Full Length Article

Global Transcriptome Analysis of Long Noncoding RNAs in Rice Seed Development

Jingai Tan[†], Peng Wang[†], Jianfeng Yu, Caijing Li, Haodong Deng, Guangliang Wu, Yanning Wang, Xin Luo, Shan Tong, Xiangyu Zhang, Qin Cheng, Haohua He[*] and Jianmin Bian[*]
Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University, Nanchang 330045, China
[*]For correspondence: jmbian81@126.com; hhhua64@163.com
[†]Contributed equally to this work and are co-first authors

Abstract

Rice seed development involves an intricate regulatory network that directly determines seed size and weight. Long noncoding RNAs (lncRNAs) have been defined as key regulators of gene expression involved in diverse biological processes. However, the function of lncRNAs in rice seed development is still poorly understood. We performed paired-end RNA sequencing of *Nipponbare* rice at 5, 10 and 15 DPA (days post anthesis) in two different environments (early and middle-season rice). A total of 382 lncRNAs were detected as differentially expressed among these stages, including 344 and 307 lncRNAs in early and middle-season rice, respectively, and 70.42% (269 of 382) of the lncRNAs were found in both environments. The results showed that environment had little effect on the expression of lncRNAs. Furthermore, there were 127, 172, and 31 DElncs (differentially expressed lncRNAs) and 154, 140, and 59 DElncs in early and middle-season rice, respectively, in comparisons of 10_DPA vs 5_DPA, 15_DPA vs 5_DPA and 15_DPA vs 10_DPA. This result indicated that the number and expression level of lncRNAs at 5 DAP were significantly different from those at 10 DAP and 15 DAP. Furthermore, GO pathway analysis of cis target genes of DElncs in 10_DPA vs 5_DPA and 15_DPA vs 5_DPA revealed that the significant GO pathways were extracellular region, nutrient reservoir activity and cell wall macromolecule catabolic process. Our study revealed dynamic expression of lncRNAs in three stages and systematically explored the differences in lncRNAs between early and middle-season rice, which could provide a valuable resource for future high-yield breeding. © 2021 Friends Science Publishers

Keywords: Rice; LncRNAs; Paired-end RNA sequencing; Seed development; Environment

Introduction

A large fraction of unexpected eukaryotic transcripts are involved in important biological processes and have been named long noncoding RNAs (lncRNAs) (Deng *et al*. 2018). LncRNAs are transcripts that are greater than 200 nucleotides in length and have no protein coding potential (Rinn and Chang 2012; Batista and Chang 2013). A large portion of the genome of eukaryotes is transcribed into noncoding RNA and noncoding sequences are far more numerous than protein-coding genes (Derrien *et al*. 2012). We now know that lncRNAs have different origins, including the intronic and exonic regions of protein-coding genes, as well as non-intergenic regions (Bonasio and Shiekhattar 2014; Deng *et al*. 2018). It is becoming clear that lncRNAs are involved in many significant biological processes and pathways (Cech and Steitz 2014; Chen *et al*. 2018). Nevertheless, our understanding of the function of lncRNAs at the molecular level is currently very incomplete (Cech and Steitz 2014; Khemka *et al*. 2016; Zou *et al*. 2016).

In plants, lncRNAs are involved in regulating diverse biological processes, such as grain yield, flowering time, and response to cold stress (Ariel *et al*. 2014; Bardou *et al*. 2014; Berry and Dean 2015; Kindgren *et al*. 2018). For example, the utilization of the intron-derived lncRNA COOLAIR during cold exposure cooperates with FLC promoter-derived lncRNA COLDWRAP to catalyze the methylation of histone H3 at Lys27 (H3K27) and silence FLC (Berry and Dean 2015; Zhao *et al*. 2018). A long noncoding RNA antisense transcript overlapping OsSOC1 named Ef-cd (early flowering-completely dominant) positively correlates with the expression of OsSOC1 and H3K36me3 deposition involved in early flowering and high yield (Fang *et al*. 2019). LAIR (LRK Antisense Intergenic RNA) regulates several LRK genes by significantly catalyzing H3K4me3 and H4K16ac in the LRK1 genomic region, contributing to increased rice grain yield (Wang *et al*. 2018). An overwhelmingly large fraction of well-characterized plant lncRNAs with established functions were researched in only a few model plants (Chen *et al*.

