The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance

Klaus Maleck[1], Aaron Levine[2], Thomas Eulgem[2], Allen Morgan[1], Jürg Schmid[1], Kay A. Lawton[1], Jeffery L. Dangl[2] & Robert A. Dietrich[1]

Infected plants undergo transcriptional reprogramming during initiation of both local defence and systemic acquired resistance (SAR). We monitored gene-expression changes in *Arabidopsis thaliana* under 14 different SAR-inducing or SAR-repressing conditions using a DNA microarray representing approximately 25–30% of all *A. thaliana* genes. We derived groups of genes with common regulation patterns, or regulons. The regulon containing *PR-1*, a reliable marker gene for SAR in *A. thaliana*, contains known *PR* genes and novel genes likely to function during SAR and disease resistance. We identified a common promoter element in genes of this regulon that binds members of a plant-specific transcription factor family. Our results extend expression profiling to definition of regulatory networks and gene discovery in plants.

Introduction

Physiological flexibility is a prerequisite for life in a changing environment and requires elaborate regulatory mechanisms to simultaneously alter the expression states of groups of genes. In prokaryotes, such regulons are often physically grouped into operons. In eukaryotes, genes coregulated under specific conditions can be dispersed throughout the genome. Commonly, a limited set of key regulators, used combinatorially, is used to trigger defined transcriptional responses throughout an entire regulon.

On recognition of avirulent pathogens, plants activate a set of transient local defence measures restricting pathogen growth (incompatible interactions)[1]. Infection with virulent pathogens that are not specifically recognized results in plant disease (compatible interactions). In addition to local defence responses, recognition of avirulent pathogens can trigger a long-lasting, broad-spectrum disease resistance throughout the entire plant, referred to as systemic acquired resistance[2] (SAR). SAR is accompanied in uninfected 'secondary' tissues by activation of a set of 'pathogenesis-related' (*PR*) marker genes. Some of these are also activated locally around infection sites. Thus, the 'pre-induction' of a suite of proteins in systemic tissue is likely to confer broad-spectrum SAR (refs 2,3).

A. thaliana and tobacco require salicylic acid accumulation to establish SAR (refs 3,4). Repression of salicylic acid accumulation by expression of a bacterial salicylic acid hydroxylase gene (*NahG*) abolishes SAR (ref. 5). Salicylic acid accumulation can be mimicked by chemical analogues including benzothiadiazol[6] (BTH) or 2,6-dichloroisonicotinic acid[7] (INA) and is stimulated in *A. thaliana* constitutive immunity (*cim*) or constitutive expression of *PR* genes (*cpr*) mutants[8,9] (K.M., unpublished data). In addition, genetic analyses identified regulators required to mediate SAR, such as NIM1/NPR-1 (refs 10,11), which appears to interact directly with transcription factors[3,12–15] (R.A.D., unpublished data). Salicylic acid accumulation functions upstream of *NIM1/NPR-1* (refs 3,11).

The transcriptional consequences of pathogen infection and SAR-inducing conditions have been tested on a limited set of marker genes in a limited number of mutant backgrounds[2,16]. We analysed concerted transcriptional reprogramming in *A. thaliana* during the SAR response using cDNA microarrays. Our results extend the fundamental roles of salicylic acid accumulation and NIM1/NPR-1 in the control of the *A. thaliana* SAR transcriptome, but suggest that this key regulator cooperates with several types of transcription factors in the execution of the SAR transcriptional program.

Results

Related SAR conditions result in similar gene-activity patterns

To examine gene-activity changes associated with the induction or maintenance of SAR in *A. thaliana*, we performed large-scale expression profiling using cDNA arrays[17,18]. We used poly(A)$^+$ mRNA isolated from age- and environment-matched plants using 14 different treatments that either induce or repress SAR (Table 1) in mixed hybridizations with mRNA from genetically matched, untreated control plants on a DNA microarray with 10,000 expressed sequence tags (ESTs). This represents approximately 7,000 or 25–30% of all *A. thaliana* genes due to redundancy in the EST set. We included two control treatments that alter plant metabolism, but are unrelated to SAR (Table 1, treatments 1 and 2). One sample was taken during SAR induction (for example, 4 h after BTH treatment). Most samples were derived from the SAR-maintenance phase (48 h after BTH treatment, *cim* mutants, or uninfected secondary tissue 44 h after infection of lower, primary leaves with avirulent *Pseudomonas syringae* bacteria). Thus, all but one of the time points for RNA collection correspond to fully induced SAR (refs 6,10,11). We required that ESTs included for analysis show differential expression equal to or greater than 2.5-fold in at least two SAR-relevant samples.

[1]*Syngenta, Research Triangle Park, North Carolina, USA.* [2]*Department of Biology and Curriculum in Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. Correspondence should be addressed to J.L.D. (e-mail: dangl@email.unc.edu).*

We found 413 ESTs to meet these criteria. A comparison of the expression changes across these ESTs for each treatment, based on the calculation of distance matrices, is shown (Fig. 1). Related experimental conditions result in similar gene-activity patterns (Fig. 1, red, purple and green). For example, all three tested SAR-constitutive *cim* mutants have gene-expression patterns similar to those caused by BTH during SAR maintenance. These sets of related expression changes are distinct from local and early defence responses. For example, expression patterns from plants inoculated 48 hours earlier with either incompatible or compatible isolates of the oomycete parasite *Peronospora parasitica* resemble the expression patterns of early responses to BTH. These three treatments are also related to the biologically induced SAR response in uninfected secondary tissue (Fig. 1, Pst avr, 2º), suggesting that biologically induced SAR is transcriptionally related to an 'early' phase of chemical induction of SAR. Overexpression of *NIM1/NPR-1*, which primes the SAR response but does not result in constitutive *PR* gene expression[19] (L. Friedrich, unpublished data), yields a distinct profile.

The expression of 86 ESTs is altered in untreated *NahG* plants, defining a regulatory role for basal salicylic acid levels or for the *NahG* product, catechol. Moreover, the expression patterns of *cim6* or *cim11* mutants are altered in a *NahG* background (Fig. 1, green).

This is consistent with the requirement for salicylic acid accumulation in the constitutive SAR phenotype of these mutants. We also confirmed a fundamental role for *NIM1/NPR-1* in pathogen-induced SAR gene activation. The transcriptional profile in SAR-activated secondary tissue from wild-type plants (Fig. 1, Pst avr, 2º) is distinct from the *nim1.4* mutant after the same treatment (Fig. 1, *nim1* Pst avr, 2º). Because the former treatment represents a classically defined SAR induction, and the latter, a mutant that disrupts SAR, the difference between these two is the SAR gene set.

