Supplementary Material

Contents:

1. Promoter Analysis

2. cDNA Microarray

Strategy for the Construction of the cDNA Array

Preparation of Subtraction Libraries

Prescreening of Library Clones

cDNA Microarray Production

cDNA Microarray Composition

3. Oligonucleotide Array

4. Microarray Data Files

5. References

1. Promoter Analysis

For promoter analysis, the web-based analysis platform Cistematic (http://cistematic.caltech.edu; S.A. Mortazavi and B. Wold, unpublished) was used for motif identification and assessment. Upstream regions of *Arabidopsis* genes were retrieved for each analysis group and analyzed using AlignACE (Roth et al., 1998), YMF (Sinha and Tompa, 2003), MEME (Bailey and Elkan, 1995), Gibbs Sampling (Thompson et al., 2003), and Co-Bind (GuhaThakurta and Stormo, 2001). Each program was run automatically with the appropriate settings for *Arabidopsis*. Resulting motifs were collected and each motif consensus was converted into Position Specific Scoring Matrices (PSSM) using the matching sites in the datasets.

Cistematic tabulated the observed number of occurrences of each motif consensus in the 1000 bp upstream regions of the genes used for the analysis as well as the number of occurrences of each motif consensus in the 1000 bp upstream region of every *Arabidopsis* gene, which was used to calculate the expected number of occurrences of

each motif in our datasets assuming random distribution. Using the expected and observed number of occurrences of a motif, MAP scores were calculated as described (Hughes et al., 2000). In addition, P-values for the enrichment of a motif in a gene group were calculated by using one-tailed X^2 or, for motifs with less than 5 occurrences, Fisher's exact tests using a 2x2 contingency table.

The presence of CArG-boxes and putative AG binding sites in the *Arabidopsis* genome was determined using Pattern Match (http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl). The presence of CArG-boxes and putative AG binding sites in different regions (as described in *Results*) of the organ-expressed genes is summarized in Tab. S1.

Tab. S1: Distribution of CArG-boxes **(A)** and putative AG binding sites **(B)** in the genome and among the organ-expressed genes. The number of sequences in the data set used for the analysis ('$N_{Data\ set}$'), the number of genes containing at least one of the sites in the indicated region ('N_{Genes}'), and the percentage of genes with the site relative to the analyzed data set is listed.

Tab. S1 A	Genome			Organ-expressed genes		
Region	$N_{Data\ set}$	N_{Genes}	Percentage	$N_{Data\ set}$	N_{Genes}	Percentage
500 bp upstream	28088	1206	4.3	1453	72	4.9
1000 bp upstream	28088	2068	7.4	1453	123	8.5
3000 bp upstream	28088	5218	18.6	1453	281	19.3
1000 bp downstream	27819	1789	6.4	1453	96	6.6
3000 bp downstream	28088	4966	17.7	1453	248	17.1
Intron	116438	1122	1.0	1453	41	2.8

Tab. S1 B	Genome			Stamen and carpel group		
Region	$N_{Data\ set}$	N_{Genes}	Percentage	$N_{Data\ set}$	N_{Genes}	Percentage
500 bp upstream	28088	220	0.8	1422	9	0.6
1000 bp upstream	28088	401	1.43	1422	19	1.3
3000 bp upstream	28088	1036	3.7	1422	47	3.3
1000 bp downstream	27819	330	1.2	1422	27	1.9
3000 bp downstream	28088	989	3.5	1422	70	4.9
Intron	116438	242	0.2	1422	10	0.7

2. Construction of the cDNA Microarray

Strategy for the Construction of the cDNA Array

Elements for the microarray were obtained from three different sources. First, for those genes that are known or suspected to be involved in flower development, we either amplified fragments of their transcribed regions from cDNA or, if available, obtained cDNA clones from EST clone collections. Second, we generated cDNA libraries using RNA extracted from different floral tissues (see below). For all libraries, the RNA preparations were subtracted with leaf RNA to reduce the abundance of clones representing ubiquitously expressed genes and thus, to enrich for flower-specific transcripts. Sequencing of several hundred library clones revealed a large portion of genes with known or presumed flower-specific expression. In order to reduce redundancy among the elements of the microarray, we pre-screened the libraries for highly abundant clones (see below). Ultimately, a total of 7366 clones were selected for the production of the cDNA array. The identity of most of these clones was not determined prior to their use for preparing the array. Lastly, since the subtraction of the libraries with leaf RNA might have led to the removal of genes that have important functions in flower development, in spite of a relatively broad or even ubiquitous expression, we added a set of 2632 non-redundant clones from a non-subtracted flower-derived library (Asamizu et al., 2000).

The total number of elements on the cDNA microarray is 10,816. So far, we have identified 2943 unique genes among 4939 sequenced array elements. Based on these results, we estimate that the total number of genes on the cDNA array is between 5000 and 6000.

