Genome-wide analysis of spatio-temporal gene expression patterns during early embryogenesis in rice

Jun-ichi Itoh\textsuperscript{a,1}, Yutaka Sato\textsuperscript{b}, Yutaka Sato\textsuperscript{c}, Ken-ichiro Hibara\textsuperscript{a}, Sae Shimizu-Sato\textsuperscript{c}, Hiromi Kobayashi\textsuperscript{a}, Hinako Takehisa\textsuperscript{b}, Karen A. Sanguinet\textsuperscript{d}, Nobukazu Namiki\textsuperscript{e}, Yoshiaki Nagamura\textsuperscript{b}

\textsuperscript{a}Graduate School of Agricultural and Life Sciences, University of Tokyo, Tokyo 113-8657, Japan
\textsuperscript{b}Genome Resource Unit, Agrogenomics Research Center, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
\textsuperscript{c}Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, 464-8601, Japan
\textsuperscript{d}Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA
\textsuperscript{e}Genome Informatics Department, Mitsubishi Space Software Co., Ltd., Takezono 1-6-1, Tsukuba, Ibaraki 305-0032, Japan

\textsuperscript{1}Corresponding author: Jun-ichi Itoh, ajunito@mail.ecc.u-tokyo.ac.jp
SUMMARY

Embryogenesis in rice is unique from most dicotolydonous plants in that it shows a non-stereotypic cell division pattern, the formation of dorsal-ventral polarity, and endogenous initiation of the radicle. To reveal the transcriptional features associated with developmental events during rice early embryogenesis, we used microarray analysis coupled with laser microdissection to obtain both spatial and temporal transcription profiles. Our results allowed us to determine spatial expression foci for each expressed gene in the globular embryo, which revealed the importance of phytohormone-related genes and a suite of transcription factors to early embryogenesis. Our analysis showed the polarized expression of a small number of genes along the apical-basal and dorsal-ventral axes in the globular embryo, which tended to fluctuate in later developmental stages. We also analyzed gene expression patterns in the early globular embryo and how it relates to expression in embryonic organs at later stages. We confirmed the accuracy of the expression patterns found via microarray analysis of embryo subdomains using in situ hybridization. Our study identified both homologous genes from Arabidopsis thaliana with known functions in embryogenesis in addition to unique and uncharacterized genes that show polarized expression patterns during embryogenesis. The results of this study are presented in a database to provide a framework for spatio-temporal gene expression during rice embryogenesis, to serve as a resource for future functional analysis of genes, and as a basis for comparative studies of plant embryogenesis.
INTRODUCTION

In higher plants, embryogenesis is a process to establish the body plan required for early vegetative growth after germination. Early embryogenesis includes critical developmental events for organizing plant architecture; namely, establishment of embryo polarity, pattern formation, and the differentiation of distinct cell types, tissues, and organs. Although these determination processes are common among plant species, there are differences in the patterns of cell division and determination processes (Natesh and Rau, 1984; Steeves and Sussex, 1989; West and Harada, 1993). The genetic and molecular mechanisms underlying developmental processes during embryogenesis have been investigated in the model eudiocot Arabidopsis thaliana. One of the advantages of Arabidopsis for such studies is its morphologically simple structure. In addition, Arabidopsis embryos undergo an invariant cell division pattern to generate diverse cell types and tissues, which allows cell lineages to be traced (Jenik et al., 2007). Embryonic studies of Arabidopsis also benefit from the abundance of genetic and genomic resources. Forward and reverse genetics studies increase our understanding of the molecular basis of embryo polarity and pattern formation, as well as cell type, tissue and organ differentiation mechanisms (Breuninger et al., 2008; Smith and Long, 2010; Ueda et al., 2011; Lau et al., 2012). Furthermore, genome-wide expression analyses using microarray and RNA-seq have been used for embryonic studies to estimate the number of genes expressed and predict the gene families associated with developmental and metabolic processes during embryogenesis (Nodine and Bartel, 2010; Xue et al., 2011; Palovaara et al., 2013). Transcriptome analyses combined with laser capture microdissection during Arabidopsis embryogenesis revealed global expression features in different parts of embryos and within embryos of different developmental stages (Casson et al., 2005; Spencer et al., 2007; Le et al., 2010; Belmonte et al., 2013).

Although it is important to determine whether the genetic programs governing embryo development during plant embryogenesis are evolutionally conserved among various plant species, molecular genetic work has mostly been conducted in Arabidopsis and is not as well understood in other species. Rice embryogenesis is a good candidate for comparative genetic studies, although the rice embryo shows a morphologically complex structure and the early embryo is deeply embedded in ovary tissues (Itoh et al., 2005). Rice embryogenesis shows biologically novel features not observed in Arabidopsis. First, since the embryo does not show stereotypic cell division patterns, embryonic cell types and tissues cannot typically be traced back to the early cell lineage. Second, the rice embryo shows an evident dorsal-ventral polarity where the shoot initiates on the ventral side and the scutellum forms dorsal side. Third, a radicle is endogenously differentiated in the center of the early embryo. Forth, complex embryo-endosperm interactions affect embryo development and growth throughout embryogenesis (Itoh et al., 2005; Nagasawa et al., 2013; Zhou et al., 2013). Thus, elucidation of the molecular mechanisms governing embryo development in rice will provide novel insights into embryogenesis of grasses and contributed to the diversity of embryonic studies in higher plants.

Functional analyses of genes involved in rice embryogenesis have been mainly performed using forward genetics.
approaches. One of the reasons why reverse genetics have not been applied to embryonic studies in rice is the limited number of knockout lines publically available. In addition, knockdown techniques such as RNAi or artificial miRNA are not successful because transgenic plants are typically generated from calli in rice and genes regulating embryo development and fertility are usually not silenced in regenerated plants. However, recent advances in genome editing techniques have improved the feasibility of the study of embryonic lethal null alleles. These techniques allow observation of null phenotypes of essential genes required for embryogenesis in rice. In particular, it has been shown that the CRISPR/Cas9 system is able to efficiently knockout multiple paralogous genes in rice for subsequent characterization in calli (Endo et al., 2015). Thus, our understanding of genetic regulation of embryo development will improve using genome editing-based reverse genetics.