Cluster analysis defines groups of coregulated genes

We applied two commonly used algorithms to analyse gene expression profiles among the 413 SAR-associated ESTs. A hierarchical 'clustergram' of genes was grouped by both related regulation patterns and expression amplitudes[20] (Fig. 2). 'Self-organizing maps' (SOMs) generate expression profiles organized by shape, essentially independent of amplitude[21] (Fig. 3). Enhanced clustergrams and enhanced SOMs listing all ESTs are available (Figs A and B, see http://genetics.nature.com/supplementary_info/). We found that genes assigned to key SOM clusters (Fig. 3) are also found in close proximity in the hierarchical clustergram, demonstrating that both algorithms generate consistent overall patterns. This is noted for SOMs c1 and c3 (Fig. 2a), where SOM c1 (Fig. 3) corresponds in large part to the hierarchical cluster node highlighted in red in Fig. 2a (38/45 ESTs from SOM c1 present, marked with asterisks in Fig. 2b). Similarly, 45 of 55 ESTs in SOM c3 (Fig. 3) group closely together in the clustergram (blue in Fig. 2a, asterisks in Fig. 2b).

Because *PR-1* activation is a reasonably robust molecular marker for SAR (ref. 22), we analysed the cluster including this gene in detail (Figs 2b, top, and 3, SOM c1). The *PR-1* regulon in SOM c1 contains 45 ESTs (from a maximum of 31 different genes). These are candidate SAR markers and the encoded proteins likely function during SAR. Several genes that clustered with *PR-1* (*PR-4*, *GST* (glutathione-S-transferase) and *PerC* (peroxidase C)) were represented on the arrays by multiple spots. Their nearly invariant clustering in the *PR-1* regulon demonstrates the internal consistency of our analysis. *PR-1* regulon members show a unique overall expression profile. They are strongly activated in secondary SAR tissue (treatment 14). These expression changes in SAR tissue are strictly dependent on *NIM1/NPR-1* (compare treatments 14 and 15). Strong activation was also observed following infection with either compatible or incompatible isolates of *P. parasitica* (treatments 12 and 16).

Table 1 • Diversity of conditions used to describe the transcriptome of *Arabidopsis* during SAR

	Sample (compared with untreated control)	Ecotype	Comments
1	Adenylsuccinate synthetase antisense	Col-0	reference inductions
2	Hydantocitidin (0.9mM)	Col-0	irrelevant to SAR
3	*cim6*	Col-0	constitutive SAR
4	*cim7*	Col-0	mutants
5	*cim11*	Col-0	
6	*Nim1 overexpresser*	Ws-0	primed SAR response
7	*NahG*	Col-0	no SA accumulation,
8	*cim6* NahG	Col-0	no SAR response
9	*cim11* NahG	Col-0	
10	BTH 4h	WS-0	SAR induction
11	BTH 48h	Col-0	SAR maitenance
12	*Peronospora parasitica* Emwa1 48h	Ws-0	compatible interaction (local)
13	*Pst* DC3000 *avrRpt2* 1º tissue 44h	Ws-0	incompatible interaction (local)
14	*Pst* DC3000 *avrRpt2* 2º tissue 44h	Ws-0	incompatible interaction (systemic)
15	*nim1+ Pst* DC3000 *avrRpt2* 2º tissue 44h	Ws-0	incompatible interaction, no SAR
16	*Peronospora parasitica* Noco2 48h	Ws-0	incompatible interaction (local)
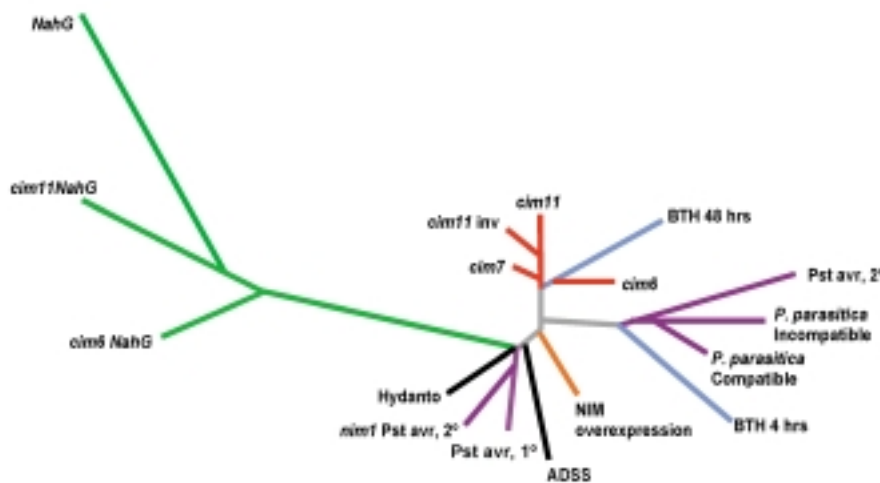
Fig. 1 Unrooted tree of relatedness of transcriptional changes across SAR-relevant conditions. The 413 ESTs that displayed a minimum of 2.5-fold differential gene expression under at least 2 of 14 SAR-relevant conditions were used to calculate similarities in gene-expression profiles. Differences in the overall gene expression patterns between the tested conditions are represented by the length of the branches within the depicted tree.

Moreover, these genes are moderately activated in the tested *cim* mutants in a salicylic-acid–dependent manner (compare treatments 3, 4 and 5 with treatments 8 and 9), and they are upregulated by BTH treatment.

Other nodes in the clustergram (Fig. 2) define regulons with different properties. For example, the node containing SOM c3 is composed of ESTs expressed strongly after infection with the incompatible *P. parasitica* isolate as well as after four hours of BTH treatment. None of the SAR-inducing conditions caused a marked activation of these genes, suggesting that they are unlikely to have a role in SAR. The rapid activation of these genes by BTH, but their consequent repression during both SAR maintenance by 48 hours of BTH and in the *cim* mutants further supports this idea. These genes may, however, function to mediate resistance to *P. parasitica* infection. Among the 55 ESTs of SOM c3, 2 senescence-associated genes, *sen1* and *sen5*, are strongly represented (11 ESTs of *sen1* and 5 ESTs of *sen5*; refs 23,24). This constitutes an additional, unbiased test of experimental reliability and reinforces emerging regulatory connections between the defence response and senescence[25]. BLAST searches revealed that the 55 ESTs of SOM c3 represent at most 32 genes, 17 of which show no significant similarity to known genes. Thus, our data set has uncovered regulons enriched in novel genes for future reverse-genetic experiments.

W boxes are highly enriched in *PR-1* regulon promoters

The coregulation of EST groups may reflect transcriptional control mechanisms shared across the different SAR-inducing treatments analysed and may be based on conserved promoter elements. The *PR-1* regulon provides a model for this analysis. Functional INA-responsive *cis* elements were previously defined on the *PR-1* promoter[7]. We searched the 26 available promoters from SOM c1 for known *cis* elements (Table 2) including those previously described[7]. We inspected 1.1 kb upstream of the predicted translation start site (ATG), because most *A. thaliana* promoters use *cis* elements located within the first 1 kb and most *A. thaliana* 5´ untranslated sequences are less than 150 bp (S.Y. Rhee, pers. comm.).

