DNA RESEARCH 7, 367–370 (2000)

Embarking on Rice Functional Genomics Via cDNA Microarray: Use of 3′ UTR Probes for Specific Gene Expression Analysis

Junshi YAZAKI,[1] Naoki KISHIMOTO,[2] Keiko NAKAMURA,[1] Fumiko FUJII,[1] Kanako SHIMBO,[1] Yoshimi OTSUKA,[1] Jianzhong WU,[1] Kimiko YAMAMOTO,[1] Katsumi SAKATA,[2] Takuji SASAKI,[2] and Shoshi KIKUCHI[2,*]

Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, 446-1 Ippaizuka, Kamiyokoba Tsukuba, Ibaraki, 305-0854, Japan[1] *and National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan*[2]

(Received 8 November 2000; revised 30 November 2000)

Because rice is the major source of food for half of the world population, a comprehensive understanding of its genome will have a tremendous impact on the status of agriculture in the Twenty-first Century. The Rice Genome Research Program (RGP) has undertaken an extensive rice genome analysis since its initial launching in 1991 that has resulted in the establishment of a catalog of rice genes,[1] a high-density linkage map,[2] and a YAC-based physical map.[3] An enormous collection of expressed sequence tags (ESTs) was generated from large-scale cDNA sequencing using cDNA libraries derived from rice callus cultured in different media and tissues such as root, shoot, leaf and panicle.[4] This included more than 9000 partial cDNA sequences corresponding to unique genes have been identified. Based on the results of large-scale cDNA analysis, microarray can be used to monitor gene expression profiles and to initiate functional analysis of the rice genome. We initiated a rice cDNA microarray project beginning in April 1999 that aims to elucidate the function of all genes in rice using a gene expression monitoring system. The National Institute of Agrobiological Resources (NIAR) and the Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries (STAFF) are jointly conducting this project in collaboration with 64 research institutions all over Japan. The research areas and investigator/institutions involved in this collaboration are listed in http://microarray.rice.dna.affrc.go.jp/. Homology search of all ESTs was performed by BLASTN and about over 9000 clones with high similarity were clustered. Clustered clones were analyzed for sequence similarity against public protein and nucleic acid databases. Then BLASTN and BLASTX were performed and clones with high similarity were identified putatively. One thousand two hundred sixty-five cDNAs in these unique clones were randomly selected and used for a pilot-scale microarray system. The identity and accession number of the 1265 rice EST clones are listed at http://microarray.rice.dna.affrc.go.jp/.

Insert cDNA clones in pBluescript II vectors were amplified by polymerase chain reaction (PCR) using M13 primer. Plasmid template (1–2 ng) was added to 50 μl of PCR mixture and inserts were amplified by PCR as follows: (1) an initial denaturation at 94°C for 1 min (2) 30 cycles of 94°C for 1 min, 60°C for 2 min, and 72°C for 2 min, and (3) a final extension at 72°C for 10 min. The 3′-untranslated region (3′-UTR) of the cDNA clones were amplified using each specific primer complementary to the specific region of each cDNA clone. The PCR was carried out as follows: (1) an initial denaturation at 94°C for 1 min, (2) 35 cycles of 94°C for 30 sec, 60°C for 1 min, and 72°C for 1 min, and (3) final extension at 72°C for 10 min. Both PCR products were purified using QIAquick 96-column (QIAGEN), and mixed with water and reagent D (Amersham Pharmacia,) for microarray fabrication. We spotted cDNA probes in duplicate on aluminum-coated and DMSO-optimized glass slide using an Array Spotter Generation III (Amersham Pharmacia).

