CellPress
PARTNER JOURNAL

Molecular Plant
Research Article

Genome-Wide Identification and Functional Analysis of Genes Expressed Ubiquitously in Rice

Ki-Hong Jung[1,2,3,*], Sung-Ruyl Kim[1,3], Hoi-Khoanh Giong[1], Minh Xuan Nguyen[1], Hyun-Jung Koh[1] and Gynheung An[1,2,*]

[1]Department of Plant Molecular Systems Biotechnology & Crop Biotech Institute, Kyung Hee University, Yongin 446-701, Korea

[2]Graduate School of Biotechnology, Kyung Hee University, Yongin 446-701, Korea

[3]These authors contributed equally to the article.

***Correspondence: Ki-Hong Jung (khjung2010@khu.ac.kr), Gynheung An (genean@khu.ac.kr)**

http://dx.doi.org/10.1016/j.molp.2014.10.013

ABSTRACT

Genes that are expressed ubiquitously throughout all developmental stages are thought to be necessary for basic biological or cellular functions. Therefore, determining their biological roles is a great challenge. We identified 4034 of these genes in rice after studying the results of Agilent 44K and Affymetrix meta-anatomical expression profiles. Among 105 genes that were characterized by loss-of-function analysis, 79 were classified as members of gene families, the majority of which were predominantly expressed. Using T-DNA insertional mutants, we examined 43 genes and found that loss of expression of six genes caused developing seed- or seedling-defective phenotypes. Of these, three are singletons without similar family members and defective phenotypes are expected from mutations. Phylogenomic analyses integrating genome-wide transcriptome data revealed the functional dominance of three ubiquitously expressed family genes. Among them, we investigated the function of *Os03g19890*, which is involved in ATP generation within the mitochondria during endosperm development. We also created and evaluated functional networks associated with this gene to understand the molecular mechanism. Our study provides a useful strategy for pheonome analysis of ubiquitously expressed genes in rice.

Key words: functional redundancy, phylogenomic analysis, rice, systematic phenotype screening, ubiquitously expressed genes

Jung K.-H., Kim S.-R., Giong H.-K., Nguyen M.X., Koh H.-J., and An G. (2015). Genome-Wide Identification and Functional Analysis of Genes Expressed Ubiquitously in Rice. Mol. Plant. **8**, 276–289.

INTRODUCTION

Because they are constitutively expressed in most tissues or organs throughout the plant life cycle, ubiquitously expressed genes are considered necessary for basic biological or cellular functions. They include housekeeping genes such as *Actin*, *Glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*), and *Ubiquitin* (Jung et al., 2005; Jain et al., 2006). These types of genes are also widely used as internal controls for quantitative RT–PCR, microarrays, and Northern analyses (Lee et al., 2007). Because they are thought to be essential, it is difficult to study their biological roles. Functional redundancy further restricts those determinations (Jung et al., 2008c). Therefore, the simplest approach is to examine singletons rather than gene families. The efficacy of this method has been demonstrated with systematic functional analyses of the genes responsible for light responses, e.g., six of 11 singletons have shown phenotypic defects in their T-DNA insertion lines (Jung et al., 2008c). Other defects have been found in four of nine lines where T-DNA was inserted into a predominant member within a gene family (Jung et al., 2008c). Those reports indicate that predominant members within a gene family are good targets for loss-of-function analyses.

Microarray is a useful tool for identifying genome-wide transcripts associated with a specific biological event or environmental condition. More than 5000 rice microarray analyses are available from public databases such as the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) and ArrayExpress (Parkinson et al., 2009; Barrett et al., 2011; Jung et al., 2011). Of these, microarray data from commercial platforms such as Affymetrix and Agilent are valuable sources. We have previously used 1150 Affymetrix and 209 Agilent 44K microarrays for meta-profiling analysis (Cao et al., 2012). Genevestigator includes 2487 Affymetrix rice array data and several meta-profiling databases applicable to anatomy,

