In Silico Biology 6 (2006) 243–252
IOS Press

AthaMap: From *in silico* Data to Real Transcription Factor Binding Sites

Lorenz Bülow, Nils Ole Steffens, Claudia Galuschka, Martin Schindler[1] and Reinhard Hehl*

Institut für Genetik, Technische Universität Braunschweig, Spielmannstr. 7, D-38106 Braunschweig, Germany

Edited by E. Wingender; received 10 January 2006; revised 23 March and 7 April 2006; accepted 9 April 2006; published 21 April 2006

ABSTRACT: AthaMap generates a map for *cis*-regulatory sequences for the whole *Arabidopsis thaliana* genome. AthaMap was initially developed by matrix-based detection of putative transcription factor binding sites (TFBS) mostly determined from random binding site selection experiments. Now, also experimentally verified TFBS have been included for 48 different *Arabidopsis thaliana* transcription factors (TF). Based on these sequences, 89,416 very similar putative TFBS were determined within the genome of *A. thaliana* and annotated to AthaMap. Matrix- and single sequence-based binding sites can be included in colocalization analysis for the identification of combinatorial *cis*-regulatory elements. As an example, putative target genes of the WRKY18 transcription factor that is involved in plant-pathogen interaction were determined. New functions of AthaMap include descriptions for all annotated *Arabidopsis thaliana* genes and direct links to TAIR, TIGR and MIPS. Transcription factors used in the binding site determination are linked to TAIR and TRANSFAC® databases. AthaMap is freely available at http://www.athamap.de.

KEYWORDS: *Arabidopsis thaliana*, database, gene expression, plant, pathogen, transcription factor

INTRODUCTION

Positional information on transcription factor binding sites in whole genomes is useful to identify target genes of specific TFs. Furthermore, such information is helpful to generate models on the regulation of genes that are investigated. AthaMap generates a positional map for TFBS in the *Arabidopsis thaliana* genome [1]. It was developed with publicly available binding sites that were mostly identified by random binding site selection experiments. The sites of these random binding site selection experiments were used to generate alignment matrices which are employed by the program PATSER to identify genomic positions of TFBS within the genome of *A. thaliana* [1,2]. The matrix-based searches were performed with transcription factors from many different plant species, based on the rationale that sequence recognition is not species-specific but similar for members of the same plant TF family. Positional information was imported into the AthaMap database and can be displayed online by entering either a specific chromosomal position or the commonly used gene model number (AGI) that can be found in the TAIR database [3]. The genomic sequence around the position entered and all putative

[1]Present address: Software Systems Engineering, Technische Universität Braunschweig, Mühlenpfordtstr. 23, D-38106 Braunschweig, Germany.
*Corresponding author. Tel.: +49 531 391 5772; Fax: +49 531 391 5765; E-mail: R.Hehl@tu-braunschweig.de.

Electronic publication can be found in *In Silico Biol.* **6**, 0023 <http://www.bioinfo.de/isb/2006/06/0023/>, 21 April 2006.

1386-6338/06/$17.00

TFBS identified in this region are displayed online. The last version of AthaMap contained more than 7.4×10^6 putative binding sites for 36 different transcription factors representing 16 different TF families [4]. Furthermore, more than 1.8×10^5 combinatorial *cis*-regulatory elements were annotated to the database [4].

A significant improvement of AthaMap constitutes a transcription factor binding site map for *Arabidopsis thaliana* that is also based on *in vivo* and experimentally verified binding sites in target genes. Towards these ends, AthaMap was now complemented with 89,416 TFBS based on publications describing experimentally determined sites for 48 *Arabidopsis thaliana* TFs that comprise 13 TF families. Furthermore, all annotated genes in AthaMap have been linked to TAIR, TIGR and MIPS [3,5] since these databases constitute the most important information resources for *A. thaliana* genes. In addition, for the study of plant-pathogen interactions and to identify target genes regulated by plant pathogens, links have also been established from the PathoPlant® database to AthaMap [6]. All transcription factors used in the binding site determination are linked to TAIR and TRANSFAC® databases [3,7,8].

NEW AthaMap DATA AND FUNCTIONALITY

For the annotation of TFBS, publications on transcription factor binding studies with *A. thaliana* factors with at least one single experimentally verified binding site were screened and sequences were extracted. In those cases where the binding site directly corresponds to an *A. thaliana* sequence, these published sequences were used to identify the sequence in the genome. The length of the employed screening sequence permitted only the detection of the single binding site within the target gene. To identify additional putative sites, all binding sites were shortened around the core sequence of the TFBS to yield sequences for genomic screenings.