Oryza sativa L. cv. Nipponbare was used for RNA extraction and target preparation. Leaf and root tissues were collected from water cultured seedlings at 28°C for 30 days post germination. Callus was induced from seeds cultured in MS medium supplemented with 2,4-dichlorophenoxy acetic acid at 25°C for 20 days. Flowers were collected at a flowering stage. Total RNA was prepared using the method of Chirgwin et al.[5] Poly(A)$^+$ RNA was purified using Oligotex-dT super (TAKARA) from 200 μg of total RNA. Each poly(A)$^+$ RNA was labeled by reverse transcript reaction with Cy-5 (Amersham Pharmacia). The reverse transcript reaction mixture was composed of 1 μg poly(A)$^+$ RNA, oligo-(dT)25, random nonamer, control cRNA, 1 × SSII reaction buffer, DTT, dATP 2 mM, dGTP 2 mM, dTTP

Communicated by Satoshi Tabata

* To whom correspondence should be addressed. National Institute of Agrobiological Resources, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan, Tel. +81-298-38-7007, Fax. +81-298-38-7007, E-mail: skikuchi@abr.affrc.go.jp

2 mM, dCTP 1 mM, Cy-5dCTP 1 mM and SSII reverse transcriptase. After incubation at 42°C for 2.5 hr, the reaction mixture was denatured at 94°C for 3 min and RNA was degraded by 2 μl of 2 N NaOH at 37°C for 15 min. Following degradation, the mixture was treated with 10 μl of 2 M HEPES buffer for neutralization. The labeled targets were purified with Qiaquick PCR Purification Kit (QIAGEN) and dried using a vacuum concentrator. The dried targets were resuspended in 6 μl of water and denatured at 95°C for 4 min. A 1.5-μl Oligo A80 (1 mg/ml) was added to the re-suspended targets and 7.5 μl of 4 × hybridization buffer containing SSC, SDS, Denhardt's solution, salmon sperm DNA, and 15 μl of formamide. A final volume of 30 μl was used for hybridization. After hybridization, the glass slides were washed in 1 × SSC/0.2% SDS for 10 min at 55°C under dark conditions, then in 0.1 × SSC/0.2% SDS for 10 min at 55°C twice in the dark condition and finally, in 0.1 × SSC for 1 min at room temperature twice. After the final wash, the slides were briefly rinsed with distilled water and air dried. The hybridized and washed microarrays were scanned using an Array Scanner Generation III (Amersham Pharmacia). A computer program, Array Vision (Imaging Research) was used to locate and delineate every spot in the array and to integrate spot intensities and volumes for each individual spot. Their values were scatter plotted by Spotfire (Spotfire).

Duplicate spots of each cDNA were applied on the same slide to evaluate the reproducibility of the key steps in the protocol such as the DNA spotting quality and homogeneous hybridization. The profiles of the gene expression volume in the duplicated arrays with one hybridization experiment are shown at the Web site http://microarray.rice.dna.affrc.go.jp/. Spots were dispersed within two fold to threefold lines suggesting that the microarrays converge into threefold fluctuation, particularly with abundant transcripts. Less abundant clones, however, showed variable results. The detection sensitivity was measured by using the control clone (mouse platelet-derived growth factor receptor, 900 bp) which showed no cross hybridization with the rice genes. This result can be accessed at http://microarray.rice.dna.affrc.go.jp/. The control cRNAs were synthesized by *in vitro* transcription and were used as quantitative controls at 1 ng, 0.5 ng, 0.1 ng and 50 pg. The ratios of the control cRNA to target RNAs were 100 : 100 000, 50 : 100 000, 10 : 100 000 and 5 : 100 000. We could detect from 1 ng (1/1000 of rice target RNA) to 50 pg of expression volume. However, at a concentration below 0.05 ng, the control cRNA showed lower signal than that of the background control (element without spotted DNA) signal.

