

# RiceXPro: a platform for monitoring gene expression in *japonica* rice grown under natural field conditions

**Yutaka Sato<sup>1</sup>, Baltazar A. Antonio<sup>1,\*</sup>, Nobukazu Namiki<sup>2</sup>, Hinako Takehisa<sup>1</sup>, Hiroshi Minami<sup>2</sup>, Kaori Kamatsuki<sup>2</sup>, Kazuhiko Sugimoto<sup>3</sup>, Yuji Shimizu<sup>2</sup>, Hirohiko Hirochika<sup>4</sup> and Yoshiaki Nagamura<sup>1</sup>**

<sup>1</sup>Genome Resource Center, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, <sup>2</sup>Mitsubishi Space Software Co. Ltd., Takezono 1-6-1, Tsukuba, Ibaraki 305-0032, <sup>3</sup>QTL Genomics Research Center, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602 and <sup>4</sup>Division of Genome Research and Biodiversity, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan

Received August 6, 2010; Revised September 30, 2010; Accepted October 15, 2010

## ABSTRACT

**Elucidating the function of all predicted genes in rice remains as the ultimate goal in cereal genomics in order to ensure the development of improved varieties that will sustain an expanding world population. We constructed a gene expression database (RiceXPro, URL: <http://ricexpro.dna.affrc.go.jp/>) to provide an overview of the transcriptional changes throughout the growth of the rice plant in the field. RiceXPro contains two data sets corresponding to spatiotemporal gene expression profiles of various organs and tissues, and continuous gene expression profiles of leaf from transplanting to harvesting. A user-friendly web interface enables the extraction of specific gene expression profiles by keyword and chromosome search, and basic data analysis, thereby providing useful information as to the organ/tissue and developmental stage specificity of expression of a particular gene. Analysis tools such as t-test, calculation of fold change and degree of correlation facilitate the comparison of expression profiles between two random samples and the prediction of function of uncharacterized genes. As a repository of expression data encompassing growth in the field, this database can provide baseline information of genes that underlie various agronomically important traits in rice.**

## INTRODUCTION

The sequencing of *Oryza sativa* L. ssp. *japonica* cv. Nipponbare genome has brought to light the genome structure of a major cereal crop that provides food for almost half the world population (1). A comprehensive annotation of the high-quality genome sequence identified approximately 32 000 genes that define the rice plant, more than half of which does not have known biological functions (2–4). Assigning function to every annotated gene has become an enormous challenge not only for complete understanding of the biology of rice, but more importantly, for efficient utilization of the most basic genome information in crop improvement. Although worldwide efforts to characterize the function of the rice genome by map-based cloning strategies (5), loss-of-function approaches through insertional mutagenesis (6–9), and gain-of-function approaches (10) have made significant progress in the last few years, there is no doubt that functional characterization of all predicted genes in rice would take much longer than anticipated earlier.

Microarray technology has now become a major strategy for analyzing the genome-wide expression of genes and could provide useful clues for functional characterization of the entire genome (11). Several microarray platforms for rice have been successfully used in characterizing tissue and organ development (12–21) as well as specific treatment conditions (22–25). More recently, global gene expression analyses encompassing

\*To whom correspondence should be addressed. Tel: +81 29 838 7456; Fax: +81 29 838 7408; Email: antonio@nias.affrc.go.jp

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

different cell types (26) and different organs/tissues representing the entire life cycle of the rice plant (27) have been reported. As a result, there is a proliferation of rice microarray datasets in public repositories based on varying platforms, varying experimental conditions, and plant samples from varying genetic backgrounds. Although intra-platform reproducibility of microarray data can be achieved among different commercial array systems (28), a more reliable biological information and interpretation of gene expression profiles can be best accomplished using a single platform and a model plant variety. In addition, for rice plants normally grown in the field, the entire cycle of growth and development is constantly subjected to changes in environmental conditions so that a number of factors that may affect subtle changes in gene expression have to be taken into consideration. This is particularly important in clarifying the dynamic nature of gene regulation that determines the expression of certain phenotypic traits and the changes in transcriptome profile for each organ or tissue as initiated by both intrinsic and extrinsic factors throughout the entire growth. Several databases of rice expression data are available including the National Science Foundation (NSF) Rice Array Database (24), Yale Rice Atlas Database (26) and Collection of Rice Expression Profiles (CREP) Database (27), thereby providing a wide range of rice gene expression data. Our goal, however, is to establish a comprehensive gene expression profile database that can be used as standard for understanding genome-wide expression of rice under normal field conditions.

We describe here the rice expression profile database (RiceXPro) of the model rice cultivar, *O. sativa* ssp. *japonica* cv. Nipponbare, with a user-friendly web interface for retrieving any expression data of interest. The overall concept of this database is to provide baseline information in terms of expression in different organs or tissues, and specificity in the stage of growth and development that can be used for elucidating the function of the predicted rice genes.