It is highly likely that shorter sequences identify additional binding sites because in many experimental setups short oligonucleotides will bind the respective TF at least *in vitro*. For example, DREB1A target sites have been identified by comparing regulatory regions in genes upregulated in *A. thaliana* overexpressing DREB1A [9]. A conserved *cis*-acting sequence was identified and experimentally verified *in vitro* as a binding site for DREB1A in the *rd29A* gene promoter. An 8 bp long double-stranded oligonucleotide (ACCGACAT) was used for competition experiments showing that this oligonucleotide can compete for binding in an electrophoretic mobility shift assay [9]. Therefore, this shorter sequence is a putative binding site at all genomic positions matching this sequence. To identify these positions, a screening sequence was employed that covers the region of the experimentally determined binding site together with two more nucleotides from the *rd29A* promoter at either side of the core sequence (CTACCGACAT, Table 1). With this screening sequence, 70 additional genomic positions were identified.

This low number of predicted binding sites in the above example demonstrates the high specificity when employing a screening sequence with a length of 10 bp. A 10 bp screening sequence with a 50% GC content theoretically detects only 151.2 sites in the *A. thaliana* genome. This binding sequence shortening was performed for all TFBS to identify additional putative binding sites. For those binding sequences that contain a 3 or 5 bp conserved core sequence, a 9 bp screening sequence was employed to maintain symmetry around the core sequence. A 9 bp screening sequence (55.6% GC-content) theoretically detects only 472.7 binding sites in the *A. thaliana* genome.

The high specificity of this screening method may not uncover all putative sites. However, using these parameters, sensitivity was still high enough to detect functional W-boxes of WRKY binding sites in many genes as demonstrated in the example given below.

Table 1
Arabidopsis thaliana transcription factors and screening sequences, with the corresponding core sequences being underlined, used for binding site determination by pattern-based screenings and numbers of predicted sites annotated to the AthaMap database

Family	Factor	Synonyms	AGI	Screening sequences	No. of sites	Reference
ABI3/VP1	ABI3		At3g24650	GCATGCATTA CCATGCAAAT GCATGCATGG	912	[18]
	FUS3		At3g26790	CCATGCATGC GCATGCATTA CCATGCAAAT GCATGCATGG	1,163	[18]
AP2/EREBP	AtERF-1		At4g17500	GAGCCGCCA TAGCCGCCA	649	[19]
	AtERF-2		At5g47220	GAGCCGCCA TAGCCGCCA	649	[19]
	AtERF-3		At1g50640	GAGCCGCCA GTGCCGCCA GAGCTGCCA GAGCCGTCA TAGCCGCCA	1,809	[19]
	AtERF-4		At3g15210	GAGCCGCCA GTGCCGCCA GAGCTGCCA GAGCCGTCA GAGCCGCTA TAGCCGCCA	1,983	[19]
	AtERF-5		At5g47230	GAGCCGCCA TAGCCGCCA	649	[19]
	DREB1A	CBF3	At4g25480	CTACCGACAT AAGCCGACAC TGGCCGACCT	213	[9,20,21]
	DREB1B	CBF1	At4g25490	TGGCCGACCT CTACCGACAT	150	[21,22]
	DREB1C	CBF2	At4g25470	TGGCCGACCT CTACCGACAT	150	[21]
	DREB2A		At5g05410	CTACCGACAT AAGCCGACAC	134	[20]
bZIP	ABI5	GIA1, EEL, DPBF1	At2g36270	CAACGTGTCA CCACGTAGCA GACACGTGGC TATACGTCAG	686	[23,24]
	AREB1	ABF2	At1g45249	CATACGTGTC	82	[20]
	AREB2	ABF4	At3g19290	CATACGTGTC	82	[20]
	bZIP12	EEL, DPBF4	At2g41070	CAACGTGTCA CCACGTAGCA	181	[23]
	HY5	TED5	At5g11260	TCCACGTGGC GACACGTGGC CCCACGTGTC	820	[25]
C2C2(Zn) GATA	GATA-1		At3g24050	GTGGATTGA GTGGATTCA ATAGATAAA AGAGATCTA TATGATAAGG ATGGATCGCG CTCGATTTCA GTGGATTTCA	9,894	[26]