EST mapping analysis revealed that primers designed from the 3′ portion (3′-UTR) of rice ESTs were more gene specific than that from the 5′ portion.[6] This observation suggests that the full-length EST insert is effective for comprehensive analysis of family gene expression while the 3′-UTR probe is useful for detecting gene-specific expression. As we have already made PCR primers for the 3′-UTR region of the EST clones, two probe systems from the full-length inserts and the 3′-UTR were fabricated and these hybridization data were compared using mRNAs derived from root tissues at 30 days after germination. The most highly expressed genes using full insert and 3′-UTR microarrays are shown in Table 1. In the full-insert microarray, the ten most highly expressed genes consist of five ubiquitin homologs (clone ID C52727, S16102, C03001, C00176, C52245), two unknown genes (S16144, C40015) and one homolog each of S-adenosyl methionine synthase (S16157), NADH dehydrogenase (C01912) and actin (E02210). In the 3′-UTR microarray, three ubiquitin homologs (C00176, C52727, S16102), four unknown genes (S16144, C40015, C12875, S10563) and one homolog each of thioredoxin (E03596), phenylalanine ammonia-lyase (S04593) and methyltransferase (E02880) showed the highest signals. Only three ubiquitin homologs (C00176, C52727, S16102) and two unknown genes (S16144, C40015), however, were highly expressed in both full-insert and 3′-UTR microarrays. The rest including two ubiquitin homologs (C03001, C52245) were selected by the 5′-portion of the full-insert probes and are probably non-specific genes in root. The total fluorescent intensity of the full-insert and 3′-UTR microarrays was about 180×10^7 and 4.5×10^7, respectively. This suggests that a high florescent signal in a full-insert microarray can be attributed to hybridization not only of labelled target but also other regions of the full-insert. Therefore, a full insert microarray can identify a relatively large number of genes having gene similarity and is effective in comprehensive screening of gene expression profiles. A 3′-UTR microarray is effective in detecting specific genes in target RNA from various tissues and at different developmental stages.

The full insert cDNA microarray was evaluated for level of gene expression in different tissues. Four poly(A)$^+$ RNA samples (panicle, callus, leaf and root) were labeled with Cy5-dCTP and hybridized to a cDNA microarray containing 1265 full inserts. We compared the signal intensity of hybridization between panicle and callus as well as between root and leaf with the scatter plot of expression profiles. The data were normalized by getting the total fluorescent signal of each target DNA and adjusting the value to obtain equal signals. The results are available at http://microarray.rice.dna.affrc.go.jp/. The results indicate that several genes showing absolute tissue specificity were identified in leaf and panicle. Homologs of leaf-specific genes such as chlorophyll *a/b* binding protein (S13320), photosystem I antenna protein (E01186), *Xa21* protein kinase (S12429) and photosystem II oxgen-envolving complex protein (E02847) showed over three fold differential expression in leaf. Ex-

Table 1. Comparison of gene expression profiles in rise root using a full insert and a 3′-UTR microarray.

Clone ID	Accession	Putative Gene Identification	Signal intensity
Full insert			
C52727	C97167	Ubiquitin (*Avena fatua*)	1.84E+07
S16157	AU065955	S-Adenosyl methionine synthetase (*Lycopersicon esculentum*)	1.47E+07
C01912	D15997	NADH dehydrogenase (*Paramecium tetraurelia*)	1.33E+07
S16102	C24892	Ubiquitin extension protein (*Lupinus albus*)	1.30E+07
E02210	C72769	Actin (*Nicotiana tabacum*)	1.16E+07
C03001	C98384	Polyubiquitin (*Arabidopsis thaliana*)	1.08E+07
C00176	C97944	Ubiquitin (*Oryza sativa*)	1.05E+07
C52245	C27568	Ubiquitin (*Gallus gallus*)	9.40E+06
S16144	AU032905	Unknown	9.37E+06
C40015	AU077483	Unknown	9.32E+06
3'UTR			
C00176	C97944	Ubiquitin (*Oryza sativa*)	2.08E+07
C52727	C97167	Ubiquitin (*Avena fatua*)	1.01E+07
S16102	C24892	Ubiquitin extension protein (*Lupinus albus*)	8.38E+06
S16144	AU032905	Unknown	3.23E+06
C40015	AU077483	Unknown	3.10E+06
E03596	AU064058	Thioredoxin M (*Zea mays*)	2.87E+06
C12875	AU068273	Unknown	2.83E+06
S04953	D41931	*ZB8*, phenylalanine ammonia-lyase (*Oryza sativa*)	2.57E+06
E02880	C73083	o-Methyltransferase (*Zea mays*)	2.09E+06
S10563	AU032465	Unknown	1.90E+06