Preparation of Subtraction Libraries

Total RNA was extracted from different floral tissues (whole inflorescences, young floral buds, and siliques) using Tri Reagent (Molecular Research Center). From the total RNA preparations, polyA$^+$ RNA was purified with the Poly(A) Pure Kit (Ambion). Using the polyA$^+$ RNA preparations as well as a sample containing RNA from whole inflorescences as well as from young floral buds in equal amounts, subtraction libraries were generated with the PCR-Select cDNA Subtraction Kit (Clontech), according to the instructions of the manufacturer. This PCR-based approach leads to the generation of cDNA fragments rather than of full-length clones. The size of the fragments ranged from about 200 to 1000 bp with an average of about 400 bp. In order to enrich for longer fragments, the cDNA populations were subjected to size exclusion chromatography using Chromaspin-400 columns (Clontech). All libraries used for the construction of the array were subtracted with leaf RNA. In addition, unsubtracted control libraries were generated and the abundance of flower-specific transcripts as well as a ubiquitously expressed transcript in the subtracted and non-subtracted libraries was determined by PCR. The following genes were tested: *SUPERMAN*, *AGAMOUS*, and *GAPDH*. The sequences of primers used for the control PCRs are listed in Supplemental Table 13.

cDNA fragments of the libraries were cloned into pCRII TOPO using the TOPO TA Cloning Kit (Invitrogen) or into the *Not*I site of pBSII-SK (Stratagene) after cutting the cDNA fragments with *Not*I/*Eag*I.

Prescreening of Library Clones

We devised a strategy to reduce redundancy among the library clones. This approach was limited to the library derived from the combined flower/inflorescence RNA since it appeared to have a good representation of flower-specific clones as well as a relatively low degree of redundancy. For this, we made probes against redundant clones as

identified by sequencing of 864 library clones. Since the protocol used to produce the subtraction libraries generates several different cDNA fragments for most transcripts, in many cases more than one probe had to be designed in order to remove highly abundant clones completely. Probes were amplified from the redundant clones by PCR using vector-specific primers. Fragments were pooled and the adaptors used in the subtraction library protocol were cleaved off by *Eag*I digestion. Radioactive labeling of the probes was done using the High Prime Labeling Kit (Roche). These probes were used simultaneously in colony hybridizations and negative (i.e., non-hybridizing) bacterial colonies were selected. Colony hybridizations were performed using standard protocols (Sambrook and Russell, 2001). The success of this procedure was assessed by sequencing of clones that passed the pre-screening process. The majority of the highly redundant clones were efficiently removed by this approach. However, an additional PCR-based step was introduced to eliminate clones that were recalcitrant to efficient removal by hybridization. For this, non-hybridizing colonies were transferred to and grown in 96-well plates and 0.5 µl of the bacterial cultures were subjected to PCR using a mixture of 34 primer pairs for abundant clones that were not efficiently removed by colony hybridization. Probes and primers used for the prescreening are listed in Supplemental Table 13.

cDNA Microarray Production

Bacteria cultures were grown in 96-well blocks and plasmids were isolated with the QiaPrep 96 Turbo Miniprep Kit (Qiagen) using a Biorobot 3000 (Qiagen). An aliquot of the plasmid DNA was diluted 500-fold in H_2O and 0.5 µl of the dilution was used as template in a standard 50 µl PCR containing 5 U of HotStarTaq DNA polymerase (Qiagen). Vector-specific primers used for amplification of vector inserts are listed in Supplemental Table 13.

PCR products were cleaned up with the QIAquick 96 PCR BioRobot Kit (Qiagen) using a Biorobot 3000 (Qiagen). DNA was eluted in H_2O. Aliquots of the purified PCR products were analyzed by gel electrophoresis to assess their quality. All array elements that were derived from low-quality PCRs were later removed from data analysis (see below). Overall, about 93% of all PCRs resulted in high quality products.

The PCR products were distributed into three sets of 384-well plates and dried to completion. The DNA was resuspended in 10 µl of printing buffer per well (150 mM Na-phosphate pH 9; 1 mM EDTA; 50% (v/v) DMSO) resulting in an estimated DNA concentration of 200 ng/µl. DNA was spotted onto poly-lysine coated glass slides (Childs et al., 2003) with a MicroGrid II spotter (Biorobotics) using Microquill 2000 pins (Majer Precision Engineering). After printing, the slides were processed and blocked following standard protocols (Childs et al., 2003).

cDNA Microarray Composition

Detailed information on the microarray platform was deposited at NCBI-GEO (http://www.ncbi.nlm.nih.gov/geo/) (accession number GPL1069).

The columns in GPL1069 are defined as follows:

'ID': contains the unique identifiers of the array elements

'Clone_ID': identifier of the gene represented by a given array element

'Name': gene name (according to TAIR)

'GB_ACC': accession numbers of ESTs used for the array

'Sequence': contains sequencing results.

