Versatile Gene-Specific Sequence Tags for Arabidopsis Functional Genomics: Transcript Profiling and Reverse Genetics Applications


Genome Res. 2004 14: 2176-2189
Access the most recent version at doi:10.1101/gr.2544504

Supplementary data
"Supplemental Research Data"
http://www.genome.org/cgi/content/full/14/10b/2176/DC1

References
This article cites 72 articles, 36 of which can be accessed free at:
http://www.genome.org/cgi/content/full/14/10b/2176#References

Article cited in:
http://www.genome.org/cgi/content/full/14/10b/2176#otherarticles

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here

Notes

To subscribe to Genome Research go to:
http://www.genome.org/subscriptions/

© 2004 Cold Spring Harbor Laboratory Press
Versatile Gene-Specific Sequence Tags for Arabidopsis Functional Genomics: Transcript Profiling and Reverse Genetics Applications

Pierre Hilson,1,2,17 Joke Allemeersch,3 Thomas Altmann,4 Sébastien Aubourg,1,2 Alexandra Avon,2 Jim Beynon,5 Rishikesh P. Bhalerao,6 Frédérique Bitton,2 Michel Caboche,2 Bernard Cannoot,1 Vasil Chardakov,7 Cécile Cognet-Holliger,8 Vincent Colot,2 Mark Crowe,9 Caroline Darimont,10 Steffen Durinck,3 Holger Eickhoff,11,16 Andéol Falcon de Longevialle,2 Edward E. Farmer,10 Murray Grant,7 Martin T.R. Kuiper,1 Hans Lehrach,11 Céline Léon,2 Antonio Leyva,12 Joakim Lundberg,13 Claire Lurin,2 Yves Moreau,3 Wilfried Nietfeld,11 Javier Paz-Ares,12 Philippe Reymond,10 Pierre Rouzé,1 Goran Sandberg,6 Maria Dolores Segura,12 Carine Serizet,1,2 Alexandra Tabrett,5 Ludivine Taconnat,2 Vincent Thareau,1,2 Paul Van Hummelen,14 Steven Vercruyssse,1 Marnik Vuylsteke,1 Magdalena Weingartner,4 Peter J. Weisbeek,15 Valtteri Wirta,13 Floyd R.A. Wittink,15 Marc Zabeau,1 and Ian Small2,17

1Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, B-9052 Gent, Belgium; 2Unité de Recherche en Génomique Végétale (INRA-CNRS-UVE), F-91057 Evry CEDEX, France; 3Department of Electrical Engineering (ESAT), Faculty of Engineering, Katholieke Universiteit Leuven, B-3001 Heverlee, Belgium; 4Universität Potsdam, Institut für Biochemie und Biologie, - Genetik-, c/o Max-Planck-Institut für molekulare Pflanzenphysiologie, D-14476 Golm, Germany; 5Horticulture Research International, Wellesbourne, Warwick CV35 9EF, United Kingdom; 6Department of Forest Genetics and Plant Physiology, The Swedish University of Agricultural Sciences, S-901 83, Umeå, Sweden; 7Department of Agricultural Sciences, Imperial College London, Wye, Ashford, Kent TN25 5AH, United Kingdom; 8Station de Génétique et Amélioration des Plantes, INRA, F-78026, Versailles CEDEX, France; 9John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, United Kingdom; 10Gene Expression Laboratory, Department of Plant Molecular Biology, University of Lausanne, CH-1015 Lausanne, Switzerland; 11Max-Planck-Institute for Molecular Genetics, Department of Vertebrate Genomics, D-14195 Berlin-Dahlem, Germany; 12Department of Plant Molecular Genetics, Centro Nacional de Biotecnología-CSIC, E-28049 Madrid, Spain; 13Department of Biotechnology, AlbaNova University Center, Royal Institute of Technology (KTH), S-106 91 Stockholm, Sweden; 14VIB MicroArray Facility, UZ Gasthuisberg, B-3000 Leuven, Belgium; 15Department of Molecular Genetics, University Utrecht, NL-3584 CH Utrecht, The Netherlands

Microarray transcript profiling and RNA interference are two new technologies crucial for large-scale gene function studies in multicellular eukaryotes. Both rely on sequence-specific hybridization between complementary nucleic acid strands, inciting us to create a collection of gene-specific sequence tags (GSTs) representing at least 21,500 Arabidopsis genes and which are compatible with both approaches. The GSTs were carefully selected to ensure that each of them shared no significant similarity with any other region in the Arabidopsis genome. They were synthesized by PCR amplification from genomic DNA. Spotted microarrays fabricated from the GSTs show good dynamic range, specificity, and sensitivity in transcript profiling experiments. The GSTs have also been transferred to bacterial plasmid vectors via recombinational cloning protocols. These cloned GSTs constitute the ideal starting point for a variety of functional approaches, including reverse genetics. We have subcloned GSTs on a large scale into vectors designed for gene silencing in plant cells. We show that in planta expression of GST hairpin RNA results in the

16Present address: Scienion AG, D-12489 Berlin, Germany.
17Corresponding authors.
E-MAIL pierre.hilson@psb.ugent.be; FAX 32 9 331 38 09.
E-MAIL small@evry.inra.fr; FAX 33 1 60874510.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2544504.
expected phenotypes in silenced Arabidopsis lines. These versatile GST resources provide novel and powerful tools for functional genomics.

[Supplemental material is available online at www.genome.org. The pAgrikola sequence data have been submitted to GenBank under accession no. (AY568055) and the transcript profiling data to Array Express under accession nos. (E-MEXP-30 and E-MEXP-65). The GST amplicons and CATMA spotted microarrays are available from the Nottingham Arabidopsis Stock Centre (NASC; http://nasc.nott.ac.uk). NASC will also distribute the cloned GST repertoires. The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: C. Helliwell, M. Salanoubat, and P. Waterhouse.]

### RESULTS

#### Design

The GSTs were selected with the Specific Primer and Amplicon Design Software, SPADS (Thareau et al. 2003), based on gene models extracted from the structural annotation of the Arabidopsis thaliana nuclear genome (January 2001) and using the eukaryotic gene finder EuGene (Schiex et al. 2001). The results of this fully automated structural annotation tool are available online (http://www.psb.ugent.be/bioinformatics/genomes_ath_index.php).

The SPADS GST selection process can be divided into four steps. Firstly, the software searched for divergent regions. For each gene model, exons were sequentially matched with BLASTn (Altschul et al. 1997) against the entire genome sequence and segments with significant homology hits were discarded. Secondly, the Primer3 software (Rozen and Skaletsky 2000) selected PCR oligonucleotide pairs in the remaining divergent exonic regions. Thirdly, these oligonucleotide pairs were filtered using BLASTn to eliminate any sequences with significant matches in the surrounding window (minimum 400 kb). This filter diminished the risk of producing spurious PCR products carrying nontarget portions of the Arabidopsis genome. Finally, the remaining potential amplicons were matched using BLASTn against the full genome sequence to determine the presence of potential paralogous genes: for every gene model, SPADS scanned sequentially each potential amplicon, starting with 3'-sequences, until one was identified that met the chosen constraints. For this study, the GST length ranges between 150 and 500 bp, large enough to be purified from contaminants such as PCR primers or primer-dimer PCR artifacts, but small enough to be amplified very efficiently and to retain a sufficient level of specificity. The selected GSTs share no more than 70% identity with any other sequence in the nuclear genome, because above that threshold cross-hybridization is detected in microarray experiments (Richmond et al. 1999; Girke et al. 2000). For other parameters in SPADS, Primer3 and BLASTn used in this study, see Methods and Thareau et al. (2003).

