Plant Biotechnology 22(1), 55–58 (2005)

Short Communication

Phylogenetic profiling of *Arabidopsis* genes by monitoring spatial gene expression using cDNA microarrays

Seiyei Yamakawa[1,2*a], Kohei Ando[2a], Kazuya Yoshida[2], Akiho Yokota[2], Atsuhiko Shinmyo[2], Takayuki Kohchi[2b]

[1] Technology Department Research Association for Biotechnology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
[2] Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan
* E-mail: s-yamaka@bs.naist.jp Tel: +81-743-72-5606 Fax: +81-743-72-5607

Received December 1, 2004; accepted January 21, 2005 (Edited by H. Sano)

Abstract *Arabidopsis* cDNA microarray analysis was performed for 5,722 genes using leaf, stem, root and flower mRNA as targets. We identified 55, 22, 21, and 131 genes expressed specifically in leaves, stems, flowers and roots, respectively. Statistical cluster analysis revealed that the roots were the most distant and the leaves and flowers the most related organs with regards to gene expression. K-means clustering revealed that the clusters for leaf-specific genes tended to mainly have genes for plastid-localized proteins, while the cluster for predominantly root-specific genes was mainly secretory proteins.

Key words: *Arabidopsis thaliana*, normalized cDNA library, expression sequence tag (EST), cDNA microarray, organ-specific gene.

In higher plants, all differentiated tissues and organs originate from meristems. In above-ground parts, apical meristems repeatedly produce leaves in the vegetative stage. Cauline leaves and inflorescence are produced following transition to the reproductive stage in *Arabidopsis*. In underground parts, the process of root growth is constituted with division of primordial cells of root meristem and longitudinal cell elongation. Lateral root primordia are initiated by pericycle cell division, and penetrate through the cortex of the main root following growth. Lateral root divergences are repeated and form complicated root systems. Differentiation and morphogenesis of plant organs are regulated by coordinated gene expressions, and many key genes involved in tissue differentiation and morphogenesis in plants have been identified by molecular genetic analysis of developmental mutants. However, it is difficult to analyze the downstream genes in signal pathways or the whole gene network by mutant approaches owing to genetic redundancy. As a method for overcoming these difficulties, microarray technology has become a useful tool for comprehensive analysis of gene expression (Schena et al. 1995; Duyk 2002). In plants, microarray technology has been widely applied for identification and characterization of genes that respond to plant hormones (Goda et al. 2004) and various stresses such as salt, drought, cold (Seki et al. 2002), ozone (Tamaoki et al. 2003), and low atmospheric pressure (Paul et al. 2004). This method should also be effective for studies of development and differentiation.

Previously we reported efficient construction of an *Arabidopsis* cDNA microarray from four different tissues (Ando et al. 2004). We identified 85 genes highly expressed in leaves using a cDNA microarray of 2,537 expressed sequence tags (ESTs), and applied these data for collection of leaf-specific promoters (Yamakawa et al. 2004). In the present study, we used the microarray for comparative gene expression analysis among four organs: leaves, flowers, stems and roots as the first step in revealing the network underlying tissue differentiation mechanisms.

Arabidopsis thaliana ecotype Colombia was used. Leaves were harvested from 10- to 14-day-old plants and flower and stems from 30-day-old plants. Roots were collected from plants grown in MS medium containing 1% sucrose for three weeks. RNA purification, construction of the cDNA library and cDNA microarray preparation were carried out as reported previously (Ando et al. 2004; Yamakawa et al. 2004, Kohchi et al. 1995; Takemura et al. 1999). In total, 6,665 cDNA clones including 5,722 independent ESTs from normalized cDNA libraries were used for construction of the microarray. Preparation of cDNA target and hybridization were carried out as reported previously

[a] These authors contributed equally to this report.
[b] Present address; Graduate School of Biostudies, Kyoto University, Kyoto 606-8502, Japan

Table 1. Classification of organ-specific genes according to MIPS functional categories.

	Leaf	Root	Flower	Stem
Metabolism	18 (13)	22 (13)	1 (1)	0
Energy	14 (11)	16 (13)	0	0
Cellular organization	7 (5)	6 (5)	0	0
Cell rescue/defense/cell death/aging	1 (1)	5 (3)	1	0
Transcription	0	7 (3)	5 (5)	2
Protein synthesis	0	2	0	0
Protein destination	0	1	0	1
Cellular communication/signal transduction	0	3 (1)	0	0
Cellular biogenesis	0	2 (1)	1 (1)	0
Transport facilitation	0	2 (2)	0	1
Cellular transport/transport mechanisms	0	1 (1)	0	0
Development	0	0	5 (5)	0
Others	6	24	5	4
Unclassified proteins	26	61	9	14

Numbers in parentheses shows the numbers of genes that coincide with the multiple functional categories.