For future functional analyses of grass embryogenesis, genome-wide expression profiling is necessary. Although expression profiling of various embryonic tissues and stages has been reported in rice and maize (Jiao et al., 2009; Takacs et al., 2012; Xu et al., 2012; Xue et al., 2012; Lu et al., 2013; Chen et al., 2014), predicting the spatial expression pattern and dynamics of gene expression during early embryogenesis remains difficult.

In this report, we performed transcriptome profiling of rice embryos using the Rice Agilent 44K microarray combined with laser microdissection. Our experimental design and data analysis allowed us to predict expression patterns and changes in the expression of specific genes during key developmental stages and in distinct embryonic organs and domains. In addition, our results point to the key role hormone-related genes and transcription factors play in early embryo development especially along the apical-basal and dorsal-ventral axes. These findings provide the foundation for future molecular analyses and for increased understanding of spatial gene regulation during rice embryogenesis.
RESULTS

Experimental design and datasets for the microarray

The developmental stages of rice embryogenesis have been previously defined (Itoh et al., 2005). After fertilization, embryos first show conspicuous morphological changes 3 days after pollination (3 DAP). However, the shape of embryos at 2 and 3 DAP is ovate and the apical-basal embryonic axis is obvious (A-B) (Fig. 1A and 1B). At 4 DAP, a protrusion of the coleoptile primordium is found on the ventral side of the embryo, which is the first morphological change along the dorsal-ventral (D-V) axis. During this stage, differentiation of other embryonic organs, including the shoot apical meristem (SAM), radical, and scutellum are also apparent (Fig. 1C). Accordingly, axis formation and pattern formation of embryos is presumably completed by 3 DAP. Organ differentiation is complete at 7-8 DAP and the embryo reaches its mature shape, although the organs continue to enlarge. At this stage, the primordium of the epiblast (a ventral protrusion from the basal part of the embryo) and a coleorhiza (a tissue enclosing the radicle) are apparent (Fig. 1D).

To acquire a comprehensive transcriptional profile of early embryogenesis in rice, we designed microarray experiments to obtain two of datasets: spatial and temporal. For temporal datasets, the entire embryo was collected at 2, 3 and 4 DAP using laser microdissection (LM) (supplementary material Fig. S1A-C). For spatial datasets, 3-DAP embryos were separated into half parts of the embryo along the A-B and D-V axis using LM (supplementary material Fig. S1D-G), which resulted in four sets of tissues derived from the apical, basal, dorsal and ventral regions of the embryo (Table 1). In addition to the spatial datasets for early embryos, we obtained embryo tissues closely associated with four embryonic organs at 7-DAP; the shoot, scutellem, root (radicle), and epiblast/coleorhizae, which were separated and collected by LM (Fig. 1D and supplementary material Fig. S2). These 11 samples were obtained in triplicate and hybridized to the 44K rice microarray (Agilent Technologies, http://www.agilent.com), which contains 36,100 independent probes corresponding to 27,370 loci published in RAP-DB (Rice Annotation Project, 2008; Sato et al., 2011) (Table 1). The processed raw signal intensity was subjected to 75th percentile normalization for inter-array comparison. The value was transformed to a log2 scale and the median expression value of the data within each dataset was subtracted for each probe. The processed value was assigned as the normalized expression value. In this study, we used 23,352 probes corresponding to 17,886 loci, with raw signal intensities above 50 in at least 3 of the 33 microarray datasets. Moreover, Pearson correlation analysis with all of the arrays showed high correlation coefficients among the replicates, indicating the reliable quality and reproducibility of the samples (supplementary material Table S1).

Expression profiles in 3-DAP embryos along the apical-basal and ventral-dorsal axes

To assess the spatial expression profiles of rice embryos prior to morphogenesis, we analyzed spatial datasets derived from four tissues of 3-DAP globular embryos. In our experimental design, relative expression values for each probe along the A-B axis were assigned by subtracting the normalized expression value of the basal region...
from that of the apical region. In the same way, relative expression values along the D-V axis were also evaluated by subtracting the ventral values from the dorsal values. Both values were plotted on a scatter diagram with the A-B axis depicted along the x-axis and D-V axis along the y-axis, which represents the relative expression foci of each probe in the entire 3-DAP embryo (Fig. 2). The pattern of data points for all probes on the scatter diagram was vertically oblong-shaped, indicating that the expression of a greater number of genes was biased along the A-B axis than the D-V axis (Fig. 2A). In addition, we observed a dispersion of the data points on the y-positive area, suggesting that genes whose expression was biased along the D-V axis are enriched in the apical part of the embryo.

To analyze the spatial expression profiles in more detail, we extracted genes preferentially expressed in the four parts of the embryo using moderated t-tests (Smyth, 2004) and fold-change analysis (FDR< 0.05). Despite the relatively lower threshold of the fold-change cut-off (FC >2), only 7% of the total expressed genes showed biased expression patterns along the A-B axis; 664 and 960 probes in the apical and basal regions, respectively (Fig. 2B and see supplementary material Table S2). Along the D-V axis, 162 and 81 probes (1% of the total number of expressed genes) showed differential expression between the dorsal and ventral regions, respectively (Fig. 2C and see supplementary material Table S3).

Along both the A-B and D-V axes, 106 probes showed significantly biased expression, and half (54 probes) were found in the apical-ventral region (Fig. 2D and see supplementary material Table S4). Among these, a few genes were detected as conspicuous outliers including: a homolog of Arabidopsis STYLISH 1 (Os09g0531600) (Eklund et al., 2010; Eklund et al., 2011) in the apical-dorsal region, OsYABBY4 (Os02g0643200) (Liu et al., 2007; Toriba et al., 2007) and an uncharacterized homeodomain containing protein (Os04g0566600) in the apical-ventral region, a hypothetical protein (Os08g0167600) in the basal-dorsal region, and OSH1 (Os03g0727000) in the basal-ventral region (Fig. 2D). The expression pattern of OSH1 has been reported based on in situ hybridization (Sato et al., 1996; Sentoku et al., 1999) (supplementary material Fig. S7) and was observed in a small domain just below the center of the ventral part of the 3-DAP embryo in agreement with our in silico prediction.