The SPADS algorithm was applied to 29,787 gene models of which 20,981 (70%) resulted in an in silico designed GST. The majority of sequences are in the lower portion of the length spectrum (Fig. 1). Starting the selection from the 3'-end of genes resulted in a distribution of 24%, 15%, and 61% of the probes in the 5', central, and 3' third of the gene models, respectively. Although all primer pairs delimiting the GST sequences lie in annotated or predicted exons, SPADS may also select GST sequences overlapping with intron(s). In this study, 606 (2.9%)
GSTs were designed with intron sequences, whereas they still contained at least 150 bp matching exons and less than half of the GST matching intron sequence. In addition to GSTs extracted from annotated protein-encoding genes, 139 tags were picked from intergenic regions distributed across the five chromosomes. These intergenic tags can serve as negative internal controls in microarray transcript profiling experiments. An example of the distribution of GST sequences along the annotated genome sequence is shown in Figure 2.

Detailed GST information is available from a database at http://www.catma.org (Crowe et al. 2003). For each GST, the CATMA database provides the sequences and corresponding PCR amplification oligonucleotides, their length, genome location, GC%, the GST primary amplification results (with associated gel images), the gene model from which each GST is derived and additional technical information. It also shows links between GSTs and corresponding Arabidopsis Genome Initiative (AGI) gene names together with the Gene Ontology (GO) information about gene function. GST information will also be available via other Arabidopsis resource Web sites (see Web Site References).

The quality of the Arabidopsis genome annotation improves continuously thanks to the refinement of in silico prediction tools and to the increasing number of experimentally documented transcription units (Haas et al. 2002; Seki et al. 2002; Yamada et al. 2003; Castelli et al. 2004). The GST collection is being enlarged incrementally to take into consideration annotation updates. The first version described above, CATMA v1, contained 21,120 in silico designed tags and was based exclusively on gene models from the 2001 EuGene genome annotation (http://www.psb.ugent.be/bioinformatics/genomes_ath_index.php). The second version, CATMA v2, contains an additional 3456 in silico tags (24,576 in total). These new tags were selected according to alternative criteria: (1) the minimum primer GC% was lowered from 40% to 30%; (2) gene models only defined by a coding sequence and that failed to yield a tag in v1 were extended with the 150-bp segment immediately following their stop codon; and (3) AGI gene models located in regions in which no genes were predicted by EuGene were included. A third additional GST set is currently in preparation.

The overlap between CATMA v2 and the latest AGI annotation performed by The Institute for Genomic Research (TIGR; release 5.0) was calculated by searching all GST sequences that aligned over a segment of at least 150 bp with at most two mismatches with the gene models including untranslated regions and introns. Out of 29,993 AGI-annotated protein-encoding genes, 21,566 matched at least one GST, and 2173 genes more than one GST. Conversely, out of 24,576 GSTs, including controls, 23,280 match at least one AGI gene.

All the data presented below on microarray transcript profiling and large-scale cloning relate to the CATMA v1 collection. All functional and structural annotation refers to the TIGR release 5.0 genome annotation.

**Synthesis**

The GSTs were PCR-amplified from genomic DNA. The gene-specific oligonucleotides used for PCR synthesis each contained a 17-nt 5'-extension added to the gene-specific portion selected by SPADS. In total, 40 arbitrary, unique extension sequences had been assigned to the 24 columns (c1–c24) and 16 rows (rA–rP) of a 384 multiwell plate (Fig. 3A; Supplemental Table S1). The column and row extensions were located upstream and downstream, respectively, of the tag sequence with respect to the direction of transcription (Fig. 3B). Based on this structured arrangement, the secondary amplification of the entire GST collection was performed with only 40 oligonucleotides, each carrying the appropriate selective extension at their 3’-end. Well-to-well cross-contaminations that commonly plague cDNA collections maintained for printing microarrays (Knight 2001) are thus resolved in secondary PCR with the set of 40 extension primers because each GST in a multiwell plate is amplified by a

---

**Figure 1** GST length distribution. The calculated length does not include the extensions added to the gene-specific portion of the tags.

**Figure 2** Graphical representation of GST distribution. The represented genomic segment is 50 kb long and centered around gene At5g57890. Light blue and dark blue boxes represent AGI mRNA and coding sequences, respectively. GSTs are shown as purple brackets. The display was extracted from a screenshot of the FLAGdb++ Arabidopsis genome database (Samson et al. 2002).
unique combination of one column primer and one row primer, depending on its position in that plate (e.g., c4 and rB in Fig. 3A).

Individual bacterial artificial chromosomes (BACs) were chosen as templates for the primary PCR amplification of the GSTs, wherever possible. Reducing the template complexity by three orders of magnitude when compared with the Arabidopsis nuclear genome increased DNA yield and diminished the synthesis of spurious products. For this purpose, we used IGF (Mozo et al. 1999) and TAMU (Choi et al. 1995) BAC clones for which bacterial strains were available, together with complete or end sequence information. Out of the 24,576 GSTs in CATMA v2, 22,614 (92.0%) were amplified from BAC templates. Besides faulty primers, failures probably represent incomplete BAC genome coverage, incorrect mapping of BAC clones on reconstructed chromosome sequences, BAC instability, or faulty BAC tracking. The 1013 remaining successful GSTs (4.1%) were amplified from purified genomic DNA (Col-4 ecotype). In total, 20,388 (96.1%) GSTs were produced as determined by the detection of a single PCR product of the expected size in agarose gel electrophoresis. All primary amplification results included gel images can be studied online via the CATMA database. As previously described (Thareau et al. 2003), a random subset of 956 (3.3%) GST amplicons were sequence-validated (579 amplified from BACs and 377 amplified from total DNA template) as an additional quality control. In all cases, the sequence determined experimentally is identical to the expected GST, indicating that the collection is of very high quality. The sequence validation data are also available in the CATMA database.

Transcript Profiling With CATMA v1 Arrays: Performance Study

GSTs can easily be synthesized in large amounts using small aliquots of the diluted primary amplicons as templates, then purified and printed on glass microarrays. Platforms using the GST collection for microarray production recovered 94.3% to 98.3% of the amplicons in secondary PCRs. We investigated the performance of arrays printed with CATMA v1 GSTs in a control experiment and in a biological experiment assessing differential expression. The first experiment focuses on the dynamic range, sensitivity, and specificity of glass microarrays printed with Arabidopsis GSTs in a series of spiked hybridizations.

The objective was to assemble a series of targets for hybridization, spiked with in vitro synthesized and purified polyadenylated RNAs, called spike RNA, covering a range of biologically relevant concentrations. The experimental design and the subsequent interpretation of the results are based on the assumption that polyadenylated messenger RNAs constitute 1% of the total RNA, that a cell contains on average 300,000 transcripts, and that the average transcript length is 1500 nt.

All hybridizations were performed with the same RNA extracted from aerial parts of germinating Arabidopsis seedlings and completed with various amounts of spike RNAs. They were chosen based on the following criteria: (1) the corresponding cDNA clones were available, and (2) their expression had not been previously detected in shoot material. A total of 14 spike RNAs were reverse-transcribed in vitro, although one of them failed quality controls and was removed from further analysis (Supplemental Table S3). The spike RNAs were added to the shoot total RNA to generate seven RNA mixes containing 10^4, 10^3, 10^2, 10, 1, 0.1 or zero copies per cell of each (Fig. 4A). The spike concentrations were staggered to obtain in each of the scaled RNAs the same total amount of spike RNAs in a weight ratio of 1:2520 versus shoot total RNA. The spike RNAs were present in seven pairs that followed the same concentration scale in the RNA mixes (each pair is represented as a separate bar in the 14 graphs; Fig. 4A). They were compared with a unique reference RNA, with all spike RNA species at a concentration of 100 copies per cell, in a total of 14 hybridizations including both dye-swap configurations (Fig. 4A).

