	Aoz
WRKY-like [16]	BBWGACYT	0.0038	635	Aoz
SA-induced (LS7) [96]	ACGTCA	0.0003	273	Aoz, SA
TGA1 [74]	TGACG	0.0067	607	Aoz, Coz, VPP

(b) Motifs over-represented in individual stress libraries

Motif name	Motif sequence	*p*-value	Promoters with motif	Library	Promoters in library
HBF [89]	CCTACC	0.00764	19	DC	108
HSE [102]	CTNGAANNTTCNA	0.03404	6	Rpm	165
AtMyb4 [101]	AMCWAMC	0.04328	134	Rpm	165
MYC [95]	CACATG	0.02338	32	SA	67

GBF, G-box factor. For other abbreviations see text and Table 2.

Table 4

Comparison of the present stress-inducible cDNA collection with stress/defense genes identified in other large-scale studies

Study	Cut-off	Induced/repressed clones	MIPS id	Represented in stress cDNA collection	Treatment
[7]	2.5-fold	413	308	50 (16%)	Benzothiadiazol treatment, bacterial and oomycete pathogen
[17]	2.5-fold	705	507	160 (32%)	Fungal pathogen, SA, MJ, ethylene
[48]	2-fold	657	281	73 (26%)	Mechanical wounding
[18]	1.5-fold	175	114	32 (28%)	Hydrogen peroxide
[49]	1.5-fold	75	69	16 (23%)	Heat treatment and senescence

Maleck *et al.* [7] studied 10,000 EST clones obtained from the *Arabidopsis* Biological Resource Center representing approximately 7,000 genes. Schenk *et al.* [17] studied a custom array containing 2,375 ESTs with a biased representation of putative defense-associated and regulatory genes. From our estimate of redundancy (approximately 30%), this array contains 1,662 distinct genes. The study of Cheong *et al.* [48] involved the Affymetrix Arabidopsis Genome GeneChip array representing 8,200 genes. Desikan *et al.* [18] studied Arabidopsis Functional Genomics Consortium microarrays containing 11,000 EST clones representing approximately 7,800 distinct genes. The custom array of Swidzinski *et al.* [49] contained 75 ESTs previously implicated in programmed cell-death responses such as senescence and hypersensitive response.

levels, many genes identified as differentially expressed in these studies would not have been included in our libraries.

Although we do not have the information to determine all of the reasons for the limited overlap between previous studies and the present one, we can identify several contributing factors. A major factor is the ability of the SSH procedure to identify genes that are expressed differentially, but at low levels. As noted earlier, more than half of the differentially expressed genes in the present collection can be classified as rarely transcribed on the basis of their representation by six or fewer ESTs among the more than 100,000 *Arabidopsis* ESTs sequenced. Indeed, almost 17% of the genes in our SSH collection are being entered in the *Arabidopsis* EST database for the first time as a result of the present work. It seems unlikely that the genes represented on the EST arrays used by Maleck *et al.* [7] and Desikan *et al.* [18] comprise an unbiased sample of *Arabidopsis* genes, as genes transcribed in normal plants and at moderate or high levels are more likely to be represented than the rarely, but differentially, transcribed genes that dominate the stress cDNA collection. In addition, there are differences in experimental conditions and in detection sensitivity that undoubtedly contribute, but whose contribution to different outcomes is difficult to assess. But it should also be noted that the overlap between microarray datasets generated using the same conditions and organisms in different laboratories can be surprisingly low, particularly when different microarray technologies are used. Enriched cDNA libraries, such as the one we have constructed, may therefore contribute to the characterization of the stress transcriptome through the construction of standardized specialized arrays.

Functional classification of genes represented in SSH libraries

We were able to assign nearly three-quarters of the genes to functional groups based on sequence similarities with known genes or motifs (see Additional data files). Although

functional assignment based only on sequence homology needs experimental verification, it nonetheless provides a measure of the diversity of the genes in the stress cDNA collection. Genes from all the major functional categories are represented in the collection (Figure 5, and see Additional data files).

Genes encoding proteins involved in stress/defense signaling comprise 13% of the genes in the collection. They include genes encoding proteins involved in signal perception (several types of receptor kinases) and signal transmission (G proteins, protein kinases, protein phosphatases, calcium-binding proteins) (see Additional data files). About 10% of the genes (79) in the collection were classified as stress/defense response genes. This includes genes encoding proteins of the antioxidant response (GSTs, peroxidases), the SAR response (pathogenesis-related genes) and cell rescue [50]. Genes induced in response to other stresses, such as drought [51], heat shock [52], dehydration [53-55], and elicitors such as *Avr9* [56], *avrRpt2* [57] and harpins [58,59], were also identified in the stress cDNA collection, underscoring the overlap between the different stress responses

Genes encoding proteins involved in moving, modifying, storing and degrading proteins constituted the third largest group (12%) in the stress cDNA collection. Nearly half of the genes (45) in this category (92) are involved in proteolysis (see Additional data files). Proteolysis of important regulatory proteins is a key aspect of cellular regulation in eukaryotes [60,61] and there is evidence that the ubiquitin-proteasome pathway is important in implementation of the plant defense response [62-64]. Proteosome subunit genes are induced in response to stresses [65,66] and several regulatory subunits were identified in the stress cDNA collection. The F-box-containing proteins of the SCF complex constitute a family of E3 ligases, key components of the ubiquitin-proteasome pathway [67], as

are many RING finger proteins [68]. The RING finger motif is thought to mediate protein-protein interactions and E3 ligase complex assembly. Certain RING finger proteins are rapidly induced by elicitors in *Arabidopsis* and may be involved in the rapid degradation of regulatory proteins during early stages of pathogen attack [69]. Four genes encoding proteins with an F-box domain and seven different genes encoding RING/RING-H2-finger proteins, including the elicitor-induced *ATL6*-like gene [69], were identified in the stress cDNA collection. In summary, the stress cDNA collection comprises a broad repertoire of stress-responsive genes encoding proteins that are involved in both the initial and subsequent stages of the physiological response to abiotic stress and pathogens.