We found only the binding site for WRKY transcription factors (W boxes; TTGAC; ref. 26) in all 26 promoters. WRKY proteins
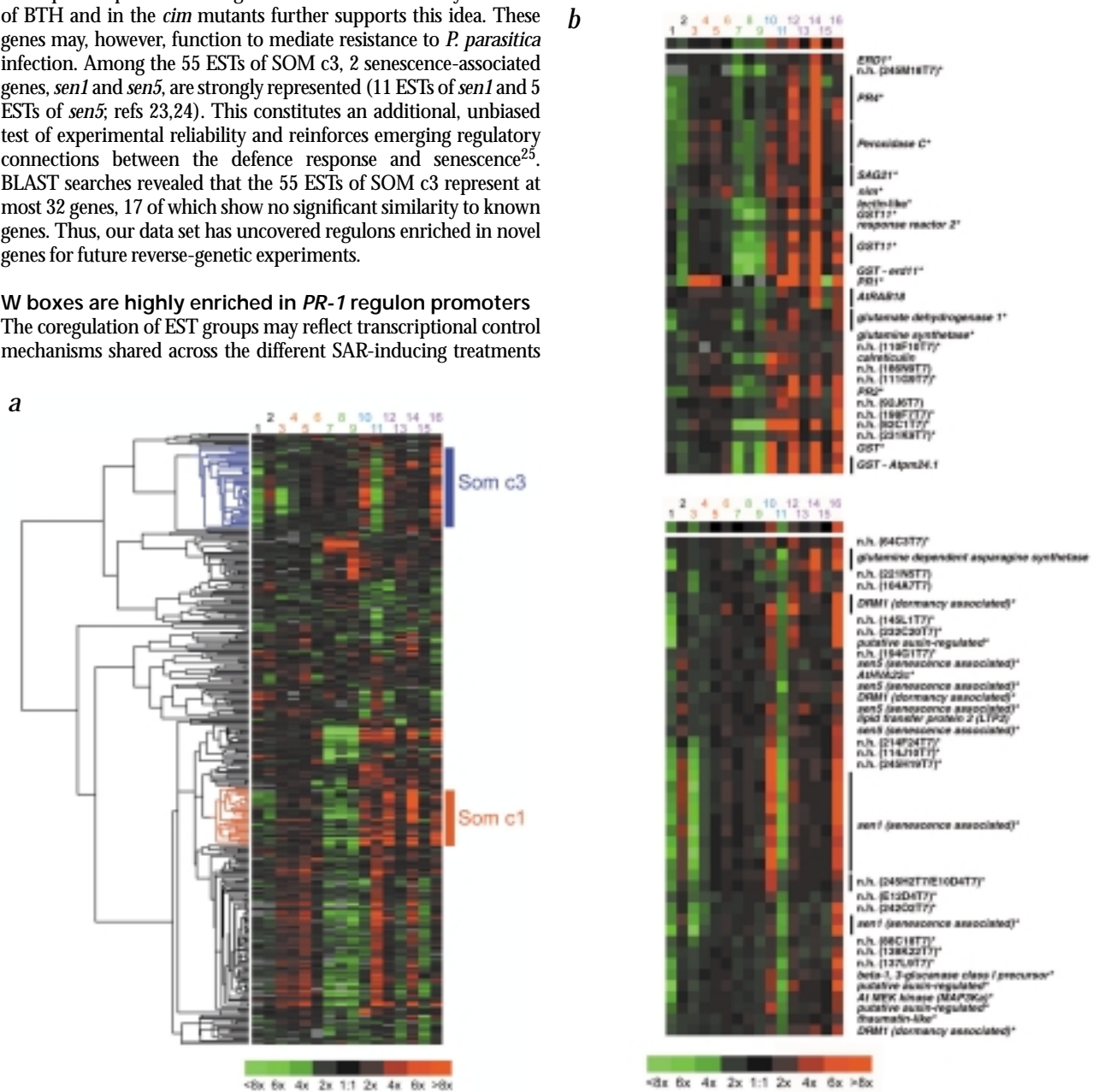
Fig. 2 *a*, Clustergram of 413 ESTs differentially expressed during SAR. Colour-coded, numbered columns correspond to experimental treatments in Table 1. Gene expression profiles are in rows: red, transcriptional activation; green, repression compared with mock-treated, ecotype-matched controls (scale at bottom); grey, missing or incomplete data. The cluster tree (left) illustrates the nodes of coregulation of gene expression over all 16 treatments. Vertical bars on right indicate the nodes of the cluster tree and the included ESTs, which are largely contained in SOMs c1 (red) and c3 (blue) in Fig. 3. *b*, Enlargement of test clusters for disease resistance and SAR. The node of the clustergram in (*a*) which largely corresponds to the *PR-1* regulon in SOM c1 (top) and the node largely corresponding to SOM c3 (bottom). Asterisks denote ESTs also found in the corresponding SOMs in Fig. 3. The average profile for each is represented by the first row below the treatment designations. Gene names are highest BLASTn hits. The usual cutoff was 100. In the cases in which only a very short sequence was almost identical to a sequence in the database, lower BLASTn scores were accepted (marked by an asterisk next to the BLASTn score in Table A, see http://genetics.nature.com/supplementary_info/).
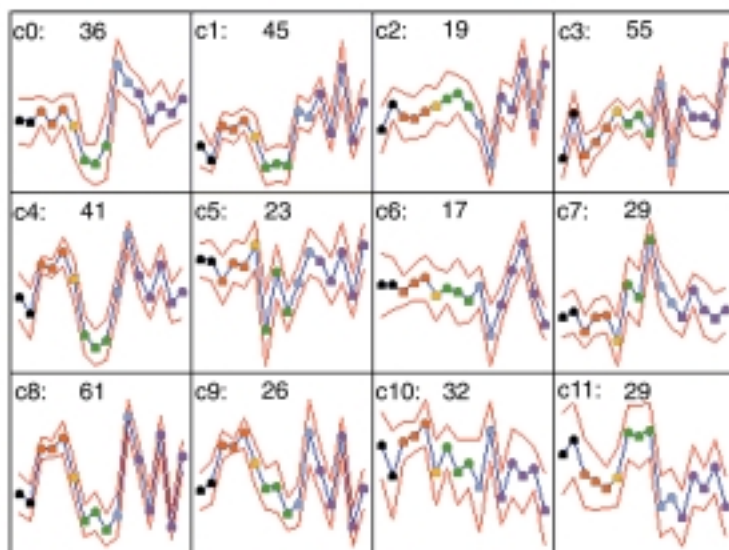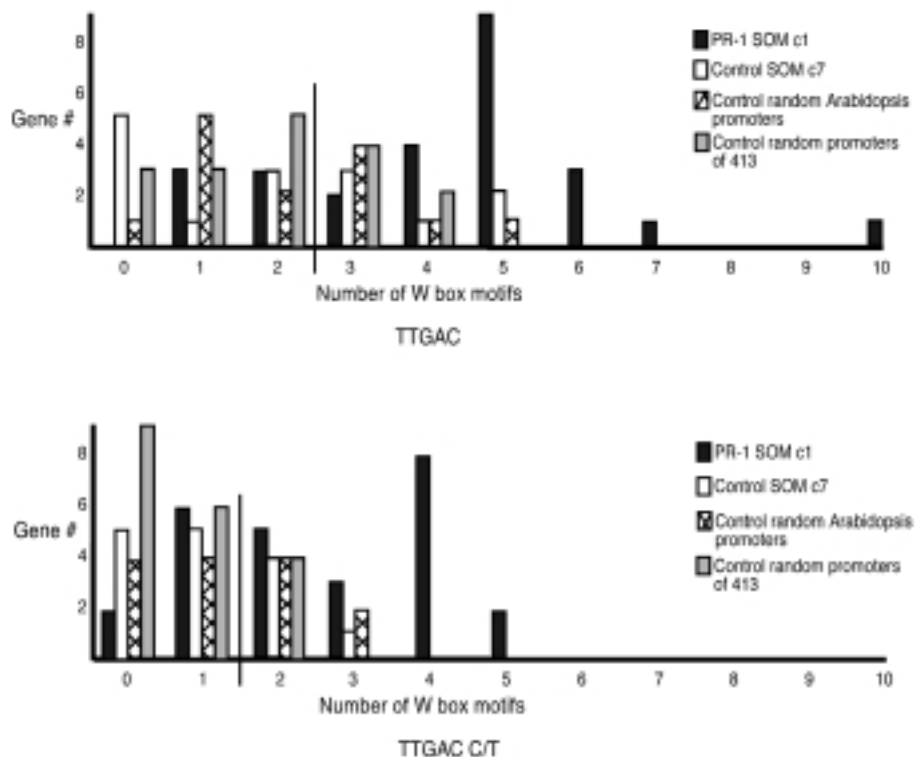
Fig. 3 Self-organizing map clusters of expression profiles. Each graph displays the mean pattern of expression (colour-coded points as in Table 1) of the ESTs in that cluster and the standard deviation of average expression (red lines). The number of ESTs in each cluster is at the top centre of each SOM. The *y* axis represents normalized gene expression levels (normalized generally to mean 0, s.d. 1).