## DATABASE COMPONENTS AND FEATURES

### Microarray platform

For microarray analysis, we used a rice  $4 \times 44\text{K}$  microarray RAP-DB (Agilent Technologies; G2519F#15241). The microarray was designed based on the manually curated annotation of the rice genome as described in the Rice Annotation Project Database (RAP-DB) (2) and consisted of 60-mer oligo sequences corresponding to 27 800 RAP loci with transcripts based on rice full-length cDNA (32 325) (29), transcripts based on expressed sequence tag (EST) support (6943), and predicted loci (2612). The oligo-sequences for all the full-length cDNA correspond to the alignment of 3'-UTR for each gene, which often show sufficient specificity, thereby producing high quality expression data. For RAP locus with several mapped full-length cDNAs, multiple 60-mer probes were designed based on the corresponding sequence. Therefore, there are probes with same sequence among the multiple probes designed for a single locus. These redundant probes

provide evidence on replicability and reliability of the expression data and are indicated in the RiceXPro database as described below. Details of our microarray platform are provided in the section of 'RICE 44K MICROARRAY' in the RiceXPro database.

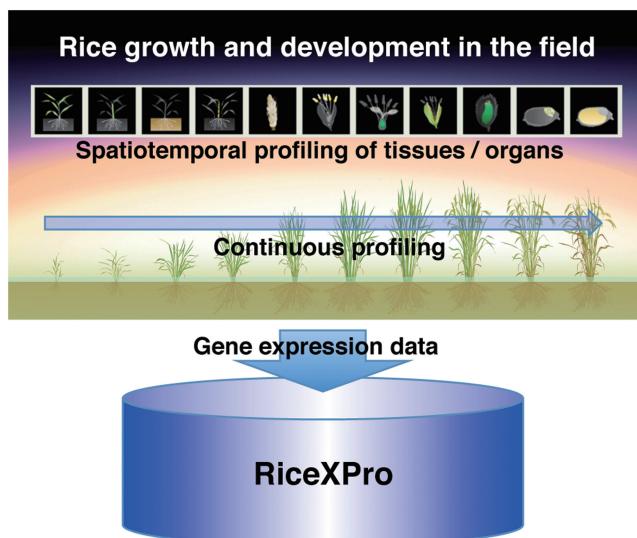
### Database content

RiceXPro currently contains gene expression data sets corresponding to spatiotemporal gene expression profiling based on 48 different tissue and organ types at various developmental stages, and continuous profiling of leaves from transplanting until harvesting in *japonica* rice cultivar, Nipponbare, grown under natural field conditions (Figure 1; manuscript submitted). The two data sets comprise 143 and 51 microarray data with three replicates, and designated as RXP\_0001 and RXP\_0003, respectively, in the list of data sets.

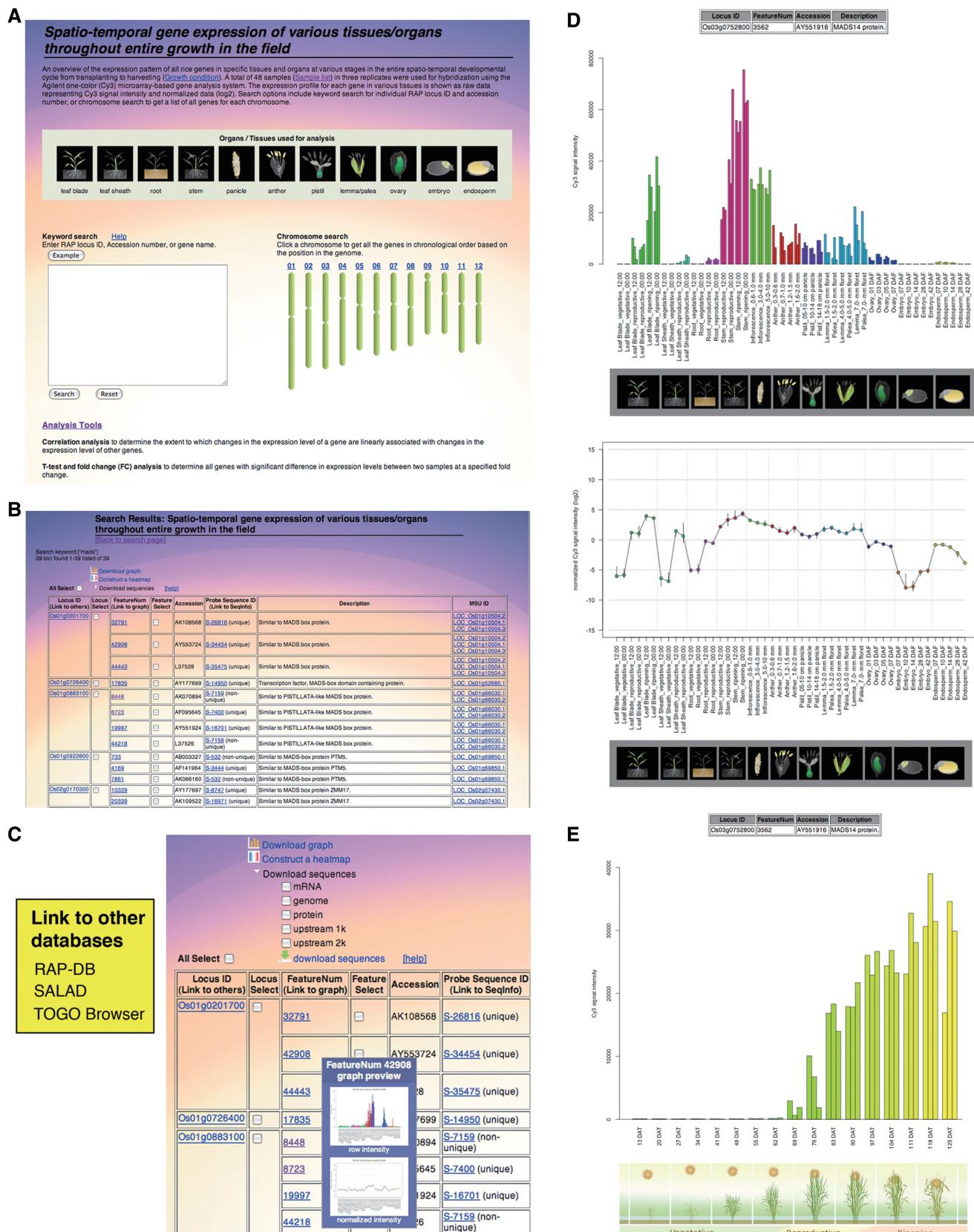
All expression data were generated by one-color (Cy3) microarray hybridization protocol. Background correction of the Cy3 raw signals was performed with the Agilent Feature Extraction software (version 9.5.3.1). The processed signal intensities of 40 121 probes corresponding to 27 800 loci as described in RAP-DB were used for construction of expression profiles in graph format. Each expression data was subjected to 75 percentile normalization and log<sub>2</sub> transformation using the R program (<http://www.R-project.org>).