Transcription factors and regulatory elements in the promoters of genes in the stress cDNA collection

The WRKY transcription factors are involved in defense, wounding, senescence and plant development [7,16,70-73]. There are 70 genes with a WRKY domain in the *Arabidopsis* genome and six of them (WRKY15, 25, 33, 46, 62, 70) are in the stress cDNA collection. These WRKY transcription factors modulate gene expression by binding to W-boxes and W-box-like motifs, which are significantly over-represented in the promoters of the genes in the stress cDNA collection (Table 3a). The bZIP transcription factors are important in the regulation of genes activated by light, UV radiation, pathogen attacks, elicitors, wounding, abscisic acid (ABA) treatment, and SA treatment [47,71,74-81]. We identified two basic zipper (bZIP) transcription factors and basic helix-loop-helix (bHLH) transcription factors in the stress cDNA collection. The promoters of a subset of genes in the stress cDNA collection are enriched in G-box and TGA motifs, suggesting that they may be regulated by these bZIP and bHLH transcription factors. Identifying insertional knock-outs or creating antisense lines for these transcription factor genes and using them for expression profiling with microarrays constructed from the stress cDNA collection will facilitate identification of the targets for these transcription factors.

Abscisic acid response elements (ABREs) are important during the plant's response to abiotic stresses such as dehydration, salinity and cold, all of which are ABA-mediated [82,83]. There is evidence that ROS are involved in ABA signaling [84]. The enrichment of the ABRE motif in the promoter sequences of the genes in the stress cDNA collection may reflect crosstalk among stress signaling pathways mediated by common second messengers, such as H_2O_2. The over-representation of the SA-inducible element in the promoter sequences of the genes in the stress cDNA collection, especially those in the SA library, further validates the enrichment of the stressor-specific genes using the SSH technique and suggests that these genes may be regulated by the changes in the levels of SA that occurs in stress responses [85].

The motifs associated with the hormone ethylene (GCC-box and EIN3) were not significantly over-represented in the promoters of genes represented in the stress cDNA collection. A plausible reason is that SA and ethylene act antagonistically to each other [5]. The over-representation of the SA-inducible genes in the collection may be correlated with the reduced representation of the ethylene-responsive genes in the stress cDNA collection. The Myb1 motif was identified in *Antirrhinum* [86], the Myb2 and Myb3 motifs in *Petunia* [87]. These sequences have not been verified as valid Myb binding motifs in *Arabidopsis*. The under-representation of the DRE-like element suggests that genes whose expression is modulated by pathogen attacks and oxidative stress signaling may have minimal overlap with genes involved in dehydration and cold responses.

There is evidence that the H-box motif is involved in developmental regulation of flowering, but there is no report of its involvement in stress/defense responses [88,89]. However, a combination of the H-box motif and the G-box motif has been shown to be important for binding of bZIP transcription factors to rapidly induce defense-related genes [47,90,91]. The significant over-representation of these two motifs together in the promoter sequences of the stress cDNA collection suggests that combinatorial interactions between these two *cis*-elements may also be important in the regulation of stress-responsive genes. It is becoming increasingly evident that a major theme underlying eukaryotic transcriptional regulation is combinatorial control [92]. Identifying the *cis*-elements and the cognate transcription factors that bind to them during stress or defense responses is the first step towards characterization of higher-order nucleoprotein complexes.

Materials and methods
Plant materials and growth conditions
A. thaliana ecotype Col-0 plants were grown in soil (Scotts-Sierra Horticultural Products Company, Marysville, OH) in 5 cm pots (50 per flat) under fluorescent light 30 W/m²/s with a 14 h light/10 h dark photoperiod for 3-4 weeks.

Ozone treatment
Plants were transferred to growth chambers for ozone (O_3) fumigation (clean air control plants were transferred to an adjacent chamber under identical conditions except for the O_3 treatment). For acute O_3 treatment, plants were transferred to the experimental chambers 4 weeks after germination, allowed to acclimate for 1-2 days, and then maintained in clean air (controls) or treated with 0.35-0.4 μl/liter O_3 for up to 6 h. For chronic O_3 treatment, plants were transferred to the experimental chambers 3 weeks after germination, allowed to acclimate 1-2 days, and then maintained in clean air (controls) or treated with 0.15 μl/liter O_3 for 6 h per day for up to 9 days. For SSH experiments, 4-10 leaves were harvested, avoiding the most immature leaves,

cotyledons, and the first two true leaves. Leaves from 1, 3, 6, 9, 12 and 24 h time points were pooled for the acute ozone library and leaves from plants exposed for 2, 4 and 6 days were pooled for the chronic ozone library. RNA was isolated as described below.