(Ando et al. 2004; Yamakawa et al. 2004). Fluorescent intensities were measured with Scan Array 4000XL (Perkin Elmer Life Sciences, Torrance, CA, USA) and analyzed using the Quant Array program (Perkin Elmer Life Sciences). The local background was subtracted from the value of each spot on the microarray. Normalization between the Cy3 and Cy5 channels was initially achieved by the global normalization method. The intensities of signals in leaves, roots, stems and flowers were measured and compared. Sixteen patterns of comparisons were performed by swapping Cy3 and Cy5 labels of target RNA. Hierarchical clustering analysis and K-means clustering were performed with the GeneSpring software package (version 5.0: Silicon Genetics, San Carlos, CA). Gene identification was performed by NCBI BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and using the MIPS *Arabidopsis thaliana* database (http://mips.gsf.de/proj/thal/index.html). For prediction of subcellular location of gene products, the Target P program (http://www.cbs.dtu.dk/services/TargetP/) was used.

Based on the results of the cDNA microarrays, we selected the genes highly expressed in leaves, stems, flowers and roots. Genes expressed in one organ at two-fold or higher than in the other three organs were defined as organ-specific genes (Table 1). Overall, 55, 21, 22 and 131 genes were selected as leaf-, flower-, stem- and root-specific genes, respectively. Of these 229 genes, 162 were functionally annotated and 67 were unknown in the MIPS *Arabidopsis thaliana* database (http://mips.gsf.de/proj/thal/db/index.html). These genes were clustered according to MIPS functional categories (Table 1). A large number of genes identified as leaf and root-specific were categorized in metabolism and energy categories, respectively. In the leaves, several genes classified as metabolism and energy genes included fructose-bisphosphate aldolase, glycelaldehyde-3-phosphate dehydrogenase, and ribulose bisphosphate carboxylase, which are required for photosynthesis. Many transcriptional factors were present in the flower-specific genes. In the roots, several genes related to metabolism were selected including alcohol dehydrogenase (ADH). In *Arabidopsis*, the ADH gene is expressed constitutively in roots, including lateral roots, but not in green aerial tissue (Chung and Ferl 1999). Transcriptional factors such as scarecrow-like gene (Sabatini et al. 2003), and stress-related proteins such as osmotin (Kononowicz et al. 1992) and HSP 90-like protein (Yabe et al. 1994), were identified as root-specific genes. In the roots, stress- or defense-related genes were also identified, reflecting that roots ordinally suffer from physical and chemical stresses.

The expression data of the 5,241 genes that gave significant signal intensities were subjected to the clustering algorithm followed by complete linkage hierarchic clustering of both the genes and experiments (Figure 1). Clustering analyses revealed that the roots had the most distant while the leaves and flowers showed the closest expression profiles. Considering the combinations of expression patterns among the four organs, expression profiles can theoretically be grouped into 16 groups: namely, they should show one of the following expression profiles: genes expressed mainly in one organ only (four groups), in two organs (six groups), in three organs (four groups), in four organs (one group), and those not expressed in any organs (one group). Since genes that did not give any detectable signals were excluded from the analysis, 5,241 genes were classified in 15 groups by K-means clustering as shown in Table 2. Cluster 6 (CL6) tended to contain many genes expressed predominantly in the flowers, CL1, CL2 and CL13 contained many genes expressed predomonantly in the stem, CL11, CL14 and CL15 included many genes expressed predominantly in the roots, and CL14 was