Relationship between temporal changes and spatial expression patterns in the globular embryo

To understand the relationship between temporal dynamics and spatial patterns of gene expression, we first categorized the expressed genes that showed fluctuating expression levels during early embryogenesis (12,232 probes) into eight groups (E1-E8) based on the similarity of expression patterns in the temporal dataset using K-means clustering algorithms (supplementary material Fig. S3A). Next, we extracted probes that were significantly biased along the A-B and D-V axes of the spatial datasets at 3 DAP (Fig. 2B and 2C) from the eight clusters. The total number of extracted probes preferentially expressed in the apical, basal, dorsal, and ventral regions was 552, 635, 67 and 130, respectively. We then examined the spatial expression patterns of the extracted probes using a scatter diagram for each cluster (supplementary material Fig. S3B and Table S5), and calculated the percentage of the extracted probe compared to the total number of preferentially expressed probes in the apical,
basal, dorsal and ventral region in each cluster (supplementary material Fig. S3C and Table S6).

Generally, the nearly 80% of genes were expressed similarly across the three stages and comprised expression groups E2, E4 and E6 yielding 38.6%, 13.4% and 27.0% of the total expressed genes (supplementary material Fig. S3D). The percentage of genes showing biased expression patterns either along the A-B or D-V axis in these groups was low (supplementary material Fig. S3D). For example, the E2 expression cluster represented 38.6% of the total expressed genes, but only a small subset were preferentially expressed, i.e. 17.0% (94/552) in the apical region, 14.0% (89/635) in the basal region, 14.9% (10/67) in the dorsal region and 11.5% (15/130) in the ventral region. In contrast, genes categorized into groups E3, E5 and E7 showed relatively larger expression changes during early embryogenesis (supplementary material Fig. S3A). These groups contained a small proportion of expressed genes, but were greatly enriched in genes expression preferentially expressed along the A-B and D-V axes. In particular, the gene cluster E5 comprised only 2.9% of the total analyzed genes, yet many of these genes showed biased expression in the apical, basal, dorsal and ventral regions; 21.4%, 8.5%, 23.9%, and 27.7%, respectively (supplementary material Fig. S3D). These results suggest a correlation between temporal changes and spatial patterns of gene expression during embryogenesis.

Spatial distribution of phytohormone-related genes

Although it has been reported that auxin and cytokinin are important phytohormones for embryonic patterning and cell specification during early embryogenesis in *Arabidopsis* (Müller and Sheen, 2008; Möller and Weijers 2009), limited information is available on the roles of auxin and cytokinin during early embryogenesis in rice. In addition, the roles of other phytohormones during rice embryogenesis remains limited. Therefore, we analyzed the spatial distributions of phytohormone-related genes in 3-DAP embryos (Fig. 3 and see supplementary material Table S7). The scatter diagrams for expression of phytohormone-related genes revealed that few genes showed biased expression patterns, although no consistent pattern for all data points was observed among the six phytohormones in 3-DAP embryos. Accordingly, the polarized biosynthesis, catabolism, and signaling of phytohormones by a limited number of genes may contribute to developmental or physiological processes in globular embryos.

Analysis of GA-related genes (Fig. 3A) revealed that OsSLRL1, a GRAS protein that functions as a repressor of GA signaling (Itoh et al., 2005), was preferentially expressed in the basal region. Although OsSLRL1 is highly expressed in inflorescence meristems and reproductive organs (Itoh et al. 2005), the function of OsSLRL1 in embryogenesis was previously unknown. We also found that OsGA20ox1, a GA biosynthesis gene, is expressed in the apical-dorsal region in 3-DAP embryo. Although bioactive GAs synthesized at the epithelium during germination are essential to induce α-amylase expression in the endosperm, the role of polarized expression of GA-related genes during early embryogenesis remains unclear.

The auxin-related genes OsIAA7 and OsARF10 were expressed in the apical and basal region (Fig. 3B). It has been shown that not only auxin biosynthesis and signaling but also polar auxin efflux is important for
auxin-dependent developmental processes during embryogenesis (Friml et al., 2003; Möller and Weijers 2009). *Arabidopsis* PIN1, PIN3, PIN4, and PIN7 are required to establish apical-basal polarity via directional auxin transport during embryogenesis (Friml et al., 2003). Thus, we examined the expression data of rice PIN-like genes and presented them in a scatter diagram (supplementary material Fig. S4). Some PIN-like genes showed polarized expression along the A-B axis. However, the rice PIN-like genes exhibiting the most polarized expression were homologous to PIN2, PIN5, and PIN8 (Wang et al., 2009), which do not have dominant functions in embryogenesis in *Arabidopsis*. Although accumulation patterns and localization of PIN1 proteins during embryogenesis were observed in maize (Forestan et al., 2010), it is important to identify the PIN-like genes that participate in auxin-dependent processes during early embryogenesis in rice.

We also identified three cytokinin (CK) A-type response regulator genes, *OsRR5*, *OsRR6* and *OsRR11* that were preferentially expressed in the apical-ventral region (Fig. 3 C). They were closely related and believed to be negative regulators of CK signaling (Hirose et al., 2007). Their closest homolog in maize is the ABPH1 gene known to play a role in the establishment of phyllotaxy by maintaining SAM size and PIN expression. It is specifically expressed in the SAM during embryogenesis (Giulini et al, 2004; Lee et al., 2009). Accordingly, the three response regulator genes detected in the apical-ventral region were required for SAM establishment before morphogenesis.