The dynamic range of the spike RNA signal ratios detected by the GSTs covered three to four logs in close-to-perfect linear dose-response curves. The sensitivity reached in this set of hybridizations was well below 10 copies per cell (Fig. 4B). However, the limit of sensitivity could not be calculated exactly because some of the genes coding for the spike RNAs might have been transcribed at low levels in the shoot sample. Probe specificity was investigated by studying whether the increase in spike RNA concentration significantly drove up the signal associated with GSTs other than the cognate probes. Such patterns could not be detected for any of the 14 genes tested, even though a spike RNA at its highest concentration represented an estimated 3% of the mRNA pool. BLAST searches indicated that the closest nonidentical GSTs that matched spike RNA sequences shared segments only of 17 ungapped nucleotides or E-values as low as 2.4 \times 10^{-13}. Taking into consideration the entire microarray data set, no correlation was observed between GST length and signal (data not shown).

Overall, these results indicate that CATMA arrays provide an excellent dynamic range, a sensitivity that is comparable to the...
Transcript Profiling With CATMA v1 Arrays: A Biological Test Case

To demonstrate the analysis of differential gene expression based on CATMA v1 arrays, we studied the changes induced by the treatment of Arabidopsis seedlings with the phytohormone indole-3-acetic acid (IAA) at physiological concentration. Seedlings were germinated and grown in liquid medium, then treated with the addition of IAA at 1 µM, 10 d after germination for zero, 30, 120, and 240 min. In total, nine microarray hybridizations were performed according to a complete experimental loop design (Supplemental Fig. S4). The statistical analysis of the hybridization data indicated that 1123 GSTs showed significant expression differences (p < 2.37 × 10−4) between time points following treatment (Supplemental Table S4). The transcript profiles of well-documented auxin-responsive genes were surveyed. Fold-induction was analyzed for members of the Aux/IAA, SAUR, and GH3 gene families represented in the array (Hagen and Guilfoyle 2002; Liscum and Reed 2002). Among these 60 genes, 16 appeared as differentially expressed, and all but one had an upward trend (Fig. 4C), confirming the transcriptional up-regulation of the three selected gene families upon application of exogenous IAA, even at the 1 µM level.

In addition, we investigated whether the global interpretation of the transcript profiles provided relevant information about the biological mechanisms at play in response to the auxin treatment. For that purpose, clusters of genes were constructed that correlated with theoretical up- or down-regulation patterns. Certain biological themes associated with these genes and defined in the controlled GO vocabulary (Rhee et al. 2003) were significantly overrepresented (Table 1): among them, as expected, the GO term “response to auxin stimuli,” but also categories indicative of an increase in protein synthesis characterized by up-regulated genes in the later time points, and of modification in energy metabolism characterized by down-regulated genes. Similarly, root segments in which lateral root formation was induced synchronously by treatment with an auxin transport inhibitor (N-1-naphthylphthalamic acid [NPA]) followed by transfer to growth medium containing auxin 1-naphthalene acetic acid (NAA) had an increased expression of genes involved in translation initiation, ribosome biogenesis, and assembly within 4 h following the transfer (Himanen et al. 2003).

As usual in genome-scale surveys of transcriptional activity, a large portion of the differentially expressed genes (334; 29.7%) had no known function. More peculiar to our case, 76 (6.8%) of the GSTs showing differential expression did not match with any gene model in the TIGR release 5.0 genome annotation, suggesting that the independent EuGene genome annotation that led to the initial GST design offers useful complementary information (http://www.psb.ugent.be/bioinformatics genomes_ath_index.php).

These observations demonstrate that CATMA v1 microarrays provide a robust platform for the genome-scale analysis of transcriptional regulation: they allow the detection of known gene expression profiles, and they are useful for the identification of relevant biological themes via the statistical analysis of functional annotations linked to differentially expressed genes. The analytical power of the GST arrays will improve with the expansion of the GST collection and the increase in genome coverage.

Large-Scale Cloning Into Plasmids

The use of GSTs as PCR amplicons fixed to a solid support is limited to hybridization with target nucleic acid in solution. The introduction of these double-strand DNA segments into bacterial vectors allows their convenient distribution as bacterial glycerol
Table 1. Functional Characterization of Gene Clusters

<table>
<thead>
<tr>
<th>GO term</th>
<th>EASE score</th>
<th>Gene symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster A. Up and early genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP response to auxin stimulus</td>
<td>1.91e-5</td>
<td>AT1G04240; AT1G15580; AT1G2830; AT3G23030; AT4G34760; AT4G34770; AT4G37390</td>
</tr>
<tr>
<td>MF heat shock protein activity</td>
<td>5.99e-3</td>
<td>AT3G09440; AT5G02490; AT5G6010; AT5G6030</td>
</tr>
<tr>
<td>BP protein folding</td>
<td>3.42e-2</td>
<td>AT3G09440; AT3G44110; AT5G02490; AT5G6010; AT5G6030</td>
</tr>
<tr>
<td>MF transcription factor activity</td>
<td>4.74e-2</td>
<td>AT1G04240; AT1G15580; AT1G2910; AT1G25440; AT2G31280; AT3G02290; AT3G15540; AT3G19360; AT3G23030; AT3G38120; AT3G60630; AT5G25190; AT5G45980; AT5G65320</td>
</tr>
<tr>
<td><strong>Cluster B. Up and late genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP response to auxin stimulus</td>
<td>3.14e-11</td>
<td>AT1G04240; AT1G09700; AT1G15580; AT1G2830; AT2G14960; AT3G07390; AT3G23030; AT4G27260; AT4G34760; AT4G34770; AT4G37390; AT4G38860; AT5G47370; AT5G54510</td>
</tr>
<tr>
<td>MF alcohol dehydrogenase activity</td>
<td>6.60e-4</td>
<td>AT2G47130; AT2G47140; AT3G26760; AT4G05530; AT4G131800</td>
</tr>
<tr>
<td>BP response to stimulus</td>
<td>4.05e-3</td>
<td>AT1G04240; AT1G05010; AT1G05530; AT1G09700; AT1G28130; AT1G70940; AT1G74670; AT2G13540; AT2G14960; AT2G22490; AT2G47860; AT3G07390; AT3G12710; AT3G23030; AT3G24500; AT4G27260; AT4G34760; AT4G34770; AT4G36610; AT4G37390; AT4G38410; AT4G38860; AT5G47370; AT5G54510</td>
</tr>
<tr>
<td>CC cytosolic small ribosomal subunit (sensu Eukarya)</td>
<td>1.42e-2</td>
<td>AT1G58380; AT4G34670; AT5G10360; AT5G35530</td>
</tr>
<tr>
<td>MF structural constituent of ribosome</td>
<td>1.60e-2</td>
<td>AT1G25260; AT1G58380; AT2G44860; AT3G30775; AT4G34670; AT5G10360; AT5G35530; AT5G45800</td>
</tr>
<tr>
<td>CC ribosome</td>
<td>1.90e-2</td>
<td>AT1G25260; AT1G58380; AT2G44860; AT3G30775; AT4G34670; AT5G10360; AT5G35530; AT5G45800</td>
</tr>
<tr>
<td>MF oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor</td>
<td>2.83e-2</td>
<td>AT2G45400; AT2G47130; AT2G47140; AT3G26760; AT4G05530; AT4G131800</td>
</tr>
<tr>
<td>MF metallooxopeptidase activity</td>
<td>3.41e-2</td>
<td>AT3G51800; AT3G59990; AT5G24800</td>
</tr>
<tr>
<td><strong>Cluster C. Down and early genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF pyruvate decarboxylase activity</td>
<td>6.59e-4</td>
<td>AT4G33070; AT5G17380; AT5G45960</td>
</tr>
<tr>
<td>BP aldehyde catabolism</td>
<td>2.93e-3</td>
<td>AT1G06130; AT4G33070; AT5G17380</td>
</tr>
<tr>
<td>MF carbon-carbon lyase activity</td>
<td>1.59e-2</td>
<td>AT4G25290; AT4G33070; AT4G38970; AT5G17380; AT5G45960</td>
</tr>
<tr>
<td>BP glyoxylate catabolism</td>
<td>2.09e-2</td>
<td>AT1G04240; AT1G05010; AT1G05530; AT1G09700; AT1G28130; AT1G70940; AT1G74670; AT2G13540; AT2G14960; AT2G22490; AT2G47860; AT3G07390; AT3G12710; AT3G23030; AT3G24500; AT4G27260; AT4G34760; AT4G34770; AT4G36610; AT4G37390; AT4G38410; AT4G38860; AT5G47370; AT5G54510</td>
</tr>
<tr>
<td>BP metal ion transport</td>
<td>2.84e-2</td>
<td>AT1G14660; AT1G70300; AT2G36950; AT3G05220; AT3G25410; AT4G37780; AT5G22830; AT5G47180</td>
</tr>
<tr>
<td>BP catabolism</td>
<td>3.07e-2</td>
<td>AT1G06130; AT1G11080; AT1G17290; AT1G26560; AT1G50250; AT1G71980; AT1G77120; AT2G35770; AT3G30775; AT4G26270; AT4G32840; AT4G33070; AT4G38970; AT5G01720; AT5G17380; AT5G23800; AT5G37930; AT5G3750</td>
</tr>
<tr>
<td>MF phosphotransferase activity, alcohol group as acceptor</td>
<td>4.04e-2</td>
<td>AT1G20650; AT1G53670; AT1G53420; AT1G53510; AT1G37510; AT2G12070; AT3G23150; AT4G23650; AT4G26270; AT4G32840; AT4G33080; AT4G33950; AT5G01810; AT5G20250; AT5G25110</td>
</tr>
<tr>
<td>BP amino acid biosynthesis</td>
<td>4.78e-2</td>
<td>AT1G17290; AT3G30775; AT3G66658; AT4G33070; AT5G10860; AT5G17380</td>
</tr>
<tr>
<td><strong>Cluster D. Down and late genes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP disaccharide biosynthesis</td>
<td>9.72e-3</td>
<td>AT1G23870; AT3G43190; AT5G20280; AT5G51460</td>
</tr>
<tr>
<td>BP aldehyde metabolism</td>
<td>2.65e-2</td>
<td>AT1G06130; AT2G36460; AT4G33070</td>
</tr>
<tr>
<td>BP glycolysis</td>
<td>3.08e-2</td>
<td>AT1G17290; AT2G36460; AT3G24170; AT4G26280; AT4G32840</td>
</tr>
<tr>
<td>BP response to wounding</td>
<td>3.95e-2</td>
<td>AT1G48420; AT2G30490; AT2G37040; AT2G43710</td>
</tr>
</tbody>
</table>