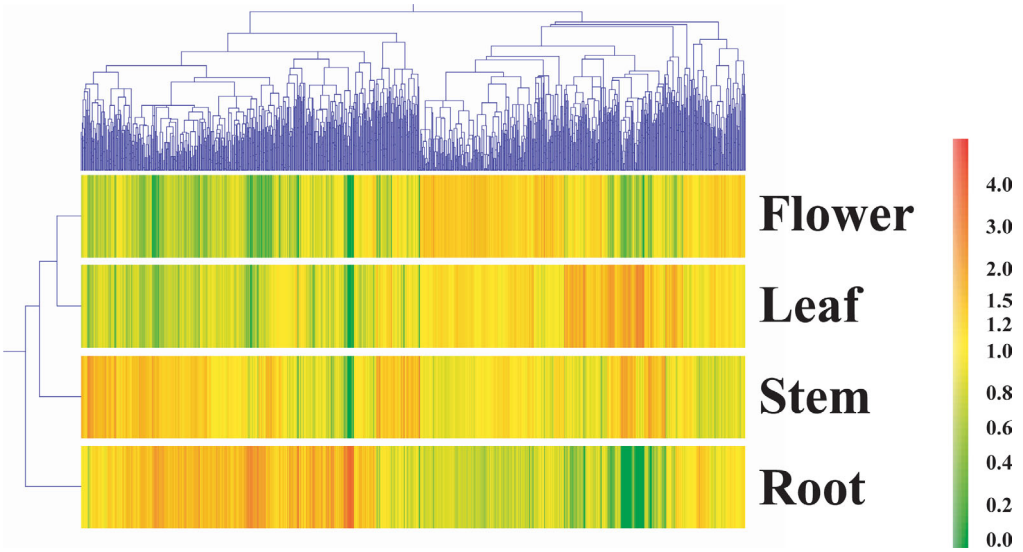

Figure 1. Hierarchical clustering displays of expression ratios in *Arabidopsis* leaves, stems, roots and flowers. In total, 5241 genes were subjected to the clustering. Each vertical line displays the expression data for one gene after normalization for the four organs as indicated. The color scale bar on the right shows the levels of gene expression. The ratios are shown under the definition that the average expression among the four organs is 1.0. The dendrograms on the upper left of the figure indicate the relationships among pairs of organ or pairs of gene in the cluster analysis.

Table 2. Target P analysis of each cluster.

Data set[a]	Number of genes	cTP[b] (%)	mTP[b] (%)	SP[b] (%)	Others[b] (%)	Expression pattern
CL1	375	38.1	9.9	13.9	38.1	Leaf, Flower/Leaf, Flower/Leaf/Stem
CL2	187	48.7	4.8	8.6	38.0	Leaf, Leaf/Stem
CL3	190	20.5	6.3	15.8	57.4	Stem, Leaf/Stem
CL4	479	12.3	8.1	15.7	63.9	Flower/Stem, Flower/Leaf/Stem
CL5	243	42.4	3.7	13.2	40.7	Leaf/Stem
CL6	679	17.2	12.2	12.4	58.2	Flower, Flower/Leaf
CL7	250	12.8	10.0	12.8	64.4	Stem, Flower/Stem
CL8	326	17.8	9.2	10.7	62.3	Root/Stem, Leaf/Root/Stem
CL9	408	11.8	8.1	9.8	70.1	Stem, Root/Stem
CL10	358	12.8	12.6	10.1	64.5	Flower/Leaf/Root/Stem
CL11	418	12.9	13.9	11.5	61.7	Root, Root/Stem
CL12	642	10.7	10.4	9.8	69.0	Root/Stem, Flower/Root/Stem
CL13	182	19.2	7.7	13.7	59.3	Leaf, Leaf/Root
CL14	86	3.5	6.9	34.9	54.7	Root
CL15	418	8.6	15.8	12.0	63.6	Root, Flower/Root
Genome	25,225	14.2	10.3	16.8	58.6	

[a] CL; Grouping of each gene by the expression organ by K-means clustering (K=15). Genome; Percentage for deduced genes in the genome was shown.

[b] Percentage of categorized sequences were given. cTP; Percentage for deduced genes that have N-terminal chloroplast transit peptide. mTP; Percentage for deduced genes that have N-terminal mitochondrial targeting peptide. SP; Percentage for deduced genes that have N-terminal secretory pathway signal peptide.

constituted by genes that showed high root-specificity. To investigate subcellular localization of the gene products in each cluster, the 5,241 genes were classified by the Target P program (http://www.cbs.dtu.dk/services/TargetP/). The results are summarized in Table 2. In classification of 25,225 genes deduced by the whole genome of *Arabidopsis* (Arabidopsis Genome Initiative 2000), the ratio of proteins with transit peptides to plastids, proteins with signal peptides to mitochondria, and secretory proteins were 14, 10, and 17%, respectively. CL1, CL2 and CL5 contained remarkably more genes that encoded plastid proteins (38, 48 and 42%, respectively) in comparison with the results of the whole genome. All these clusters had many genes that were highly expressed in the leaf or stem. CL4, which was composed of highly root-specific genes tended to be rich in secretory proteins (35%), and CL14 was composed of gene with high root-specificity. This is consistent with the fact that a large number of secretory proteins are needed for absorbing various nutrients from

the rhizosphere such as acid phosphatase (Haran et al. 2000). In the case of proteins that are transported to the mitochondria, no clusters had a conspicuous tendency.