Bacterial infections

Pseudomonas syringae pv. *tomato* (*Pst* DC3000), which causes bacterial speck disease in Col-0 plants, was used to infect plants for construction of the virulent bacterial library. The *Pst* DC3000 expressing *avrRpm1* gene induces hypersensitive cell death in Col-0 plants and was used to infect plants for constructing the avirulent bacterial library. The bacteria were grown on King's agar plates at 28°C. Bacterial cultures were prepared by resuspending the cells from overnight cultures in 10 mM $MgCl_2$ to the required optical density ($OD_{600} = 0.05$; $1 \ OD_{600} = 10^9$ colony-forming units (CFU)/ml). A titer of 5×10^7 CFU/ml was used for infiltrations. Leaves were infiltrated on the abaxial side using a 1-ml syringe. The inoculated leaves were harvested 1, 8 and 24 h after inoculation. Equal amounts of tissue were pooled from each time point for library construction.

Inoculations with *Peronospora parasitica*

The Ahco strain of *P. parasitica* was maintained on Col-0 plants and the Emwa strain on Ws plants in a growth chamber with 10-h day at 16°C and 14-h dark period at 20°C. Fresh spores were collected in water and the inoculum was adjusted to 2×10^4 conidiosporangia per milliliter and applied as a fine mist to the seedlings using an airbrush sprayer (Paasche Air Brush Company, Harwood Heights, IL). The sprayed plants were placed in a tray containing water-soaked paper towels, covered with a lid sprayed on the inside with a fine mist of water, and sealed using plastic wrap to maintain the humidity. The sprayed plants were transferred to the growth chamber in which the pathogen is regularly maintained. Control plants were sprayed with water using the airbrush sprayer and were maintained under the same conditions. Seedlings were harvested 1, 3 and 5 days after inoculation and pooled for RNA extraction and identification of genes induced in the compatible interaction between Col-0 plants and the Ahco strain of the pathogen. Seedlings were harvested 12 h, 1, 2 and 3 days after inoculation for the incompatible interaction between Col-0 plants and the Emwa strain of *P. parasitica*.

Salicylic acid (SA) and methyl jasmonate (MJ) treatments

Four-week-old plants were sprayed with 1 mM sodium salicylate dissolved in water (Sigma, St. Louis, MO) and control plants were treated with water. Leaves were harvested 1, 8 and 24 h after the treatment. Equal amounts of tissues were pooled from each time point for RNA isolations. Four-week-old plants were sprayed with 50 mM methyl jasmonate (MJ) (Sigma) in 0.001% ethanol and control plants were treated with 0.001% ethanol. Leaves were harvested 1, 8, 24 and

48 h after the treatment. Equal amounts of tissue were pooled from each time point for RNA isolation.

RNA isolation

Total RNA was isolated from frozen leaf tissue using Trizol (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Briefly, tissue was ground to a fine powder in liquid N_2 and homogenized in Trizol (1 ml reagent/0.1 g fresh weight tissue), then allowed to stand at room temperature for 5-10 min. Chloroform:isoamyl alcohol (24:1) was added using 0.2 ml/ml Trizol. Samples were mixed by vortexing and allowed to stand at room temperature for 2-5 minutes more, then centrifuged for 15 min at 10,000*g*. The upper aqueous layer was removed to a clean tube and RNA was precipitated with an equal volume of isopropanol. After 10 min at room temperature, samples were centrifuged for 15 min at 10,000*g* to pellet the RNA. The pellet was washed with 75% ethanol and resuspended in RNase-free water. The typical yield of total RNA was 50-80 μg/100 mg leaf tissue. Poly(A)$^+$ RNA was purified from total RNA using PolyTract oligo(dT) columns (Promega, Madison, WI).

Suppression subtractive hybridization (SSH)

SSH was carried out using the PCR-Select Subtractive Hybridization kit (Clontech, Palo Alto, CA). Experimental and control samples for each treatment were processed simultaneously to reduce false positives. We increased the amount of mRNA to 3-4 μg from the 2 μg recommended by the manufacturer to compensate for the loss of mRNA during the phenol chloroform extractions. cDNA prepared from the treated samples was used as the 'tester' and that from the control sample as 'driver' for the forward subtraction to isolate fragments corresponding to genes whose expression level was increased following the treatment. The reverse subtraction was carried out with the control sample as tester to isolate fragments corresponding to genes whose expression level decreased following the treatment. The PCR-based enrichment of differentially expressed sequences depends on the number of tester molecules with adaptors ligated to their ends. If the fraction of tester cDNA with the adaptors was less than 25%, ligations were repeated. We designed plant-specific glyceraldehyde-3-phosphate dehydrogenase (*G3PDH*) primers to test the ligation efficiency as recommended by the manufacturer. A *G3PDH* gene fragment of approximately 300 bp was amplified with *G3PDH* 3′ and 5′ primers (GAPA-F: GGTAGGATCGGGAGGAAC; GAPA-R: GATAACCTTCTTGGCACCAG) using the adaptor-ligated cDNA as template. The tester cDNA was also amplified with G3PDH 3′ primer and an adaptor-specific primer, which yields a fragment that is 200 bp bigger than the fragment generated with the gene-specific primers. Samples from the *G3PDH* gene fragment amplifications using the subtracted and unsubtracted cDNA pools were analyzed after 15, 20, 25 and 30 cycles of PCR. Primers to amplify regions without an *Rsa*I site were designed for two stress-induced genes, the

