Journal of Plant Physiology 271 (2022) 153663

Contents lists available at ScienceDirect

Journal of Plant Physiology

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

Transcriptomic identification of long noncoding RNAs and their hormone-associated nearby coding genes involved in the differential development of caryopses localized on different branches in rice

Hanmeng Wang [a], Zhilin Chu [a], Shu Chang [a], Shenghua Jia [a], Lu Pang [a], Chao Xi [a], Jin Liu [a], Heping Zhao [a], Yingdian Wang [a,b,**], Shengcheng Han [a,b,*]

[a] Beijing Key Laboratory of Gene Resources and Molecular Development, College of Life Sciences, Beijing Normal University, Beijing, 100875, China
[b] Academy of Plateau Science and Sustainability of the People's Government of Qinghai Province & Beijing Normal University, Qinghai Normal University, Xining, 810008, Qinghai, China

ARTICLE INFO

Keywords:
Caryopses located on apical primary branches
Caryopses located on proximal secondary branches
Long noncoding RNA
Caryopsis differential development
Transcriptome-wide analysis

ABSTRACT

Long noncoding RNAs (lncRNAs) play important regulatory roles in caryopsis development and grain size in rice. However, whether there exist differences in lncRNA expression between caryopses located on primary branches (CPB) and caryopses located on secondary branches (CSB) that contribute to their differential development remains elusive. Here, we performed transcriptome-wide analysis to identify 2,273 lncRNAs expressed in CPB and CSB at 0, 5, 12, and 20 days after flowering (DAF). Although these lncRNAs were widely distributed, the majority were located in intergenic regions of the 12 rice chromosomes. Based on gene expression cluster analysis, lncRNAs expressed in CPB and CSB were clustered into two subtypes in a position-independent manner: one includes 0- and 5-DAF CPB and CSB, and 12-DAF CSB; the second includes 12-DAF CPB and 20-DAF CPB and CSB. Furthermore, according to the expression value of each lncRNA, *K*-means cluster analysis revealed 135 early-stage, 116 middle-stage, and 114 late-stage expression-delayed lncRNAs in CSB. Then, we analyzed the expression values of the expression-delayed lncRNAs and nearby coding genes (100 kb upstream and downstream of the lncRNAs), and found 631 lncRNA–mRNA pairs, including 258 lncRNAs and 571 nearby coding genes, some of which are related to hormone-regulated grain development. These results suggested that expression-delayed lncRNAs in CSB may regulate the development of CPB and CSB, providing insight into the mechanism underlying the developmental differences between CPB and CSB, and the differences in grain yield.

1. Introduction

Long noncoding RNAs (lncRNAs) are defined as RNAs with lengths greater than 200 base pairs (bp) that appear to have no coding potential. They were considered inessential transcriptional "noise" until the discovery of noncoding RNA possessing definite biological functions (Ponjavic et al., 2007). Recent studies have shown that lncRNAs play important regulatory roles in numerous biological processes (Wilusz et al., 2009). Based on their roles, lncRNAs can be grouped into four classes: signals, decoys, guides, and scaffolds (Ponting et al., 2009). Swiezewski et al. (2009) showed that COOLAIR, an antisense transcription product located in the 3′ terminal heterochromatin region of the *FLC* gene, acts as a "signal" for the regulation of histone methylation of FLC, thus inhibiting *FLC* expression and promoting flowering in *Arabidopsis*. Furthermore, lncRNA HOTAIR acts as a scaffold allowing or complex assembly of two or more proteins: the 5′ functional domain of HOTAIR binds PRC2 and methylates H3K27 to enhance gene silencing, whereas the 3′ functional domain of HOTAIR binds the LSD1–CoREST

Abbreviations: CPB, caryopses located on apical primary branches; CSB, caryopses located on proximal secondary branches; DAF, days after fertilization; lncRNA, long noncoding RNA.

* Corresponding author. Beijing Key Laboratory of Gene Resource and Molecular Development College of Life Sciences Beijing Normal University, Beijing, 100875, China.
** Corresponding author. Beijing Key Laboratory of Gene Resource and Molecular Development College of Life Sciences Beijing Normal University, Beijing, 100875, China.
E-mail addresses: ydwang@bnu.edu.cn (Y. Wang), schan@bnu.edu.cn (S. Han).

https://doi.org/10.1016/j.jplph.2022.153663
Received 14 January 2022; Received in revised form 27 February 2022; Accepted 27 February 2022
Available online 1 March 2022

complex to mediate histone 3 lysine 4 (H3K4) demethylation and promote target gene expression (Tsai et al., 2010). Martin et al. (Martín et al., 2000) showed that the *Arabidopsis* non-protein coding gene *IPS1* (*INDUCED BY PHOSPHATE STARVATION 1*, *AtIPS1*) is induced in leaves and roots by phosphate starvation. Twenty-three nucleotides in the conserved region of *AtIPS1* were shown to be complementary to part of the sequence of miR399, functioning as a "decoy" to dismiss its inhibition of *PHO2* transcription (Franco-Zorrilla et al., 2007). Campalans et al. (2004) showed that lncRNA *Enod40* functions as a "guide" to directly bind *MtPBP1* and regulate its translocation from the nucleus to the cytoplasm in root nodules of *Medicago truncatula*. In addition, according to their locations relative to protein-coding genes in the genome, lncRNAs can be further grouped into sense, antisense, intergenic, intronic, and bidirectional lncRNAs (Rinn and Chang, 2012). There are two types of lncRNA-regulated gene expression at the transcriptional level: *cis* and *trans*. Simply, *cis*-regulation refers to the transcriptional activation of noncoding RNA on adjacent mRNAs, while *trans*-regulation refers to downregulation of the transcription of remote mRNAs (Kopp and Mendell, 2018). *Arabidopsis MAS*, a NAT-lncRNA produced from the *AtMAF4* locus and induced by cold stress, interacts with and recruits *WDR5a* to *MAF4* to enhance H3K4me3, which activates the transcription of *AtMAF4* in a *cis*-regulated manner (Zhao et al., 2018). *Drosophila HOXAIR* was reported to regulate the expression of HOX in *cis* (Petruk et al., 2006); however, interaction of the human homolog of *HOXAIR* with the PRC2 complex is essential for H3K27me3 expression at *PRC2* and *HOXD* sites in *trans* (Rinn et al., 2007).

Rice (*Oryza sativa*) is a monocotyledon model plant and important cereal crop for food security. In recent years, many lncRNAs have been identified by high-throughput sequencing in different tissues and developmental stages of rice. A total of 2,224 lncRNAs were identified by transcriptome sequencing of 14-day seedlings, anthers, and pistils of rice, and by sequencing the caryopses of rice 5 days after flowering (DAF) (Zhang et al., 2014). lncRNA *XLOC_057324* was highly expressed in young panicles and pistils, and the T-DNA insertion mutant of *XLOC_057324* exhibited earlier blossoming and lower fertility than wild-type rice (Zhang et al., 2014). In addition, 540 lncRNAs were identified by RNA sequencing (RNA-Seq) from pistils and caryopses at 3 and 7 DAF, in which *TCONS_00023703*-RNAi plants showed reduced grain length and grain weight (Zhao et al., 2020). These results suggested that lncRNAs play important roles in caryopsis development, which can be described as follows in rice. Caryopses located on primary branches (CPB), or superior caryopses, flower earlier, fill faster, and produce larger and heavier grains, while caryopses located on secondary branches (CSB), usually referred to as inferior caryopses, flower late and are either sterile or fill slowly to produce small grains, which greatly limits production potential (Peng et al., 2011, 2014; Wang et al., 2020). Dynamic differences in gene expression patterns between CPB and CSB, during development and after fertilization, were examined by RNA-Seq (Jia et al., 2021; Sun et al., 2015) and PCR-based suppression subtractive hybridization (SSH) (Sekhar et al., 2015). Short noncoding RNAs and microRNAs (miRNAs) were also characterized in superior and inferior spikelets, which are involved in maintaining hormone homeostasis, carbohydrate metabolic pathways, and cell division during grain development (Peng et al., 2011, 2013, 2014). However, research on lncRNA regulation of the developmental difference between CPB and CSB is limited.