bind to sequences with an invariant TGAC core, which is in most cases preceded by a T. This TTGAC pentamer appears 111 times on both strands in 26 *PR-1* regulon promoters, an average of 4.3 copies per promoter (Fig. 4). The statistical expectation for a randomly distributed pentamer is 2.1 copies per promoter. A control scan with all possible permutations of the TTGAC motif gave an average of 1.6±0.4 per promoter. This enrichment of W boxes in *PR-1* regulon promoters is highly significant compared with the distribution of W boxes in three control promoters sets (15 genes from SOM c7, 17 random genes selected from the 413 analysed ESTs and 14 random *A. thaliana* genes not included in our 413 ESTs). We generated a theoretical distribution of W box copies on any 1.1-kb promoter (Fig. C, see http://genetics.nature.com/supplementary_info/). We used this to describe both the number of promoters in each set expected (*E*) to have a given number of W boxes and *P* values, expressing the probability that the observed W box distributions occur by chance. This confirmed the significance of W box enrichment within *PR1* regulon promoters (Table 3).

Transcription factors of the TGA subgroup of bZIP proteins (TGA-bZIPs) are positive regulators of INA-induced *PR-1* expression. They recognize a pentameric element, TGACG, that potentially overlaps the W box consensus[27] (Table 2). The TGAC W box core, however, is followed by C or T in all available reports of WRKY/W box interactions[26]. Thus, a distinction between binding sites for these two regulatory factor families can be predicted. In fact,

only 12 of 111 TTGAC motifs found within the 26 available *PR-1* regulon promoters are followed by a G. Moreover, the minimum TGA recognition motif, TGACG, is lacking in 9 of the 26 *PR-1* regulon promoters. This site occurs in the 26 *PR-1* regulon promoters only 24 times, or an average of once per promoter and less than 50% of the statistical expectation. Although a single motif might be sufficient to mediate a regulatory event on any promoter, it is unlikely that the TGA-bZIP factors constitute a common regulator of *PR-1* regulon genes across the conditions analysed.

In contrast, using the most stringent definition for WRKY binding sites (a TTGACC/T hexamer[26]), we still detected a significant over-representation of 68 potential binding sites, an average of 2.6 copies per promoter, in the *PR-1* regulon (Fig. 4). We note that the *P* value for distribution of W-box hexamers is also highly significant (Table 3). Only 2 of 26 *PR-1* regulon promoters do not contain a copy of the TTGACC/T motif. One of these promoters, however, contains the W-box–like hexamer TTGACA. This motif is found 31 times within the 26 *PR-1* regulon promoters, which is twice the statistically expected frequency. Clustered occurrences of W boxes are common among promoters of WRKY-regulated

Fig. 4 W box motifs in *PR-1* cluster genes. Promoters for each gene represented by an EST in SOM c1 (the *PR-1* cluster) and three control sets of promoters (15 genes from SOM c7, 17 random genes selected from the 413 analysed ESTs and 14 random *A. thaliana* genes not included in our 413 ESTs) were searched for W boxes using either the core WRKY binding consensus (TTGAC; top) or a more stringent consensus (TTGACC/T; bottom). The number of genes found (*y* axis) versus the number of W boxes found in a given gene's promoter (*x* axis) is displayed. Thin vertical bar differentiates random W box occurrences (left) from non-random (right).
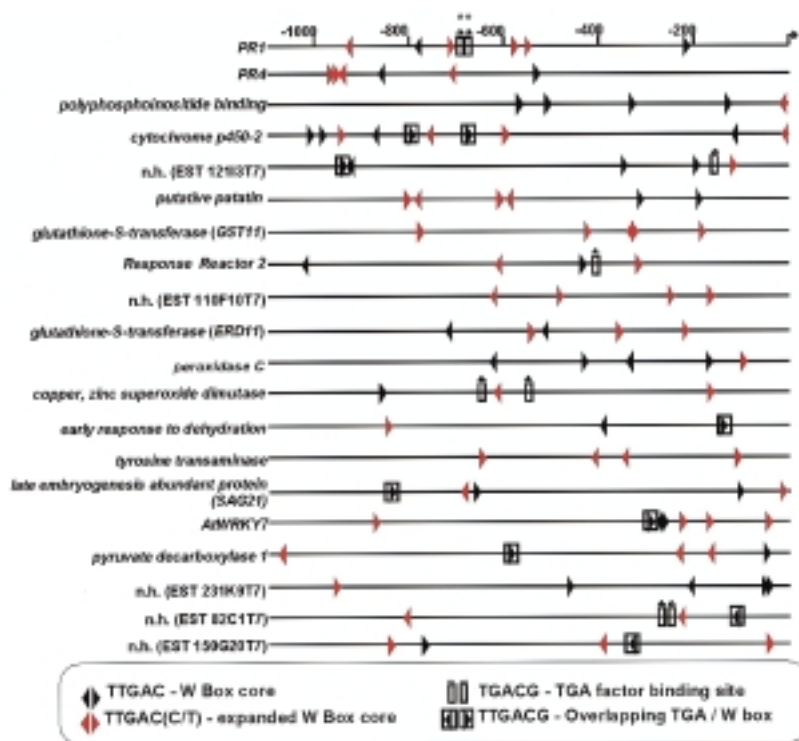
Fig. 5 *Cis* element architecture in the *PR-1* cluster. Shown is 1.1 kb of promoter sequence for the 20 genes in the *PR-1* cluster with sequenced promoter regions containing at least 3 W boxes. Black triangles, W box core (TTGAC); red triangles, the stringent W box consensus (TTGACC/T); black boxes, TGA-bZIP binding sites (TGACG); forward-facing arrows and triangles, motifs on the forward strand pointing toward the translational start site. An asterisk indicates that the motif has been shown to function as a binding site in a previous study. Numbering is from presumptive translation start codons, not transcriptional starts, most of which are unmapped. Hence the W box mapped on the *PR-1* promoter at –676 (top strand) appears here at –712.