Table 1, continued

Family	Factor	Synonyms	AGI	Screening sequences	No. of sites	Reference
				TATTATCGTC		
				GGGTATCGAA		
	GATA-2		At2g45050	GTGGATTGA	4,290	[26]
				GTGGATTCA		
				AGAGATCTA		
				TATGATAAGG		
	GATA-3		At4g34680	GTGGATTGA	4,290	[26]
				GTGGATTCA		
				AGAGATCTA		
				TATGATAAGG		
	GATA-4		At3g60530	GTGGATTGA	4,290	[26]
				GTGGATTCA		
				AGAGATCTA		
				TATGATAAGG		
C2H2(Zn)	SUP	FLO10, FON1	At3g23130	GACAGTGTC	501	[27]
E2F/DP	E2Fa	E2F3	At2g36010	TTTTCCCGCG	396	[28,29]
				AGCGGGAAAA		
				ATTCCCGCCAAT		
	E2Fb	E2F1	At5g22220	ATTTCCCGCT	605	[28–30]
				ATTTCCCGCC		
				TTTTCCCGCG		
				ATTCCCGCCAAT		
	E2Fc	E2F2	At1g47870	CGCGCCAAA	2,752	[28,29,31]
				CCCGCCAAA		
				TTTTCCCGCG		
				AGCGGGAAAA		
				ATTCCCGCCAAT		
	E2Fd	E2L1, DEL2	At5g14960	CGCGCCAAA	2,754	[28,29,31]
				CCCGCCAAA		
				TTTTCCCGCG		
				AGCGGGAAAA		
				ATTCCCGCCAAT		
	E2Fe	E2L3, DEL1	At3g48160	TTTTCCCGCG	396	[28,29]
				AGCGGGAAAA		
				ATTCCCGCCAAT		
	E2Ff	E2L2, DEL3	At3g01330	TTTTCCCGCG	114	[28]
GARP/ARR-B	ARR1		At3g16857	TANGATTGT	8,752	[32]
				TAGGATYGT		
	ARR2		At4g16110	TANGATTGT	13,767	[32,33]
				TAGGATYGT		
				TTTGATTGT		
HD-Zip	ATML1		At4g21750	GTAAATGCAC	130	[34]
	PDF2		At4g04890	GTAAATGCAC	130	[35]
MYB	AtMYB44	AtMYBR1	At5g67300	TCAGTTAGGG	485	[36]
				AGTTAGTTAC		
	MYB1		At3g09230	CCTAACTGA	962	[37]
				TCTAACTGC		
	MYB2		At2g47190	GAAAACCAA	5,400	[36,37,38]
				AGCAACGCC		
				CCTAACTGA		
				TCTAACTGC		
NAC	ANAC019		At1g52890	TAACACGCAT	104	[39]
	ANAC055	NAC3	At3g15500	TAACACGCAT	104	[39]
	ANAC072	RD26	At4g27410	TAACACGCAT	104	[39]
	NAM		At1g52880	AAGGGATGA	982	[40]

Table 1, continued

Family	Factor	Synonyms	AGI	Screening sequences	No. of sites	Reference
SBP	SPL1		At2g47070	CCGTACAAT	382	[41]
	SPL3		At2g33810	CCGTACAAT TCGTACAAC	772	[41,42]
	SPL4		At1g53160	CCGTACAAC CCGTACAAT	717	[41,43]
	SPL5		At3g15270	CCGTACAAT	382	[41]
	SPL7		At5g18830	CCGTACAAC	335	[43]
Trihelix	GT-1		At1g13450	TGGTTAATA AGGTAAATC AATGATATAG	3,702	[44]
	GT-2		At1g76890	CGGTAATTA	513	[45]
	GT-3b		At2g38250	AAGAAAAATA	4,914	[46]
WRKY(Zn)	WRKY18		At4g31800	TTTTGACAG CATTGACGA CCTTGACTT TTGACTTGAC TTGACNNTTGAC	5,063	[12,16,47]
	WRKY6		At1g62300	GTTGACTAT	1,122	[48]
Total:					89,416	

To detect binding sites, a Perl script was written to perform pattern-based screenings of the *Arabidopsis thaliana* genome (TIGR release 5.0, January 21, 2004). Both strands of the annotated genome were screened resulting in records harboring absolute positional information and orientation. Table 1 shows a compilation of all *A. thaliana* TFs with experimentally verified binding sites that have been annotated to the AthaMap database. The sequences used in the pattern-based screening are indicated with the corresponding core sequences being underlined. The most current name and earlier synonyms for the factors are displayed. All factors were assigned to a specific TF family according to Riechmann *et al.* [10]. The number of sites detected in the *Arabidopsis thaliana* genome, the AGI number and the reference are listed. All positional information determined with the TFBS of these factors was imported into the AthaMap database. It is important to note that overlapping sites were not eliminated. All TFs that bind or putatively bind a site are shown on the AthaMap web site [4]. This is very important because TFs themselves are regulated and expression of two factors that recognize the same sequence can be spatially or temporally different. For example, DREB1A and DREB2A bind to the same target site but are either upregulated by low temperature (DREB1A) or by NaCl (DREB2A) [11]. This illustrates the importance to identify all TFs that can potentially bind to the same target site.