In the present study, we performed high-throughput RNA-Seq to analyze the expression dynamics of lncRNAs during whole rice caryopsis development, characterized the expression profiles of lncRNAs in CPB and CSB at four developmental stages, and predicted the function of expression-delayed lncRNAs in CSB based on expression patterns and their adjacent coding genes. Our work highlights the differences in expression profiles of lncRNAs between CPB and CSB, and indicates potential lncRNA targets for understanding the genetic regulation of rice caryopsis development.

2. Materials and methods

2.1. Plant materials and growth conditions

The rice (*Oryza sativa* L. subspecies *japonica*) variety Zhonghua 15 was planted in the greenhouse and paddy fields of Beijing Normal University, Beijing, China from April to October on an annual basis. In the middle of April, the seeds were soaked overnight with 1% carbendazim, spread into the seedling tray, covered with thin soil, and watered twice per day. After 3–4 weeks, 3.5–4.0 leaf-age seedlings were inserted into paddy fields with a planting density of 22.1 holes per square meter (three plants per hole). Conventional paddy field management was performed until rice grain maturity. In July, open caryopses of the glume sheath on primary branches (CPB) or secondary branches (CSB) were marked at 10–11 a.m. every day and labeled as flowering on day 0. Then, CPB and CSB were collected at 0, 5, 12, and 20 days after flowering (DAF), rapidly frozen in liquid nitrogen, and stored at −80 °C for further experiments.

2.2. RNA extraction and quantitative real-time PCR (qRT-PCR)

Total RNA was extracted from CPB and CSB at different developmental stages using TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions. Approximately 2 μg RNA was reverse-transcribed using a Reverse-Aid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, USA) and used as template for qRT-PCR. PCR reactions were performed using a 7500 Fast Real-Time PCR System (Applied Biosystems, USA) with Power SYBR Green PCR Master Mix (Applied Biosystems). PCR was performed as follows: 2 min at 50 °C and 10 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 60 s at 60 °C. Three independent biological replicates were performed for qRT-PCR analyses. *OsActin1* was used as an internal control to standardize the results. Primers were designed using Primer5 software and are listed in Supplementary Table S9.

2.3. Data mining and identification of lncRNAs

The RNA-Seq data of rice CPB and CSB at 0, 5, 12, and 20 DAF were deposited in the NCBI Sequence Read Archive (SRA) with the following accession numbers: SRR4111270–4111278 (Jia et al., 2021). Using the TopHat program, the clean reads were searched against the rice genome (MSU Rice Genome Annotation Project Release 7, http://rice.uga.edu/) to obtain the transcripts expressed in CPB and CSB at 0, 5, 12, and 20 DAF. Then, the transcripts were assembled using StringTie (Version 1.0.4, http://ccb.jhu.edu/software/stringtie/) and generated with the Cuffmerge program. High-quality assemblies with lengths ≥200 bp and fragments per kilobase of transcript per million mapped reads (FPKM) values > 0.5 (calculated using Cuffdiff v2.1.1) were retrieved. Three programs, Coding Potential Calculator (CPC; http://www.mybiosoftware.com/cpc-0-9r2-assess-protein-coding-potential-transcripts.html), txCdspredict (http://hgdownload.soe.ucsc.edu/admin/jksrc.zip), and Coding-Non-Coding Index (CNCI; https://github.com/www-bioinfo-org/CNCI), were used to predict the coding potential of the transcripts. Score thresholds were set to distinguish lncRNA from mRNA (CPC threshold: ≥ 0 = mRNA, < 0 = lncRNA; txCdspredict threshold: ≥ 500 = mRNA, < 500 = lncRNA; CNCI threshold: ≥ 0 = mRNA, < 0 = lncRNA) (Kong et al., 2007). This analysis was combined with information from the Pfam protein database (http://pfam.xfam.org/) to ensure that the predicted lncRNA transcripts did not contain protein-coding domains. Transcripts reported as lncRNAs by at least three of the four above prediction methods were identified as lncRNAs.

2.4. Characterization of lncRNAs expressed in caryopses

The datasets of lncRNAs expressed in rice anthers, pistils, 5-DAP (days after pollination) seeds, and 14-DAG (days after germination)

shoots were downloaded from NCBI SRA (Zhang et al., 2014). Unique and shared lncRNAs expressed in CPB and CSB were classified by Venn diagrams (http://bioinformatics.psb.ugent.be/software/details/Venn-Diagrams).

2.5. Expression analysis of lncRNAs in CPB and CSB

The correlation of lncRNA expression between each pair of samples ($q < 0.05$) was calculated using IBM SPSS Statistics 22.0 (IBM Corp., Armonk, NY, USA). Hierarchical clustering was performed with Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/), and a heatmap was generated by Java TreeView-1.1.6r4-win (https://sourceforge.net/projects/jtreeview/files/jtreeview/). Log transformation and mean centering were applied, and the Euclidean distance was calculated. *K*-means clustering performed with MEV software (http://www.tm4.org/mev), was used to cluster lncRNAs with similar expression patterns. The results were displayed using Excel.

2.6. Interaction networks of lncRNAs and adjacent coding genes

lncRNAs with nearby coding genes within 100 kb upstream and downstream formed lncRNA–mRNA pairs. The co-expression relationship of lncRNAs and coding genes was assessed based on FPKM values and visualized using Cytoscape 3.8.1 (http://www.cytoscape.org) in the Java 11 environment. The maximum FPKM value of eight samples from CPB and CSB at 0, 5, 12, and 20 DAF was set as 1.0, and the relative expression levels of seven other samples were calculated as FPKM/maximum FPKM.

3. Results

3.1. Identification and chromosome distribution analysis of lncRNAs in caryopses

To explore the role of lncRNAs in the developmental differences between CPB and CSB, we first identified the lncRNAs expressed in CPB and CSB at four representative developmental stages (0, 5, 12, and 20 DAF). Using previously reported RNA-Seq data deposited in NCBI SRA (accession numbers SRR4111270–4111278) (Jia et al., 2021), we mapped the clean reads to the rice genome (MSU Rice Genome Annotation Project Release 7) using the TopHat program and generated final transcript numbers separately for eight samples by cufflinks assembly: 30,598, 30,353, 30,141, 29,362, 31,847, 28,614, 31,416, and 31,766 (Supplementary Table S1). The transcripts with FPKM $\geq$0.5 and length $\geq$200 nt were selected for determination of coding potential using the CPC, txCdsPredict, and CNCI programs, and BLAST search of the Pfam database was performed to identify lncRNAs. Finally, 2,273 lncRNAs were identified in rice caryopses.