Clone ID: the original clone identification number of the sequence submitted to DDBJ. Putative gene identification: gene annotation obtained for sequences identified from homology search of the public databases. The gene with the highest score was used as the putative gene identification for the sequence. The organism in which the gene was identified is indicated in parenthesis. Signal intensity: The abundance of transcripts from target sample is reflected by signal intensity.

cept for cytochrome P-450 (C30225) the other highly expressed genes in leaf correspond to shoot cDNAs (S10109, S11976, S16431, S14062, S16149). In panicle, the major allergen homolog *Cynod1* (E01196) and the pollen allergen *PhlpII* homolog (E03507) were highly abundant and expressed over 2000-fold. The calcium-binding pollen allergen gene (E01969) was also expressed over 200-fold in the panicle. Although the other three homologs (E02862, E01143, E03522) were not panicle-specific genes, the clones correspond to cDNA libraries derived from flowering stage panicle. The rest of the highly expressed genes in panicle (C52093, S04987, S11020, S16157) correspond to clones derived from callus and shoot cDNA libraries. On the other hand, among highly expressed genes in root and callus, only one or two clones correspond to cDNA libraries derived from either root or callus. The rest of the clones in each tissue correspond to cDNA libraries from shoot and panicle.

We have already constructed a rice cDNA microarray with approximately 9000 ESTs and embarked on a rice functional genomics project. Information on the cDNA clones including identity and accession number can be accessed at http://microarray.rice.dna.affrc.go.jp/. Further cDNA cloning and structural analysis are needed to cover those entire expressing genes. For such purpose the rice full-length cDNA project was launched at the beginning of 2000. This project will be useful in establishing a more comprehensive rice cDNA microarray.

Acknowledgements: We thank Kohei Suzuki, Hideaki Suzuki (STAFF-Institute) and Takumi Koyama (National Institute of Animal Health) for providing the animal cDNA clones; Hideki Nagasaki, Yoshiyuki

Mukai and Kazunori Waki (STAFF-Institute) for technical assistance and many useful suggestions; Bal Antonio (STAFF-Institute) for helpful discussions; and Dr. Elizabeth Dennis (CSIRO Plant Industry) for critical reading of the manuscript.

This work was supported by a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Rice Genome Project MA-1001).

References

1. Sasaki, T., Song, J., Koga-Ban, Y. et al. 1994, Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library, *Plant J.*, **4**, 615–624.
2. Harushima, Y., Yano, M., Shomura, A. et al. 1998, A high-density rice genetic linkage map with 2275 markers using a single F2 population, *Genetics*, **48**, 479–494.
3. Sasaki, T., Yano, M., Kurata, N. et al. 1996, The japanese rice genome research program, *Genome Res.*, **6**, 661–666.
4. Yamamoto, K. and Sasaki, T. 1997, Large-scale EST sequencing in rice, *Plant Mol. Biol.*, **35**, 135–144.
5. Chirgwin, J. M., Przybyla, A. E., and MacDonald, R. J. 1979, Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease, *Biochemistry*, **18**, 5294–5299.
6. Wu, J., Yamamoto, S., Maehara, T. et al. 2000, A comprehensive map of rice expressed sequence tags (ESTs): its characteristics and application to genome analysis. 6th International Congress of Plant Molecular Biology, *Supplement to reporter*, **18**, S1-92.