All listed GO terms have an EASE score indicative of biased representation below 5 × 10⁻² and are ordered according to decreasing score values. To limit redundancy, parent categories synonymous of child categories were omitted when gene content overlapped by two-thirds or more, within the same GO main category (BP, biological process; MF, molecular function; CC, cellular component).

stocks in multiwell plates. More importantly, it is the first step for subcloning into vectors designed for functional assays based on gene-specific tags, such as RNAi. For that purpose, the GST collection was introduced into a bacterial entry plasmid using the Gateway recombinational cloning technology (Hartley et al. 2000; Walhout et al. 2000). This system is advantageous because it permits the efficient cloning of large numbers of DNA fragments regardless of their sequences. Furthermore, fragments can be transferred straightforwardly from these entry clones into any destination vectors of interest, provided the appropriate recombination sites and selectable cassettes are combined.

The attB1 and attB2 recombination sites were added to the respective 5'-end and 3'-end of the original primary amplicons by PCR. To this end, each secondary amplicon was synthesized with one of 24 upstream column oligonucleotides made from the attB1 sequence fused to a “t” extension, and one of 16 downstream row oligonucleotides made of the attB2 sequence fused to an “r” extension (Fig. 3C; Supplemental Table S1). The attB1-GST-attB2 amplicons were then recombined with the attP1-cd-B-attP2 cassette in the pDONR207 vector (Fig. 3D). The successful insertion of a GST as an attL1-GST-attL2 cassette into an entry clone was confirmed by the analysis of the insert size: tertiary validation amplicons were produced by PCR with the DNR5 and DNR3 oligonucleotides anchored on the backbone of the vector, beyond the attL sites, and sized by agarose gel electrophoresis (Fig. 3E). For this purpose, template DNA was extracted from pooled *Escherichia coli* clones selected in gentamycin liquid medium after the BP clonase reaction and bacterial transformation.
Because this reaction is highly efficient, it is not necessary to pick individual colonies grown on selective solid medium to recover the expected entry clones. All reactions (secondary PCR, BP clonase, E. coli transformation and growth, tertiary PCR) were performed in 96-well plates maintaining the same format as for the primary GST amplifications. As this paper goes into press, we confirmed entry clone pools for 17,304 different GSTs. The pooled E. coli strains carrying the GST entry plasmids will be distributed by the Nottingham Arabidopsis Stock Centre (NASC).

Subcloning in hpRNA Expression Vectors

The most efficient method to knock down a plant gene by post-transcriptional silencing is to express in transgenic individuals an hpRNA made of a double-strand stem with a fragment of the transcript targeted for degradation and a loop with intron sequences that are spliced out from the RNA molecule (Smith et al. 2000). Systematic reverse genetics screens based on this method required the construction of large series of recombinant DNA molecules in which, for each gene of interest, a fragment was cloned as an inverted repeat separated by the intron spacer. Because the GSTs were selected to avoid cross-hybridization between distinct genes, they were an ideal substrate for the production of a genome-scale Arabidopsis hpRNA expression plasmid collection. Conventional cloning protocols using restriction and ligation enzymes for the production of such hairpin transgenes are laborious and not conducive to high-throughput approaches. Therefore, vectors were designed for the one-step introduction of two identical inverted repeats into a destination vector, taking advantage of the Gateway technology (Wesley et al. 2001). The vector we developed for this purpose was a bicistronic expression plasmid containing the expression construct and a bar gene facilitating herbicide selection of transformed plants.

Entry clones carrying an attL1-GST-attL2 cassette were recombined with both pAGRIKOLA at attT1-cclB-attT2 cassettes in a double LR clonase reaction. Note that the DNA template carrying the attL1-GST-attL2 cassette recombined in the double LR reaction could either be the purified entry plasmid (Fig. 3D) or the tertiary validation PCR product (Fig. 3E). Both templates yielded >95% subcloning efficiency in 96-well-plate LR clonase reactions.

For each GST, the following steps included (1) heat-shock transformation of E. coli (DH5α) with the double LR reaction products; (2) selection in kanamycin liquid medium; (3) validation of the resulting E. coli bacterial pool containing the hpRNA expression constructs; (4) purification of the pool plasmid DNA; (5) freeze-thaw transformation of Agrobacterium tumefaciens (GV3101) with plasmid DNA; (6) selection of transformed agrobacteria in kanamycin liquid medium; (7) plating of transformants on kanamycin solid medium; (8) picking of individual A. tumefaciens colonies carrying hpRNA plasmids; and (9) PCR verification that these colonies contained a correct hpRNA expression construct. Steps 1 through 6 were all performed in 96-well plates with the same format as that for the primary GST amplifications.