Next, we characterized the distribution of lncRNAs in the rice genome. First, the length distribution of lncRNAs was analyzed. We found that 200–400-nt lncRNAs were the most abundant (n = 1,159), accounting for more than half of all lncRNAs; 552 lncRNAs between 401 and 1,000 nt and 331 lncRNAs >2,000 nt in length were identified (Fig. 1A and Supplementary Table S2). Although the number of lncRNAs on each chromosome differed, these lncRNAs were almost evenly distributed across the 12 chromosomes in rice, with no obvious location preferences (Fig. 1B). In addition, the distribution of lncRNAs on the Crick and Watson strands of each chromosome was similar (Fig. 1C and Supplementary Table S3). Compared with the base numbers of each chromosome, the number of lncRNAs was highest on chromosome 3, and lowest on chromosome 11 (Supplementary Table S3). Furthermore, we compared the relative position of lncRNAs with nearby coding genes in the genome, which were classified into the following types: 32 overlap, 319 overlap-anti, 162 intragenic, 200 intragenic-anti, and 1,560 intergenic lncRNAs (Fig. 1D). These results indicated that the majority of lncRNAs were located in intergenic regions. The GC content of these lncRNAs was 43.61% on average, while the maximum value was 49.56% (for overlap lncRNAs) and the minimum value was 38.41% (for intragenic lncRNAs) (Fig. 1D).

3.2. Expression of lncRNAs in different caryopses is development-dependent, but not position-dependent

Among the 2,273 lncRNAs, 566 were specifically expressed in CPB, and 625 in CSB; 1,082 were co-expressed in CPB and CSB (Supplementary Fig. S1A). In addition, compared with 2,224 lncRNAs

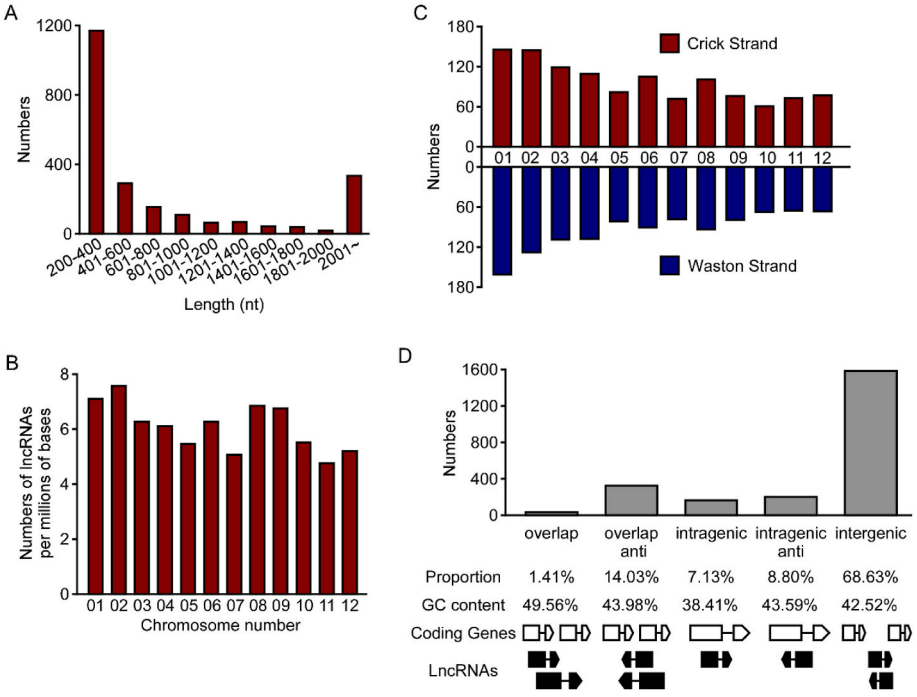

Fig. 1. Characteristics of rice lncRNAs expressed in caryopses.
A. Length distribution of lncRNAs expressed in rice caryopses. B. Number of lncRNAs per million bases in 12 rice chromosomes. C. Distribution of lncRNAs from caryopses in Crick and Watson strands of 12 rice chromosomes. D. Classification of lncRNAs expressed in caryopses according to the genomic position relative to protein-coding genes. The numbers of lncRNAs for each of the five main classes are shown on the columns (overlap, overlap-anti, intragenic, intragenic-anti, and intergenic lncRNAs). The proportions of the five types of lncRNAs were calculated. A scheme showing the position of the lncRNA (black box) relative to neighboring genes (black empty box) is provided at the bottom.

previously characterized in rice (Zhang et al., 2014), 915 lncRNAs were common and 1,358 were unique in this study (Supplementary Fig. S1B).

To further explore the relationship between the developmental stages of CPB and CSB and the expression patterns of lncRNAs in CPB and CSB at 0, 5, 12, and 20 DAF, we carried out cluster analysis based on the FPKM values of lncRNAs at these four developmental stages using Cluster 3.0. The expression patterns of lncRNAs clustered into two subtypes: the first included 0-DAF CPB and CSB, 5-DAF CPB and CSB, and 12-DAF CSB; the second included 12- and 20-DAF CPB, and 20-DAF CSB (Fig. 2A). We further analyzed the unique and shared lncRNAs in these two subtypes using a Venn diagram, and found 432 co-expressed lncRNAs in subtype I (Figs. 2B) and 376 lncRNAs in subtype II (Fig. 2C). In addition, many unique genes were found (Fig. 2B and C). These results revealed that the expression patterns of IncRNAs in CPB and CSB are consistent with their developmental stages, but independent of their position.

3.3. Classification and qRT-PCR validation of expression-delayed lncRNAs in CPB and CSB

To further characterize the differential expression of lncRNAs in CPB and CSB at four developmental stages, we performed *K*-means cluster analysis of 1,082 lncRNAs co-expressed in CPB and CSB (Supplementary Fig. 1A) using MEV software; these lncRNAs were classified into 18 categories according to lncRNA expression in CPB at four developmental stages (Supplementary Table S4). Then, we compared the expression patterns of these 18 types of lncRNAs in CSB at the same developmental stages (Supplementary Table S4). K01–06 were dominant in terms of the expression of lncRNAs (Fig. 3A); the 12 other categories are shown in Supplementary Fig. S2. As shown in Fig. 3A, the K01 classifier included the lncRNAs with the highest expression levels at 0-DAF in both CPB and CSB, while K06 included lncRNAs with the highest expression levels at 20-DAF in both CPB and CSB. More interestingly, the K02 classifier contained lncRNAs expressed at higher levels in 0-DAF and 5-DAF CPB