Figure 1 displays a web interface screen shot showing binding sites of one of the new AthaMap database entries, WRKY18, within the sequence window in the region of the *NPR1* gene. AthaMap identifies three WRKY18 binding sites that had previously been determined experimentally [12]. The transcribed region is underlined. The gene is encoded on the bottom strand. As a new feature, a short description of the gene shown in the sequence window together with links for additional information leading to the corresponding records in the external databases TIGR, TAIR, and MIPS are provided below the sequence window (Fig. 1).

Exact positional information of the individual binding site is shown in a tool tip box that opens by moving the mouse over the arrow heads which indicate the orientation of the sites (Fig. 1). General information on the transcription factor is provided in a separate pop-up window that opens by clicking on the factor's name. In this window (Fig. 1), the factor family, binding and screening sequences, and references are displayed. For further information, external links (AGI, TRANSFAC ID) to the corresponding records in the TAIR and TRANSFAC databases are provided [3,8].

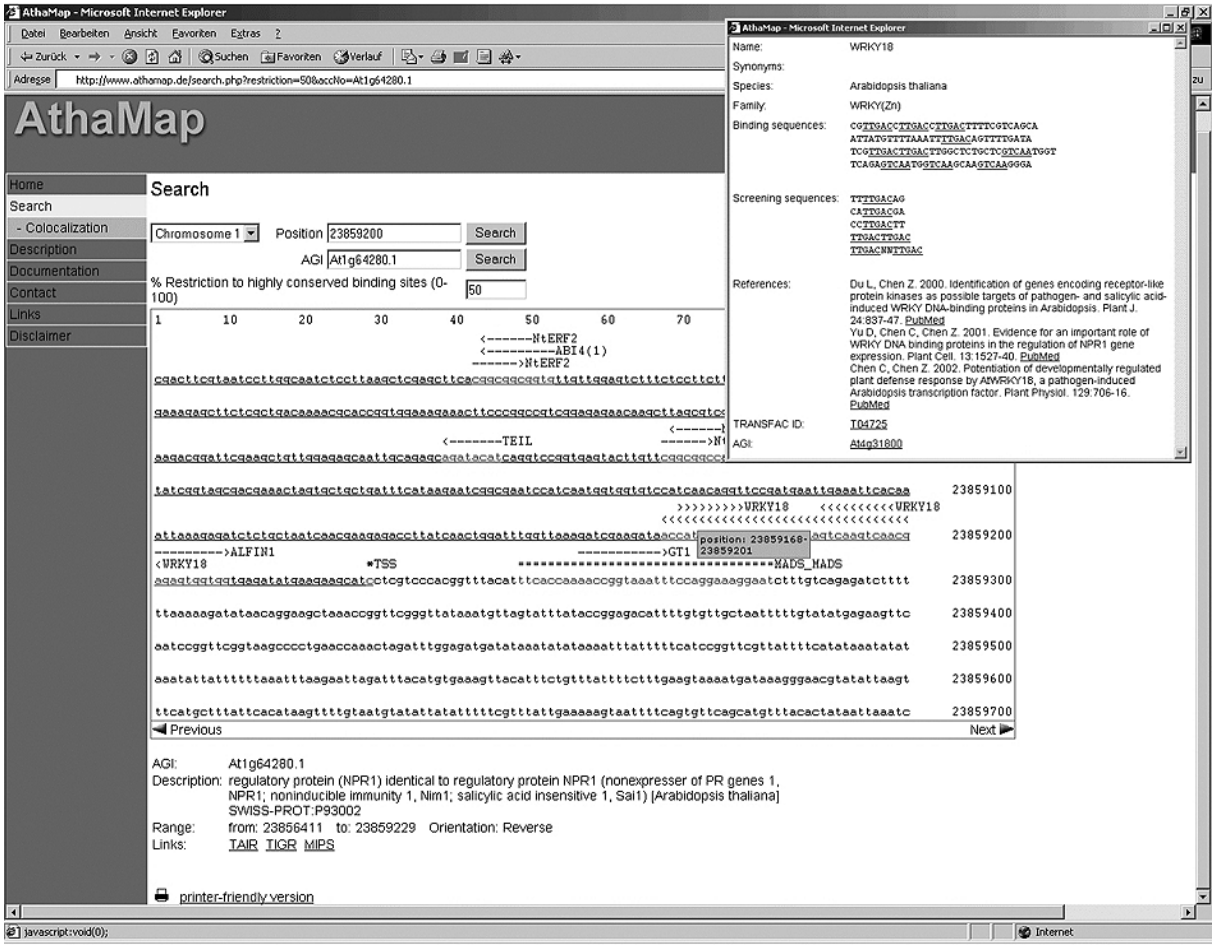

Fig. 1. Screen shot of the AthaMap web interface showing a specific search result together with selected links. The screen shot was generated by entering the AGI of the *NPR1* gene (At1g64280) in the search field and by entering 50 to restrict the display to highly conserved binding sites [4]. The screen shot includes a tool tip box displaying the position of binding sites and a pop-up window for a transcription factor database entry (WRKY18).