pathogen-inducible *PR1* gene (PR1-F: ATGAATTTTACTG-GCTATTC; PR1-R: AACCCACATGTTCACGGCGGA), the O_3-inducible amino-cyclopropane-carboxylate (ACC) synthase gene, *ACS6* (ACS6-F: CATAAGTGTTGCGGAAGTAA; ACS6-R: GGCAATGGAACGAACC) and the jasmonate-inducible defensin gene, *PDF1.2* (PDF-F: ATGGCTAAGTTTGCTTCCAT; PDF-R: ACATGGGACGTAACAGATAC) [9,19,35]. These were used to test the subtraction efficiency of the corresponding libraries before cloning.

Cloning and differential screening

Secondary PCR products were cloned into the T/A cloning vector pCR2.1TOPO (Invitrogen; acute O_3 library) or the Advantage PCR cloning vector (Clontech; all the other libraries) according to manufacturer's instructions. About 500-1000 colonies were picked and grown in 96-well microtiter plates in LB medium with 100 mg/l ampicillin. The clones were amplified using the nested primers 1 and 2R (Clontech manual) to check for the presence and size of individual inserts. The PCR products were run on high-density agarose gels in duplicate (200 wells/gel) and transferred to nylon filters. The membranes were hybridized under stringent conditions with equivalent amounts of ^{32}P-labeled probes generated from unsubtracted and subtracted cDNAs.

Sequencing and sequence analysis

Cycle sequencing reactions were prepared using BigDye dye terminator, modified for smaller reactions with Half-term (GenPak, Stony Brook, NY) to conserve reagents. Sequencing electrophoresis was carried out by the PSU Nucleic Acid Facility using the nested 1 or 2R adaptor primers. Each sequence was edited to correct sequencing ambiguities and remove the primer sequence. The edited sequences were used to query the *Arabidopsis* Genome Initiative (AGI) database using the BLAST sequence comparison algorithms at the TAIR website [93]. Homologies exceeding 50 nucleotides that showed more than 90% identity to sequences in the database were considered significant. Sequences that failed to show significant homologies were used to query the GenBank (minus EST and BAC ends) database using the BLAST algorithms. The MIPS website [36] was searched for each sequence that exceeded the significance threshold to gather information, including the genomic location of the clone, the name of the gene, and the number of ESTs. The entire predicted protein-coding sequence of each gene obtained from the MIPS website was then used to search the InterPro database [37] for identifying protein domains.

Northern-blot analysis

Total RNA was fractionated on a 1.2% agarose/0.4 M formaldehyde RNA gel and transferred to Hybond N$^+$ nylon membrane (Amersham-Pharmacia, UK). Probes were made from PCR-amplified fragments of selected clones using the ReadyPrime random primed DNA labeling kit (Amersham-Pharmacia) with [α-^{32}P] (ICN Biomedicals, Irvine, CA). Blots were hybridized and washed according to standard procedures [94].

Stress motifs for promoter analysis

Sixteen stress-related *cis*-elements were selected on the basis of their identification in other studies on *Arabidopsis* stress/defense-modulated genes [7,16,71]. The frequency of W-box (TTGACY) and W-box-like (BBWGACYT) elements was tested as these are binding sites for plant-specific WRKY transcription factors involved in plant development and plant responses to environmental stresses [7,16,70-73]. G-boxes (CACGTG), H-boxes (CCTACC) and TGA motifs (TGACG) are binding sequences for bZIP transcription factors, which have an important role in the regulation of genes activated by environmental cues [71,74-81]. We also tested some of the well-characterized stress hormone-responsive motifs, including the ABA-response element (BACGTGKM) [95], SA-inducible motif (ACGTCA) [96], two ethylene-related motifs, GCC-box (GCCGCC) [97] and the EIN3 motif (GGATGTA) [98]. The *Arabidopsis* Myc and Myb homologs are important in the plant's responses to pathogens, low temperatures and dehydration [82,99,100]. We tested for the frequency of occurrence of the Myc-element (CACATG) and four different Myb motifs (AtMyb1, MTCCWACC; AtMyb2, TAACSGTT; AtMyb3, TAACTAAC; AtMyb4, AMCWAMC) [71,101] in the promoters of genes identified by cDNAs in the stress collection. We also tested the frequency of occurrence of two other known stress motifs, the DRE (DRCCGACNW) [95], and the HSE (CTNGAANNTTCNA) [102] to determine the extent of overlap in genes activated during different stresses.

Probability and significance calculations for promoter analysis

To determine whether a particular motif is over-represented among the promoters of a given SSH library, we first determined the frequency of each motif in the total complement of *Arabidopsis* promoters using the 1,000 bp upstream promoter files from TAIR [103]. We then calculated the probability of finding m promoter regions having one or more motifs in the set of n promoters in each SSH library, as well as in the total SSH library collection. We considered a motif to be significantly over-represented if this probability was less than 0.05. These calculations were implemented using Perl scripts that are available from the StresDB [104].