results have been reported[31]: cDNA-AFLP analysis revealed that *Cf-9*–triggered defence responses in cultured cells affected the expression of approximately 1% of all tobacco genes.

Although we did not repeat individual microarray hybridizations, the high similarity of overall gene expression changes among related treatments argues for the reliability of our data. Similarly, the reproducibility of expression data for genes represented by multiple ESTs on the arrays confirms a low variability among individual ratio measurements. To further confirm the quality of our experiments, we re-tested 40 data pairs by RNA blots and found good correlation to the microarray data. Two different *A. thaliana* ecotypes (Ws-0 and Col-0) were used here and the transcriptional responses to the various treatments may in each case be influenced partly by the specific genetic background. Yet the profound differences in overall gene expression we observed are unlikely to be based significantly on ecotype differences.

Gene identity can suggest metabolic processes induced along with SAR. For example, several *PR-1* regulon genes encode proteins involved in redox regulation. Reactive oxygen intermediates (ROI) act as local secondary messengers upstream of salicylic acid accumulation[32,33] and are implicated in defence responses, as direct cellular and microbial toxins, or in cell-wall cross-linking and other oxidative processes[34]. Of the several *A. thaliana* peroxidase genes, *PerC* appears to be specifically activated under plant defence conditions, as are some *GST* genes. This is consistent with the previous observation that both *PerC* and *GST* are transcriptionally activated in a superoxide-dependent manner in the *A. thaliana* mutant *lsd1*, which lacks the ability to halt hypersensitive-response-like lesions[35,36]. We note that nine genes of the *PR-1* regulon encode proteins of unknown function. Strict coregulation of genes encod-

genes. Association of these *cis* elements as palindromes or direct repeats with spacing of less than 20 bp results in binding cooperativity and synergistic transcriptional activation[26,28–30]. We found both the core and stringent W-box elements (TTGAC and TTGACC/T) arranged in a palindromic or directly repeated manner on many promoters from the *PR-1* regulon (Fig. 5; and Table B, see http://genetics.nature.com/supplementary_info/).

Discussion

We visualized gene-expression states over approximately 25–30% of the *A. thaliana* genome under various SAR-related conditions using cDNA arrays. We focused on identification of coregulated genes, as opposed to identification of genes activated during single SAR or defence-response conditions, and thus set conservative criteria for inclusion into our data set of 2.5-fold transcriptional change in 2 or more of 14 relevant conditions. In fact, most ESTs are transcriptionally regulated above 2.5-fold in several treatments (213 were differentially expressed in 4 or more conditions). Our selection for events occurring at least twice identified genes whose regulation is generally altered across this set of conditions. Because lower fold changes in transcript abundance may still have biological relevance, we probably underestimated the total number of biologically relevant genes for defence and SAR. We also will have missed regulatory events unique to any single treatment. Our inclusion criteria, however, should minimize inclusion of false positives and maximize the identification of coregulated groups of genes broadly involved in SAR and response to infection. We identified approximately 300 genes using these criteria, or approximately 1.5% of the total *A. thaliana* gene set. Similar

Table 2• Frequencies of potential binding motifs recognized by defence-associated transcription factors in *PR-1* regulon promoters

Factor type	Binding site[a]	Average frequency in *PR1* regulon promoters[b]	Statistically expected frequency per promoter[c]
WRKY[26]	TTGAC	4.3	2.1
WRKY[26]	TTGACC/T	2.6	1.1
TGA (bZIP)[27]	TGACG	0.9	2.1
GBF (bZIP)[44]	CACGTG	0.2	0.5
G/HBF-1 (bZIP)[45]	CCTACC	0.1	0.5
ERF (AP2-like)[46]	GCCGCC	0.2	0.5
ORCA (AP2-like)[47]	ACCGCC	0.1	0.5
EIN3/EIL[48]	GGATGTA	0.1	0.1
Mybl[49]	GG/TTA/TGG/TT	1.2	1.1

Core motifs are displayed for the defined factor binding sites that were considered in our search. [a]Core motifs of factor binding sites. [b]1.1 kb upstream of translation start was examined. [c]Within 1.1 kb of promoter.

Table 3 • Distribution of W boxes

TTGAC

	Total	n^a=2 or more			n^a=3 or more			n^a=4 or more		
		N^b	E^c	p^d	N^b	E^c	p^d	N^b	E^c	p^d
SOMc1[e]	26	23	16.5	3.46e−3	20	9.5	2.48e−5	18	4.4	5.20e−9
SOMc7[e]	15	9	9.5	0.20	6	5.5	0.20	3	2.6	0.24
Random (s)[f]	17	11	10.8	0.20	6	6.2	0.20	2	2.9	0.24
Random (g)[g]	14	8	8.9	0.19	6	5.1	0.19	2	2.4	0.28

TTGACC/T

	Total	n^a=2 or more			n^a=3 or more			n^a=4 or more		
		N^b	E^c	p^d	N^b	E^c	p^d	N^b	E^c	p^d
SOMc1[e]	26	18	7.6	2.29e−5	13	2.5	1.35e−7	10	0.6	2.15e−10
SOMc7[e]	15	5	4.4	0.20	1	1.4	0.35	0	0.4	0.70
Random (s)[f]	17	2	5.0	0.07	0	1.6	0.19	0	0.4	0.66
Random (g)[g]	14	6	4.1	0.12	2	1.3	0.25	0	0.3	0.71

Distribution of W box motifs, TTGAC or TTGACC/T, is shown in promoters of SOMc1, SOMc7, randomly chosen genes from the set of 413 analysed ESTs and randomly chosen *A. thaliana* genes not included in the set of 413 ESTs. [a]W boxes per promoter. [b]Number of promoters within the specified set that have n W boxes. [c]Number of promoters that would be expected to have n W boxes by chance. [d]Probability of seeing the observed number of promoters with n W boxes by chance. [e]Refers to the promoters of genes in SOM c1 or c7 of Fig. 3. [f]Refers to the promoters of genes selected randomly from the set of 413 ESTs (not including the ESTs in SOM c1 or c7). [g]Refers to promoters of genes selected randomly from *A. thaliana* genomic sequence.

ing enzymes catalysing successive steps in plant secondary metabolism has been demonstrated[37] and gene-expression profiling data have been successfully used to infer gene function in yeast[38]. Thus, coregulation of these novel genes with genes of known biochemical function during SAR will focus future studies aimed at functional characterization of novel SAR-associated proteins.