Because the double LR recombination could produce different intermediates depending on which repeats recombined with each other, the intron spacer in the expression clone either retained its original orientation or was flipped. Because the spacer contained two head-to-head introns, a portion of the hpRNA loop would be spliced out in plant cells regardless of its final orientation, which guarantees efficient silencing (Helliwell and Waterhouse 2003). The integrity of the artificial GST hairpin gene was monitored via multiplex PCR (Steps 3 and 9) to confirm that both GST inserts had the expected size and to determine the orientation of the intron spacer. The four primers (Agri51, Agri56, Agri64, and Agri69; Fig. 5C) were positioned in such a way that the GST subunits present in hpRNA expression plasmids

Figure 5 Construction of hpRNA expression plasmids. (A) Map of the pAGRIKOLA destination vector. (B) Schematic representation of the multiplex analysis of the expression plasmids. Numbers on the left show how many base pairs were added to the primary GSTs upon amplification with the Agri primers indicated at the top of each lane. The largest band corresponds to the amplification of both GST cassettes with the external Agri51 and Agri69 primers. Profiles in lanes 7 and 8 correspond to results in panels D and E, respectively, and are generated via amplification with all four Agri primers. The profile in lane 8 only shows results obtained with a plasmid in which the spacer intron is initially in its original orientation, as pictured in C. (C) Structure of recombinant hairpin cassette with the inverted GST repeats. GST elements are depicted as in Figure 3. Bracket numbers indicate the length of the fragment delimited by the 5′-end of each Agri primer and the 3′-end of the closest ortb site, on either strand. Configuration with the flipped spacer intron is not shown. (D) Multiplex PCR validation of double LR clonase reaction products. Each lane shows the amplification results for one E. coli pool after liquid selection (see Supplemental Table S2). (E) Multiplex PCR verification of A. tumefaciens transformants. For each initial pool, four individual Agrobacterium colonies were picked. Most tetrad contain both types of colonies, with the expression plasmid carrying the intron spacer initially either in the original or flipped orientation (see Supplemental Table S2).
were easily distinguished by size in agarose gel electrophoresis. Whether the intron spacer was in its original or flipped orientation, the amplified paired DNA fragments were Agri51-attB1-GST-attB2-Agrif6 (adding 245 bp to the primary GST length) and Agri64-attB2-GST-attB1-Agrif69 (plus 342 bp) or Agri51-attB1-GST-attB2-Agrif64 (plus 442 bp) and Agri56-attB2-GST-attB1-Agrif69 (plus 145 bp; Fig. 5B). In Step 3, the pool of E. coli transformants selected in liquid medium contained both constructs with either the original or the flipped intron spacer and gave rise to the four possible fragments in the multiplex PCR (Fig. 5D). In Step 9, individual Agrobacterium colonies originating from the acquisition of a single hpRNA expression plasmid molecule yielded prominently one of the two combinations, but often also the other in lower amounts (Fig. 5E). The latter might result from template-switching during PCR or from recombination events between the inverted GSTs occurring in Agrobacterium. As the plasmids are equally effective with the introns in either orientation, this “intron-flipping” is no more than a cosmetic nuisance. The fifth larger band generally observed in Steps 3 and 9 corresponded to the complete hairpin cassette amplified by the outside primers Agri51 and Agrif69. As this paper goes to press, we confirmed GST hpRNA expression clones for 81.36 different GSTs. The pooled E. coli strains carrying the hpRNA expression plasmids will be distributed by the Nottingham Arabidopsis Stock Centre (NASC).

Phenotypes of Silenced Arabidopsis Lines
As proof of concept, a medium-scale transformation project is currently underway to analyze randomly selected hpRNA clones. A. tumefaciens containing the hpRNA constructs was used to transform Arabidopsis thaliana (Col-0) by the conventional floral dip method (Clough and Bent 1998). Seeds from transformed plants were germinated under BASTA selection, and a minimum of 10 individual T1 transformants per construct were selected to provide a collection of T2 RNAI lines. Preliminary results were encouraging and suggested that the hpRNA constructs could be effectively deployed to generate an “interference” series in T1 Arabidopsis lines. We noted a varied degree of penetrance in phenotypes from individual lines that might result from different T-DNA copy numbers, positional effects, or from effective interference initiating at different stages of plant development. T2 lines have yet to be tested to evaluate the heritability and reproducibility of the observed phenotypes. We report here three illustrative examples in which phenotypic series were obtained useful for determining the function of the target gene product.

The first example indicated successful silencing of the target gene CesA7 (Atsg17420; GST ID: CATMA5a15680) encoding a component of cellulose synthases involved in the production of secondary cell walls. The cesA7 knockout mutant is smaller and grows more slowly than wild type. The cellulose content of secondary cell walls is greatly reduced in the mutant plants, resulting in less rigid stems unable to keep a normal upright position, which leads to a characteristic drooping phenotype (Turner and Somerville 1997; Taylor et al. 1999). The same characteristic phenotype was observed to different degrees of penetrance in each of the 10 T1 transformants obtained with this hpRNA clone (Fig. 6A). This example demonstrates that GST-induced silencing could effectively mimic a null mutant and give the phenotype expected when the target gene is not expressed.

The second example was the effect of silencing At1g20260 (CATMA1a19260), a gene encoding the vacuolar-type H+-ATPase subunit B3 (VHA-B3). V-type H+-ATPases are ubiquitous eukaryotic heteromeric enzyme complexes that acidify endomembrane compartments and are essential for many transport processes involved in development and tolerance to environmental stress (Sze et al. 2002). Classical forward genetics studies in Arabidopsis have characterized a single mutation in V-type ATPases, the det3 mutant (Schumacher et al. 1999). det3 plays an important role in the control of growth and morphogenesis of Arabidopsis seedlings and exhibits severe dwarfing resulting from a reduction in cell expansion. Interestingly, these developmental effects are
more pronounced in cells of the hypocotyl, petioles, and inflorescence stems. Despite the severe phenotype, det3 is a weak allele and, consequently, it is likely that lesions in other subunits of the Arabidopsis V-type ATPases have not been described because they lead to lethality, as is the case for null mutations in Neurospora (Ferea and Bowman 1996) and Drosophila (Davies et al. 1996; Guo et al. 1996). The det3 mutant may interfere with signal transduction pathways controlling meristem activity, possibly by regulation through multiple phytohormones (Schumacher et al. 1999). Here, we observed severe dwarfing in the rosette leaves in transgenic plants specifying hpRNA to the H+–ATPase subunit B3 (Fig. 6B, insets). Transgenic plants displayed increased serrations and some leaf and laminar distortions. The majority of the transformants showed some degree of dwarfing that became more pronounced as the plants developed. The range of whole plant phenotypes is shown in Figure 6B. The severity of the symptoms was in agreement with those predicted from the det3 studies and supported the hypothesis that vascular ATPases are essential enzymes for maintaining cellular homeostasis. This example demonstrates that GST-induced silencing can be used to obtain mutants deficient in expression of an essential gene, difficult or impossible to obtain by classical insertional mutagenesis.

The third illustration of targeted silencing concerns At1g20300 (CATMA1a19300), encoding a gene product with 11 pentatricopeptide repeats (PPRs) that is predicted to be imported into mitochondria. The function of this gene is completely unknown, but homologs from other plant species are involved in the control of male fertility by regulating expression of mitochondrial "sterility" genes (Bentolila et al. 2002; Brown et al. 2003; Desloire et al. 2003; Kazama and Toriyama 2003; Koizuka et al. 2003; Komori et al. 2004). Eight out of the 10 selected Arabidopsis lines expressing the At1g20300 hpRNA construct developed strong growth phenotypes (representative individuals are shown in Fig. 6C). In general, the suppressed plants were smaller in stature and showed a trend toward early flowering. We observed a marked decrease in fertility correlated with this growth suppression, as illustrated by the reduced or aborted siliques of six individual lines (Fig. 6C, inset). Even in plants with little apparent reduction in phenotype, seed set was markedly reduced relatively to wild-type plants. This example demonstrates that GST-induced silencing can and will give invaluable information and the fact that constructs yielding gross morphological alterations had no obvious size bias (data not shown) suggest that GSTs across the chosen length range may induce gene silencing when expressed as hairpin RNA.

