Gene 530 (2013) 236–247

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

# Genome-wide analysis of endosperm-specific genes in rice

Dong-Ming Nie [a], Yi-Dan Ouyang [a], Xin Wang [a], Wei Zhou [a], Chun-Gen Hu [b], Jialing Yao [a,*]

[a] College of Life Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
[b] College of Horticulture and Forestry Science, Huazhong Agricultural University, Wuhan 430070, China

ARTICLE INFO

*Article history:*
Accepted 26 July 2013
Available online 13 August 2013

*Keywords:*
Endosperm-specific expression
Genome
*Cis*-element
Expression profile
Rice

ABSTRACT

The endosperm of the cereal crop is an important nutrient source for humans. It also acts as a critical integrator of plant seed growth and development. Despite its importance, the comprehensive understanding in regulating of endosperm development in rice remains elusive. Here, we performed a genomic survey comprising the identification and functional characterization of the endosperm-specific genes (*OsEnS*) in rice using Affymetrix microarray data and Gene Ontology (GO) analysis. A total of 151 endosperm-specific genes were identified, and the expression patterns of 13 selected genes were confirmed by qRT-PCR analysis. Promoter regions of the endosperm-specific expression genes were analyzed by PLACE Signal Scan Search. The results indicated that some motifs were involved in endosperm-specific expression regulation, and some *cis*-elements were responsible for hormone regulation. The bootstrap analysis indicated that the RY repeat (CATGCA box) was over-represented in promoter regions of endosperm-specific expression genes. GO analysis indicated that these genes could be classified into 12 groups, namely, transcription factor, stress/defense, seed storage protein (SSP), carbohydrate and energy metabolism, seed maturation, protein metabolism, lipid metabolism, transport, cell wall related, hormone related, signal transduction, and one unclassified group. Taken together, our results provide informative clues for further functional characterization of the endosperm-specific genes, which facilitate the understanding of the molecular mechanism in rice endosperm development.



## 1. Introduction

The endosperm, a terminally differentiated tissue that nourishes the embryo during seed development, acts as a critical integrator of seed growth and development. The endosperm of cereal is a highly specialized storage organ for nutrients, including starch, proteins and lipids. It is the most important source of food and feed for humans and animals, which also provides raw materials for manufacturing countless industrial goods such as biofuels. Explorations in endosperm development are of significant importance in understanding the molecular mechanisms involved in plant seed development. These findings also have important implications for crop genetic improvement.

The endosperm may have an ephemeral existence, particularly in dicots such as *Arabidopsis thaliana*, which is consumed by the developing embryo. On the contrary, in monocotylous cereals such as maize and rice, the endosperm is a prominent feature of the mature seed (Brown and Lemmon, 2007; Brown et al., 1999; Mu et al., 2010). Rice endosperm develops and matures within a short period. Besides, it contains several specialized cell types, which is a good model for developmental biology study (Nguyen et al., 2007). Rice endosperm undergoes the *ab initio* nuclear type of development, which is the most prevalent and best characterized type of endosperm formation (Olsen, 2004). After fertilization, the triploid primary endosperm nucleus undergoes mitotic divisions without cell-wall formation, resulting in a multi-nucleated syncytium (Sabelli and Larkins, 2009b). Generally, cell wall formation occurs at three days after pollination (Xi and Zheng, 2011). Soon afterwards, endosperm cells differentiate into four major types of tissue: the starchy endosperm, the aleurone layer, transfer cell and embryo-surrounding region (Nguyen et al., 2007; Sabelli and Larkins, 2009b). The starchy endosperm and the aleurone layer accumulate storage compounds such as starch, proteins, and lipids (Ishimaru et al., 2007). The transfer cells absorb and transport nutrients from mother tissues to endosperm (Royo et al., 2007; Sabelli and Larkins, 2009b). Finally, the endosperm enters into the maturation stage and the starchy endosperm initiates programmed cell death (PCD), leaving the cells of the aleurone layer alive in the mature endosperm (Sabelli, 2012; Young and Gallie, 2000).