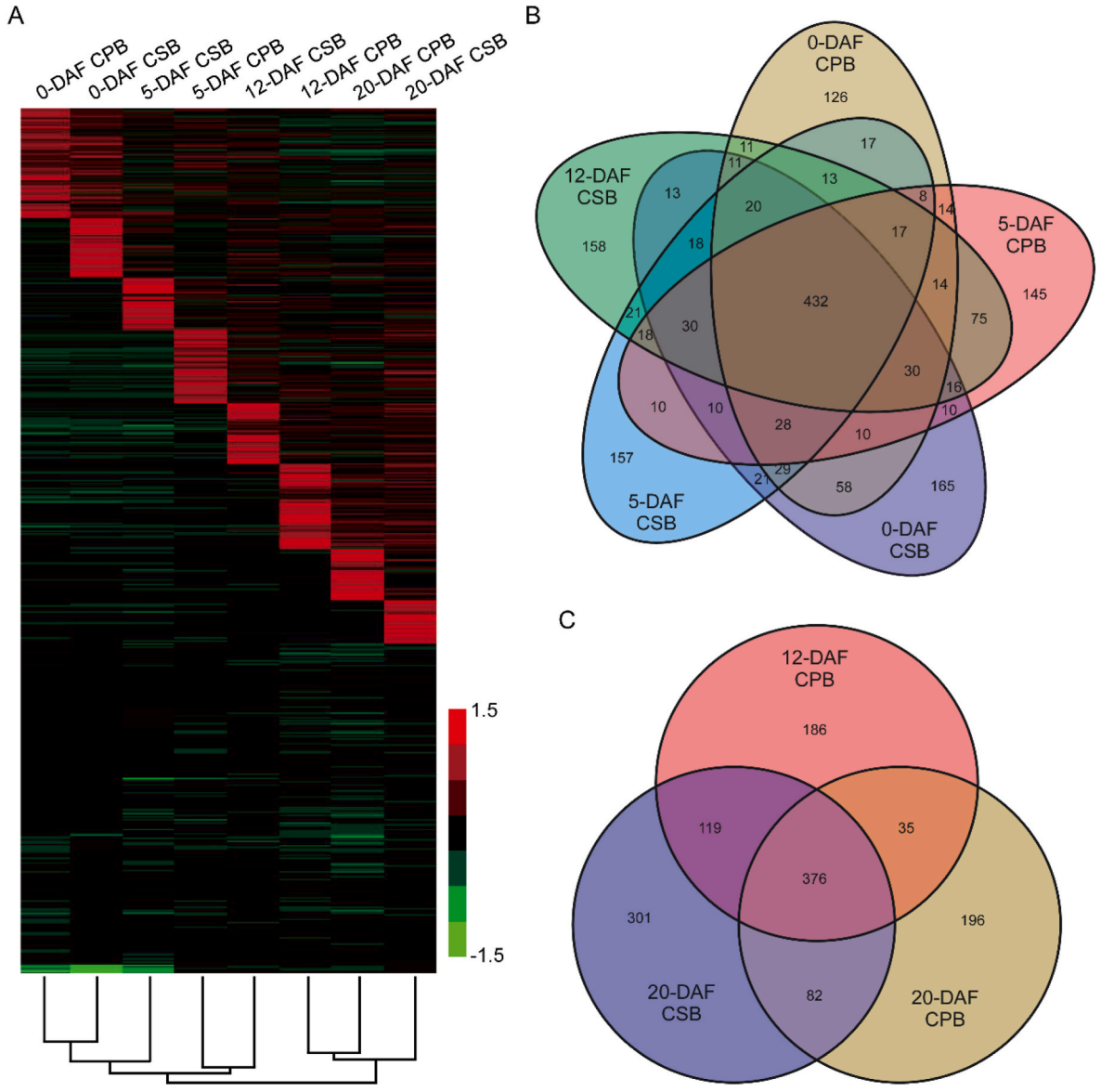

Fig. 2. Cluster analysis of lncRNAs expressed in CPB and CSB at 0, 5, 12, and 20 days after fertilization (DAF).
A. Hierarchical cluster analysis of lncRNAs expressed in CPB and CSB at 0, 5, 12, and 20 DAF was performed with Cluster 3.0 software. Pseudo-color was used to label expression patterns: red, high expression; green, low expression. B and C. A Venn diagram was used to show unique and shared genes expressed in subtype I, including 0- and 5-DAF CPB, and 0-, 5-, and 12-DAF CSB (B), and in subtype II including 12- and 20-DAF CPB, and 20-DAF CSB (C). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