Additional data files

Additional data files including a table of the digital northern data showing the number of ESTs for each clone in the SSH library, the microarray data for the response to 1 hour of ozone treatment and five days after virulent oomycete infection, and details of how the microarrays were printed and hybridized are available as with the online version of this paper.

Higgins and Lynn Lindstrom for their help with the differential screening and sequencing.

References

1. Ryals J, Uknes S, Ward E: **Systemic acquired resistance.** *Plant Physiol* 1994, **104:**1109-1112.
2. Lamb CJ: **Plant disease resistance genes in signal perception and transduction.** *Cell* 1994, **76:**419-422.
3. Lamb C, Dixon RA: **The oxidative burst in plant disease resistance.** *Annu Rev Plant Physiol Plant Mol Biol* 1997, **48:**251-275.
4. Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar SP: **Signaling in plant-microbe interactions.** *Science* 1997, **276:**726-733.
5. Dong X: **SA, JA, ethylene, and disease resistance in plants.** *Curr Opin Plant Biol* 1998, **1:**316-323.
6. Glazebrook J: **Genes controlling expression of defense responses in *Arabidopsis*.** *Curr Opin Plant Biol* 1999, **2:**280-286.
7. Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA: **The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance.** *Nat Genet* 2000, **26:**403-410.
8. Pieterse CM, van Wees SC, Hoffland E, van Pelt JA, van Loon LC: **Systemic resistance in *Arabidopsis* induced by biocontrol bacteria is independent of salicylic acid accumulation and pathogenesis-related gene expression.** *Plant Cell* 1996, **8:**1225-1237.
9. Penninckx IA, Eggermont K, Terras FR, Thomma BP, De Samblanx GW, Buchala A, Metraux JP, Manners JM, Broekaert WF: **Pathogen-induced systemic activation of a plant defensin gene in *Arabidopsis* follows a salicylic acid-independent pathway.** *Plant Cell* 1996, **8:**2309-2323.
10. Penninckx IA, Thomma BP, Buchala A, Metraux JP, Broekaert WF: **Concomitant activation of jasmonate and ethylene response pathways is required for induction of a plant defensin gene in *Arabidopsis*.** *Plant Cell* 1998, **10:**2103-2113.
11. Chrispeels MJ, Holuigue L, Latorre R, Luan S, Orellana A, Pena-Cortes H, Raikhel NV, Ronald PC, Trewavas A: **Signal transduction networks and the biology of plant cells.** *Biol Res* 1999, **32:**35-60.
12. Pastori GM, Foyer CH: **Common components, networks, and pathways of cross-tolerance to stress. The central role of "redox" and abscisic acid-mediated controls.** *Plant Physiol* 2002, **129:**460-468.
13. Xiong L, Zhu JK: **Abiotic stress signal transduction in plants: molecular and genetic perspectives.** *Physiol Plant* 2001, **112:**152-166.
14. Reymond P, Weber H, Damond M, Farmer EE: **Differential gene expression in response to mechanical wounding and insect feeding in *Arabidopsis*.** *Plant Cell* 2000, **12:**707-720.
15. Hasegawa P, Bresson R, Zhu JK, Bohnert HJ: **Plant cellular and molecular responses to salinity.** *Annu Rev Plant Physiol Plant Mol Biol* 2000, **51:**463-499.
16. Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, *et al.*: **Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses.** *Plant Cell* 2002, **14:**559-574.
17. Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM: **Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis.** *Proc Natl Acad Sci USA* 2000, **97:**11655-11660.
18. Desikan R, A-H-Mackerness S, Hancock JT, Neill SJ: **Regulation of the *Arabidopsis* transcriptome by oxidative stress.** *Plant Physiol* 2001, **127:**159-172.
19. Crute I, Beynon J, Dangl JL, Holub E, Mauch-Mani B, Slusarenko A, Staskawicz B, Ausubel F: **Microbial pathogenesis of *Arabidopsis*.** In *Arabidopsis 1994.* Edited by Meyerowitz EM, Somerville CR. New York: Cold Spring Harbor Laboratory Press; 1994:705-747.
20. Debener T, Lehnackers H, Arnold M, Dangl JL: **Identification and molecular cloning of a single *Arabidopsis* locus conferring resistance against a phytopathogenic *Pseudomonas* isolate.** *Plant J* 1991, **1:**289-302.
21. Koch E, Slusarenko A: ***Arabidopsis* is susceptible to infection by a downy mildew fungus.** *Plant Cell* 1990, **2:**437-445.
22. Reymond P, Farmer EE: **Jasmonate and salicylate as global signals for defense gene expression.** *Curr Opin Plant Biol* 1998, **1:**404-411.
23. Conklin PL, Last RL: **Differential accumulation of antioxidant mRNAs in *Arabidopsis thaliana* exposed to ozone.** *Plant Physiol* 1995, **109:**203-212.
24. Sharma YK, Leon J, Raskin I, Davis KR: **Ozone-induced responses in *Arabidopsis thaliana*: the role of salicylic acid in the accumulation of defense-related transcripts and induced resistance.** *Proc Natl Acad Sci USA* 1996, **93:**5099-5104.
25. Schlagnhaufer CD, Glick RE, Arteca RN, Pell EJ: **Molecular cloning of an ozone-induced 1-aminocyclopropane-1-carboxylate synthase cDNA and its relationship with a loss of rbcS in potato (*Solanum tuberosum* L.) plants.** *Plant Mol Biol* 1995, **28:**93-103.
26. Ruzsa SM, Mylona P, Scandalios JG: **Differential response of antioxidant genes in maize leaves exposed to ozone.** *Redox Rep* 1999, **4:**95-103.
27. Jabs T, Tschöpe M, Colling C, Hahlbrock K, Scheel D: **Elicitor-stimulated ion fluxes and O$_2^-$ from the oxidative burst are essential components in triggering defense gene activation and phytoalexin synthesis in parsley.** *Proc Natl Acad Sci USA* 1997, **94:**4800-4805.
28. Runeckles V, Chevonne B: **Crop responses to ozone.** In *Surface Level Ozone Exposures and Their Effects on Vegetation.* Edited by Lefohn AS. Chelsea, MI: Lewis; 1992:189-270.
29. Liang P, Pardee AB: **Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction.** *Science* 1992, **257:**967-971.
30. Lisitsyn N, Wigler M: **Cloning the differences between two complex genomes.** *Science* 1993, **259:**946-951.
31. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270:**484-487.
32. Watson JB, Margulies JE: **Differential cDNA screening strategies to identify novel stage-specific proteins in the developing mammalian brain.** *Dev Neurosci* 1993, **15:**77-86.
33. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD: **Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries.** *Proc Natl Acad Sci USA* 1996, **93:**6025-6030.
34. Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov KA, Vagner LL, Ermolaeva OD, Lukyanov SA, Sverdlov ED: **Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate.** *Anal Biochem* 1996, **240:**90-97.
35. Schlagnhaufer CD, Arteca RN, Pell EJ: **Sequential expression of two 1-aminocyclopropane-1-carboxylate synthase genes in response to biotic and abiotic stresses in potato (*Solanum tuberosum* L.) leaves.** *Plant Mol Biol* 1997, **35:**683-688.
36. **MATDB entry page** [http://mips.gsf.de/proj/thal/db]
37. **InterPro** [http://www.ebi.ac.uk/interpro/scan.html]
38. Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, Van Staveren M, Stiekema W, *et al.*: **Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*.** *Nature* 1998, **391:**485-488.
39. Ewing R, Poirot O, Claverie JM: **Comparative analysis of the *Arabidopsis* and rice expressed sequence tag (EST) sets.** *In Silico Biol* 1999, **1:**197-213.
40. Mekhedov S, de Ilarduya OM, Ohlrogge J: **Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis.** *Plant Physiol* 2000, **122:**389-402.
41. Miernyk JA: **The J-domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones.** *Cell Stress Chaperones* 2001, **6:**209-218.
42. Audic S, Claverie JM: **The significance of digital gene expression profiles.** *Genome Res* 1997, **7:**986-995.
43. Ohlrogge J, Benning C: **Unraveling plant metabolism by EST analysis.** *Curr Opin Plant Biol* 2000, **3:**224-228.
44. Bishop JO, Morton JG, Rosbash M, Richardson M: **Three abundance classes in HeLa cell messenger RNA.** *Nature* 1974, **250:**199-204.
45. Axel R, Feigelson P, Schutz G: **Analysis of the complexity and diversity of mRNA from chicken liver and oviduct.** *Cell* 1976, **7:**247-254.
46. Wan JS, Sharp SJ, Poirier GM, Wagaman PC, Chambers J, Pyati J, Hom YL, Galindo JE, Huvar A, Peterson PA, *et al.*: **Cloning differentially expressed mRNAs.** *Nat Biotechnol* 1996, **14:**1685-1691.
47. Droge-Laser W, Kaiser A, Lindsay WP, Halkier BA, Loake GJ, Doerner P, Dixon RA, Lamb C: **Rapid stimulation of a soybean**