2018; Deng *et al*. 2018; Wang *et al*. 2018). In contrast, the regulatory mechanisms of lncRNAs in rice remain fragmentary. Therefore, it is necessary to further investigate the function of lncRNAs in rice.

Rice (*Oryza sativa* L.) is a staple and important cereal crop worldwide (Song *et al*. 2007). Seed size is an important agronomic trait that affects potential yield, and it is determined by three indicators: length, width and thickness. Previous studies have identified several genes involved in regulating grain size, grain weight and grain length (Mao *et al*. 2010; Tong *et al*. 2012; Liu *et al*. 2017; Hu *et al*. 2018). For example, GW5 encodes a calmodulin binding protein and can physically interact with and repress glycogen synthase kinase 2 (GSK2), which positively regulates the brassinosteroid (BR)-responsive gene to increase rice yield (Liu *et al*. 2017). GS3 is a major quantitative trait locus for grain yield that can negatively regulate grain size and seed size (Mao *et al*. 2010). OsGSK5 is a member of the glycogen synthase kinase 3/shaggy-like family and can interact with and phosphorylateauxin response factor 4 (OsARF4), which may be involved in regulating auxin-responsive genes to affect rice yield (Hu *et al*. 2018). Seed growth and development play a pivotal role in the physiological process of seed maturation that directly determines yield and quality (Finnie *et al*. 2002). The three major components of the seed are the endosperm, embryo and seed coat. The endosperm accounts for 85% of a mature seed and contains a great deal of nutrition to support embryo development and seed germination in angiosperms (Zhan *et al*. 2015). As seeds are the reproductive organ of rice, it is of great significance to elaborate the genetic networks that regulate seed growth and development, and this could provide a theoretical basis for increasing rice yield.

The function of lncRNAs during seed development in rice remains poorly understood. In this study, we used a deep RNA sequencing (RNA-Seq) strategy to comprehensively profile the lncRNAs expressed during early grain development, the grain filling stage and the grain mature stage in two different environments (early and middle-season rice), which could clarify the potential functions of these lncRNAs in mediating seed development.

Materials and Methods

Plant materials and growing conditions

The experiments all used *Nipponbare* (*Oryza sativa*). Rice plants were cultivated in the experimental field of Jiangxi Agricultural University, Jiangxi Province, China (28°45′36″N, 115°22′58″E), in the early (March to July) and middle (May to September) seasons with a transplant spacing of 13.3 cm × 26.7 cm during the 2018 crop season. Grain samples were collected at 5, 10- and 15-days post anthesis (DPA). These three stages mainly cover the cellularization and maturation of endosperm in early and middle-season rice. All eighteen samples (each containing three biological replicates) were immediately frozen in liquid nitrogen and stored at -80°C.

RNA extraction and sequencing

Total RNA was extracted from each sample using the TRK1001 Total RNA Purification Kit (LC Science, Houston, TX, USA). The construction of transcriptome libraries and deep sequencing were performed by the Lianchuan Biological Company (Hangzhou). Total RNA was quality controlled and quantified using a Bioanalyzer 2100 and RNA 1000 Nano LabChip Kit (Agilent, CA, USA) with an RIN >7.0. RNA purity was checked using aNanoPhotometer spectrophotometer (Implen, Los Angeles, CA, USA). Following purification, the RNA was fragmented by the addition of divalent cations under high temperature conditions. Fragments of suitable size were selected with AMPure XP beads for PCR amplification to create a cDNA library. All the RNA-seq data have been deposited in the NCBI SRA database.

lncRNA identification

Low quality and adaptor sequences in raw RNA-Seq reads from 18 samples were trimmed using Trimmomatic (v0.36) (Bolger *et al*. 2014). Then, clean reads were mapped to the rice reference genome (IRGSP-1.0) using TopHat2 (v2.1.0) (Kawahara *et al*. 2013; Kim *et al*. 2013). Mapped reads (bam file) from 18 samples were merged into a single bam file. Genome-guided transcript assembly was performed using Cufflinks (v2.1.0) based on the merged bam file (Trapnell *et al*. 2010).