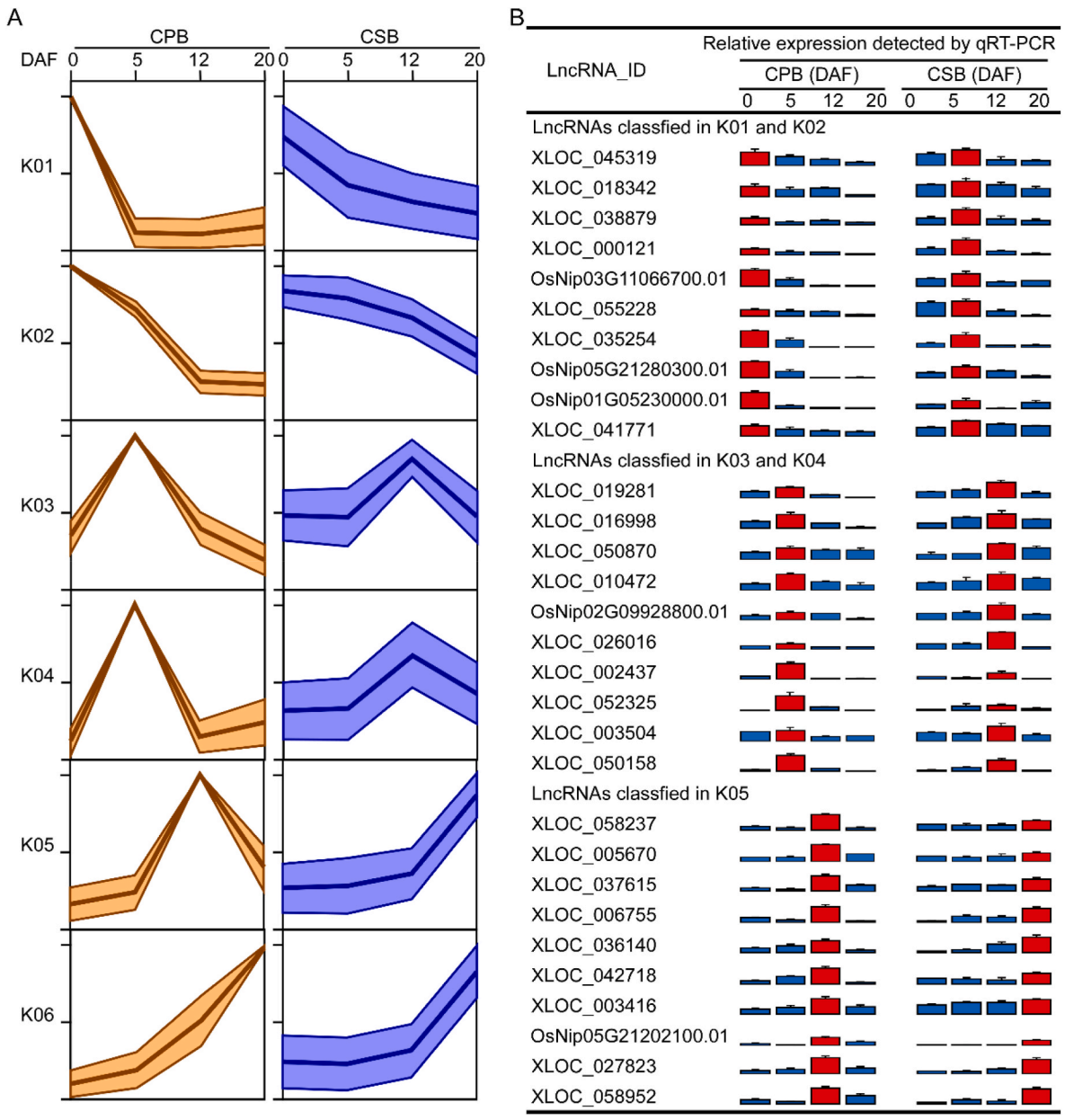

Fig. 3. Expression analyses and verification of lncRNAs in CPB and CSB at four different DAF.
A. *K*-means clustering was used to classify lncRNA expression patterns in CPB at 0, 5, 12, and 20 DAF using the MEV program, according to the fragments per kilobase of transcript per million mapped reads (FPKM) value of each lncRNA. Expression patterns of the same genes in CSB are also presented. Six major clusters (K01–K06) are shown, in orange for CPB (left) and blue for CSB (right). The average expression level of genes in each cluster is indicated by a dark line, and the ranges of standard deviation are represented by shadows. B. Verification of the expression of lncRNAs in CPB and CSB at 0, 5, 12, and 20 DAF by qRT-PCR. Rice *Actin 1* was used as an internal control. Three biological replicates were examined and had similar results; the data are presented as means ± standard deviation. The lowest expression value of each lncRNA among 0-, 5-, 12-, and 20-DAF CPB and CSB was set as 1.0. Red represents the highest expression among four developmental stages in CPB and CSB. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and CSB, the K03 and K04 classifiers included lncRNAs expressed at higher levels in 5-DAF CPB and 12-DAF CSB, and the K05 classifier included lncRNAs expressed at higher levels in 12-DAF CPB and 20-DAF CSB. These results indicated that, compared with the same developmental stage of CPB, the expression of some lncRNAs in the K02–05 classifiers was delayed in CSB; this is consistent with the delayed development phenotype of CSB. Therefore, we analyzed the expression patterns of lncRNAs at different DAF of CPB and CSB and obtained three groups of lncRNAs. The first group included 135 lncRNAs, which were expressed at higher levels in 0-DAF CPB and 5-DAF CSB compared with the other stages of CPB and CSB (Supplementary Table S5); we refer to these as early-stage expression-delayed lncRNAs. The second group included 116 lncRNAs, which were expressed at higher levels in 5-DAF CPB and 12-DAF CSB (Supplementary Table S6), and are termed middle-stage expression-delayed lncRNAs. The third group included 114 lncRNAs, which were expressed at higher levels in 12-DAF CPB and 20-DAF CSB (Supplementary Table S7), and are termed late-stage expression-delayed lncRNAs. We randomly selected 10 lncRNAs from each of these three groups and performed qRT-PCR to verify their expression levels in CPB and CSB at 0, 5, 12, and 20 DAF. Compared with CPB at the same stage, the expression of these 30 lncRNAs was consistently delayed in CSB (Fig. 3B). Therefore, we inferred that these expression-delayed lncRNAs in CSB could be important in the developmental differences between CPB and CSB.

3.4. Regulatory network of expression-delayed lncRNAs and hormone-associated nearby coding genes in CPB and CSB

To explore the regulatory roles of these 365 expression-delayed lncRNAs in CSB, we analyzed the expression patterns of nearby coding genes localized within 100 kb upstream and downstream of the lncRNAs, according to the transcriptomic data of CPB and CSB at 0, 5, 12, and 20 DAF (Jia et al., 2021). We identified nearby coding genes with the same expression patterns of lncRNAs as potential target genes, and found 631 lncRNA–mRNA pairs, composed of 258 lncRNAs and 571 nearby coding genes. These lncRNA–mRNA pairs separately included 198 early-stage expression-delayed lncRNA-mRNA pairs composed of 98 lncRNAs and 180 nearby coding genes, 281 middle-stage expression--delayed lncRNA-mRNA pairs composed of 92 lncRNAs and 255 nearby coding genes, and 152 late-stage expression-delayed lncRNA-mRNA pairs composed of 70 lncRNAs and 136 nearby coding genes (Supplementary Table S8).

Firstly, four early-stage expression-delayed lncRNAs formed lncRNA–mRNA pairs with different hormone-associated nearby coding genes, respectively: *OsNip03G11193300.01* co-expressed with *LOC_Os03G03070* and *LOC_Os03G03080*; *XLOC_03046* with *LOC_Os03G63074*, *LOC_Os03G63260*, *LOC_Os03G63290*, and *LOC_Os03G63320*; *OsNip09G38820200.01* with *LOC_Os09G37344*, *LOC_Os09G37410*, and *LOC_Os09G37420*; and *XLOC_014416* with

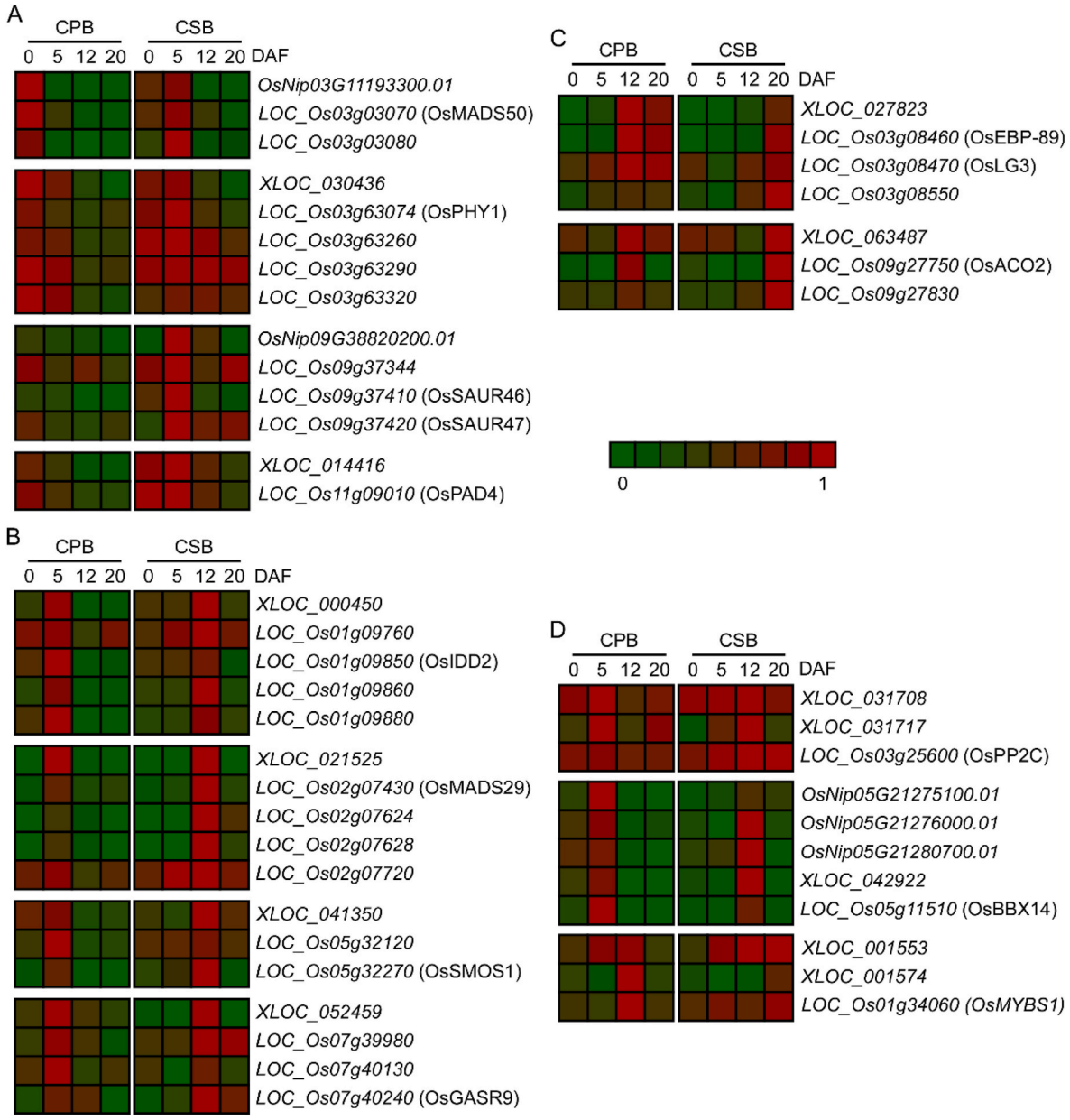

Fig. 4. Expression pattern of the expression-delayed lncRNAs and their hormone-associated nearby coding genes in CPB and CSB at 0, 5, 12, and 20 DAF.
A. Expression of the early-stage expression-delayed lncRNAs and their hormone-associated nearby coding genes in CPB and CSB at 0, 5, 12, and 20 DAF. B. Expression of the middle-stage expression-delayed lncRNAs and their hormone-associated nearby coding genes in CPB and CSB at 0, 5, 12, and 20 DAF. C. Expression of the late-stage expression-delayed lncRNAs and their hormone-associated nearby coding genes in CPB and CSB at 0, 5, 12, and 20 DAF. D. Expression of the multiple expression-delayed lncRNAs with one hormone-associated nearby coding gene pairs in CPB and CSB at 0, 5, 12, and 20 DAF. The FPKM values for different lncRNAs and their nearby coding genes were obtained by RNA-Seq, and the highest expression value of each gene among 0-, 5-, 12-, and 20-DAF CPB and CSB was set as 1. Red, high expression; green, low expression. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