Brassinosteroid (BR-related) genes *OsDET2* and *OsBASIL1* are possibly involved in BR biosynthesis and inactivation, respectively, and were found to be expressed in the basal region (Fig. 3 D). The BR receptor-like protein, *OsBRL1*, which is partially associated with BR perception, was also preferentially expressed in the basal region. Because *OsBRL1* was mainly expressed in roots (Nakamura et al., 2006), the expression of *OsBRL1* in the basal region may be associated with radicle development during early embryogenesis.

Amongst ethylene and ABA-related genes (Fig. 3E and 3F), few showed biased expression patterns. For example, biosynthesis genes for ethylene and ABA were expressed in the basal region of the embryo (Fig. 3E). However, it is unknown whether these plant hormones are associated with early embryogenesis. Future studies will explore how these genes are involved in the regulation of early embryogenesis.

**Spatial distribution of transcription factors and transcriptional regulator genes**

Transcription factors (TF) and transcriptional regulator (TR) genes are known to play critical roles in various developmental and physiological processes during embryogenesis. Thus, we analyzed the spatial expression of TF and TR genes in 3-DAP embryos. TF and TR genes in rice have been categorized into dozens of families (Pérez-Rodríguez et al., 2010) (Table 2). Scatter diagrams of the 43 TF and TR families containing more than 10 probes were analyzed (Table S8). We found that data point patterns were considerably different among the TF and TR families. We then categorized TF and TR families into three groups according to the degree of dispersion of the data points, and three examples of each category with the plots of several outliers are shown in Fig. 4. Type I represents TF and TR families where greater than 20% of genes show higher absolute relative expression values.
(AREV = (x^2+y^2)^{1/2}>2.0); or more specifically more genes showing biased expression from the center of the embryo (Table 2). OFP (ovate family protein), LOB (lateral organ boundaries), C2C2-Dof, C2C2-YABBY, GRF (growth regulating factor), HB (homeobox), MADS, TCP, AUX/IAA and bHLH families were categorized in this group. For example, more than half of the OFP genes showed polarized expression patterns with AREVs greater than 2 (Fig. 4A). Similarly, about half of C2C2-Dof genes showed biased expression patterns, and their expression was preferentially detected in the apical region of the embryo (Fig. 4B). HB genes belong to the second-largest gene family, and 22 HB probes showed AREVs greater than 2. Among them, OSH1 and an uncharacterized WOX gene (Os07g0684900, Os05g0564500), which are closely related to WOX11/12 and WOX8/9 in Arabidopsis (Dai et al., 2007; Nardmann et al., 2007), showed more deviated expression patterns (Fig. 4C).

The TF and TR families in Type II contain only a few genes with biased expression patterns (Table 2 and Fig. 4D to 4F). A total of 19 TF and TR families, such as AP2-EREBP (which is the largest TF family), bZIP, and NAC were included in these families. In AP2-EREBP, a homolog of ESR/DRN that is known to be a developmental regulator of embryogenesis in Arabidopsis (Chandler et al., 2007; Cole et al., 2009) and an uncharacterized ERF/AP2 gene (Os02g0657000) exhibited deviated expression patterns (Fig. 4D). Similarly, two genes, OsNAM, a rice homolog of Arabidopsis CUC1, which is involved in shoot meristem formation (Hibara and Nagato, 2005), and OsNAC7, a homolog of SOMBRERO (SMB) involved in root cap development in Arabidopsis (Bennett et al., 2010), were expressed with polarized patterns, while the majority of other NACs were not (Fig. 4F).

A total of 13 TF and TR families were categorized into Type III (Table 2 and supplementary material Fig. S5). Data points of the genes in this group showed a concentrated pattern towards the center of the scatter diagram, suggesting that the expression of these genes was uniform. Six of eight TR families were categorized as Type III, suggesting that spatial regulation of TR families may be less important than that of TF families, excluding AUX/IAA and TRAF (BTB/POZ domain containing protein) families (Table 2).

These analyses revealed that the spatial distribution of gene expression was significantly different among the TF families. TF families in Type I, such as HB and YABBY, include putative homologs of well-known developmental regulators expressed in specific parts of early embryos in Arabidopsis (Capron et al., 2009). On the other hand, Type I TF families also contained genes whose functions in embryogenesis remain unknown, even in Arabidopsis. In particular, more than half of the OFP genes showed highly biased expression patterns in 3-DAP embryos (Fig. 4A). Although OFP family members play multiple roles in developmental processes as a transcriptional repressor, and few members of Arabidopsis OFP are known to interact with KNOX proteins (Li et al., 2011), further analysis of OFP genes is required to examine their involvement in embryo development. C2C2-Dof proteins were involved in various aspects of development, growth, and environmental response. The majority of Arabidopsis Dof genes were expressed in vascular tissues (Le Hir and Bellini, 2013). Some rice C2C2-Dof genes were expressed in the center of the apical region (Fig. 4B), suggesting that they are associated with early vascular development or specification in rice embryos.

Only a few genes in the Type II TF families showed biased expression patterns. In the case of NAC genes,
OsNAM and OsNAC7 showed highly polarized expression patterns (Fig. 4F), whose homologs in Arabidopsis are known to be involved in shoot and root development, respectively. It is interesting that genes showing biased expression patterns in rice are orthologs of important developmental regulators in Arabidopsis, such as CUC1, SMB, and ESR/DRN. Accordingly, spatial expression-based functional analysis will be used to explore the roles of uncharacterized TF genes in embryonic development.

**Relationship between expression profiles in globular embryo and later embryonic stages**

To examine the relationship between the expression patterns of genes in globular embryos and embryos in the later stages, we investigated expression profiles in four embryonic organs; i.e., shoot, scutellum, root (radicle), and epiblast/coleorhizae-containing tissues at 7 DAP. We identified genes whose expression values in one region of 7-DAP embryos were two-fold higher than in the other three regions; these were defined as tissue-specific upregulated genes. ANOVA (FDR< 0.05) revealed that 12,403 probes were differentially expressed among shoot, scutellum, root (radicle) and epiblast/coleorhiza-containing tissues at 7 DAP. Among the 12,403 probes, 393, 593, 401, and 453 were detected as tissue-specific upregulated genes in shoot, scutellum, root, and epiblast tissues, respectively (supplementary material Table S9).