'Q-PCR': summarizes the results of the quality control of PCR products. The following codes were used:

0:	single band
1:	empty (on purpose)
2:	double band (or multiple bands)
3:	faint (but single band)
4:	smear
5:	band - but supposed to be empty
6:	empty - but supposed to be not empty
7:	chimeric clone
8:	questionable
9:	control element ('spike')

'Vector Type': the vectors used to subclone the cDNA fragments are indicated. The following abbreviations were used:

ZL:	lambdaZipLox (ESTs)
PCRII:	pCRII TOPO (Invitrogen)
pBS:	pBLUESCRIPT II SK (Stratagene)

'Remarks': contains the clone identifiers for the non-subtracted, non-redundant flower-derived library (Asamizu et al., 2000). Clone sequences can be retrieved from the web site of the Kazusa DNA Research Institute (http://www.kazusa.or.jp/en/plant/arabi/EST/).

3. Oligonucleotide Array

The oligonucleotides of the *Arabidopsis* Genome Oligo Set Version 1.0 (Operon) were provided in 384-well plates. The dried DNA was resuspended in 15 µl of printing buffer per well (150 mM Na-phosphate pH 8.5) resulting in an estimated DNA concentration of 40 µM. DNA was spotted onto poly-lysine coated glass slides (Childs et al., 2003) with a MicroGrid II spotter (Biorobotics) using Microspot 10K quill pins (Biorobotics). After printing, the slides were processed and blocked following standard protocols (Childs et al., 2003).

Further information on the microarray platform was deposited at NCBI-GEO (http://www.ncbi.nlm.nih.gov/geo/) (accession number GPL1077).

4. Microarray Data Files

Microarray data after preprocessing and normalization as well as microarray raw data (in Axon Genepix format) were deposited at NCBI-GEO (http://www.ncbi.nlm.nih.gov/geo/). Accession numbers are listed in Tab. S2.

Tab. S2: NCBI-GEO accession numbers for the presented data sets. The data are summarized in the series GSE1275.

Array type	Platform	Mutant	Samples	Series
cDNA	GPL1069	*ap1*	GSM21001-21003	GSE1265
cDNA	GPL1069	*ap2*	GSM21004-21006	GSE1266
cDNA	GPL1069	*ap3*	GSM21007-21009	GSE1267
cDNA	GPL1069	*pi*	GSM21010-21012	GSE1268
cDNA	GPL1069	*ag*	GSM21013-21015	GSE1269
Operon	GPL1077	*ap1*	GSM21016-21019	GSE1270
Operon	GPL1077	*ap2*	GSM21020-21023	GSE1271
Operon	GPL1077	*ap3*	GSM21024-21027	GSE1272
Operon	GPL1077	*pi*	GSM21028-21031	GSE1273
Operon	GPL1077	*ag*	GSM21032-21035	GSE1274

Cy3 and Cy5 dyes were used in the experiments as follows.

For the cDNA array:

Set 1:	wild type-Cy5	mutant-Cy3
Set 2:	wild type-Cy3	mutant-Cy5
Set 3:	wild type-Cy5	mutant-Cy3

For the oligonucleotide array:

Set 1:	wild type-Cy3	mutant-Cy5
Set 2:	wild type-Cy5	mutant-Cy3
Set 3:	wild type-Cy3	mutant-Cy5
Set 4:	wild type-Cy5	mutant-Cy3

5. References

Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. (2000). A large scale analysis of cDNA in *Arabidopsis thaliana*: generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. DNA Res. **7,** 175-180.

Bailey, T.L. and Elkan, C. (1995) The value of prior knowledge in discovering motifs with MEME. Proc. Int. Conf. Intell. Syst. Mol. Biol. **3**, 21–29.

Childs, G., DeRisi, J., Harris, T., Holloway, A., Hou, B.-H., Massimi, A., Murphy, M., and Somerville, S. (2003). Printing spotted glass microarrays. In DNA Microarrays: A Molecular Cloning Manual, D. Bowtell and J. Sambrook, eds (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

GuhaThakurta, D. and Stormo, G.D. (2001) Identifying target sites for cooperatively binding factors. Bioinformatics **17**, 608–621.

Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database. Nucleic Acids Res. **27**, 297-300.

Hughes, J.D., Estep, P.W., Tavazoie, S., and Church, G.M. (2000) Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. J. Mol. Biol. **296**, 1205–1214.

Kepler, T.B., Crosby, L., and Morgan, K.T. (2002). Normalization and analysis of DNA microarray data by self-consistency and local regression. Genome Biol. **3**, research0037.1–0037.12.

Roth, F.P., Hughes, J.D., Estep, P.W., and Church, G.M. (1998) Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat. Biotechnol. **16**, 939–945.

Sambrook, J. and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sinha, S. and Tompa, M. (2003) YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. Nucleic Acids Res. **31**, 3586-3588.

Thompson, W., Rocuhka, E.C., Lawrence, C.E. (2003) Gibbs Recursive Sampler: finding transcription factor binding sites. Nucleic Acids Res. **31**, 3580-3585.