In addition to the newly annotated binding sites determined by pattern search, 638,144 predicted matrix-based transcription factor binding sites for 4 new transcription factors representing the NAC, MYB, GARP/ARR-B, and AP2/EREBP families were determined and have been imported into AthaMap. Matrix-based searches were performed as described earlier [1]. Table 2 lists the factors, the factor families and the references from which the sequences were extracted.

The new data presented here increases the number of transcription factors in the database from previously 36 to 88. These belong to 21 different families and detect more than 8×10^6 TFBS in the *Arabidopsis thaliana* genome.

IDENTIFYING TARGET GENES OF TFs

The screen shot in Fig. 1 shows *in vivo* binding sites of WRKY18, a member of the WRKY transcription factor family. Several plant WRKY transcription factor genes are known to be induced upon pathogen

Table 2
New transcription factor binding sites predicted by matrix-based screenings annotated to the AthaMap database

Factor	Family	Species	No. of sites	Reference for alignment matrix
TaNAC69	NAC	*Triticum aestivum*	114	[49]
TaMYB80	MYB	*T. aestivum*	19,023	[49]
ARR10[a]	GARP/ARR-B	*A. thaliana*	153,308	[50]
NtERF2	AP2/EREBP	*Nicotiana tabacum*	465,699	[51]
Total:			638,144	

[a]AGI: At4g31920.

Table 3
Putative target genes of WRKY18 determined by colocalization analysis of WRKY18 binding sites

AGI	Function	No. of colocalizing WRKY18 binding sites
At1g07530	Scarecrow-like transcription factor 14 (SCL14)	2
At1g29720	Protein kinase family protein	2
At1g43150	Non-LTR retrotransposon family	2
At1g52680	Late embryogenesis abundant protein-related/LEA protein-related	2
At1g63740	Disease resistance protein (TIR-NBS-LRR class)	2
At1g63750	Disease resistance protein (TIR-NBS-LRR class)	2
At1g64280	Regulatory protein (NPR1), nonexpresser of PR genes 1	3
At1g64440	UDP-glucose 4-epimerase	2
At1g66910	Protein kinase, putative similar to receptor serine/threonine kinase PR5K	2
At1g68740	EXS family protein/ERD1/XPR1/SYG1 family protein	2
At1g76260	Transducin family protein/WD-40 repeat family protein contains 6 WD-40 repeats	2
At2g22490	Cyclin delta-2 (CYCD2)	2
At2g29010	Pseudogene, receptor protein kinase	2
At3g24065	Expressed protein; expression supported by MPSS	3
At3g46280	Protein kinase-related	2
At3g50150	Expressed protein, plant protein of unknown function; expression supported by MPSS	2
At3g60630	Scarecrow transcription factor family protein scarecrow-like 6	2
At4g06631	Pseudogene, hypothetical protein	2
At4g15520	tRNA/rRNA methyltransferase (SpoU) family protein	2
At4g23000	Calcineurin-like phosphoesterase family protein	2
At4g23180	Receptor-like protein kinase 4 (RLK4)	4
At4g31800	WRKY 18, WRKY family transcription factor	3
At4g34180	Cyclase family protein	2
At4g35310	Calcium-dependent protein kinase, putative/CDPK	2
At5g39480	F-box family protein	2
At5g41140	Expressed protein	2
At5g45730	DC1 domain-containing protein	2
At5g54230	Myb family transcription factor (MYB49)	2
At5g55040	DNA-binding bromodomain-containing protein	2
At5g64360	DNAJ heat shock N-terminal domain-containing protein	2

infection, elicitors, or by treatment with salicylic acid (SA) [13–16]. WRKY18 from *A. thaliana* is involved in the induction of defense-related genes like *NPR1* [12]. NPR1 is a key regulator of SA-dependent systemic PR-protein induction and is regulated by binding of WRKY18 to multiple W-boxes present in the *NPR1* gene (At1g64280) (Fig. 1). Therefore, a colocalization analysis of WRKY18