Genetic evidence has identified a series of mutants that affect endosperm development. To date, some *Arabidopsis* mutants with defects in endosperm development have been identified, which seem to have common phenotype of aborted or small seeds, such as *s5*, *exs*, *tor*, *haiku*/*iku*, *mimi3*, and *agl62* (Pagnussat et al., 2005; Sabelli and Larkins,

*Abbreviations:* NAA, 1-Naphthaleneacetic acid; ABA, Abscisic acid; BAC, Bacterial artificial chromosomes; bZIP, Basic leucine zipper; DAP, Days after pollination; DOF, DNA binding with one finger; *OsEns* gene, Endosperm-specific gene; FAA, Formalin–acetic acid–alcohol; GO, Gene ontology; GA, Gibberellin; IAA, Indole acid; KT, kinetin; KD, Knock-down; KEGG, Kyoto Encuclopedia of Genes and Genomes; LEA, Late embryogenesis abundant; nr, Non-redundant; ORF, Open reading frame; PAC, Phage artificial chromosome; PCD, Programmed cell death; qRT-PCR, Real-time quantitative polymerase chain reaction; RNAi, RNA interference; SSP, Seed storage protein; TF, Transcription factor.

* Corresponding author. Tel./fax: +86 27 87282866.
*E-mail address:* yaojlmy@mail.hzau.edu.cn (J. Yao).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.07.088

2009a; Sabelli et al., 2007). In cereals, similar phenomenon was also observed. Maize *des5* mutant seeds are smaller than wild-type seeds with a defective outcome, due to the development arrest of endosperm and a prematurely aborted embryo (Olsen et al., 2008); Maize *emp4* mutants show more profound phenotypic defects on the endosperm than on the embryo. And at the maturity period, homozygous *emp4* seeds are unable to germinate (Gómez-Porras et al., 2007). In maize, the *MEG1* with a maternal patent-of-origin expression in endosperm transfer cell may regulate nutrient trafficking, which forms the maternal tissue into the developing seed (Gutiérrez-Marcos et al., 2004). A maize transcriptional activator *MRP-1* is expressed in the transfer cell. It regulates several transfer cell-specific genes, which might be a key regulator of the transfer cell differentiation process (Gómez et al., 2009). After fertilization, the endosperm of rice *enl1* mutant starts to develop but degenerates at an early stage, resulting in the production of a very large embryo (Miyoshi et al., 2000). Recently, two endosperm transfer cell-specific genes of *AL1* and *AL2* were identified in rice. The result suggested that only the promoter region of *AL1* exhibited transfer cell-specific activity. However, the function of *AL1* gene is still unclear (Kuwano et al., 2011).

Many efforts have been made to characterize the synthesis and accumulation mechanism of endosperm storage components. A large number of mutants with starch and seed storage proteins synthesis and accumulation have been reported. In rice, 11 starch synthesis related enzymes are specifically expressed in developing endosperm, which are exclusively involved in amylopectin biosynthesis (Fujita et al., 2007; Hwang et al., 2010; Li et al., 2009; Ohdan et al., 2005; Yu et al., 2011). Among them, OsPUL is important for synthesis of short-chain maltooligosaccharides during the initial stage of starch synthesis (Hwang et al., 2010); SSIIIa affects the structure of amylopectin (Fujita et al., 2007). The *GluD-1* gene encoding a glutelin protein is maximally expressed in the starchy endosperm (Kawakatsu et al., 2008). Fu et al. reported that *RSR1* negatively regulated the starch synthesis genes in developmental rice seeds (Fu and Xue, 2010); OsVPE1 is required for glutelin maturation and OsRab5a protein plays an important role in storage protein trafficking (Y. Wang et al., 2009, 2010).

The analysis of spatial and temporal expression of tissue-specific genes is of great value to investigate the molecular mechanism of plant development. Systemic transcriptomic analysis of seed development has been performed in *Arabidopsis* (Day et al., 2008), maize (Xiong et al., 2011) and rice (Xue et al., 2012). Although the expression analyses in endosperm-related genes were involved, these studies mainly focused on the metabolic pathways of endosperm such as nutrient partitioning. Thus the understandings in regulatory mechanism of the endosperm development remain limited. Analysis of endosperm-specific gene will be helpful for cloning and identifying novel genes involved in endosperm development and nutrient synthesis. In addition, it will shed light on illuminating the metabolism mechanism of cell-fate determination and differentiation, which improves the yield and quality in rice and facilitates the cereal crop improvement. Here, microarray data in CREP database (http://crep.ncpgr.cn/crep-cgi/home.pl) (L. Wang et al., 2010) was used to determine the genes specifically expressed in rice endosperm. Real-time quantitative PCR analyses of 13 selected genes were performed to confirm the microarray data. We focus on the identification and GO analysis of the *OsEnS* genes (endosperm-specific genes) and their corresponding *cis*-elements in promoter regions. These results will provide further clues to investigate the functions and molecular mechanism of these candidate genes in rice endosperm development.