**DISCUSSION**

**From Sequence to Function**

Initial analyses of the Arabidopsis genome sequence described fewer than 25,000 genes (The Arabidopsis Genome Initiative 2000). Experimental evidence of gene structure (from full-length cDNAs) and expression was available for only a small proportion of these putative genes and experimental evidence of function available for even fewer. In the past three years, giant strides have been made in compiling cDNA sequences (Haas et al. 2002; Seki et al. 2002; Castelli et al. 2004) and expression data (Schaffer et al. 2000; Kim et al. 2003). Together with improvements in automatic annotation routines and better use of expert manual annotation, these data have enabled us to have a much better view of the coding potential of the Arabidopsis genome (Wortman et al. 2003; http://www.psb.ugent.be/bioinformatics/genomes_ath_index.php). However, these improvements in structural annotation have not been matched by equivalent advances in functional annotation. Functional analysis of Arabidopsis genes or proteins remains a long and painstaking task, but one that can be considerably facilitated by constructing and widely distributing shared functional genomics resources such as collections of insertion mutants (Arzibech-Leehan and Feldmann 1997; Parinov et al. 1999; Speelman et al. 1999; Sussman et al. 2000; Samson et al. 2002; Sessions et al. 2002; Alonso et al. 2003; Rosso et al. 2003) or cDNA clones (Yamada et al. 2003). The GST collection described here is a novel, versatile resource for functional genomics with applications in expression profiling, reverse genetics, and any other approach that requires gene-specific hybridization.

**The GST Collection**

The GSTs are formatted to constitute a resource that is durable, versatile, and easy to multiply and distribute. Because of the standard amplicon structure, a set of only 40 universal primers is sufficient for the reamplification of the entire collection, facilitating the iterative production, at limited cost, of the large amount of GST DNAs necessary for the printing of glass microarrays. The extensions that distinguish all rows and all columns help resolve the well-to-well cross-contamination unavoidable during the manipulation of DNA and clone repertoires stored in multwell plates. As illustrated here with the addition of Gateway utiB sites, the extension primers are also useful for adding any sequence to either side of the GSTs simply by PCR amplification, making them compatible with a wide range of cloning techniques and vectors. GST cloning into Gateway entry vectors has established the collection as a set of bacterial clones, which greatly facilitates distribution of the resource.

Functional genomics projects have to keep up with evolving genome annotation. The flexible GST format makes it possible to add amplicons, plasmids, and expression vectors as additional transcription units are described. Because the GSTs are amplified from genomic DNA templates, any sequence/gene of interest is straightforwardly accessible. Thus, the GST repertoire provides a better genome representation than the best combined collection of cDNA clones, in which for multicellular eukaryotes up to half of the predicted genes may be missing. In that respect, it is interesting to note that transcript profiling data obtained with CATMA v1 microarrays and mutant phenotypes observed in lines expressing GST hpRNA indicate that genes not documented in the latest genome annotation (TIGR release 5.0), but identified in an independent structural annotation effort (http://www.psb.ugent.be/bioinformatics/genomes_ath_index.php), are transcriptionally functional. This observation is in agreement with a study based on a whole-genome oligonucleotide tiling array showing that many transcription units have been overlooked in the annotation process (Yamada et al. 2003). Altogether, the experimental data stress the necessity to integrate the results obtained from independent prediction tools together with information about function, such as expression profiles and mutant phenotypes, to continuously improve the genome annotation.

Much thought went into designing the GST resource to achieve maximum versatility and ease of use. In addition to the applications in transcript profiling and reverse genetics described below, other uses for the resource can be imagined. For example, the amplicons could be synthesized with either 24-column or 16-row biotinylated-extension primers, making it possible to purify the sense or antisense single-strand GST DNAs, respectively. Such single-strand DNAs could be used as probes for microarrays.
Alternatively, they may serve as specific probes to capture corresponding clones from cDNA libraries. We envisage using the resource for recovering full-length cDNAs from genes for which these are not currently available.

**Application of the GST Collection to Transcript Profiling**

We have shown that known transcript profile changes can be reproduced in experiments performed with CATMA microarrays. We also illustrate that the integration of genome-scale functional annotation and transcriptome data obtained with the GST collection yields valuable information for the study of biological mechanisms in *Arabidopsis*.

Multiple studies, often supported by commercial providers, argue that oligonucleotide arrays offer the best solution for microarray transcript profiling experiments (e.g., Zhu and Wang 2000; Hughes et al. 2001). In our hands, CATMA v1 microarrays perform at least as well in terms of dynamic range, sensitivity, and specificity as oligonucleotide microarrays distributed by commercial providers (J. Allemeersch, S. Durinck, R. Vanderhaeghen, P. Alard, R. Maes, K. Seeuw, T. Bogaert, K. Coddens, K. Deschouwer, P. Van Hummelen, et al., in prep.). Moreover, because the creation of the collection was supported by academic laboratories, all information relative to CATMA arrays, including DNA feature design and sequence, is freely available, and GST resources are distributed for a nominal fee.

The CATMA array format permits applications incompatible with other platforms. The collection is flexible and GST amplicons can be cherry-picked for the production of dedicated arrays or combined with other probe sets. Because they are shared by all DNA features, oligonucleotides carrying the extension sequences and labeled with fluorophores may also constitute calibrated references for the study of mRNA abundance (Dudley et al. 2002).

**Application of the GST Collection to Reverse Genetics**

Posttranscriptional gene silencing is known to be helpful for the study of gene function in *Arabidopsis* and in many other plant and metazoan species. Yet, no systematic attempts to carry out gene silencing at the genome scale have yet been completed in plants. With the GST collection described here, the technical procedures have been put in place to generate the tens of thousands of transformants required for such an endeavor.

We have shown that the introduction of a transgene expressing homologous GST hairpin RNAs into wild-type plants may result in phenotypes that mimic the effects of known null mutations or that indicate the function of essential or unknown genes. Pursuing this effort with a significant portion of *Arabidopsis* genes should provide useful information on the overall efficacy of the approach, the phenotypic classes that can be obtained, and the specificity and stability of the knockdown effects.

We anticipate that the described RNAi tools will complement the numerous *Arabidopsis* insertion mutant collections that have been built up over the last decade. Insertional mutants constitute a valuable resource, especially since the systematic determination of the sequences flanking the T-DNA or transposable element inserts and the publication of that information via online databases. Yet, insertional mutagenesis in plants suffers from significant drawbacks inherent to the approach: (1) the introduction of inserts into the genome cannot be directed and hundreds of thousands of transformants must be generated to reach saturation; (2) therefore, only a very few genotypes have been used as recipients in *Arabidopsis* reverse genetics programs; (3) most informative insertions result in null mutations, and they rarely yield hypomorphic alleles displaying a range of phenotypes; (4) embryo-lethal mutants are difficult to recover and gamete-lethal mutants are lost; and (5) combining multiple insertion mutations is cumbersome. In addition, available flanking sequence tags are often hundreds of base pairs away from the actual insertion site and the majority of phenotypes in T-DNA insertion lines do not result from a mutation tagged with a selectable marker. Consequently, obtaining validated homozygous insertion lines in a locus of interest is not trivial. In contrast, targeted gene silencing only requires a few independent transgenics for the study of any given gene, and the same silencing construct can easily be introduced into multiple genetic backgrounds, including mutants of particular interest to investigate potential interactions involving the target gene. Most knockdown lines are not null mutants and generally show diverse phenotypes typical of allelic series. Finally, essential genes may be studied via partial or conditional silencing.

A genome-scale GST collection is attractive for any species in which RNAi silencing is a practical reverse genetics method. In plants it could form the basis of systematic knockdown programs with a variety of approaches: hpRNA expression as illustrated here, virus-induced gene silencing (Lu et al. 2003) or transitive silencing (Van Houdt et al. 2003). Because the GSTs are available as Gateway entry clones, the same basic resource is useful regardless of the strategy, provided the transgene inducing silencing can be built via recombinational cloning. The enhanced plant silencing vectors that will undoubtedly be developed in the coming years will be compatible with the CATMA collection.