binding sites harboring W-boxes was performed in AthaMap. This analysis results in 61 colocalizations of at least two WRKY18 binding sites with a maximum distance of 50 bp (data not shown). The colocalizations are in the vicinity of 51 individual genes. In 30 of these genes, colocalizations are present in the upstream region of the translation start. Many of these genes are directly involved in plant defence responses and/or signal transduction and gene regulation. Table 3 shows a list of these genes with WRKY18 binding site colocalizations upstream of the translation start. Four genes contained more than two colocalizing WRKY18 binding sites in their upstream regions, i.e. *NPR1* (At1g64280), *RLK4* (At4g23180), an undefined expressed protein (At3g24065), and the *WRKY18* gene itself (At4g31800). The *RLK4* gene had previously been shown to be induced by bacterial pathogens and SA treatment and to be regulated by WRKY18 [17]. This example demonstrates the use of the AthaMap database resource as a tool to predict putative target genes of specific TFs in *A. thaliana*.

AVAILABILITY

The AthaMap resources are freely available for non-commercial users at http://www.athamap.de.

ACKNOWLEDGEMENTS

We would like to thank Gülsen Okunakul for help with the literature screening and data extraction. This work was supported by the German Ministry of Education and Research (BMBF grant no. 031U110C/031U210C) and was carried out in the Intergenomics Center at Braunschweig.

REFERENCES

[1] Steffens, N. O., Galuschka, C., Schindler, M., Bülow, L. and Hehl, R. (2004). AthaMap: an online resource for *in silico* transcription factor binding sites in the *Arabidopsis thaliana* genome. Nucleic Acids Res. **32**, D368-D372.

[2] Hertz, G. Z. and Stormo, G. D. (1999). Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics **15**, 563-577.

[3] Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L. A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D. C., Wu, Y., Xu, I., Yoo, D., Yoon, J. and Zhang, P. (2003). The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res. **31**, 224-228.

[4] Steffens, N. O., Galuschka, C., Schindler, M., Bülow, L. and Hehl, R. (2005). Web tools for database-assisted identification of combinatorial *cis*-regulatory elements and the display of highly conserved transcription factor binding sites in *Arabidopsis thaliana*. Nucleic Acids Res. **33**, W397-W402.

[5] Schoof, H., Ernst, R., Nazarov, V., Pfeifer, L., Mewes, H. W. and Mayer, K. F. (2004). MIPS *Arabidopsis thaliana* Database (MAtDB): an integrated biological knowledge resource for plant genomics. Nucleic Acids Res. **32**, D373-376.

[6] Bülow, L., Schindler, M., Choi, C. and Hehl, R. (2004). PathoPlant®: A database on plant-pathogen interactions. In Silico Biol. **4**, 0044.

[7] Hehl, R. and Wingender, E. (2001). Database-assisted promoter analysis. Trends Plant Sci. **6**, 251-255.

[8] Matys, V., Fricke, E., Geffers, R., Gößling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A. E., Kel-Margoulis, O. V., Kloos, D. U., Land, S., Lewicki-Potapov, B., Michael, H., Münch, R., Reuter, I., Rotert, S., Saxel, H., Scheer, M., Thiele, S. and Wingender, E. (2003). TRANSFAC®: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. **31**, 374-378.

[9] Maruyama, K., Sakuma, Y., Kasuga, M., Ito, Y., Seki, M., Goda, H., Shimada, Y., Yoshida, S., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2004). Identification of cold-inducible downstream genes of the *Arabidopsis* DREB1A/CBF3 transcriptional factor using two microarray systems. Plant J. **38**, 982-993.

[10] Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K. and Yu, G. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science **290**, 2105-2110.

[11] Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., and Shinozaki, K. (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in Arabidopsis. Plant Cell **10**, 1391-1406

[12] Yu, D., Chen, C. and Chen, Z. (2001). Evidence for an important role of WRKY DNA binding proteins in the regulation of NPR1 gene expression. Plant Cell **13**, 1527-1540.

[13] Eulgem, T., Rushton, P. J., Schmelzer, E., Hahlbrock, K. and Somssich, I. E. (1999). Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors. EMBO J. **18**, 4689-4699.

[14] Eulgem, T., Rushton, P. J., Robatzek, S. and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. Trends Plant Sci. **5**, 199-206.

[15] Chen, C. and Chen, Z. (2000). Isolation and characterization of two pathogen- and salicylic acid-induced genes encoding WRKY DNA-binding proteins from tobacco. Plant Mol. Biol. **42**, 387-396.

[16] Dellagi, A., Heilbronn, J., Avrova, A. O., Montesano, M., Palva, E. T., Stewart, H. E., Toth, I. K., Cooke, D. E. L., Lyon, G. D. and Birch, P. R. J. (2000). A potato gene encoding a WRKY-like transcription factor is induced in interactions with *Erwinia carotovora* subsp. *atroseptica* and *Phytophthora infestans* and is coregulated with class I endochitinase expression. Mol. Plant Microbe Interact. **13**, 1092-1101.