protein-serine kinase that phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses.** *EMBO J* 1997, **16:**726-738.

48. Cheong YH, Chang HS, Gupta R, Wang X, Zhu T, Luan S: **Transcriptional profiling reveals novel interactions between wounding, pathogen, abiotic stress, and hormonal responses in *Arabidopsis*.** *Plant Physiol* 2002, **129:**661-677.

49. Swidzinski JA, Sweetlove LJ, Leaver CJ: **A custom microarray analysis of gene expression during programmed cell death in *Arabidopsis thaliana*.** *Plant J* 2002, **30:**431-446.

50. Dietrich RA, Richberg MH, Schmidt R, Dean C, Dangl JL: **A novel zinc finger protein is encoded by the *Arabidopsis LSD1* gene and functions as a negative regulator of plant cell death.** *Cell* 1997, **88:**685-694.

51. Gosti F, Bertauche N, Vartanian N, Giraudat J: **Abscisic acid-dependent and -independent regulation of gene expression by progressive drought in *Arabidopsis thaliana*.** *Mol Gen Genet* 1995, **246:**10-18.

52. Yabe N, Takahashi T, Komeda Y: **Analysis of tissue-specific expression of *Arabidopsis thaliana* HSP90-family gene *HSP81*.** *Plant Cell Physiol* 1994, **35:**1207-1219.

53. Kiyosue T, Yamaguchi-Shinozaki K, Shinozaki K: **Characterization of cDNA for a dehydration-inducible gene that encodes a CLP A, B-like protein in *Arabidopsis thaliana* L.** *Biochem Biophys Res Commun* 1993, **196:**1214-1220.