**Conclusion**

To our knowledge, this is the first attempt to construct a clonable GST resource covering the majority of genes in an organism. Previously constructed sequence collections are either oligonucleotide-based and thus unclonable, or are EST, cDNA, or ORF collections and thus not designed to be gene-specific. This novel and versatile resource has great promise for large-scale transcript profiling and reverse genetics projects, and is being actively used in two major projects of this type, CAGE and AGRIKOLA. However, its uses are not limited to these applications. We hope that this resource will be a cornerstone of other future projects. The principles and technologies described here are in no way specific to *Arabidopsis* or plants in general. Similar resources should prove equally valuable for functional genomics projects in other organisms.

**METHODS**

**SPADS GST Design**

All GSTs were selected with the Specific Primer and Amplicon Design Software (SPADS; Thareau et al. 2003). The database used as the *A. thaliana* reference nuclear genome contains the five chromosome pseudomolecule sequences (TIGR ASMBL_ID 68170, 51595, 68173, 68164, and 68172) available from TIGR on January 28, 2001 (ftp://ftp.tigr.org/pub/data/a_thaliana/ath1). Parameters for the design of the gene-specific PCR oligonucleotide sequence (without its 5′-extension) were 18 to 25 bases *Tm* of 50° to 65°C with a maximum Δ*Tm* of 4°C between paired primers. All other parameters were as described in Results or elsewhere (Thareau et al. 2003). All gene (At) codes listed as GST matches refer to the TIGR release 5.0 genome annotation.

**DNA Synthesis**

The 40 arbitrary 5′-extensions contain 17 nt, eight to nine being C or G. To avoid hairpin structures, their sequences have no self-complementary repeats of four or more bases separated by three or more bases. They do not contain stretches of more than four contiguous G or C, or any ATG or stop codon in the sense orientation. To avoid PCR primer dimers, the last four bases of an extension (column or row) do not match any sequence in that same extension or in the opposite subset of extensions (rows or
columns). All were tested in silico to minimize priming against the BAC plasmid backbone and the *E. coli* genome.

The primary amplification products were synthesized as follows. A master mix base of 10 µL of 10× Invitrogen PCR buffer (50 mM KCl, 20 mM Tris-HCl at pH 8.3), 3 µL of 50 mM MgCl₂ and 75 µL of double-distilled water (ddH₂O) was made and stored at 4°C. On day of use, 2 µL of a 10 mM dNTP stock and 0.4 µL of Invitrogen Platinum Taq (5 U/µL) were added to the master mix base. Then, 91 µL of the complete master mix was dispensed per well into an ABgene semiskirted PCR plate. GST-specific primers were provided by Sigma-Genosys as mixed pairs in 96-well plate format. These were diluted with 10 mM Tris (pH 8.0) to make a 10 µM working stock, and 4 µL was dispensed into the matching well of the PCR plate containing the complete master mix. Finally, 2.5 ng of a BAC DNA miniprep, containing the gene to which the specific GST primers were designed, was added to the appropriate well. When a BAC clone was not available, 5 µL of a 1 ng/µL Columbia genomic (Col-0) DNA prep was substituted. PCR amplification was carried out using the following program: hold for 5 min at 94°C; 11 cycles of 30 sec at 94°C, 30 sec at 63°C (temperature reduced by 1°C per cycle), and 30 sec at 72°C; 30 cycles of 30 sec at 94°C, 30 sec at 63°C, 30 sec at 72°C; 5 min at 72°C; and 1 min at 4°C for 1 min. To assess quality of product, 1 µL from each well was run on a 2% agarose gel. No purification of the primary product was carried out.

**Microarray Spiking Experiment**

*Arabidopsis* Col-0 seeds were sown on plates containing agar-solidified culture medium (1× MS [Duchefa], 0.5 g/L MES at pH 6.0, 1 g/L sucrose, and 0.6% plant tissue culture agar [LabM]). After germination, plates were cold-stratified for 7 d at 4°C and subsequently transferred to a growth chamber kept at 22°C (24 h photoperiod with 16 h of light at 65 µE/m² sec PAR). Whole shoots were harvested at growth stage 1.04 (Boyes et al. 2001), frozen immediately in liquid nitrogen, and stored at −70°C until needed. Frozen seedling material was ground, and total RNA was extracted using TRIzol reagent (Invitrogen). The spiking experiment protocols will be detailed further elsewhere with the complete experimental loop design. They included eight hybridizations per slide. Fg intensity measurements below background threshold (i.e., Fg < Bg + 2Bg standard deviation) were replaced by random values, or channel effects with corresponding to the intergenic regions were selected, and the median of these control Fg intensities was computed for each slide. Fg intensity measurements below background threshold (i.e., Fg < 8B + 2 × 8B standard deviation) were replaced by the corresponding slide median Fg intensity value to reduce the impact of background variability in the statistical analysis of differential expression. The leveled data were normalized as described above, in which some effects are considered fixed and others random, were used to identify genes that were significantly differentially expressed in the experimental time course (Wolfinger et al. 2001). First, array and channel effects were removed from the expression responses by a linear normalization ANOVA model of the form Yijkl = µ + Ak + ADi + kl + (i (i = 1,2,120, the number of microarray features in the array) indexing the selected GSTs, j (j = 1...4) indexing the time points (treatments), Ak representing the random array effects (k = 1...9), (ADi)jk representing the random array × dye combinations, or channel effects with k = 1...9 arrays and l = 1...2 dyes, and eijkl representing the molar. The fully automated protocol is highly efficient for 50% dimethyl sulfoxide (DMSO) solution on Ultra-GAPS slides (Corning) with a QArray arrayer (Genetix), then linked to the slide surface by UV treatment (250 mJ/cm²).

The biological samples were prepared as follows. *Arabidopsis* seeds (Col-0) were germinated and grown in 0.5× MS (Duchefa) liquid medium supplemented with 0.5% sucrose at 22°C (24 h photoperiod with 16 h of light at 75 µE/m² sec PAR). After 10 d (2 h after dawn), seedlings were treated with 1 µM indole-3-acetic acid for a period of 0, 30, 120, and 240 min. At the end of the treatment, the seedlings were washed once with an excess of MS medium with 0.5% sucrose for 5 min, then immediately frozen in liquid nitrogen and stored at −70°C until needed. Each time-point sample was made of a pool of seedlings grown simultaneously in three independent vials. Frozen seedling material was ground and total RNA was extracted using the QIAGEN RNAeasy kit.

For each biological sample, 20 µg of total RNA was annealed to 10 µg of anchored oligo(dt) primer (dT20VN; MWG Biotech AG) and reverse-transcribed to cDNA at 42°C during a 105-min reaction. The 30-µL reaction contained 2 mM dNTPs (dTTP:ami-noallyl-dUTP in 1:4; unmodified Amersham Biosciences, modified Sigma-Aldrich), first-strand buffer (50 mM Tris-HCl at pH 8.3, 75 mM KCl, 3 mM MgCl₂), 10 mM DTT, and 400 U SuperScript II (Invitrogen). RNA strand hydrolysis (15 min at 70°C in the presence of 150 mM NaOH) was followed by a purification using the Epicentre spin columns (QIAgen) with the provided wash and elution buffers replaced by 80% ethanol and 100 mM NaHCO₃ (pH 9.0), respectively. The monofunctional NHS-ester Cy3 or Cy5 fluorophores (Amersham Biosciences) were coupled to the purified cDNA DNA at a 90-min incubation at room temperature. Prior to pooling and spin column purification, ester groups on the unincorporated fluorophores were inactivated using 730 mM hydroxylamine.

Slides were hybridized for 16–18 h using a two-step protocol in the GeneTec hybridization station (Genomic solutions). Prehybridization for 45 min at 42°C (5× SSC, 1% BSA [Sigma-Aldrich], 0.1% SDS, 40 µg of poly(dA) [Sigma-Aldrich], and 20 µg of tRNA [Sigma-Aldrich]) was followed by a hybridization at 42°C with the labeled material in a buffer containing 5× SSC, 25% formamide, 0.1% SDS, 40 µg of poly(dA), and 20 µg of tRNA. Hybridized slides were washed with 2× SSC and 0.1% SDS at 42°C, followed by 0.1× SSC and 0.1% SDS at room temperature, and finally by three washes with 0.1× SSC at room temperature. Slides were scanned at a 10-µm resolution using the G2565BA DNA microarray scanner (Agilent Technologies). The acquired TIFF images were processed using the GenePix 4.1 software (Axiom Instruments).