### Web interface

A gene expression database must be structured in a way that facilitates both visualization and analysis of a large amount of data. The RiceXPro database can be accessed through a user-friendly web interface (<http://ricexpro.dna.affrc.go.jp/>) that provides two search options: keyword search and chromosome search, and two analysis



**Figure 1.** Overview of data source used in database construction. RiceXPro currently contains data sets corresponding to spatiotemporal profiling based on 48 different tissues and organs at various developmental stages, and continuous profiling of leaves from transplanting until harvesting. The database was designed to provide baseline information on gene expression of rice grown under natural field conditions.



**Figure 2.** Retrieval of gene expression profile in RiceXPro. (A) A snap shot of the web-interface for RXP\_0001 data set. Data search can be initiated by entering one or more keywords such as RAP-DB locus ID, MSU Osa1 Rice Loci, accession number or gene name, or by selecting a particular chromosome to get the expression data of all genes based on the position in the chromosome. Correlation and *t*-test/FC analysis allow the user to perform basic data analysis. (B) Tabular list of genes generated from keyword search with ‘MADS’. (C) Links and other information that can be

Continued

tools: correlation analysis and *t*-test/fold change (FC) analysis to evaluate the expression pattern based on the signal intensity of genes in different organs/tissues under natural field conditions (Figure 2A).

Keyword search can be used for RAP-DB locus ID, MSU Osa1 Rice Loci, accession number, gene name or any word included in the gene description. In the search option, multiple keyword search can be performed by (i) entering two or more keywords with one keyword per line in the search box to initiate an OR search or (ii) entering several keywords per line with a space in between to initiate an AND search. The chromosome search option provides expression data of genes based on the position in the genome. A graphical representation of the 12 chromosomes allows the user to select a particular chromosome. This search system can provide valuable information in assessing the spatiotemporal specificity of a gene of interest to facilitate an efficient approach in forward genetics.

The search options provide a tabular list of the gene/genes with locus ID (RAP-DB), 'feature number' representing the ID number of the probes in the Agilent microarray platform, accession number, probe sequence ID, gene description and MSU Osa1 Rice Loci (Figure 2B). For example, the list of genes generated by the keyword search with 'MADS' consisted of 39 loci represented by 97 probes. Mouse over on the locus ID opens a pop-up window with direct links to RAP-DB (4), SALAD database (30) and Rice TOGO Browser (<http://agri-trait.dna.affrc.go.jp/>) to obtain detailed information on annotation, a genome-wide comparative analysis of motifs and genetic mapping data (Figure 2C). In addition, the MSU Osa1 Rice Loci provides direct links to the Rice Genome Annotation Project database (<http://rice.plantbiology.msu.edu/>). A probe sequence ID is assigned based on the alignment of the probe sequence, so that probes with same sequence would have the same identifier. For example, two probes with feature number (8448 and 44218) representing Os01g0883100 have same sequence (indicated as 'non-unique' in probe sequence ID) (Figure 2C). The expression profiles for each gene are shown as two graphs with the raw signal intensity for each sample in triplicate and the normalized signal in log<sub>2</sub> scale (Figure 2D). A line graph representing normalized data shows median value in triplicate with error bars to indicate maximum and minimum values, respectively. These graphical data can be accessed from the tabular list generated from the search via the feature number column. Putting the mouse on a feature number gives an overview of the expression profile for each gene through a pop-up window (Figure 2C). In association with

multiple keyword search and chromosome search, this function allows the user to identify and select a specific gene of interest. Furthermore, an image of each tissue and organ represented in the graphical data helps in understanding the expression signature of a particular gene. Figure 2D shows the gene expression data of Os03g0752800 that encodes OsMADS14 protein based on spatiotemporal gene expression profiling of organs and tissues at various developmental stages (RXP\_0001). *OsMADS14* is homologous to an *Arabidopsis* floral identity gene *APETALA1*, and has been reported to be induced by *Hd3a* and *RFT1*, rice orthologs of *Arabidopsis* florigen gene *FLOWERING LOCUS T (FT)* (31,32). Our data provided from RiceXPro indicate that the *OsMADS14* is expressed in reproductive organs as well as vegetative organs after the reproductive phase. This is further confirmed by the expression data of *OsMADS14* based on continuous profiling (RXP\_0003) as shown in Figure 2E. A high-resolution gene expression data as provided in RiceXPro could, therefore, be useful in characterizing genes in rice based on the combined expression pattern in different tissues and organs, and at different stages of growth and development under natural field conditions.