LOC_Os11g09010 (Supplementary Fig. S3A and Fig. 4A). Previous studies showed that *LOC_Os03G03070* encodes OsMADS50, which is directly activated by OsmiR156–*OsSPL3/OsSPL12* module and involved in auxin-regulated crown root development in rice (Shao et al., 2019); *LOC_Os03G63074* encodes OsPHY1, which is associated with phosphate homeostasis in seeds and strongly induced by phytohormones abscisic acid (ABA), gibberellin (GA_3), and indole-3-acetic acid (IAA), as well as osmotic stresses of salt, drought, and cold (Guo et al., 2013); *LOC_Os09G37410* and *LOC_Os09G37420* encode OsSAUR46 and OsSAUR47, respectively, which belong to auxin-responsive SAUR gene family members (Jain et al., 2006); and *LOC_Os11g09010* encodes OsPAD4, which is involved in JA-dependent defense signaling (Ke et al., 2014).

Secondly, four middle-stage expression-delayed lncRNAs constructed lncRNA–mRNA pairs with multiple hormone-associated nearby coding genes (Supplementary Fig. S3B and Fig. 4B). *XLOC_000450* was co-expressed with *LOC_Os01g09760*, *LOC_Os01g09850*, *LOC_Os01g09860*, and *LOC_Os01g09880*. *LOC_Os01g09850* encodes a zinc finger and INDETERMINATE DOMAIN transcription factor (OsIDD2), which is involved in GA-mediated cell proliferation in rice (Lu et al., 2020). *XLOC_021525* formed lncRNA–mRNA pairs with four coding genes: *LOC_Os02g07430*, *LOC_Os02g07624*, *LOC_Os02g07628*, and *LOC_Os02g07720*. Chang et al. (2020) previously reported that *LOC_Os02g07430* encodes the transcription factor OsMADS29, which is involved in auxin-regulated development difference between CPB and CSB. *XLOC_041350* formed lncRNA–mRNA pairs with two coding genes, *LOC_Os05g32120* and *LOC_Os05g32270*, which encodes an unusual APETALA2 (AP2)-type transcription factor, SMOS1/SHB/RLA1/NGR5, involved in the crosstalk among auxin, GA, and brassinosteroid signaling pathway for cell elongation and proliferation (Aya et al., 2014; Li et al., 2015; Qiao et al., 2017; Wu et al., 2020). *XLOC_052459* formed lncRNA–mRNA pairs with three coding genes, *LOC_Os07g39980*, *LOC_Os07g40130*, and *LOC_Os07g40240*. Li et al. (2019) showed that *LOC_Os07g40240* encodes OsGASR9, which is involved in GA-regulated grain size and yield in rice.

Thirdly, *XLOC_027823* and *XLOC_063487* are two late-stage expression-delayed lncRNAs that form lncRNA–mRNA pairs with *LOC_Os03g08460*, *LOC_Os03g08470*, and *LOC_Os03g08550*, and *LOC_Os09g27750* and *LOC_Os09g27830*, respectively (Supplementary Fig. S3C and Fig. 4C). Previous studies have shown that *LOC_Os09g27750* encodes 1-aminocyclopropane-1-carboxylate oxidase 2 (ACO2) for ethylene synthesis (Iwai et al., 2006), while *LOC_Os03g08460* and *LOC_Os03g08470* encodes OsEBP-89 and OsLG3, two putative ethylene response factors that participate in ethylene-dependent grain development and maturation in rice (Yang et al., 2002; Yu et al., 2017).

In addition, we found multiple lncRNAs corresponding to one nearby coding gene that is also hormone-associated. *XLOC_031708* and *XLOC_031717*, two middle-stage expression-delayed lncRNAs form lncRNA–mRNA pairs with *LOC_Os03g25600* (Supplementary Fig. S3D and Fig. 4D). *LOC_Os03g25600* encodes protein phosphatase 2C, which may be involved in the crosstalk between abscisic acid and brassinosteroid hormones (Wang et al., 2017). Four middle-stage expression-delayed lncRNAs, *XLOC_042922*, *OsNip05G21275100.01*, *OsNip05G21276000.01*, and *OsNip05G21280700.01*, had one nearby coding gene, *LOC_Os05g11510* (Supplementary Fig. S3D and Fig. 4D), which encodes the b-box protein OsBBX14 that in turn regulates anthocyanin biosynthesis in rice (Kim et al., 2018). Furthermore, *LOC_Os01g34060* is a nearby coding gene of two late-stage expression-delayed lncRNAs, *XLOC_001553* and *XLOC_001574* (Supplementary Fig. S3B and Fig. 4B), and encodes an MYB transcription factor that integrates diverse nutrient starvation and gibberellin signaling pathways during rice grain germination (Hong et al., 2012). These results suggest that the different hormone-associated expression-delayed lncRNA–mRNA pairs play important roles on regulating the differential development of CPB and CSB in rice.

4. Discussion

In rice, the development of caryopses can be divided into three stages according to developmental phase: the early stage of caryopsis fertilization, in which the initiation of cell proliferation of the embryo and endosperm occurs at 0–3 DAF; the middle stage, in which cell proliferation and differentiation of the embryo and endosperm, and the beginning of grain filling, occur at 3–10 DAF; and the late stage, in which material accumulation in the endosperm and seed maturity occur at 10–40 DAF (Brown et al., 1996; Ishimaru et al., 2003; Lopes and Larkins, 1993). Studies have shown that the proliferation rate and number of endosperm cells, as well as the material accumulation rate, of CSB were significantly lower than those of CPB, thus limiting rice yield (Ishimaru et al., 2003; Wang et al., 2020; Yang and Zhang, 2010). Therefore, understanding the mechanism underlying the developmental differences between CPB and CSB will be important. While lncRNAs have been reported to regulate caryopsis development and grain formation in rice (Zhang et al., 2014; Zhao et al., 2020), the roles of lncRNAs in the differential development of CPB and CSB remain unclear. Here, a total of 2,273 lncRNAs were identified in CPB and CSB by high-throughput RNA-Seq. We found 135 early-stage, 116 middle-stage, and 114 late-stage expression-delayed lncRNAs in CSB, as well as potential target genes that may be involved in the developmental differences between CPB and CSB.

In the present study, we identified lncRNAs expressed in CPB and CSB at 0, 5, 12, and 20 DAF. Cluster analysis based on the FPKM values of lncRNAs at these developmental stages clustered the lncRNAs into two subtypes according to their expression patterns: the first cluster included 0-DAF CPB and CSB, 5-DAF CPB and CSB, and 12-DAF CSB, and the second included 12- and 20-DAF CPB, and 20-DAF CSB, which was consistent with the expression pattern of coding genes at these stages (Jia et al., 2021). These results suggested that the different developmental stage of caryopsis after pollination, but not the position information between CPB and CSB, is the key regulating factor for the expression of lncRNAs and coding genes in rice, which is the foundation for exploring the essential lncRNA and coding genes for regulating the developmental difference between CPB and CSB. Moreover, *K*-means cluster analysis revealed expression-delayed lncRNAs in CSB at three different stages; these lncRNAs could play important regulatory roles in the developmental differences between CPB and CSB, by activating or inhibiting the transcription of target genes. A previous study indicated that the distribution of lncRNAs on chromosomes, especially the relative position of coding genes, determines their regulatory effects and functions: lncRNAs activate the expression of nearby coding genes in a *cis* manner, and inhibit the transcription of remote target genes in a *trans* manner (Kopp and Mendell, 2018). Because it is difficult to predict remote target genes of lncRNAs, we identified 631 lncRNA–mRNA pairs with 258 expression-delayed lncRNAs and 571 nearby coding genes according to similar expression patterns between CPB and CSB. An in-depth analysis of the regulation of expression of lncRNA–mRNA pairs during caryopsis development will be important to better understand the developmental differences between CPB and CSB.