[17] Du, L. and Chen, Z. (2000). Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in *Arabidopsis*. Plant J. **24**, 837-847.

[18] Reidt, W., Wohlfarth, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L. and Bäumlein, H. (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. Plant J. **21**, 401-408.

[19] Fujimoto, S. Y., Ohta, M., Usui, A., Shinshi, H. and Ohme-Takagi, M. (2000). Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. Plant Cell **12**, 393-404.

[20] Narusaka, Y., Nakashima, K., Shinwari, Z. K., Sakuma, Y., Furihata, T., Abe, H., Narusaka, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2003). Interaction between two *cis*-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis rd29A* gene in response to dehydration and high-salinity stresses. Plant J. **34**, 137-148.

[21] Gilmour, S. J., Zarka, D. G., Stockinger, E. J., Salazar, M. P., Houghton, J. M. and Thomashow, M. F. (1998). Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. Plant J. **16**, 433-442.

[22] Stockinger, E. J., Gilmour, S. J. and Thomashow, M. F. (1997). *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc. Natl. Acad. Sci. USA **94**, 1035-1040.

[23] Bensmihen, S., Rippa, S., Lambert, G., Jublot, D., Pautot, V., Granier, F., Giraudat, J. and Parcy, F. (2002). The homologous ABI5 and EEL transcription factors function antagonistically to fine-tune gene expression during late embryogenesis. Plant Cell **14**, 1391-1403.

[24] Carles, C., Bies-Etheve, N., Aspart, L., Leon-Kloosterziel, K. M., Koornneef, M., Echeverria, M. and Delseny, M. (2002). Regulation of *Arabidopsis thaliana Em* genes: role of ABI5. Plant J. **30**, 373-383.

[25] Cluis, C. P., Mouchel, C. F. and Hardtke, C. S. (2004). The *Arabidopsis* transcription factor HY5 integrates light and hormone signaling pathways. Plant J. **38**, 332-347.

[26] Teakle, G. R., Manfield, I. W., Graham, J. F. and Gilmartin, P. M. (2002). *Arabidopsis thaliana* GATA factors: organisation, expression and DNA-binding characteristics. Plant Mol. Biol. **50**, 43-57.

[27] Dathan, N., Zaccaro, L., Esposito, S., Isernia, C., Omichinski, J. G., Riccio, A., Pedone, C., Di Blasio, B., Fattorusso, R. and Pedone, P. V. (2002). The *Arabidopsis* SUPERMAN protein is able to specifically bind DNA through its single Cys_2-His_2 zinc finger motif. Nucleic Acids Res. **30**, 4945-4951.

[28] Mariconti, L., Pellegrini, B., Cantoni, R., Stevens, R., Bergounioux, C., Cella, R. and Albani, D. (2002). The E2F family of transcription factors from *Arabidopsis thaliana*. Novel and conserved components of the retinoblastoma/E2F pathway in plants. J. Biol. Chem. **277**, 9911-9919.

[29] Egelkrout, E. M., Mariconti, L., Settlage, S. B., Cella, R., Robertson, D. and Hanley-Bowdoin, L. (2002). Two E2F elements regulate the proliferating cell nuclear antigen promoter differently during leaf development. Plant Cell **14**, 3225-3236.

[30] de Jager, S. M., Menges, M., Bauer, U.-M. and Murra, J. A. H. (2001). Arabidopsis E2F1 binds a sequence present in the promoter of S-phase-regulated gene AtCDC6 and is a member of a multigene family with differential activities. Plant Mol. Biol. **47**, 555-568.

[31] Stevens, R., Mariconti, L., Rossignol, P., Perennes, C., Cella, R. and Bergounioux, C. (2002). Two E2F sites in the *Arabidopsis MCM3* promoter have different roles in cell cycle activation and meristematic expression. J. Biol. Chem. **277**, 32978-32984.