#### Download option of expression data and sequence information

In RiceXPro, the graph images for expression data of multiple locus IDs or feature IDs can be downloaded from the tabular list generated by two search options: keyword search and chromosome search, and two analysis tools: correlation analysis and *t*-test, and FC analysis. Additionally, the user is also provided with an option to display the expression data in a heatmap format for a maximum of 100 genes. For these two options, the user must first specify the genes by checking 'Select all', 'Locus Select' or 'Feature Select' buttons, and then select either 'Download graph' or 'Construct a heatmap' option (Figure 2C). The heatmap can be constructed based on a customized ordering, or alternatively, based on a clustering method of selected genes. For customized ordering, the locus\_feature IDs can be rearranged by drag-and-drop (Figure 3A) before initiating the heatmap construction. The clustering method is based on the correlation distance and complete linkage of each gene using heatmap.2 in the 'gplots' package of R program (Figure 3A). For the construction of a heatmap, the expression level for each gene across all the data within a data set is normalized by shifting the baseline of median value to zero. A title can be added and the heatmap can be downloaded as PDF with the designated title.

Figure 2. Continued

obtained from the tabular list of searched data. The locus ID column provides a pull-down menu for direct links to other databases and results of correlation analysis within a dataset. The expression profile is obtained via the 'FeatureNum' column, and putting the mouse on a feature number opens a pop-up window of the profile for each gene. The 'Select all' or the 'Locus Select' or 'Feature Select' buttons are used to specify genes for downloading the graph, heatmap construction and sequence retrieval options. In order to retrieve the sequence of selected genes, user must check 'mRNA', 'genome', 'protein', 'upstream 1 k' or 'upstream 2 k' button. Each probe is assigned with a sequence ID based on the alignment of the sequence and provides a link to the sequence information. (D) Graphical representation of expression profile for each gene. The bar graph and the line graph represent raw signal intensity and normalized signal intensity for 'Os03g0752800', respectively, derived from spatiotemporal profiling of various tissues and organs (RXP\_0001). (E) The bar graph representing raw signal intensity for 'Os03g0752800' derived from continuous profiling of leaves (RXP\_0003).

**A**

## Construct a heatmap

This option facilitates the construction of a heatmap for selected multiple locus IDs or feature IDs. The heatmap is constructed based on a customized order, or alternatively, based on clustering of correlation distance and complete linkage of each gene using heatmap.2 in the "gplots" package of R program. The expression level for each gene is normalized by shifting the baseline of median value to zero for each gene across all data within a dataset. The heatmap can be downloaded as PDF.

### Heatmap construction option

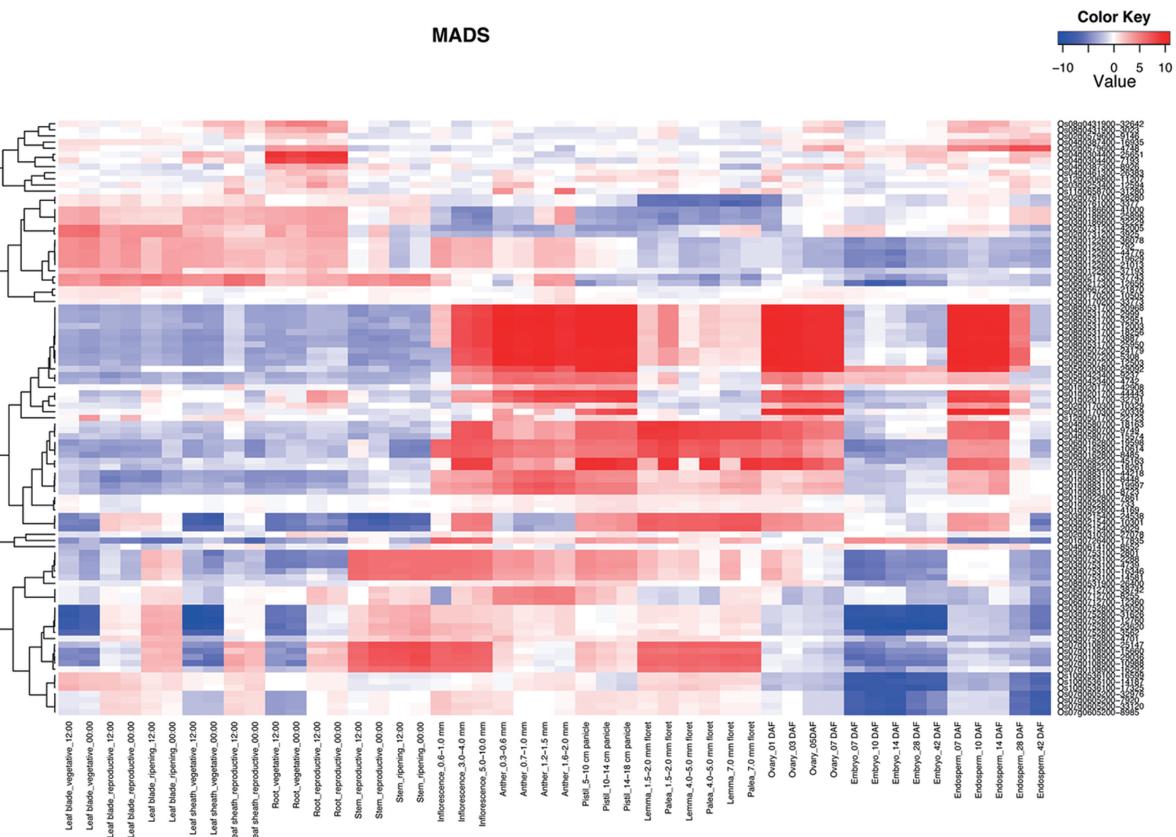
- Customized ordering
- Clustering

Enter title MADS

download

For customized ordering, rearrange the locus\_feature IDs by drag and drop.