54. Koizumi M, Yamaguchi-Shinozaki K, Tsuji H, Shinozaki K: **Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*.** *Gene* 1993, **129:**175-182.

55. Nylander M, Svensson J, Palva ET, Welin BV: **Stress-induced accumulation and tissue-specific localization of dehydrins in *Arabidopsis thaliana*.** *Plant Mol Biol* 2001, **45:**263-279.

56. Durrant WE, Rowland O, Piedras P, Hammond-Kosack KE, Jones JD: **cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles.** *Plant Cell* 2000, **12:**963-977.

57. Reuber TL, Ausubel FM: **Isolation of *Arabidopsis* genes that differentiate between resistance responses mediated by the *RPS2* and *RPM1* disease resistance genes.** *Plant Cell* 1996, **8:**241-249.

58. Varet A, Parker J, Tornero P, Nass N, Nurnberger T, Dangl JL, Scheel D, Lee J: **NHL25 and NHL3, two NDR1/HIN1-like genes in *Arabidopsis thaliana* with potential role(s) in plant defense.** *Mol Plant Microbe Interact* 2002, **15:**608-616.

59. Dong H, Delaney TP, Bauer DW, Beer SV: **Harpin induces disease resistance in *Arabidopsis* through the systemic acquired resistance pathway mediated by salicylic acid and the *NIM1* gene.** *Plant J* 1999, **20:**207-215.

60. del Pozo JC, Estelle M: **Function of the ubiquitin-proteosome pathway in auxin response.** *Trends Plant Sci* 1999, **4:**107-112.

61. Wei Y, Waltz DA, Rao N, Drummond RJ, Rosenberg S, Chapman HA: **Identification of the urokinase receptor as an adhesion receptor for vitronectin.** *J Biol Chem* 1994, **269:**32380-32388.

62. Azevedo C, Sadanandam A, Kitagawa K, Freialdenhoven A, Shirasu K, Schulze-Lefert P: **The RAR1 interactor SGT1, an essential component of R gene-triggered disease resistance.** *Science* 2002, **295:**2073-2076.

63. Austin MJ, Muskett P, Kahn K, Feys BJ, Jones JD, Parker JE: **Regulatory role of SGT1 in early R gene-mediated plant defenses.** *Science* 2002, **295:**2077-2080.

64. Liu Y, Schiff M, Serino G, Deng XW, Dinesh-Kumar SP: **Role of SCF ubiquitin-ligase and the COP9 signalosome in the N gene-mediated resistance response to tobacco mosaic virus.** *Plant Cell* 2002, **14:**1483-1496.

65. Shirley BW, Goodman HM: **An *Arabidopsis* gene homologous to mammalian and insect genes encoding the largest proteasome subunit.** *Mol Gen Genet* 1993, **241:**586-594.

66. Peng Z, Staub JM, Serino G, Kwok SF, Kurepa J, Bruce BD, Vierstra RD, Wei N, Deng XW: **The cellular level of PR500, a protein complex related to the 19S regulatory particle of the proteasome, is regulated in response to stresses in plants.** *Mol Biol Cell* 2001, **12:**383-392.

67. del Pozo JC, Estelle M: **F-box proteins and protein degradation: an emerging theme in cellular regulation.** *Plant Mol Biol* 2000, **44:**123-128.

68. Saurin AJ, Borden KL, Boddy MN, Freemont PS: **Does this have a familiar RING?** *Trends Biochem Sci* 1996, **21:**208-214.

69. Salinas-Mondragon RE, Garciduenas-Pina C, Guzman P: **Early elicitor induction in members of a novel multigene family coding for highly related RING-H2 proteins in *Arabidopsis thaliana*.** *Plant Mol Biol* 1999, **40:**579-590.

70. Eulgem T, Rushton PJ, Robatzek S, Somssich IE: **The WRKY superfamily of plant transcription factors.** *Trends Plant Sci* 2000, **5:**199-206.

71. Rushton PJ, Somssich IE: **Transcriptional control of plant genes responsive to pathogens.** *Curr Opin Plant Biol* 1998, **1:**311-315.

72. Robatzek S, Somssich IE: **A new member of the *Arabidopsis* WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes.** *Plant J* 2001, **28:**123-133.

73. Robatzek S, Somssich IE: **Targets of AtWRKY6 regulation during plant senescence and pathogen defense.** *Genes Dev* 2002, **16:**1139-1149.

74. Schindler U, Beckmann H, Cashmore AR: **TGA1 and G-box binding factors: two distinct classes of *Arabidopsis* leucine zipper proteins compete for the G-box-like element TGACGTGG.** *Plant Cell* 1992, **4:**1309-1319.

75. Menkens AE, Schindler U, Cashmore AR: **The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins.** *Trends Biochem Sci* 1995, **20:**506-510.

76. Manjunath S, Sachs MM: **Molecular characterization and promoter analysis of the maize cytosolic glyceraldehyde 3-phosphate dehydrogenase gene family and its expression during anoxia.** *Plant Mol Biol* 1997, **33:**97-112.

77. Kusano T, Berberich T, Harada M, Suzuki N, Sugawara K: **A maize DNA-binding factor with a bZIP motif is induced by low temperature.** *Mol Gen Genet* 1995, **248:**507-517.