## 2. Materials and methods

### 2.1. Identifying of the OsEnS genes in rice

The tissue-specific expressed gene data was used to identify the endosperm-specific genes in the whole life cycle of rice (callus were excluded) (Table S1) (L. Wang et al., 2010). Three endosperm developmental stages, 7 days after pollination (DAP), 14DAP and 21DAP, were investigated in two rice varieties of Zhenshan 97 and Minghui 63. The genes that expressed in at least one endosperm tissue were selected, with the removal of those ones expressed in vegetative tissues. By eliminating the genes that were not annotated in Release 7 of the MSU Rice Genome Annotation Project, the remaining ones were considered as endosperm-specific genes. The PFam (http://www.sanger.ac.uk/Software/Pfam/) and SMART database (http://smart.embl-heidelberg.de/smart/batch.pl) were finally used to search the conserved domains of predicted *OsEnS* proteins. Detail information in *OsEnS* genes was procured from MSU (Ouyang et al., 2007), KOME (Satoh et al., 2007) and Pfam (Finn et al., 2010), including accession number, Pfam domain, chromosomal location, ORF length, introns number, BAC/PAC accessions, and protein parameters.

### 2.2. Chromosomal localization and GO annotation of OsEnS genes

*OsEnS* genes were mapped on rice chromosomes using the software MapInspect (http://www.dpw.wau.nl/pv/pub/MapComp/) according to their positions given in the MSU rice database. The positions of *OsEnS* genes on the rice chromosomes were modified manually. For functional categories of *OsEnS* genes, we got the GO annotation of *OsEnS* genes from GO Slim of MSU. In addition, the BLAST2GO was used to annotate the *OsEnS* genes without GO annotation with the cutoff of E value $< 10^{-3}$ against NCBI non-redundant (nr) protein database.

### 2.3. Promoter analysis of OsEnS genes

To elucidate the possible *cis*-regulator elements that responsible for gene expression in developing endosperm, 2000-bp of genomic DNA sequences upstream from initiation codon (ATG) of the *OsEnS* genes was analyzed. The PLACE database (http://www.dna.affrc.go.jp/htdocs/PLACE) was adopted to identify the *cis*-regulator elements in the promoter regions.

### 2.4. Expression analysis of OsEnS genes using microarray data

Expression profile of *OsEnS* genes was extracted from the Affymetrix microarray data in CREP database (http://crep.ncpgr.cn), which composed of 39 tissues that covered the whole-life cycle of rice. The developmental stages and tissues used in this work were listed in Table S2. Expression values of each gene were logarithmized and cluster analyses were performed using J-Express 2011 with Euclidean distances and hierarchical cluster method of "complete linkage". The average signal value of two biological replicates for each sample, except for tissues 2, 3, 15, 16 and 17 (six biological replicates) was used for analysis. When more than one probe set was available for one gene, the higher signal value of the probe sets was used for analysis. For tissues 11 and 12, we used the normal signal value instead of average signal value for analysis if there is only one signal value.

### 2.5. Real-time quantitative PCR analysis

Total RNA from roots, stems, leaves, sheaths, and endosperms at 5, 7, 10, 14, and 21 DAP without embryo from Zhonghua 11 (*Oryza sativa* L. ssp. *japonica*) was extracted using TRIzol reagent (TransGen) or RNAiso with RNAiso-mate (TaKaRa) according to the manufacturer's instructions. First-stand cDNA was reversed transcribed from RNase-free DNase I-treated (TaKaRa) RNA with an oligo $(dT)_{15}$ primer, using M-MLV reverse transcriptase (Promega). To avoid DNA contaminate, PCR was performed with 5′-CGACCCGTTCATCACCACCGAC-3′ and 5′-AGCTAGCAGCCCTTCCACCTCTCCA-3′ as the forward and reverse primers under the following conditions: 94 °C for 2 min, followed by 40 cycles of 94 °C for 45 s, 60 °C for 45 s and 72 °C for 90 s, 72 °C for 10 min and 4 °C for 10 min.