In this study, we systematically screened organ-specific genes with a microarray of 5,241 cDNA clones. The total number of *Arabidopsis* genes was estimated as about 26,000 (Arabidopsis Genome Initiative 2000). Considering that we excluded the genes showing no expression in any of the organs estimated, our microarray only covered one quarter or more the total expressed genes. Statistical and functional cluster analyses by bioinformatic methods have provided much information about the identity, possible industrial use, and physiological functions of selected genes. Moreover, many novel genes have been screened in bioinformatic studies. More studies will provide further efficient information for clarification of the network of gene expressions that regulate the differentiation and morphogenesis of plant organs.

Our obtained information about gene annotations and expression specificity will be applicable in various fields of biology and biotechnology. In molecular breeding using recombinant DNA technology, many types of promoters are needed for improving plant characters, and in particular, organ-specific promoters will be valuable in the production of useful plant materials. For example, leaf specific promoters will allow mass production of materials in photosynthetically active leaves, and root-specific promoters enabling regulation of root systems will be important for industrial applications such as phytoremediation and rhizosecretion (Yoshida and Shinmyo 2000).

Acknowledgements

This work was supported by the Innovative Technology for the Earth Program, the "Department of Transgenic Plants for Production of Industrial Materials" from the New Energy and Industrial Technology Development Organization (NEDO) to T.K. and A.S.

References

Ando K, Yamakawa S, Miyashita K, Yoshida K, Yokota A, Shinmyo A, Kohchi, T (2004) Efficient construction of cDNA microarrays utilizing normalized cDNA libraries of *Arabidopsis thaliana*. *J Biosci Bioeng* 97: 85–88

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815

Chung HJ, Ferl RJ (1999) *Arabidopsis* alcohol dehydrogenase expression in both shoots and roots is conditioned by root growth environment. *Plant Physiol* 121: 429–436

Duyk GM (2002) Sharper tools and simpler methods. *Nat Genet* 32 Suppl: 465–468

Goda H, Sawa S, Asami T, Fujioka, S, Shimada, Y, Yoshida, S (2004) Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in *Arabidopsis*. *Plant Physiol* 134: 1555–1573

Haran S, Logendra S, Seskar M, Bratanova M, Raskin I (2000) Characterization of *Arabidopsis* acid phosphatase promoter and regulation of acid phosphatase expression. *Plant Physiol* 124: 615–626

Kohchi T, Fujishige K, Ohyama K (1995) Construction of an equalized cDNA library from *Arabidopsis thaliana*. *Plant J* 8: 771–776

Kononowicz AK, Nelson DE, Singh NK, Hasegawa PM, Bressan RA (1992) Regulation of the osmotin gene promoter. *Plant Cell* 4: 513–524

Paul AL., Schuerger AC, Popp MP, Richards JT, Manak MS, Ferl RJ (2004) Hypobaric biology: Arabidopsis gene expression at low atmospheric pressure. *Plant Physiol* 134: 215–223

Sabatini S, Heidstra R, Wildwater M, Scheres B (2003) SCARECROW is involved in positioning the stem cell niche in the Arabidopsis root meristem. *Genes Dev* 17: 354–358

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467–470

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002) Monitoring the expression profiles of 7,000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31: 279–292

Takemura M, Fujishige K, Hyodo H, Ohashi Y, Kami C, Nishii A, Ohyama K, Kohchi T (1999) Systematic isolation of genes expressed at low levels in inflorescence apices of *Arabidopsis thaliana*. *DNA Res* 6: 275–282

Tamaoki M, Nakajima N, Kubo A, Aono M, Matsuyama T, Saji H (2003) Transcriptome analysis of O_3-exposed *Arabidopsis* reveals that multiple signal pathways act mutually antagonistically to induce gene expression. *Plant Mol Biol* 53: 443–456

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

Yamakawa S, Ando K, Chisaka A, Yoshida K, Shinmyo A, Kohchi T (2004) Systematic transient assays of promoter activities for leaf-specific genes identified by gene-expression profiling with cDNA microarrays in *Arabidopsis thaliana*. *J. Biosci. Bioeng* 98: 140–143

Yoshida K, Shinmyo A (2000) Transgene expression systems in plant, a natural bioreactor. *J Biosci Bioeng* 90: 353–362