A previous study showed that auxin helps promote replication and division of the central cell, which is required for endosperm development in *Arabidopsis* (Figueiredo et al., 2016). Jia et al. (2021) revealed that the application of exogenous indole-3-acetic acid (IAA) to CSB before fertilization expedited the transcription of express-delayed coding genes, promoted caryopsis development and increased grain weight to levels closer to that observed in CPB. Chang et al. (2020) further showed that higher levels of auxin in CPB compared with CSB are maintained by the IAA–MADS29–OsAsp1–OsTIF1–OsTAA1–IAA regulatory cycle, which is important in the distinct development of CPB and CSB. Here, we found that one middle-stage expression-delayed lncRNA, *XLOC_021525*, formed separate lncRNA–mRNA pairs with the coding gene *LOC_Os02g07430*, which encodes OsMADS29, *LOC_Os02g07624*, *LOC_Os02g07628*, and *LOC_Os02g07720*. In addition, we showed that

early-stage expression-delayed lncRNA–mRNA pairs, *OsNip03G11193300.01* with *LOC_Os03G03070* (*OsMADS50*), *XLOC_03046* with *LOC_Os03G63074* (*OsPHY1*), and *OsNip09G38820200.01* with *LOC_Os09G37410* (*OsSAUR46*) and *LOC_Os09G37420* (*OsSAUR46*), were involved in the auxin signaling pathway in rice (Guo et al., 2013; Jain et al., 2006; Shao et al., 2019). These results indicated that auxin likely regulates the transcription of expression-delayed coding genes by activating their paired lncRNA, ultimately leading to developmental differences between CPB and CSB.

Besides auxin, we also found that the early-stage expression-delayed lncRNA–mRNA pair of *XLOC_014416* with *LOC_Os11g09010* (*OsPAD4*) is involved in the JA signaling pathway (Ke et al., 2014), the middle-stage expression-delayed lncRNA-coding gene pairs *XLOC_000450* with *LOC_Os01g09850* (*OsIDD2*) and *XLOC_052459* with *LOC_Os07g40240* (*OsGASR9*) are involved in the GA signaling pathway (Li et al., 2019; Lu et al., 2020), and the late-stage expression-delayed lncRNA–mRNA pairs *XLOC_063487* with *LOC_Os09g27750* (OsACO2) and *XLOC_027823* with *LOC_Os03g08460* (*OsEBP-89*) and *LOC_Os03g08470* (*OsLG3*) are involved in ethylene biosynthesis (Iwai et al., 2006) or ethylene signaling pathway (Yang et al., 2002; Yu et al., 2017). These results suggested that different expression-delayed lncRNAs are involved in hormone-regulated caryopsis development. However, the detailed mechanism of how lncRNAs regulate the developmental differences between CPB and CSB requires further study in the future.

5. Conclusions

Using high-throughput RNA-Seq, we characterized 2,273 lncRNAs expressed in rice CPB and CSB at 0, 5, 12, and 20 DAF; more than two-thirds were located in intergenic regions of rice chromosomes. *K*-means cluster analysis further revealed 135 early-stage, 116 middle-stage, and 114 late-stage expression-delayed lncRNAs in CSB. In addition, we analyzed the expression data of expression-delayed lncRNAs and nearby coding genes, and found 631 lncRNA–mRNA pairs, including 258 lncRNAs and 571 nearby coding genes, some of which are related to hormone-regulated caryopsis development. These results provide insight into the mechanism underlying the developmental differences between CPB and CSB, and the differences in grain yield.

CRediT authorship contribution statement

Hanmeng Wang: Conceptualization, Data curation, Investigation, Methodology, Writing – original draft. **Zhilin Chu:** Conceptualization, Data curation, Methodology, Validation. **Shu Chang:** Formal analysis, Methodology. **Shenghua Jia:** Methodology, Validation. **Lu Pang:** Investigation, Methodology. **Chao Xi:** Investigation. **Jin Liu:** Methodology. **Heping Zhao:** Resources. **Yingdian Wang:** Funding acquisition, Supervision, Writing – review & editing. **Shengcheng Han:** Conceptualization, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Key Basic Research Program of China (973 Program, 2013CB126902) and the National Natural Science Foundation of China (Grant No. 30570148 and 31370307).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jplph.2022.153663.