- ↔  Os01g0922800\_733 Similar to MADS-box protein PTM5.
- ↔  Os03g0122600\_792 Transcription factor, MADS-box domain containing protein.
- ↔  Os03g0215400\_2029 Box protein (MADS box protein MADS1).

**B**

**Figure 3.** Download option for heatmap of expression data. **(A)** Web interface for construction of heatmap. Users can select 'Customized ordering' or 'Clustering' option to construct a heatmap. **(B)** The heatmap of 97 probes corresponding to 39 'MADS' loci generated with the clustering method.

Figure 3B shows the heatmap obtained from clustering of 97 probes representing 39 loci generated by keyword search with 'MADS'. The heatmap option facilitates the visualization and comparison of the expression profiles of any number of genes.

The RiceXPro further provides a retrieval system for the sequence of the genes generated from the search

options and the analysis tools. The mRNA sequence, genome sequence, protein sequence and 1 kb or 2 kb sequence upstream of the transcription start site of the genes specified by checking the 'All Select' or 'Locus Select' button can be downloaded in fasta format via the 'Download sequence' option (Figure 2C). This function helps the user to further examine in detail the function

**A**

## Analysis Tools

[\[Back to search page\]](#)

### Correlation analysis

Enter a RAP locus ID or accession number and click submit. Please wait for about a few seconds for the results.

Submit

### T-test & fold change (FC) analysis

Select a data as control and another data to be used for comparison by clicking the corresponding radio buttons from the list of samples. Then specify the p-value and fold change from the pull-down menu, and click submit. Please wait for a few seconds for the results.

p-value  FC  = 3.32 (log<sub>2</sub>scale)

Sample ID	Control	Test	Sample ID	Control	Test
Leaf blade_vegetative_12:00	<input checked="" type="radio"/>	<input type="radio"/>	Anther_1.6-2.0 mm	<input type="radio"/>	<input type="radio"/>
Leaf blade_vegetative_00:00	<input type="radio"/>	<input checked="" type="radio"/>	Pistil_5-10 cm panicle	<input type="radio"/>	<input type="radio"/>
Leaf blade_reproductive_12:00	<input type="radio"/>	<input type="radio"/>	Pistil_10-14 cm panicle	<input type="radio"/>	<input type="radio"/>
Leaf blade_reproductive_00:00	<input type="radio"/>	<input type="radio"/>	Pistil_14-18 cm panicle	<input type="radio"/>	<input type="radio"/>

**B**

Search: FeatureNum:17549(Os04g0179700|AY347880) correlation analysis  
1-100 listed of 40121

[First](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [>](#) [Last](#)

Download graph

All Select  Download sequences [\[help\]](#)

sort by ascending correlation

Locus ID (Link to others)	FeatureNum (Link to graph)	Select	Accession	Probe Sequence ID (Link to SeqInfo)	Description	correlation coefficient
Os04g0179700	17549	<input type="checkbox"/>	AY347880	S-14706 (unique)	Syn-pimara-7,15-diene synthase.	1.00000
Os04g0179700	24667	<input type="checkbox"/>	AK119327	S-20458 (unique)	Syn-pimara-7,15-diene synthase.	0.97381
Os04g0180400	5260	<input type="checkbox"/>	AK071546	S-4396 (unique)	Similar to Cytochrome P450 CYP99A1 (EC 1.14.-.) (Fragment).	0.93253
Os04g0179200	23077	<input type="checkbox"/>	AK106087	S-19187 (non-unique)	Similar to Stem secoisolariciresinol dehydrogenase (Fragment).	0.89146
Os02g0570700	32657	<input type="checkbox"/>	AK070167	S-26712 (unique)	Cytochrome P450 family protein.	0.88982
Os04g0178400	35117	<input type="checkbox"/>	AK071864	S-28612 (unique)	Similar to Cytochrome P450 CYP99A1 (EC 1.14.-.) (Fragment).	0.88045
Os04g0179200	33410	<input type="checkbox"/>	AK120852	S-19187 (non-unique)	Similar to Stem secoisolariciresinol dehydrogenase (Fragment).	0.87927
Os02g0569400	8706	<input type="checkbox"/>	AK060722	S-7384 (unique)	Cytochrome P450 family protein.	0.87499
Os02g0569900	44168	<input type="checkbox"/>	AK105913	S-35296 (unique)	Cytochrome P450 family protein.	0.87466
Os02g0569400	27126	<input type="checkbox"/>	AK069701	S-22412 (unique)	Cytochrome P450 family protein.	0.87068
Os04g0178300	12640	<input type="checkbox"/>	AK121319	S-10685 (unique)	Similar to Copalyl diphosphate synthetase (Fragment).	0.87038
Os12g0555200	38780	<input type="checkbox"/>	AK099157	S-24364 (non-unique)	Similar to Probenazole-inducible protein PBZ1.	0.86795
Os07g0663900	25280	<input type="checkbox"/>		S-20939 (unique)	Short-chain dehydrogenase/reductase SDR family protein.	0.86694
Os1g0561600	23247	<input type="checkbox"/>	AK069494	S-19321 (unique)	Cytochrome P450 family protein.	0.86682
Os02g0570400	15383	<input type="checkbox"/>	AB089272	S-12966 (unique)	Similar to Ent-kaurene synthase 1A.	0.86318
Os04g0178300	4621	<input type="checkbox"/>	AY530101	S-3837 (unique)	Similar to Copalyl diphosphate synthetase (Fragment).	0.85724
Os12g0555200	29649	<input type="checkbox"/>	AK070762	S-24364 (non-unique)	Similar to Probenazole-inducible protein PBZ1.	0.85293
Os04g0179100	24278	<input type="checkbox"/>	C1027926	S-20146 (unique)	Similar to Stem secoisolariciresinol dehydrogenase (Fragment).	0.85202
Os02g0571900	25023	<input type="checkbox"/>	AK101003	S-20742 (unique)	Cytochrome P450 family protein.	0.84465