We next analyzed the spatial pattern of expression of these tissue-specific upregulated genes and calculated the barycenter of the data points in 3-DAP embryos (supplementary material Fig. S6). Data points of shoot-specific were broadly scattered in the apical-ventral region. The barycenter of shoot-specific upregulated genes was located in the x-negative and y-positive area. Thus, genes specifically expressed in the shoot at 7 DAP tended to be expressed in the apical-ventral region of 3-DAP embryos. The pattern of data points of scutellum-specific upregulated genes was similar to that of shoot-specific upregulated genes, but more genes were detected in the apical-dorsal region. In contrast, global expression bias along the D-V axis was not observed in root-specific upregulated genes, and the barycenter of root-specific upregulated genes was relatively close to the origin of the diagram, indicating that many genes specifically expressed in the root tissue showed uniform or centered expression in 3-DAP embryos. The epiblast-specific upregulated genes were more concentrated in the basal region with a narrow range along the D-V axis, and the barycenter was located in the y-negative area near the x-axis. The relative positions of the barycenters for the four tissues in 3-DAP embryos corresponded to the relative positions of the organs in 7-DAP embryos.

**Confirmation of gene expression patterns using in situ hybridization**

To verify whether our microarray analyses of expression patterns were consistent with actual expression patterns, in situ hybridization experiments were performed with 55 genes expressed in a polarized pattern along the A-B and D-V axes. These genes were randomly selected using the following criteria: raw signal intensity greater than 200, annotated gene functions, and cloned full-length transcripts (supplementary material Table S10). In addition to the expression pattern at 3 DAP, we also examined the pattern at 4 DAP to evaluate changes from 3
DAP to 4 DAP. Expression patterns of most genes in 3-DAP embryos were consistent with the expression domain based on the microarray data, although some genes showed weak and uniform expression patterns (Fig. 5 and supplementary material Fig. S7). We selected examples of genes showing specific spatial expression patterns (Fig. 5). The Os06g0130100 gene was similar to ERL1 and ERL2 in Arabidopsis (Shpak et al., 2004), and its expression was strongly detected in the apical region of both 3- and 4-DAP embryos (Fig. 5A). The microarray data indicated that this gene should be expressed in the apical region of the 3-DAP embryo (scatter diagram in Fig. 5A). Similarly, Os03g0244600 encoding an AUX1-like protein was predicted to be in the apical region of the 3-DAP embryo; however, expression of the gene was detected in the center of the 3-DAP embryo and in the pre-vascular domain of the 4-DAP embryo (Fig. 5B). The microarray data indicated Os02g0614300 (Fig. 5C), Os09g0451000 (Fig. 5D), Os06g0520600 (Fig. 5E), and Os01g0691100 (Fig. 5F) were similarly expressed in the basal region of the 3-DAP embryo. In situ hybridization revealed these genes were expressed in the center (Fig. 5C), a small domain (Fig. 5D), a broad domain (Fig. 5E), as well as a few layers of the basal region (Fig. 5F) in 3-DAP embryos. The difference in expression patterns among the genes was more apparent in 4-DAP embryos. Os06g0265400 (Fig. 5G) was expressed in a few cells on the ventral side of the apical embryo and in the tip of presumptive coleoptile primordium based on the expression pattern of 4-DAP embryos. Although Os01g0954500 (Fig. 5H) was annotated as an unknown protein, it encodes a protein homologous to LATERAL ROOT PRIMORDIUM 1 in Arabidopsis (Smith and Fedoroff, 1995). Expression of this gene was detected on the dorsal side of 3-DAP embryos and in the scutellum of 4-DAP embryos. Taken together, these results indicate that our transcriptional profiling of early embryogenesis was reliable, and that clustering based on the expression domains may facilitate screening for marker genes for embryogenesis in rice.
DISCUSSION

In summary, we report herein the spatiotemporal transcriptome of embryogenesis in rice. The spatial data sets of 3-DAP embryo allowed us to predict the spatial bias of gene expression along the A-B and D-V axes in the globular embryo. Our data also showed temporal changes in the expression levels and sites of individual genes from globular to pre-mature embryos. The analysis suggests that the expression of a small number of genes is polarized along the A-B and D-V axes in the globular embryo, and that the expression level of these polarized genes tends to fluctuate depending on the developmental stage. This indicates that only a small number of expressed genes is polarized in the rice globular embryo, and a greater number of those genes show biased expression patterns along the A-B axis than the D-V axis. Considering that the globular embryo at 3 DAP showed no morphological changes along the D-V axis, the small number of genes showing polarized expression along the D-V axis must be required for successful morphogenesis and establishment of the D-V axis.

The spatial distribution of phytohormone-related genes suggests that a few genes related to each phytohormone show embryonic axis-dependent expression. For example, our finding that OsGA20ox1 is expressed in the apical-dorsal region of 3-DAP embryos is consistent with a previous report that OsGA20ox1 is expressed in the epithelium of the scutellum (Kaneko et al., 2003). Strikingly, our analysis of auxin efflux PIN genes revealed that rice homologs of PIN2, PIN5, and PIN8 were the only PINs showing polarized expression in the rice embryo, which differs from Arabidopsis embryogenesis. Interestingly, a rice homolog of the maize ABPH1 gene was preferentially expressed in the apical ventral region of the embryo suggesting a role for cytokinin A-type response regulators in regulating rice embryonic SAM formation and perhaps PIN expression. We also identified an OsBRL1 gene preferentially expressed in the basal region and BR perception is known to be important for root development. We did not detect many ABA or ethylene-related genes with polarized or biased expression in the embryo with the exception of a few biosynthesis genes. While the roles of most of the phytohormone-related genes in embryogenesis remains unknown, future functional characterization of some of the genes described in this study should confirm their importance during rice early embryogenesis.