We found that SAR-associated gene expression changes overlapped with those observed during a compatible interaction. The reason for this paradox may be that the host initiates its defence while the pathogen is spreading during compatible interactions. This initiation of defence responses is ultimately too slow to stop disease progression. By contrast, a potential pathogen infecting SAR-induced tissue is immediately confronted with a pre-activated defence system hindering its spread in the plant. Nonetheless, SAR-associated genes activated locally during compatible interactions may limit pathogen spread, as suggested by the existence of *eds* mutants which show enhanced susceptibility to virulent pathogens and are partly compromised in SAR-induction[39].

The SAR-associated expression changes we characterized were either strongly attenuated or reversed in the *nim1-4* mutant. Previous investigations established a functional link between NIM1/NPR-1 and TGA-bZIP transcription factors. Direct interactions between NIM1/NPR-1 and TGA-bZIPs were demonstrated[12,14,15]. A TGA-bZIP factor binding site, LS7 (TGACGT), within the *PR-1* promoter was shown to be important for INA activation of *PR-1* (ref. 7). A second TGA-bZIP factor binding site, directly upstream of the LS7 element, LS5, might function as a weak silencer element[7]. *In vitro* binding of one TGA-bZIP family member, TGA2, to both LS5 and LS7 was weakly enhanced in the presence of NIM1/NPR-1 (ref. 12). We did identify TGA-bZIP factor sites in 17 of 26 *PR-1* regulon promoters, but their overall frequency is not enriched above that expected at random. We also asked whether the 17 of 26 TGA-bZIP occurrences was higher in the *PR-1* regulon than in the sets of control promoters described above, and it is not (data not shown). Hence, although it is clear that TGA-bZIP factors can regulate *PR-1* expression under the limited set of inductive conditions analysed so far, they are probably not the common factor that interacts with NIM/NPR to broadly regulate SAR gene expression under a diversity of SAR-inducing conditions.

Our study identifies a second family of transcription factors as the probable common regulators of genes in the *PR-1* regulon across the treatments analysed. The significant over-representation of W-box motifs, and their clustering, on *PR-1* regulon gene promoters suggests that WRKY factors are crucial in coregulation of these genes. WRKY factors constitute a large group of plant-specific transcriptional regulators implicated in pathogen and stress responses[26]. Their cognate W box was originally described as a positively acting *cis* element[30], but one of the potential WRKY binding sites in the *PR-1* promoter is a negative *cis* element for INA response[7] (LS4). Thus, one explanation for our data is that *PR-1* regulon genes are commonly repressed by WRKY proteins, and that removal of this repression is a common step in a variety of plant defence conditions. This scenario is further supported by the recent discovery of *sni1*, a suppressor of *nim1/npr1* (ref. 13). SNI1 acts as a negative regulator of *PR* gene expression, is localized to the nucleus and has no obvious DNA-binding domain. One possible mode of its action may involve interactions with WRKY transcription factors bound to negatively acting W-box elements. This model is also consistent with de-repression of basal defence responses in the *A. thaliana lsd1* mutant[36] and the barley *mlo* mutant[40].

The strict *NIM1/NPR-1* dependence of SAR-associated transcriptional activation throughout the *PR-1* regulon can be explained by a functional linkage between NIM1/NPR-1 and WRKY factors. NIM1/NPR1 may mediate WRKY-dependent de-repression of *PR-1* regulon genes. Alternatively, it may drive immediate early expression of a limited set of WRKY proteins that subsequently regulate a broader set of WRKY-dependent SAR target genes. WRKY protein binding to a tight palindromic arrangement of W boxes leads to rapid activation of parsley *WRKY1*, a possible regulator of *PR* genes[28]. In *A. thaliana*, the *PR-1* regulon in SOM c1 contains *WRKY7* (ref. 26), whose promoter contains tightly clustered W box elements similar to the parsley *WRKY1* promoter (Fig. 5). *A. thaliana WRKY7* may thus be a primary, NIM1/NPR-1 and WRKY-dependent target of SAR signalling, encoding a regulator of secondary SAR target genes in the *PR-1* regulon.

Finally, we note that the W-box–like hexamer TTGACA has not been described as a binding site of WRKY proteins. Its frequent occurrence among *PR-1* regulon promoters, however, suggests functional relevance for this motif in SAR gene regulation. A recent comparative structural analysis of the WRKY superfamily showed that family members can be assigned to distinct groups and subgroups[26]. Representatives of most of these (sub)groups can recognize the conventional W-box motif TTGACC/T. One possibility is that atypical WRKY proteins may participate in SAR gene regulation by interacting with the TTGACA hexamer.

We have described the first map of the plant defence transcriptome during SAR in *A. thaliana*. Complex signalling networks are best defined using multiple biologically relevant conditions, as opposed to traditional pair-wise comparisons. Subsequent analyses of additional mutants and infection conditions, using the soon to be completed *A. thaliana* genomic sequence, will both enlarge the complexity and refine the resolution of the *A. thaliana* innate immunity transcriptome. These will serve as a springboard from which both diagnostic and disease control tools can be designed in crops.

Methods

Plant growth and RNA sample preparation. Age-matched *A. thaliana* plants were all grown under the same conditions (21 °C during the day, 17.5 °C at night, 60% relative humidity, 9-h day length, 15-h night). All treatments, infections and preparation of RNA samples were performed on leaves from age-matched (4 weeks), non-bolting plants at the same time of day. RNA was prepared in parallel by two people. Hence, the results may potentially be slightly influenced by preparation of these samples. Poly(A) RNA was isolated using the PolyAttract System II (Promega).

cDNA microarrays. We PCR-amplified 10,000 ESTs from the Arabidopsis Biological Resource Center (ABRC; Ohio State University) using M13 universal reverse and –21 forward primer, with modified 5′ amino end for spotting onto glass slides. All 413 ESTs included in our data set were re-sequenced. Roughly 20% of the EST sequences did not correspond to the original sequence as represented in the AATDB. An Excel file containing all EST names, BLAST matches and fluorescence ratio data is available (Table A, see http://genetics.nature.com/supplementary_info/). Redundancy in the 10,000 ESTs is based on our finding that 215 of 413 ESTs had significant homology to known genes or proteins. Discounting the 27 deliberately spotted ESTs (representing 12 genes), this leaves 188 ESTs that are identical to 138 distinct genes. Using this as an estimate of redundancy, we expect that the 198 ESTs showing no homology will represent ~145 genes. Thus, our 413 ESTs should represent ~300 genes. Further extrapolation suggests that our 10,000 EST arrays represent around 7,000 genes, and if the estimate of 25,000 *A. thaliana* genes is correct, then we sample 25–30% of the total. This approximation is consistent with other estimations of redundancy in this EST set[41].