**C**

Search: p-value<0.01, FC>10, control:Leaf blade\_vegetative\_12:00(No:1-3), test:Leaf blade\_vegetative\_00:00(No:4-6)  
1-100 listed of 180

[First](#) [1](#) [2](#) [Last](#)

Download graph

Construct a heatmap

All Select  Download sequences [\[help\]](#)

Result download

Locus ID (Link to others)	FeatureNum (Link to graph)	Select	Accession	Probe Sequence ID (Link to SeqInfo)	Description	Fold Change	T-Test p-value
Os04g0583900	41367	<input type="checkbox"/>	AK101209	S-33320 (unique)	Similar to LHY protein.	767.409	0.0000112
Os08g0520700	34368	<input type="checkbox"/>	AK060548	S-28044 (unique)	Conserved hypothetical protein.	602.480	0.0004010
Os10g0565200	17363	<input type="checkbox"/>	CI411567	S-14560 (unique)	Similar to Beta-amylase PCT-BMYI (EC 3.2.1.2).	577.648	0.0000706

**Figure 4.** Analysis tools provided by RiceXPro database. (A) Interface for correlation analysis and t-test/FC analysis. (B) Tabular list of genes generated by correlation analysis for 'Os04g0179700'. (C) Tabular list of genes generated by t-test and FC analysis with P-value (<0.01) and FC (>10) setting between a control (leaf blade\_vegetative\_12:00) and a test sample (leaf blade\_vegetative\_00:00).

of genes of interest or the regulatory sequences associated with gene expression.

### Analysis tools

RiceXPro provides information of coexpressed genes based on calculation of Pearson's correlation coefficients (*r*-values) of the normalized signal intensities ( $\log_2$ ) in each data set (Figure 4A). The coexpression data can be used in speculating the function of uncharacterized genes of interest and finding of new genes which maybe associated to a particular gene function (33–35). In data set RXP\_0001 that contains expression profiles derived from a broad range of organs and tissues at various developmental stages, data mining based on coexpression data may facilitate identification of functionally related genes. Both locus ID and accession number can be used as a query in the search function of the coexpression analysis tool (Figure 4A). The search option provides a tabular list of genes in descending order of *r*-values, with the corresponding locus ID, feature number, accession number, probe sequence ID and description (Figure 4B). The list can be changed to ascending order to identify the genes with negative correlation. Figure 4B shows the gene list generated by search with 'Os04g0179700' as a query. The gene corresponding to Os04g0179700 encodes 9 $\beta$ H-pimara-7,15-diene synthase (OsKS4), which is related to the biosynthesis of momilactones, diterpenoid phytoalexins (36,37). The genes coexpressed with OsKS4 include Os04g0180400 (*CYP99A2*; *r* = 0.9325), Os04g0179200 (*OsMAS*; *r* = 0.89146), Os04g0178400 (*CYP99A3*; *r* = 0.88045) and Os04g0178300 (*OsCPS4*; *r* = 0.87038) (37). In addition to OsKS4, these genes present as a gene cluster of chromosome 4 and involved in momilactone biosynthesis (37). Although RiceXPro basically provides experimental condition-dependent coexpression data within a data set, the coexpression analysis tool may be used not only to extract genes based on similarity of expression pattern but also to identify various functionally related gene networks in rice.

The *t*-test (Welch's *t*-test) and FC analysis tools in RiceXPro can be used to extract genes with significant difference in expression levels between two samples (Figure 4A). Options are available for *P*-values of <0.01 and <0.05 significant levels. For FC analysis, options to identify up or downregulated genes can be selected from 2- to 10-fold. Any two samples can be compared with a list of genes with the corresponding FC values and *P*-value in descending order of FC values. Figure 4C shows the gene list generated by *t*-test and FC analysis with a setting of *P*-value (<0.01) and FC (>10) between leaf blade\_vegetative\_12:00 as a control sample and leaf blade\_vegetative\_00:00 as a test sample. As shown in the list, Os04g0583900 shows 767-fold level of nighttime expression as compared with daytime. The tabular list generated by this analysis can be downloaded as a text file. This analysis tool can be useful in extracting organ/tissue- or developmental stage-specific expressed genes, finding novel genes that maybe associated with organ/tissue development, and searching for candidate promoters that can be used for transgenic approaches.