The spatial gene expression patterns of TF and TR families differed significantly, indicating that specific TF families may contribute to early patterning of the embryo. The expression bias in the globular embryo was weakly related to the expression level of the embryonic organs at later stages, suggesting that organ-specific programs are initiated before morphogenesis.

Overall these analyses indicated that genetic programs required for embryonic organ function are initiated in globular embryos, and the relative positions of the embryonic organs are determined during this stage. It is not surprising that gene expression patterns in 3 DAP are associated with those at 7 DAP, but the spatial correlation between them was not as tight as expected. This suggested that the spatial expression of a number of genes can be significantly changed after morphogenesis (e.g., Fig. 5D). In addition, root-specific upregulated genes were observed around the center of 3-DAP embryo. Accordingly only a few genes are known to be involved with root development and function, e.g. OsNAC7 and OsBRL1, and were expressed in the basal region (Fig. 4F and 3D).
This indicates that the majority of genes associated with root function are likely activated in the center of the embryo at 3 DAP. This is consistent with the fact that the root primordium is formed endogenously near the center of the early embryo in grasses, unlike the Arabidopsis embryonic root which differentiates at the basal part of the embryo proper during the globular stage (Chandler et al., 2008).

Furthermore, we confirmed the accuracy of the expression patterns identified by microarray analysis using *in situ* hybridization, and found several genes expressed in a specific domain during early embryogenesis. However, we also found that some genes showing preferential expression in the microarray analysis did not always show similar expression patterns via *in situ* hybridization (e.g., Os06g0520600 and Os01g0691100) (Fig. 5E and Fig. 5F). One reason for this is that our experimental design cannot distinguish between gradients of expression levels in the radial direction. Further analysis using centric tissues of the 3-DAP embryos along the radial direction can be used to increase the accuracy of expression domains, which will be used to establish a “digital *in situ* expression” database in 3-DAP embryos of rice similar to, previous high-resolution expression analyses performed with Arabidopsis roots (Brady et al., 2007). Finally, expression data were deposited in the expression profile database, RiceXPro (http://ricexpro.dna.affrc.go.jp/RXP_4002/index.php) (Sato et al, 2011; Sato et al, 2013), and can be accessed in the form of raw and normalized expression graphs and expression site graphs similar to electronic Fluorescent Pictograph seen in other database (Winter et al., 2007) (supplementary material Fig. S8).

Our transcriptome study serves as a foundation for functional analysis of genes regulating rice embryogenesis and for diversity and conservation of genetic programs of embryogenesis by establishing common and different genetic landmarks for embryogenesis in various plant species.
MATERIALS AND METHODS

Plant materials and growth conditions
Rice (Oryza sativa L. japonica cv. Nipponbare) plants were grown in a paddy field under normal conditions. Ovaries at 2, 3, 4 and 7 days after pollination (DAP) were collected.

Sample preparation
For plastic sections, the collected ovaries were fixed in a 1:1:18 solution of formaldehyde, acetic acid, and 50% ethanol for 24 h at 4°C. After dehydration in a graded ethanol series, samples were embedded in Technovit 7100 (Heraeus Kulzer) and cut into 3-µm-thick sections with a microtome. Sections were stained with toluidine blue.

For LM, ovaries at 2, 3, 4 and 7 days after pollination (DAP) were fixed in pre-chilled solution of 75% ethanol/25% acetate with vacuum infiltration. Paraffin embedding was performed using a microwave processor (Energy Beam Sciences, http://www.ebsciences.com) according to Takahashi et al. (2010). Paraffin-embedded samples were cut into 10-µm-thick sections using a microtome (RM2255; Leica). Laser microdissection was performed using the Veritas Laser Microdissection System LCC1704 (Arcturus, now Life Technologies).

For in situ hybridization, samples were fixed in 4% paraformaldehyde in 0.1 M sodium phosphate buffer for 24 h at 48°C, and then dehydrated in a graded ethanol series. The dehydrated samples in 100% ethanol were replaced with xylene and embedded in Paraplast Plus (Leica Biosystems Richmond Inc., http://www.leica-microsystems.com/). Paraffin sections (8 mm thick) were applied to microscope slides coated with 3-aminopropyl triethoxysilane (Matsunami Glass). To generate probes for 55 genes (see supplementary material Table S10), the corresponding full-length cDNA clones deposited in the NIAS Genebank (http://www.dna.affrc.go.jp/distribution/) were ordered and used as templates. The full-length cDNA fragments were amplified by PCR and then transcribed using the Digoxigenin-labeled antisense riboprobes using a MAXIscript® In Vitro Transcription Kit (Life Technologies) with Digoxigenin-11-UTP (Roche).

RNA extraction and microarray analysis
RNA extraction and microarray analysis were performed according to Takehisa et al., 2011. Briefly, total RNA was extracted from the collected samples with a Pico-Pure RNA isolation kit (Arcturus, now Molecular Devices), and the Cy3-labeled cRNA was purified using the RNeasy Mini Kit (Qiagen, http://www.qiagen.com). A total of 1,000-ng Cy3-labeled cRNA was fragmented and hybridized on a slide of rice 4x44K microarray RAP-DB (G2519F#15241; Agilent Technologies). After washing, slides were scanned on an Agilent G2505B DNA microarray scanner, and background correction of the raw Cy3 signals was performed using the Feature Extraction 10.5.1.1 software (Agilent Technologies).
Statistical analysis

Normalization procedures were performed using the 75th percentile normalization for inter-array comparison and baseline correlation to the median of data within each dataset using GeneSpringGX12. The processed value was used as a normalized expression value. In addition, Pearson correlation analysis with all of the arrays using normalized data for looking at similarity between replicates was performed in EXCEL (Microsoft, http://www.microsoft.com). In this study, we used 23,352 probes corresponding to 17,886 loci, with raw signal intensities above 50 in at least 3 of the 33 microarray datasets. Differentially expressed genes between the apical and basal and dorsal and ventral tissue in 3-DAP embryos were statistically identified using Moderated t-tests (Smyth, 2004), and fold-change analysis was performed using GeneSpringGX12. P values were adjusted for multiple testing using Benjamini and Hochberg’s method for correcting FDR. The relative expression values along the A-B axis and the D-V axis were calculated by subtracting the normalized expression values of the basal region from those of the apical region, and those of the ventral region from the dorsal region, respectively. Scatter diagrams of spatial expression profiles in 3-DAP embryos were constructed in EXCEL. Differentially expressed genes among the temporal data set were categorized into eight clusters using K-means cluster analyses based on a Euclidean algorithm with GeneSpringGX12. For extraction of differentially expressed genes in the temporal data set (whole embryo at 2, 3, and 4 DAP) and the sub-spatial data sets (shoot, scutellem, root (radicle) and epiblast/coleorhizae related tissues at 7-8 DAP), we performed ANOVA (FDR < 0.05) and fold-change analyses using the relative expression values in each dataset.