78. Dolferus R, Jacobs M, Peacock WJ, Dennis ES: **Differential interactions of promoter elements in stress responses of the *Arabidopsis Adh* gene.** *Plant Physiol* 1994, **105:**1075-1087.

79. Despres C, DeLong C, Glaze S, Liu E, Fobert PR: **The *Arabidopsis* NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors.** *Plant Cell* 2000, **12:**279-290.

80. Kim JC, Lee SH, Cheong YH, Yoo CM, Lee SI, Chun HJ, Yun DJ, Hong JC, Lee SY, Lim CO, Cho MJ: **A novel cold-inducible zinc finger protein from soybean, SCOF-1, enhances cold tolerance in transgenic plants.** *Plant J* 2001, **25:**247-259.

81. Zhou JM, Trifa Y, Silva H, Pontier D, Lam E, Shah J, Klessig DF: **NPR1 differentially interacts with members of the TGA/OBF family of transcription factors that bind an element of the PR-1 gene required for induction by salicylic acid.** *Mol Plant Microbe Interact* 2000, **13:**191-202.

82. Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K: **Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression.** *Plant Cell* 1997, **9:**1859-1868.

83. Cheng SH, Sheen J, Gerrish C, Bolwell GP: **Molecular identification of phenylalanine ammonia-lyase as a substrate of a specific constitutively active *Arabidopsis* CDPK expressed in maize protoplasts.** *FEBS Lett* 2001, **503:**185-188.

84. Pei ZM, Murata Y, Benning G, Thomine S, Klusener B, Allen GJ, Grill E, Schroeder JI: **Calcium channels activated by hydrogen peroxide mediate abscisic acid signalling in guard cells.** *Nature* 2000, **406:**731-734.

85. Dempsey DA, Shah J, Klessig DF: **Salicylic acid and disease resistance in plants.** *Crit Rev Plant Sci* 1999, **18:**547-575.

86. Grotewold E, Drummond BJ, Bowen B, Peterson T: **The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset.** *Cell* 1994, **76:**543-553.

87. Solano R, Nieto C, Avila J, Canas L, Diaz I, Paz-Ares J: **Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from *Petunia hybrida*.** *EMBO J* 1995, **14:**1773-1784.

88. Sablowski RW, Moyano E, Culianez-Macia FA, Schuch W, Martin C, Bevan M: **A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes.** *EMBO J* 1994, **13:**128-137.

89. Sablowski RW, Baulcombe DC, Bevan M: **Expression of a flower-specific Myb protein in leaf cells using a viral vector causes ectopic activation of a target promoter.** *Proc Natl Acad Sci USA* 1995, **92:**6901-6905.

90. Faktor O, Kooter JM, Loake GJ, Dixon DP, Lamb C: **Differential utilization of regulatory *cis*-elements for stress induced and tissue-specific activity of a French bean chalcone synthase promoter.** *Plant Sci* 1997, **124:**175-182.

91. Faktor O, Loake GJ, Dixon DP, Lamb C: **The G-box and H-box in a 39 bp region of a French bean chalcone synthase promoter constitutes a tissue-specific regulatory element.** *Plant J* 1997, **11:**1105-1113.

92. Singh KB, Foley RC, Onate-Sanchez L: **Transcription factors in plant defense and stress responses.** *Curr Opin Plant Biol* 2002, **5:**430-436.

93. **TAIR Blast 2.0** [http://www.arabidopsis.org/Blast/]

94. Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual* 2nd edn. New York: Cold Spring Harbor Laboratory Press; 1989.

95. Shinozaki K, Yamaguchi-Shinozaki K: **Molecular responses to dehydration and low temperature: Differences and cross-talk between two stress signalling pathways.** *Curr Opin Plant Biol* 2000, **3:**217-223.

96. Lebel E, Heifetz P, Thorne L, Uknes S, Ryals J, Ward E: **Functional analysis of regulatory sequences controlling *PR-1* gene expression in *Arabidopsis*.** *Plant J* 1998, **16:**223-233.

97. Hao D, Ohme-Takagi M, Sarai A: **Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant.** *J Biol Chem* 1998, **273:**26857-26861.

98. Solano R, Stepanova A, Chao Q, Ecker JR: **Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1.** *Genes Dev* 1998, **12:**3703-3714.

99. Urao T, Katagiri T, Mizoguchi T, Yamaguchi-Shinozaki K, Hayashida N, Shinozaki K: **Two genes that encode Ca(2+)-dependent protein kinases are induced by drought and high-salt stresses in *Arabidopsis thaliana*.** *Mol Gen Genet* 1994, **244:**331-340.

100. da Costa e Silva O, Klein L, Schmelzer E, Trezzini GF, Hahlbrock K: **BPF-1, a pathogen-induced DNA-binding protein involved in the plant defense response.** *Plant J* 1993, **4:**125-135.

101. Martin C, Paz-Ares J: **MYB transcription factors in plants.** *Trends Genet* 1997, **13:**67-73.

102. Sorger PK, Pelham HR: **Purification and characterization of a heat-shock element binding protein from yeast.** *EMBO J* 1987, **6:**3035-3041.

103. **TAIR blast datasets** [ftp://ftp.arabidopsis.org/Sequences/blast_datasets/]

104. ***Arabidopsis* microarray database at Penn State** [http://arastressdb.cac.psu.edu/]