Array preparation and hybridizations were performed by Synteni (now Incyte) as described[42]. Technical controls are available (http://www.genomesystems.com/support/gem/controls/gsicontrolplate.html), as well as fluorescence ratio data for these ESTs (Table A, see http://genetics.nature.com/supplementary_info/). According to a reproducibility study by INCYTE (http://www.incyte.com/gem/technology/reproducibility.shtml), 0.5% of the individual expression ratios are spuriously altered more than ±1.74-fold. To estimate the probability for erroneous inclusion of a given EST in our data set, each measurement was modelled as a Bernoulli trial and analysed with the binomial distribution. The calculation was performed as follows: $p(2 \text{ or more errors}) = 1 - p(\text{at most 1 error}) = 1 - (_{14}C_1)pq^{13} + q^{14} = 0.0022$; where $P = 0.005$ and $q = 1 - p = 0.995$. Thus, at most 0.22% of our ESTs (~22/10,000 ESTs) should be erroneously included in our data set of 413 ESTs. Given these calculations for spurious inclusion at a threshold of ±1.74-fold, we consider our choice of a threshold for inclusion of ±2.5-fold to be conservative.

We estimated spot-to-spot variability using 27 known cDNAs spotted in triplicate as less than 20%. Inverse labelling of sample *cim11* (*cim11 inv* in Fig. 1) demonstrates reproducibility of independent experiments (linear regression coefficient $r = 0.83$), and labelling symmetry. We also compared more than 40 DNA microarray data points with RNA blot results and found a linear correlation of $r = 0.83$. Both are consistent with published data[17].

Data analysis. Analysis of differences between conditions based on expression profiles required the calculation of distance matrices in S-Plus (S-PLUS 2000 Professional, Release 2 MathSoft) using the dist function under the Euclidean metric. To indicate a scale of overall difference in expression profiles between 2 conditions, we summed the Euclidean distance between expression levels in the 2 conditions over all 413 ESTs. For example, this resulted in a difference of 19.6 units compared with 159.4 difference units of the comparison (*cim11*, *NahG*). The unrooted tree depicting similarity

in gene-expression patterns was produced using the fitch and drawtree programs distributed in the Phylip suite version 3.57c (ref. 43).

"Cluster" and "Tree view" software[20] were used (http://rana.Stanford.EDU/software/) to group and display genes with similar expression profiles, and to confirm the tree described above. We used the default options of hierarchical clustering using the uncentred correlation similarity metric. We performed this analysis using both normalized and non-normalized data, and the outcomes were essentially the same. We also clustered our data into a 3x4 SOM geometry[21] using the "Gene Cluster" program (Version 1.0-default settings; http://waldo.wi.mit.edu/MPR). To focus on the shape of expression patterns rather than the magnitude of individual expression changes, we normalized the expression level of each EST to have mean=0 and s.d.=1 across the 16 samples. Normalization was performed by subtracting the mean across all 16 treatments from each data point and then dividing the result by the s.d. This standard technique was used as implemented in the Gene Cluster program.

Probability and expected value calculations. To calculate the probability (p) of seeing m promoters with n or more W boxes from a set of N promoters, we modelled each of the N promoters as a Bernoulli trial and applied the binomial distribution. So if the probability (q) of seeing n or more W boxes in a single promoter is known, then $P = {}_NC_m * q^m * (1-q)^{N-m}$. For example, the probability of finding 20 promoters with 3 or more W boxes from a set of 26 promoters is calculated as follows: $P = {}_{26}C_{20} * q^{20} * (1-q)^6$. The value of q is calculated by assuming that each base of a hypothetical promoter can either be part of a W box or not and modelling this situation as a Bernoulli trial. Thus q is calculated as follows:

$$q = 1 - \sum_{i=0}^{n-1} {}_{2200}C_i \, p^i (1-p)^{2200-i}$$

where $P = 1/1024$ (the probability of finding a W box starting at a given nucleotide) and 2200 is the number of base pairs in both strands in a 1.1-kb promoter. Expected values are calculated by summing the probability of a given outcome multiplied by the actual outcome over all possible values in a distribution. Thus, the expected value E which represents the number of promoters expected by chance to have n or more W boxes from a set of N promoters is:

$$E = \sum_{i=0}^{N} i \left({}_NC_i * q^i * (1-q)^{N-i} \right)$$

where q is the probability of seeing n or more W boxes in a single promoter as calculated above.