Gene annotation used for analyses

The gene annotation used in this study was obtained from the RAP-DB (Rice Annotation Project, 2008). The list of genes associated with biosynthesis, catabolism, and signaling for the six phytohormones described here was derived from Hirano et al., 2008. For extraction of transcription factor and transcriptional regulator genes from the rice genome, the list in the PLANT TRANSCRIPTION FACTOR DATABASE (Pérez-Rodríguez et al., 2010, http://plntfdb.bio.uni-potsdam.de/v3.0/) was used.

Accession numbers

The data obtained in this study have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), and are accessible through GEO Series accession number (GSE63110).

ACKNOWLEDGEMENTS

We would like to thank Ms Ritsuko Motoyama (NIAS) for microarray analysis and Ms Minami Omae for help with in situ hybridization experiments. We also thank Ms Kaori Kamatsuki (MSS) for helping with the database construction and Ms Hisae Kamakura (NIAS) for help with laser microdissection experiments.
FOOTNOTES

Competing interests

The authors declare no competing financial interests.

Author contributions


Funding

This work and the database construction of this work was supported mainly by a project of the Ministry of Agriculture, Forestry and Fisheries of Japan (Genomics for Agricultural Innovation, RTR0002, RTR0006, and Development of Genome Information Database System for Innovation of Crop and Livestock Production), and partly by a grant from a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (24380005 to J.I.).
REFERENCES


Table 1. Embryo stages and tissues collected using LMD.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tissue (abbreviation)</th>
<th>DAP</th>
<th>Stage (*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal dataset</td>
<td>Whole</td>
<td>2</td>
<td>Early globular (Em3)</td>
</tr>
<tr>
<td></td>
<td>Whole</td>
<td>3</td>
<td>Late globular (Em4)</td>
</tr>
<tr>
<td></td>
<td>Whole</td>
<td>4</td>
<td>Coleoptiler (Em5)</td>
</tr>
<tr>
<td>Spatial dataset</td>
<td>Apical half</td>
<td>3</td>
<td>Late globular (Em4)</td>
</tr>
<tr>
<td></td>
<td>Basal half</td>
<td>3</td>
<td>Late globular (Em4)</td>
</tr>
<tr>
<td></td>
<td>Dorsal half</td>
<td>3</td>
<td>Late globular (Em4)</td>
</tr>
<tr>
<td></td>
<td>Ventral half</td>
<td>3</td>
<td>Late globular (Em4)</td>
</tr>
<tr>
<td></td>
<td>Shoot (ST)</td>
<td>7</td>
<td>Completion of organogenesis (Em7)</td>
</tr>
<tr>
<td></td>
<td>Root (RT)</td>
<td>7</td>
<td>Completion of organogenesis (Em7)</td>
</tr>
<tr>
<td></td>
<td>Scutellum (SC)</td>
<td>7</td>
<td>Completion of organogenesis (Em7)</td>
</tr>
<tr>
<td></td>
<td>Epiblast/Coleorhiza (EP)</td>
<td>7</td>
<td>Completion of organogenesis (Em7)</td>
</tr>
</tbody>
</table>

* Stage symbol defined by Itoh et al. (2005)
Table 2. Spatial distribution of transcription factor and transcriptional regulator genes in 3-DAP embryos.