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

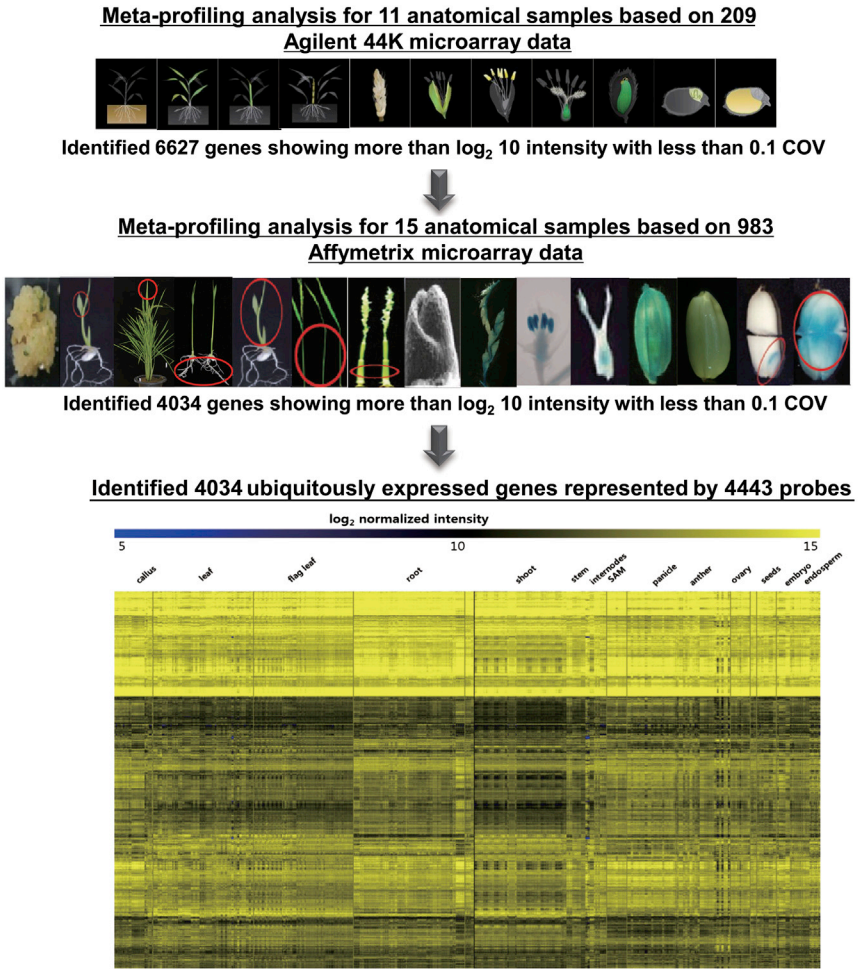

Figure 1. Procedures for Evaluating Ubiquitously Expressed Genes in Rice Using the Platform of Both Affymetrix and Agilent 44K Arrays.

A total of 6627 genes showing intensities greater than $\log_2 10$ with less than 0.1 CV were identified from Agilent 44K anatomical meta-profiles and further refined with Affymetrix anatomical meta-profiles. These independent meta-profiles revealed 4034 genes as ubiquitously expressed. In the heatmap, blue indicates a low level of expression based on microarray data; yellow indicates high expression.

developmental stage, stress, and mutations (Zimmermann et al., 2008). A large collection of human microarray databases related to various tissues, disease states, and cell lines have also been used to identify genes that are consistently expressed (Lee et al., 2007) . However, no reports of a genome-wide analysis and identification of ubiquitously expressed genes are available for plants.

In this study, we identified 4034 ubiquitously expressed genes. Functional analysis of 43 genes via T-DNA insertional mutagenesis led to the identification of six genes associated with defective phenotypes.

RESULTS

Identification of Ubiquitously Expressed Genes in Rice

To identify ubiquitously expressed genes in rice, we used microarray data from public databases, which included NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/) and ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) (Parkinson et al., 2009; Barrett et al., 2011). Anatomical meta-profiling databases are supported by 983 Affymetrix microarray data divided into 15 anatomical organ categories and 209 Agilent 44K microarray data assigned to 11 organ categories (Cao et al., 2012). From these analyses, we found 4034 loci with intensity values that were greater than $\log_2 10$ and with coefficient of variation (CV) values less than 0.1 (Supplemental Table 1). Our overall strategy is presented in Figure 1.

Verification of Expression Patterns Using RT–PCR and GUS Assays

Using RT–PCR analysis, we examined the ubiquitous expression patterns of 31 genes (Supplemental Figure 1). We also monitored the *in vivo* expression of *Os03g01910*, *Os03g08010*, *Os06g30750*, and *Os04g42090* by using promoter trap lines that expressed a fusion protein between the gene product in front of T-DNA insertion and GUS reporter (Figures 2 and 3). We have previously described this technique (Jeon et al., 2000; Jeong et al., 2006). *Os03g01910*-encoding RNA polymerase B transcription factor 3 (BTF3) is thought to have a role in the formation ofa stable complex with RNA polymerase II to initiate transcription (Tanaka et al., 2010). *Os03g08010* encoding elongation factor 1α (OsEF1α) has been identified as a ubiquitously expressed gene (Jain et al., 2006). Our *GUS* expression analyses also demonstrated their ubiquitous functioning (Figure 2). Whereas *Os06g30750* is known to encode reticulon domain-containing protein and *Os04g42090* encodes conserved peptide upstream of open reading frame-containing transcript 7 (CPuORF7), annotated functions have been less clear than those for the other two genes. Nevertheless, measurements of GUS activity in our T-DNA tagged lines indicated that those two are ubiquitously expressed in seedling roots and leaves, mature leaves, floral organs, and geminating seeds (Figure 3). RT–PCR analyses of all four genes also supported the belief that they are ubiquitously expressed (Supplemental Figure 1 and Supplemental Table 2). Schematic diagrams of T-DNA insertions in the promoter trap lines are presented in Supplemental Figure 2.

Biological Processes Associated with the Ubiquitously Expressed Genes

To identify their functional roles, we performed Gene Ontology (GO) enrichment analysis with the Rice Oligonucleotide Array Database (ROAD; http://www.ricearray.org/) (Jung et al.,