The performed microarray hybridizations constituted a complete experimental loop design. They included eight hybridizations for the comparison of T0, T30, T120, and T240 samples, with reciprocal dye swap, and a T0 self-hybridization (Supplemental Fig. S1). The raw data were preprocessed as follows. Spots corresponding to the intergenic regions were selected, and the median of these control Fg intensities was computed for each slide. Fg intensity measurements below background threshold (i.e., Fg < 8B + 2 × 8B standard deviation) were replaced by the corresponding slide median Fg intensity control value to reduce the impact of background variability in the statistical analysis of differential expression. The leveled data were normalized as described above, in which some effects are considered fixed and others random, were used to identify genes that were significantly differentially expressed in the experimental time course (Wolfinger et al. 2001). First, array and channel effects were removed from the expression responses by a linear normalization ANOVA model of the form Yijkl = µ + Ak + (ADi)jk + eijkl with i (i = 1–2120, the number of microarray features in the array) indexing the selected GSTs, j (j = 1...4) indexing the time points (treatments), Ak representing the random array effects (k = 1...9), (ADi)jk representing the random array × dye combinations, or channel effects with k = 1...9 arrays and l = 1...2 dyes, and eijkl representing the
random error. Second, residuals, denoted as $r_{ijkl}$, and computed by subtracting the fitted values for the effects from the observed values $y_{ijkl}$ were then subjected to 21,120 gene-specific models of the form $r_{ijkl} = \mu + (GD)_{ijkl} + (AT)_{ijkl} + (GA)_{ijkl} + \gamma_{ijkl}$ partitioning gene-specific variation into fixed gene-specific dye effects $(GD)_{ijkl}$, fixed time-point effects $(AT)_{ijkl}$, random spot effects $(GA)_{ijkl}$, and random error $\gamma_{ijkl}$. In these gene-specific models, the spot effect serves to account for the spot-to-spot variability inherent in spotted microarray data. To test differences between time points, we used the Wald statistic, which should follow approximate $y_{ijkl}$-value cutoff at the Bonferroni value of 5% divided by 21,120, or $2.37 \times 10^{-6}$.

Transcript profiles close to theoretical patterns were identified with the template-matching module of the MultiExperiment Viewer (MeV) software package (Saeed et al. 2003: the four patterns defined up-regulated genes reaching a maximum plateau at 30 and 120 min (up and early) or 120 and 240 min (up and late) or down-regulated genes with a minimum plateau at 30 and 120 min (down and early) or 120 and 240 min (down and late). The biological theme representation analysis performed with the EASE software (Hosack et al. 2003) was based on the TAIR Arabidopsis GO term list (ATH_GO_20040217.txt available at ftp://tairpub.tairpub@ftp.arabidopsis.org/home/tair/Gene_Ontology/).

Recombinant Cloning and Plant Transformation

For the addition of the attB sites to the GSTs, primary ampiclons were used as templates and amplified by Taq polymerase (NEB) with 16 pmoles of the relevant row and column primers (Supplemental Table S1) in 20-µL reactions. This step and all subsequent procedures were carried out in 96-well microtiter plates. The PCR conditions were as follows: 5 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 55°C, 2 min at 72°C, and terminated by 5 min at 72°C. Plasmid DNA was prepared from the transformed bacteria by alkaline lysis (Montage Plasmid Miniprep 96 Kit; Millipore). Arabidopsis plants were transformed using the floral dip method (Clough and Bent 1998).

For the addition of the attP sites to the GSTs, two 2-mL 96-well plates were set up to contain 0.5 µL of the transformed cells and shaken for 20 h to an OD of 1.4. The plates were then incubated at 25°C for 2 min after transformation, except that 50 µg/mL kanamycin was used to select transformed cells. The suspension of transformed cells (0.5 µL) was verified by PCR using 20 pmoles of each of the primers Agri51/Agri56/Agri64/Agri69 (Fig. S1; Low DNA Mass Ladder; Invitrogen). The PCR conditions were as follows: 5 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 55°C, 2 min at 72°C, and terminated by 5 min at 72°C. Plasmid DNA was prepared from the transformed bacteria by alkaline lysis (Montage Plasmid Miniprep 96 Kit; Millipore). Arabidopsis plants were transformed using the floral dip method (Clough and Bent 1998).

For the addition of the attB sites to the GSTs, two 2-mL 96-well plates were set up to contain 0.5 µL of the transformed cells and shaken for 20 h to an OD of 1.4. The plates were then incubated at 25°C for 2 min after transformation, except that 50 µg/mL kanamycin was used to select transformed cells. The suspension of transformed cells (0.5 µL) was verified by PCR using 20 pmoles of each of the primers Agri51/Agri56/Agri64/Agri69 (Fig. S1; Low DNA Mass Ladder; Invitrogen). The PCR conditions were as follows: 5 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 55°C, 2 min at 72°C, and terminated by 5 min at 72°C. Plasmid DNA was prepared from the transformed bacteria by alkaline lysis (Montage Plasmid Miniprep 96 Kit; Millipore).

ACKNOWLEDGMENTS

The CATMA Consortium efforts were supported by multiple grants, namely, No-2000-009 from the French Génopole program to P. Hilson and P. Rouzé, BMBF-312274 to H. Lehrach, BI02000/3094/E from the Spanish Biotechnology program–MCYT to J. Paz-Ares, grants from the Wallenberg Consortium Nord (RNA platform 1b) to R.P. Bhalaria and J. Lundeberg, from the University of Lausanne (Genomics Project) to P. Raymond and E.E. Farmer, and from the BBSRC as part of the GARNET consortium (UK Investigating Gene Function Initiative) to J. Beynon. The construction of a Compendium of Arabidopsis Gene Expression profiles based on CATMA arrays is supported by the CAGE demonstration project (QLRT-2001-02035). The transfer of the CATMA GSTs into bacterial plasmids and the production of silenced Arabidopsis lines is supported by the AGRIKOLA grant (QLRT-2001-01741). The CAGE and AGRIKOLA projects are funded in the 5th European Framework Programme. We are grateful to Stefanie Albrecht, Caroline Buysschaert, Beate Hoffman, Yolanda Leo del Puerto, Anders Holmberg, Antje Krüger, Morten Lukacs, Ruth Maes, Peter Nilsson, Jennie Rowley, Kurt Sætue, Rudy Vanderhaegen, and Raimundo Villarroel for technical assistance in GST synthesis, microarray production, and GST cloning; to Philippe Alard, Eric Bonnet, Björn De Meyer, David Lonsdale, and Martin Trick for assistance with the CATMA database; to Philippe Grevet for assistance with the AGRIKOLA database and Web site; to Mark A. Bennett for establishing the plant transformation database; to Wendy Byrne, Bertrand DuBreucq, Mathilde Grelen, Melanie Lück, Raphael Mercier, Hakim Mireau, Jean-Christophe Palaqui, and Matthieu Simon for assistance with production and analysis of transgenic plants; and to Uli Schneider and Gerit Beemster for helpful discussions. We thank Peter Waterhouse and Chris Hellwell for sharing with us materials prior to publication and for very useful advice on the use of plant hRNA vectors and Martine De Cock for help in preparing the manuscript. The TAMU BAC clones were graciously provided by Marcel Salanoubat (Génoscope). T. Moreau contributed to this work while on leave at the Center for Biologi- cal Sequence Analysis, Danish Technical University, Lyngby, Denmark.

REFERENCES


Downloaded from www.genome.org on December 22, 2007 - Published by Cold Spring Harbor Laboratory Press