<table>
<thead>
<tr>
<th>Type(^a)</th>
<th>Name of TF and TR(^b)</th>
<th>Number of TF probes expressed in 3DAP embryo</th>
<th>Number of probes with AREV more than 2</th>
<th>Percentage of probes with AREV more than 2</th>
<th>TF or TR(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I</td>
<td>OFP</td>
<td>18</td>
<td>10</td>
<td>55.6</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>LOB</td>
<td>16</td>
<td>8</td>
<td>50.0</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>C2C2-Dof</td>
<td>17</td>
<td>8</td>
<td>47.1</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>C2C2-YABBY</td>
<td>12</td>
<td>5</td>
<td>41.7</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>GRF</td>
<td>11</td>
<td>4</td>
<td>36.4</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>HB</td>
<td>110</td>
<td>37</td>
<td>33.6</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>MADS</td>
<td>49</td>
<td>14</td>
<td>28.6</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>TCP</td>
<td>19</td>
<td>5</td>
<td>26.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>AUX/IAA</td>
<td>32</td>
<td>8</td>
<td>25.0</td>
<td>TR</td>
</tr>
<tr>
<td></td>
<td>bHLH</td>
<td>73</td>
<td>17</td>
<td>23.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>ABI3VP1</td>
<td>26</td>
<td>5</td>
<td>19.2</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>AP2-EREBP</td>
<td>119</td>
<td>22</td>
<td>18.5</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>C2H2</td>
<td>57</td>
<td>9</td>
<td>15.8</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>Trihelix</td>
<td>19</td>
<td>3</td>
<td>15.8</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>WRKY</td>
<td>45</td>
<td>7</td>
<td>15.6</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>C2C2-GATA</td>
<td>21</td>
<td>3</td>
<td>14.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>Orphans</td>
<td>46</td>
<td>5</td>
<td>10.9</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>G2-like</td>
<td>29</td>
<td>3</td>
<td>10.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>TRAF</td>
<td>21</td>
<td>2</td>
<td>9.5</td>
<td>TR</td>
</tr>
<tr>
<td>Type II</td>
<td>MYB-related</td>
<td>66</td>
<td>6</td>
<td>9.1</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>MYB</td>
<td>60</td>
<td>5</td>
<td>8.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>HSF</td>
<td>37</td>
<td>3</td>
<td>8.1</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>ARF</td>
<td>38</td>
<td>3</td>
<td>7.9</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>NAC</td>
<td>52</td>
<td>4</td>
<td>7.7</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>bZIP</td>
<td>82</td>
<td>6</td>
<td>7.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>GRAS</td>
<td>42</td>
<td>3</td>
<td>7.1</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>SBP</td>
<td>14</td>
<td>1</td>
<td>7.1</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>FHA</td>
<td>19</td>
<td>1</td>
<td>5.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>CCAAT</td>
<td>47</td>
<td>2</td>
<td>4.3</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>C3H</td>
<td>74</td>
<td>1</td>
<td>1.4</td>
<td>TF</td>
</tr>
<tr>
<td>Type III</td>
<td>Jumonji</td>
<td>13</td>
<td>0</td>
<td>0.0</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>mTERF</td>
<td>28</td>
<td>0</td>
<td>0.0</td>
<td>TR</td>
</tr>
<tr>
<td></td>
<td>PHD</td>
<td>47</td>
<td>0</td>
<td>0.0</td>
<td>TR</td>
</tr>
<tr>
<td></td>
<td>SET</td>
<td>34</td>
<td>0</td>
<td>0.0</td>
<td>TR</td>
</tr>
<tr>
<td></td>
<td>SNF2</td>
<td>38</td>
<td>0</td>
<td>0.0</td>
<td>TR</td>
</tr>
<tr>
<td></td>
<td>Tify</td>
<td>12</td>
<td>0</td>
<td>0.0</td>
<td>TF</td>
</tr>
<tr>
<td></td>
<td>TUB</td>
<td>22</td>
<td>0</td>
<td>0.0</td>
<td>TF</td>
</tr>
</tbody>
</table>

AREV: Absolute relative expression values.

\(^a\)The types are categorized based on the percentage of probes with AREV greater than 2. Type I, II, and III are the TF and TR families containing more than 20%, between 20% and 0%, and 0% sites with AREV of greater than 2, respectively.

\(^b\)TF and TR families are categorized using the PLANT TRANSCRIPTION FACTOR DATABASE (http://plntfdb.bio.uni-potsdam.de/v3.0/).
Figures

Figure 1. Embryogenesis of early embryos in rice.

(A-E) Morphology of rice embryos during embryogenesis. The putative embryonic axes are indicated in the panels. (A) Embryos at 2 days after pollination (2 DAP). (B) 3 DAP. (C) 4 DAP. (D) 8 DAP. Colored outlines in red, green, blue and yellow indicate the regions dissected by LM as the shoot, scutellum, root, and epiblast/coleorhiza-containing tissues, respectively. (E) Mature embryo. Ct: coleoptile, Sm: shoot meristem, Sc: scutellum, Rd: radicle, St: shoot, Eb: epiblast, Cr: coleorhiza.
Figure 2. Spatial expression pattern of genes in 3-DAP embryos.

(A) Expression sites of all genes expressed in 3-DAP embryos depicted in scatter diagrams. The position of a dot indicates the virtual expression site of each probe in 3-DAP embryos. X- and Y-axes indicate the relative expression values (log2) along the dorsal-ventral and apical-basal axes, respectively. (B) Expression sites of genes showing significantly biased expression along the apical-ventral axis. (C) Expression sites of genes showing significantly biased expression along the dorsal-ventral axis. (D) Expression sites of genes showing significantly biased expression along both the apical-basal and dorsal-ventral axes. Red circles indicate probes showing conspicuously biased expression along the apical-basal and/or dorsal-ventral axes as outliers. The RAP-ID and annotation of the outliers are presented. The number of probes in each group is indicated in the diagram.
Figure 3. Spatial distribution of phytohormone-related genes.

Expression sites of phytohormone-related genes expressed in 3-DAP embryos in the scatter diagram. X- and Y-axes indicate the relative expression values (log2) along the dorsal-ventral and apical-basal axes, respectively. Red circles indicate probes that showed conspicuously biased expression along the dorsal-ventral and/or apical-basal axes as outliers. The RAP-ID and annotations of the outliers are represented. (A) Gibberellin, (B) auxin, (C) cytokinin, (D) brassinosteroid, (E) ethylene, and (F) abscisic acid.
Figure 4. Spatial distribution of transcription factor and transcriptional regulator genes.

Sites of transcription factor (TF) and transcriptional regulator (TR) genes expression in 3-DAP embryos in the scatter diagram. X- and Y-axes indicate the relative expression values (log2) along the dorsal-ventral and apical-basal axes, respectively. Data points inside the green circle represent probes with absolute relative expression values of less than 2. Red circles indicate probes that showed conspicuously biased expression along the dorsal-ventral and/or apical-basal axes as outliers. The RAP-ID and annotations of outliers are presented. (A) to (C) and (D) to (F) TF and TR categorized as Type I and II, respectively (See text and Table 2). (A) Ovate family protein, (B) C2C2-Dof, (C) Homeobox, (D) AP2-EREBP, (E) bZIP and (F) NAC.
Figure 5. Spatial expression patterns of several potential marker genes.

(A) Os06g0130100  (B) Os03g0244600  (C) Os02g0614300  (D) Os09g0451000  (E) Os06g0520600  (F) Os01g0691100  (G) Os06g0265400, and (H) Os01g0954500. On the left of the upper and lower images of each panel are the expression patterns of 3- and 4-DAP embryos, respectively. The right scatter diagram indicates the predicted expression site of the corresponding probe based on microarray analysis. The RAP-ID, probe number, annotations and marker ID (see supplementary material Table S10) of the genes are indicated. Scale bars = 50 µm.