[32] Sakai, H., Aoyama, T. and Oka, A. (2000). *Arabidopsis* ARR1 and ARR2 response regulators operate as transcriptional activators. Plant J. **24**, 703-711.
[33] Lohrmann, J., Sweere, U., Zabaleta, E., Bäurle, I., Keitel, C., Kozma-Bognar, L., Brennike, A., Schäfer, E., Kudla, J. and Harter, K. (2001). The response regulator ARR2: a pollen-specific transcription factor involved in the expression of nuclear genes for components of mitochondrial complex I in *Arabidopsis*. Mol. Genet. Genomics **265**, 2-13.
[34] Abe, M., Takahashi, T. and Komeda, Y. (2001). Identification of a *cis*-regulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. Plant J. **26**, 487-494.
[35] Abe, M., Katsumata, H., Komeda, Y. and Takahashi, T. (2003). Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in *Arabidopsis*. Development **130**, 635-643.
[36] Kirik, V., Kölle, K., Miséra, S. and Bäumlein, H. (1998). Two novel MYB homologues with changed expression in late embryogenesis-defective Arabidopsis mutants. Plant Mol. Biol. **37**, 819-827.
[37] Urao, T., Yamaguchi-Shinozaki, K., Urao, S. and Shinozaki, K. (1993). An Arabidopsis *myb* homolog is induced by dehydration stress and its gene product binds to the conserved MYB recognition sequence. Plant Cell **5**, 1529-1539.
[38] Hoeren, F. U., Dolferus, R., Wu, Y., Peacock, W. J. and Dennis, E. S. (1998). Evidence for a role for AtMYB2 in the induction of the Arabidopsis alcohol dehydrogenase gene (*ADH1*) by low oxygen. Genetics **149**, 479-490.
[39] Tran, L. S., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., Fujita, M., Seki, M., Shinozaki, K. and Yamaguchi-Shinozaki, K. (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive *cis*-element in the *early responsive to dehydration stress 1* promoter. Plant Cell **16**, 2481-2498.
[40] Duval, M., Hsieh, T. F., Kim, S. Y. and Thomas, T. L. (2002). Molecular characterization of *AtNAM*: a member of the *Arabidopsis* NAC domain superfamily. Plant Mol. Biol. **50**, 237-248.
[41] Cardon, G., Höhmann, S., Klein, J., Nettesheim, K., Saedler, H. and Huijser, P. (1999). Molecular characterisation of the *Arabidopsis* SBP-box genes. Gene **237**, 91-104.
[42] Cardon, G. H., Höhmann, S., Nettesheim, K., Saedler, H. and Huijser, P. (1997). Functional analysis of the *Arabidopsis thaliana* SBP-box gene SPL3: a novel gene involved in the floral transition. Plant J. **12**, 367-377.
[43] Yamasaki, K., Kigawa, T., Inoue, M., Tateno, M., Yamasaki, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Nunokawa, E., Ishizuka, Y., Terada, T., Shirouzu, M., Osanai, T., Tanaka, A., Seki, M., Shinozaki, K. and Yokoyama, S. (2004). A novel zinc-binding motif revealed by solution structures of DNA-binding domains of *Arabidopsis* SBP-family transcription factors. J. Mol. Biol. **337**, 49-63.
[44] Hiratsuka, K., Wu, X., Fukuzawa, H. and Chua, N. H. (1994). Molecular dissection of GT-1 from Arabidopsis. Plant Cell **6**, 1805-1813.
[45] Kuhn, R. M., Caspar, T., Dehesh, K. and Quail, P. H. (1993). DNA binding factor GT-2 from Arabidopsis. Plant Mol. Biol. **23**, 337-348.
[46] Park, H. C., Kim, M. L., Kang, Y. H., Jeon, J. M., Yoo, J. H., Kim, M. C., Park, C. Y., Jeong, J. C., Moon, B. C., Lee, J. H., Yoon, H. W., Lee, S. H., Chung, W. S., Lim, C. O., Lee, S. Y., Hong, J. C. and Cho, M. J. (2004). Pathogen- and NaCl-induced expression of the SCaM-4 promoter is mediated in part by a GT-1 box that interacts with a GT-1-like transcription factor. Plant Physiol. **135**, 2150-2161.
[47] Chen, C. and Chen, Z. (2002). Potentiation of developmentally regulated plant defense response by AtWRKY18, a pathogen-induced Arabidopsis transcription factor. Plant Physiol. **129**, 706-716.
[48] Robatzek, S. and Somssich, I. E. (2002) Targets of *At*WRKY6 regulation during plant senescence and pathogen defense. Genes Dev. **16**, 1139-1149.
[49] Xue, G. P. (2005). A CELD-fusion method for rapid determination of the DNA-binding sequence specificity of novel plant DNA-binding proteins. Plant J. **41**, 638-649.
[50] Hosoda, K., Imamura, A., Katoh, E., Hatta, T., Tachiki, M., Yamada, H., Mizuno, T. and Yamazaki, T. (2002). Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the Arabidopsis response regulators. Plant Cell **14**, 2015-2029.
[51] Hao, D., Ohme-Takagi, M. and Yamasaki, K. (2003). A modified sensor chip for surface plasmon resonance enables a rapid determination of sequence specificity of DNA-binding proteins. FEBS Lett. **536**, 151-156.