References

Aya, K., Hobo, T., Sato-Izawa, K., Ueguchi-Tanaka, M., Kitano, H., Matsuoka, M., 2014. A novel AP2-type transcription factor, SMALL ORGAN SIZE1, controls organ size downstream of an auxin signaling pathway. Plant Cell Physiol. 55 (5), 897–912.
Brown, R.C., Lemmon, B.E., Olsen, O.-A., 1996. Development of the endosperm in rice (*Oryza sativa* L.): cellularization. J. Plant Res. 109 (3), 301–313.
Campalans, A., Kondorosi, A., Crespi, M., 2004. Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in *Medicago truncatula*. Plant Cell 16 (4), 1047–1059.
Chang, S., Chen, Y., Jia, S., Li, Y., Liu, K., Lin, Z., Wang, H., Chu, Z., Liu, J., Xi, C., Zhao, H., Han, S., Wang, Y., 2020. Auxin apical dominance governed by the OsAsp1-OsTIF1 complex determines distinctive rice caryopses development on different branches. PLoS Genet. 16 (10), e1009157.
Figueiredo, D.D., Batista, R.A., Roszak, P.J., Hennig, L., Köhler, C., 2016. Auxin production in the endosperm drives seed coat development in *Arabidopsis*. Elife 5, e20542.
Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., García, J.A., Paz-Ares, J., 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat. Genet. 39 (8), 1033–1037.
Guo, C., Guo, L., Li, X., Ma, C., Duan, W., Gu, J., Xu, Z., Li, R., Lu, W., Xiao, K., 2013. Transcriptional regulation of the rice phytase gene OsPHY1 by several phytohormones and osmotic stresses using promoter-gus analysis. Plant Mol. Biol. Rep. 31 (6), 1461–1473.
Hong, Y.-F., Ho, T.-H.D., Wu, C.-F., Ho, S.-L., Yeh, R.-H., Lu, C.-A., Chen, P.-W., Yu, L.-C., Chao, A., Yu, S.-M., 2012. Convergent starvation signals and hormone crosstalk in regulating nutrient mobilization upon germination in cereals. Plant Cell 24 (7), 2857–2873.
Ishimaru, T., Matsuda, T., Ohsugi, R., Yamagishi, T., 2003. Morphological development of rice caryopses located at the different positions in a panicle from early to middle stage of grain filling. Funct. Plant Biol. 30 (11), 1139–1149.
Iwai, T., Miyasaka, A., Seo, S., Ohashi, Y., 2006. Contribution of ethylene biosynthesis for resistance to blast fungus infection in young rice plants. Plant Physiol. 142 (3), 1202–1215.
Jain, M., Tyagi, A.K., Khurana, J.P., 2006. Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (*Oryza sativa*). Genomics 88 (3), 360–371.
Jia, S., Chang, S., Wang, H., Chu, Z., Xi, C., Liu, J., Zhao, H., Han, S., Wang, Y., 2021. Transcriptomic analysis reveals the contribution of auxin on the differentially developed caryopses on primary and secondary branches in rice. J. Plant Physiol. 256, 153310.
Ke, Y., Liu, H., Li, X., Xiao, J., Wang, S., 2014. Rice OsPAD4 functions differently from Arabidopsis AtPAD4 in host-pathogen interactions. Plant J. 78 (4), 619–631.
Kim, D.-H., Park, S., Lee, J.-Y., Ha, S.-H., Lee, J.-G., Lim, S.-H., 2018. A rice B-Box protein, OsBBX14, finely regulates anthocyanin biosynthesis in rice. Int. J. Mol. Sci. 19 (8), 2190.
Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., Gao, G., 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 35, W345–W349.
Kopp, F., Mendell, J.T., 2018. Functional classification and experimental dissection of long noncoding RNAs. Cell 172 (3), 393–407.
Li, J., Zhao, Y., Chu, H., Wang, L., Fu, Y., Liu, P., Upadhyaya, N., Chen, C., Mou, T., Feng, Y., Kumar, P., Xu, J., 2015. SHOEBOX modulates root meristem size in rice through dose-dependent effects of gibberellins on cell elongation and proliferation. PLoS Genet. 11 (8), e1005464-e1005464.
Li, X., Shi, S., Tao, Q., Tao, Y., Miao, J., Peng, X., Li, C., Yang, Z., Zhou, Y., Liang, G., 2019. OsGASR9 positively regulates grain size and yield in rice (*Oryza sativa*). Plant Sci. 286, 17–27.
Lopes, M.A., Larkins, B.A., 1993. Endosperm origin, development, and function. Plant Cell 5 (10), 1383–1399.
Lu, Y., Feng, Z., Meng, Y., Bian, L., Xie, H., Mysore, K.S., Liang, J., 2020. SLENDER RICE1 and *Oryza sativa* INDETERMINATE DOMAIN2 regulating OsmiR396 are involved in stem elongation. Plant Physiol. 182 (4), 2213–2227.
Martín, A.C., del Pozo, J.C., Iglesias, J., Rubio, V., Solano, R., de La Peña, A., Leyva, A., Paz-Ares, J., 2000. Influence of cytokinins on the expression of phosphate starvation responsive genes in Arabidopsis. Plant J. 24 (5), 559–567.
Peng, T., Du, Y., Zhang, J., Li, J., Liu, Y., Zhao, Y., Sun, H., Zhao, Q., 2013. Genome-wide analysis of 24-nt siRNAs dynamic variations during rice superior and inferior grain filling. PLoS One 8 (4), e61029.
Peng, T., Lv, Q., Zhang, J., Li, J., Du, Y., Zhao, Q., 2011. Differential expression of the microRNAs in superior and inferior spikelets in rice (*Oryza sativa*). J. Exp. Bot. 62 (14), 4943–4954.
Peng, T., Sun, H., Qiao, M., Zhao, Y., Du, Y., Zhang, J., Li, J., Tang, G., Zhao, Q., 2014. Differentially expressed microRNA cohorts in seed development may contribute to poor grain filling of inferior spikelets in rice. BMC Plant Biol. 14, 196.
Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., Mazo, A., 2006. Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx *in cis* by transcriptional interference. Cell 127 (6), 1209–1221.

Ponjavic, J., Ponting, C.P., Lunter, G., 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res. 17 (5), 556–565.
Ponting, C.P., Oliver, P.L., Reik, W., 2009. Evolution and functions of long noncoding RNAs. Cell 136 (4), 629–641.
Qiao, S., Sun, S., Wang, L., Wu, Z., Li, C., 2017. The RLA1/SMOS1 transcription factor functions with OsBZR1 to regulate brassinosteroid signaling and rice architecture. Plant Cell 29 (2), 292–309.
Rinn, J.L., Chang, H.Y., 2012. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166.
Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L. H., Helms, J.A., Farnham, P.J., Segal, E., Chang, H.Y., 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129 (7), 1311–1323.
Sekhar, S., Gharat, S.A., Panda, B.B., Mohaptra, T., Das, K., Kariali, E., Mohapatra, P.K., Shaw, B.P., 2015. Identification and characterization of differentially expressed genes in inferior and superior spikelets of rice cultivars with contrasting panicle-compactness and grain-filling properties. PLoS One 10 (12), e0145749.
Shao, Y., Zhou, H.-Z., Wu, Y., Zhang, H., Lin, J., Jiang, X., He, Q., Zhu, J., Li, Y., Yu, H., Mao, C., 2019. OsSPL3, an SBP-domain protein, regulates crown root development in rice. Plant Cell 31 (6), 1257–1275.
Sun, H., Peng, T., Zhao, Y., Du, Y., Zhang, J., Li, J., Xin, Z., Zhao, Q., 2015. Dynamic analysis of gene expression in rice superior and inferior grains by RNA-Seq. PLoS One 10 (9), e0137168.
Swiezewski, S., Liu, F., Magusin, A., Dean, C., 2009. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. Nature 462 (7274), 799–802.
Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., Chang, H.Y., 2010. Long noncoding RNA as modular scaffold of histone modification complexes. Science 329 (5992), 689–693.
Wang, L., Lin, G., Yu, X., Wu, Y., Chen, G., Xiong, F., 2020. Endosperm enrichment and physicochemical properties of superior and inferior grain starch in super hybrid rice. Plant Biol. 22 (4), 669–678.
Wang, Y., Tong, X., Qiu, J., Li, Z., Zhao, J., Hou, Y., Tang, L., Zhang, J., 2017. A phosphoproteomic landscape of rice (*Oryza sativa*) tissues. Physiol. Plantarum 160 (4), 458–475.
Wilusz, J.E., Sunwoo, H., Spector, D.L., 2009. Long noncoding RNAs: functional surprises from the RNA world. Gene Dev. 23 (13), 1494–1504.
Wu, K., Wang, S., Song, W., Zhang, J., Wang, Y., Liu, Q., Yu, J., Ye, Y., Li, S., Chen, J., Zhao, Y., Wang, J., Wu, X., Wang, M., Zhang, Y., Liu, B., Wu, Y., Harberd Nicholas, P., Fu, X., 2020. Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. Science 367 (6478), eaaz2046.
Yang, H.J., Shen, H., Chen, L., Xing, Y.Y., Wang, Z.Y., Zhang, J.L., Hong, M.M., 2002. The OsEBP-89 gene of rice encodes a putative EREBP transcription factor and is temporally expressed in developing endosperm and intercalary meristem. Plant Mol. Biol. 50 (3), 379–391.
Yang, J., Zhang, J., 2010. Grain-filling problem in 'super' rice. J. Exp. Bot. 61 (1), 1–5.
Yu, J., Xiong, H., Zhu, X., Zhang, H., Li, H., Miao, J., Wang, W., Tang, Z., Zhang, Z., Yao, G., Zhang, Q., Pan, Y., Wang, X., Rashid, M.A.R., Li, J., Gao, Y., Li, Z., Yang, W., Fu, X., Li, Z., 2017. OsLG3 contributing to rice grain length and yield was mined by Ho-LAMap. BMC Biol. 15 (1), 28.
Zhang, Y.C., Liao, J.Y., Li, Z.Y., Yu, Y., Zhang, J.P., Li, Q.F., Qu, L.H., Shu, W.S., Chen, Y. Q., 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. Genome Biol. 15 (12), 512.
Zhao, J., Ajadi, A.A., Wang, Y., Tong, X., Wang, H., Tang, L., Li, Z., Shu, Y., Liu, X., Li, S., Wang, S., Liu, W., Zhang, J., 2020. Genome-wide identification of lncRNAs during rice seed development. Genes 11 (3), 243.
Zhao, X., Li, J., Lian, B., Gu, H., Li, Y., Qi, Y., 2018. Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nat. Commun. 9 (1), 5056.