### FUTURE DIRECTIONS

RiceXPro is designed to provide a comprehensive gene expression profile of all rice genes under different experimental conditions. The current version focusing on growth and development under natural field conditions will be further expanded to include gene expression profiles of various tissues and organs at different intervals during the growth process. Microarray analysis involving rice plants treated with various hormones, and rice plants subjected to biotic and abiotic stresses are also being undertaken. The field transcriptome data could then be used as reference for comparative expression data analysis thereby enhancing our understanding of many biological processes that determine productivity of an important agricultural crop.

### ACKNOWLEDGEMENTS

We thank Hajime Ohyanagi and Hiroshi Ikawa (Mitsubishi Space Software Co. Ltd) for critical reading of the manuscript, Ritsuko Motoyama for microarray analysis, and the members of the Genome Resource Center for preparation of samples.

### FUNDING

Genomics for Agricultural Innovation (RTR0002), Ministry of Agriculture, Forestry and Fisheries (MAFF) of Japan to Y. Nagamura. Funding for open access charge: National Institute of Agrobiological Sciences.

*Conflict of interest statement.* None declared.

### REFERENCES

- International Rice Genome Sequencing Project. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.
- Ohyanagi,H., Tanaka,T., Sakai,H., Shigemoto,Y., Yamaguchi,K., Habara,T., Fujii,Y., Antonio,B.A., Nagamura,Y., Imanishi,T. et al. (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.*, **34**, D741–D744.
- Rice Annotation Project. (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res.*, **17**, 175–183.
- Rice Annotation Project. (2008) The rice annotation project database (RAP-DB): 2008 update. *Nucleic Acids Res.*, **36**, D1028–D1033.
- Jung,K.H., An,G. and Ronald,P.C. (2008) Towards a better bowl of rice: assigning function to tens of thousands of rice genes. *Nat. Rev. Genet.*, **9**, 91–101.
- Miyao,A., Tanaka,K., Murata,K., Sawaki,H., Takeda,S., Abe,K., Shinozuka,Y., Onosato,K. and Hirochika,H. (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell*, **15**, 1771–1780.
- An,G., Lee,S., Kim,S.H. and Kim,S.R. (2005) Molecular genetics using T-DNA in rice. *Plant Cell Physiol.*, **46**, 14–22.
- Jeong,D.H., An,S., Park,S., Kang,H.G., Park,G.G., Kim,S.R., Sim,J., Kim,Y.O., Kim,M.K., Kim,S.R. et al. (2006) Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. *Plant J.*, **45**, 123–132.
- Hsing,Y.I., Chern,C.G., Fan,M.J., Lu,P.C., Chen,K.T., Lo,S.F., Sun,P.K., Ho,S.L., Lee,K.W., Wang,Y.C. et al. (2007) A rice

- gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.*, **63**, 351–364.
10. Nakamura,H., Hakata,M., Amano,K., Miyao,A., Toki,N., Kajikawa,M., Pang,J., Higashi,N., Ando,S., Toki,S. *et al.* (2007) A genome-wide gain-of-function analysis of rice genes using the FOX-hunting system. *Plant Mol. Biol.*, **65**, 357–371.
  11. Schmid,M., Davison,T.S., Henz,S.R., Pape,U.J., Demar,M., Vingron,M., Schölkopf,B., Weigel,D. and Lohmann,J.U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.*, **37**, 501–506.
  12. Zhu,T., Budworth,P., Chen,W., Provart,N., Chang,H.S., Guimil,S., Su,W., Estes,B., Zou,G. and Wang,X. (2003) Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnol. J.*, **1**, 59–70.
  13. Ma,L., Chen,C., Liu,X., Jiao,Y., Su,N., Li,L., Wang,X., Cao,M., Sun,N., Zhang,X. *et al.* (2005) A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*. *Genome Res.*, **15**, 1274–1283.
  14. Wang,Z., Liang,Y., Li,C., Xu,Y., Lan,L., Zhao,D., Chen,C., Xu,Z., Xue,Y. and Chong,K. (2005) Microarray analysis of gene expression involved in anther development in rice (*Oryza sativa* L.). *Plant Mol. Biol.*, **58**, 721–737.
  15. Furutani,I., Sukegawa,S. and Kyozuka,J. (2006) Genome-wide analysis of spatial and temporal gene expression in rice panicle development. *Plant J.*, **46**, 503–511.
  16. Jain,M., Nijhawan,A., Arora,R., Agarwal,P., Ray,S., Sharma,P., Kapoor,S., Tyagi,A.K. and Khurana,J.P. (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol.*, **143**, 1467–1483.
  17. Li,M., Xu,W., Yang,W., Kong,Z. and Xue,Y. (2007) Genome-wide gene expression profiling reveals conserved and novel molecular functions of the stigma in rice. *Plant Physiol.*, **144**, 1797–1812.
  18. Suwabe,K., Suzuki,G., Takahashi,H., Shiono,K., Endo,M., Yano,K., Fujita,M., Masuko,H., Saito,H., Fujioka,T. *et al.* (2008) Separated transcriptomes of male gametophyte and tapetum in rice: validity of laser microdissection (LM) microarray. *Plant Cell Physiol.*, **49**, 1407–1416.
  19. Hirano,K., Aya,K., Hobo,T., Sakakibara,H., Kojima,M., Shim,R.A., Hasegawa,Y., Ueguchi-Tanaka,M. and Matsuoka,M. (2008) Comprehensive transcriptome analysis of phytohormone biosynthesis and signaling genes in microspore/pollen and tapetum of rice. *Plant Cell Physiol.*, **49**, 1429–1450.
  20. Hobo,T., Suwabe,K., Aya,K., Suzuki,G., Yano,K., Ishimizu,T., Fujita,M., Kikuchi,S., Hamada,K., Miyano,M. *et al.* (2008) Various spatiotemporal expression profiles of anther-expressed genes in rice. *Plant Cell Physiol.*, **49**, 1417–1428.
  21. Wei,G., Tao,Y., Liu,G., Chen,C., Luo,R., Xia,H., Gan,Q., Zeng,H., Lu,Z., Han,Y. *et al.* (2009) A transcriptomic analysis of superhybrid rice *LYP9* and its parents. *Proc. Natl Acad. Sci. USA*, **106**, 7695–7701.
  22. Yazaki,J., Shimatani,Z., Hashimoto,A., Nagata,Y., Fujii,F., Kojima,K., Suzuki,K., Taya,T., Tonouchi,M., Nelson,C. *et al.* (2004) Transcriptional profiling of genes responsive to abscisic acid and gibberellin in rice: phenotyping and comparative analysis between rice and *Arabidopsis*. *Physiol. Genomics*, **17**, 87–100.
  23. Cheng,C., Yun,K.Y., Ressom,H.W., Mohanty,B., Bajic,V.B., Jia,Y., Yun,S.J. and de los Reyes,B.G. (2007) An early response regulatory cluster induced by low temperature and hydrogen peroxide in seedlings of chilling-tolerant japonica rice. *BMC Genomics*, **8**, e175.
  24. Jung,K.H., Dardick,C., Bartley,L.E., Cao,P., Phetsom,J., Canlas,P., Seo,Y.S., Shultz,M., Ouyang,S., Yuan,Q. *et al.* (2008) Refinement of light-responsive transcript lists using rice oligonucleotide arrays: evaluation of gene-redundancy. *PLoS One*, **3**, e301.
  25. Yun,K.Y., Park,M.R., Mohanty,B., Herath,V., Xu,F., Maulene,R., Wijaya,E., Bajic,V.B., Bruskiewich,R. and de los Reyes,B. (2010) Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. *BMC Plant Biol.*, **10**, e16.
  26. Jiao,Y., Tausta,S.L., Gandotra,N., Sun,N., Liu,T., Clay,N.K., Ceserani,T., Chen,M., Ma,L., Holford,M. *et al.* (2009) A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.*, **41**, 258–263.
  27. Wang,L., Xie,W., Chen,Y., Tang,W., Yang,J., Ye,R., Liu,L., Lin,Y., Xu,C., Xiao,J. *et al.* (2010) A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J.*, **61**, 752–766.
  28. Larkin,J.E., Frank,B.C., Gavras,H., Sultana,R. and Quackenbush,J. (2005) Independence and reproducibility across microarray platforms. *Nat. Methods*, **2**, 337–344.
  29. Rice Full-Length cDNA Consortium. (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science*, **301**, 376–379.
  30. Mihara,M., Itoh,T. and Izawa,T. (2010) SALAD database: a motif-based database of protein annotations for plant comparative genomics. *Nucleic Acids Res.*, **38**, D835–D842.
  31. Komiyama,R., Ikegami,A., Tamaki,S., Yokoi,S. and Shimamoto,K. (2008) *Hd3a* and *RFT1* are essential for flowering in rice. *Development*, **135**, 767–774.
  32. Komiyama,R., Yokoi,S. and Shimamoto,K. (2009) A gene network for long-day flowering activates *RFT1* encoding a mobile flowering signal in rice. *Development*, **136**, 3443–3450.
  33. Obayashi,T., Hayashi,S., Saeki,M., Ohta,H. and Kinoshita,K. (2009) ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucleic Acids Res.*, **37**, D987–D991.
  34. Usadel,B., Obayashi,T., Mutwil,M., Giorgi,F.M., Bassel,G.W., Tanimoto,M., Chow,A., Steinhäuser,D., Persson,S. and Provart,N.J. (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant Cell Environ.*, **32**, 1633–1651.
  35. Obayashi,T. and Kinoshita,K. (2010) Coexpression landscape in ATTED-II: usage of gene list and gene network for various types of pathways. *J. Plant Res.*, **123**, 311–319.
  36. Sakamoto,T., Miura,K., Itoh,H., Tatsumi,T., Ueguchi-Tanaka,M., Ishiyama,K., Kobayashi,M., Agrawal,G.K., Takeda,S., Abe,K. *et al.* (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol.*, **134**, 1642–1653.
  37. Shimura,K., Okada,A., Okada,K., Jikumaru,Y., Ko,K.W., Toyomasu,T., Sassa,T., Hasegawa,M., Kodama,O., Shibuya,N. *et al.* (2007) Identification of a biosynthetic gene cluster in rice for momilactones. *J. Biol. Chem.*, **282**, 34013–34018.