1. Somssich, I.E. & Halbrock, K. Pathogen defense in plants—a paradigm of biological complexity. *Trends Plant Sci.* **3**, 86–90 (1998).
2. Ryals, J.L. *et al.* Systemic acquired resistance. *Plant Cell* **8**, 1809–1819 (1996).
3. Dong, X. SA, JA, ethylene, and disease resistance in plants. *Curr. Opin. Plant Biol.* **1**, 316–323 (1998).
4. Vernooij, B. *et al.* Salicylic acid is not the translocated signal responsible for inducing systemic acquired resistance, but is required in signal transduction. *Plant Cell* **6**, 959–965 (1994).
5. Delaney, T. *et al.* A central role of salicylic acid in plant disease resistance. *Science* **266**, 1247–1250 (1994).
6. Lawton, K. *et al.* Benzothiadiazole induces disease resistance in *Arabidopsis* by activation of the systemic acquired resistance signal transduction pathway. *Plant J.* **10**, 71–82 (1996).
7. Lebel, E. *et al.* Functional analysis of regulatory sequences controlling PR-1 gene expression in Arabidopsis. *Plant J.* **16**, 223–233 (1998).
8. Bowling, S.A. *et al.* A mutation in Arabidopsis that leads to constitutive expression of systemic acquired resistance. *Plant Cell* **6**, 1845–1857 (1994).
9. Dangl, J.L., Dietrich, R.A. & Richberg, M.H. Death don't have no mercy: cell death programs in plant-microbe interactions. *Plant Cell* **8**, 1793–1807 (1996).
10. Cao, H., Bowling, S.A., Gordon, S. & Dong, X. Characterization of an Arabidopsis mutant that is non-responsive to inducers of systemic acquired resistance. *Plant Cell* **6**, 1583–1592 (1994).
11. Delaney, T.P., Friedrich, L. & Ryals, J.A. *Arabidopsis* signal transduction mutant defective in chemically and biologically induced disease resistance. *Proc. Natl Acad. Sci. USA* **92**, 6602–6606 (1995).
12. Depres, D., DeLong, C., Glaze, S., Liu, E. & Fobert, P.R. The Arabidopsis NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. *Plant Cell* **12**, 279–290 (2000).
13. Li, X., Zhang, Y., Clarke, J.D., Li, Y. & Dong, X. Identification and cloning of a negative regulator of systemic acquired resistance, *SNI1*, through a screen for suppressors of npr1-1. *Cell* **98**, 329–339 (1999).
14. Zhang, Y., Fan, W., Kinkema, M., Li, X. & Dong, X. Interaction of NPR1 with basic leucine zipper protein transcription factors that bind sequences required for salicylic acid induction of the PR-1 gene. *Proc. Natl Acad. Sci. USA* **96**, 6523–6528 (1999).
15. Zhou, J.-M. *et al.* 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.* **13**, 191–202 (2000).
16. Maleck, K. & Dietrich, R.A. Defense on multiple fronts: how do plants cope with diverse enemies? *Trends Plant Sci.* **4**, 215–219 (1999).
17. White, K., Rifkin, S., Hurban, P. & Hogness, D. Microarray analysis of Drosophila development during metamorphosis. *Science* **286**, 2179–2184 (1999).
18. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
19. Cao, H., Li, X. & Dong, X. Generation of broad spectrum disease resistance by over-expression of an essential regulatory gene in systemic acquired resistance. *Proc. Natl Acad. Sci. USA* **95**, 6531–6536 (1998).
20. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
21. Tamayo, P. *et al.* Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl Acad. Sci. USA* **96**, 2907–2912 (1999).
22. Uknes, S. *et al.* Biological induction of systemic acquired resistance in *Arabidopsis*. *Mol. Plant Microbe Interact.* **6**, 692–698 (1993).
23. Oh, S.A., Lee, S.Y., Chung, I.K., Lee, C.H. & Nam, H.G. A senescence-associated gene of Arabidopsis thaliana is distinctively regulated during natural and artificially induced leaf senescence. *Plant Mol. Biol.* **30**, 739–754 (1996).
24. Park, J.H., Oh, S.A., Kim, Y.H., Woo, H.R. & Nam, H.G. Differential expression of senescence-associated mRNAs during leaf senescence induced by different senescence-inducing factors in Arabidopsis. *Plant Mol. Biol.* **37**, 445–454 (1998).
25. Quirino, B.F., Normanly, J. & Amasino, R.M. Diverse range of gene activity during Arabidopsis thaliana leaf senescence includes pathogen independent induction of defense-related genes. *Plant Mol. Biol.* **40**, 267–278 (1999).
26. Eulgem, T., Rushton, P.J., Robatzek, S. & Somssich, I.E. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* **5**, 199–206 (2000).
27. Schindler, U., Beckmann, H. & Cashmore, A.R. TGA1 and G-box binding factors: two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG. *Plant Cell* **4**, 1309–1319 (1992).
28. Eulgem, T., Rushton, P.J., Schmelzer, E., Hahlbrock, K. & Somssich, I.E. Early nuclear events in plant defence signalling: rapid activation by WRKY transcription factors. *EMBO J.* **18**, 4689–4699 (1999).
29. Fukuda, Y. Interaction of tobacco nuclear proteins with an elicitor responsive element in the promoter of a basic class I chitinase gene. *Plant Mol. Biol.* **34**, 81–87 (1997).
30. Rushton, P.J. *et al.* Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *EMBO J.* **15**, 5690–5700 (1996).
31. Durrant, W.E., Rowland, O., Piedras, P., Hammond-Kossak, K.E. & Jones, J.D.G. cDNA-AFLP reveals a striking overlap in the race-specific resistance and wound response expression profiles. *Plant Cell* **12**, 963–977 (2000).
32. Neuenschwander, U. *et al.* Is hydrogen peroxide a second messenger of salicylic acid in systemic acquired resistance? *Plant J.* **8**, 227–234 (1995).
33. Bi, Y.-M., Kenton, P., Mur, L., Darby, R. & Draper, J. Hydrogen peroxide does not function downstream of salicylic acid in the induction of PR protein expression. *Plant J.* **8**, 235–246 (1995).
34. Lamb, C. & Dixon, R.A. The oxidative burst in plant disease resistance. *Annu. Rev. Physiol. Plant Mol. Biol.* **48**, 251–275 (1997).
35. Jabs, T., Dietrich, R.A. & Dangl, J.L. Initiation of runaway cell death in an Arabidopsis mutant by extracellular superoxide. *Science* **273**, 1853–1856 (1996).
36. Dietrich, R.A. *et al.* Arabidopsis mutants simulating disease resistance response. *Cell* **77**, 565–578 (1994).
37. Batz, O., Logemann, E., Reinold, S. & Hahlbrock, K. Extensive reprogramming of primary and secondary metabolism by fungal elicitor or infection in parsley cells. *Biol. Chem.* **379**, 1127–1135 (1998).
38. Marcotte, E.M., Pellegrini, M., Thompson, M.J., Yeates, T.O. & Eisenberg, D. A combined algorithm for genome-wide prediction of protein function. *Nature* **402**, 83–86 (1999).
39. Glazebrook, J., Rogers, E.E. & Ausubel, F.M. Use of Arabidopsis for genetic dissection of plant defense responses. *Annu. Rev. Genet.* **31**, 547–569 (1997).
40. Wolter, M., Hollricher, K., Salamini, F. & Schulze-Lefert, P. The *mlo* resistance alleles to powdery mildew infection in barley trigger a developmentally controlled defence mimic phenotype. *Mol. Gen. Genet.* **239**, 122–128 (1993).
41. Delseny, M., Cooke, R., Raynal, M. & Grellet, F. The Arabidopsis thaliana cDNA sequencing projects. *FEBS Lett.* **403**, 221–224 (1997).
42. Ruan, Y., Gilmore, J. & Conner, T. Towards Arabidopsis genome analysis: monitoring exprssion profiles of 1400 genes using cDNA microarrays. *Plant J.* **15**, 821–833 (1997).
43. Felsenstein, J. PHYLIP, Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
44. Menkens, A.E., Schindler, U. & Cashmore, A.R. The G-box: a ubiquitious regulatory DNA element in plants bound by the GBF family of bZIP proteins. *Trends Biochem. Sci.* **20**, 506–510 (1995).
45. Dröge-Laser, W. *et al.* Rapid stimulation of a soybean protein serine-threonine kinase which phosphorylates a novel bZIP DNA-binding protein, G/HBF-1, during the induction of early transcription-dependent defenses. *EMBO J.* **16**, 726–738 (1997).
46. 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.* **273**, 26857–26861 (1998).
47. Menke, F.L.H., Champion, A., Kijne, J.W. & Memelink, J. A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO J.* **18**, 4455–4463 (1999).
48. Solano, R., Stepanova, A., Chao, Q. & Ecker, J.R. Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* **12**, 3703–3714 (1998).
49. Yang, Y. & Klessig, D.F. Isolation and characterization of a tobacco mosaic virus-inducible *myb* oncogene homolog from tobacco. *Proc. Natl Acad. Sci. USA* **93**, 14972–14977 (1996).