LncRNAs were identified from the above assembled transcripts using a modified pipeline (Wang *et al*. 2017). In summary, transcripts shorter than 200 nt were removed first. Long transcripts (> 200 nt) overlapping with reference genes in the rice genome were discarded, and the remaining transcripts were classified into 3 categories based on their locations with respect to reference genes in the reference genome: 1) intergenic transcripts, 2) intronic transcripts and 3) antisense transcripts. Then, potential protein-coding transcripts in these three types of transcripts were removed based on a similarity search against the SWISS_PROT protein database and prediction of the longest open reading frame (ORF) (Bairoch and Apweiler 2000). Finally, the expression values (raw read counts) of the remaining transcripts were checked, and only transcripts with expression values (raw read counts) greater than 10 in at least 4 of 18 samples were kept and considered robust lncRNAs in this study.

Quantitative real-time PCR (qRT-PCR) validation of lncRNAs

Four lncRNAs were randomly selected for validation by

quantitative real-time PCR (qRT-PCR). RNA used for RNA-seq samples was reverse transcribed with the PrimeScript™ RT reagent kit with gDNA Eraser (PrimeScript™ Reverse Transcription System, Takara, Dalian, China). qRT-PCR was performed in triplicate on an ABI 7500 with three biological replicates.The following cycling conditions were used for qRT-PCR: 50°C for 2 min; 95°C for 2 min; 40 cycles of 15 s at 95°C and 30 s at 60°C; and a final step for melting curve determination (15 s at 95°C, 1 min at 60°C and 15 s at 95°C). *GAPDH* was used as an internal control. lncRNA expression was calculated based on the 2-ΔΔCt method.

Identification of differentially expressed lncRNAs

The R package "edgeR" was used to perform lncRNA differential expression analysis (Robinson *et al*. 2010). Library sizes of 18 samples were calculated by adding all reads mapped to protein-coding genes as well as lncRNAs together. Meanwhile, we chose the expressed transcripts (for multiple-exon transcripts, FPKM $\geq$0.5; for single-exon transcripts, FPKM $\geq$2) for further research. Statistically significantly DE lncRNAs were selected according to an FDR (false discovery rate) threshold of $P < 0.05$.

LncRNA cis target gene prediction and functional enrichment analysis

We tested the correlation of expression between lncRNAs and their putative cis target genes, which were spaced 100 kb upstream and downstream of these lncRNAs. To understand the function of the neighboring target genes of lncRNAs, Gene Ontology (GO) term enrichment was used to perform the classification analysis with" hypergeometric" as the statistical test method and "Hochberg FDR" to correct for multiple testing (Du *et al*. 2010).

Coexpression network analysis of lncRNAs and cis target genes

To illustrate the potential regulatory interactions between lncRNAs and mRNAs during the seed development process, a coexpression network for seed development was constructed based on WGCNA, which is a comprehensive collection of R functions for weighting correlation network analysis. LncRNAs regulate gene expression through cis-acting interactions, which were spaced 100 k upstream and downstream of these lncRNAs, which are likely to be potential targets and subjected to coexpression analysis. Raw counts of lncRNAs and their neighboring genes were normalized using VST (variance stabilized transformation). Coexpression networks were reconstructed using WGCNA with soft power "10" (Langfelder and Horvath 2008). Coexpression networks were visualized using Cytoscape (v3.8.0) (Shannon *et al*. 2003).

Results

Genome-wide identification of candidate lncRNAs responding to seed development in rice

To uncover rice noncoding transcripts responding to seed development, we performed paired-end RNA-seq of 18 samples. In total, we obtained 55,432 assembled long transcripts (> 200 nt) after transcriptome reconstruction using cufflinks. Subsequently, we discarded long transcripts overlapping with reference gene annotations, and the remaining transcripts were classified into 3 categories according to their locations, including 9,448 (85.5%) intergenic transcripts, 1,018 (9.2%) intronic transcripts and 590 (5.3%) antisense transcripts. After removing protein-coding transcripts based on a similarity search against the SWISS_PROT protein database and prediction of the longest ORF, we were able to identify 1,305 lncRNAs. To obtain a more stringent lncRNA dataset, we filtered low-expressed lncRNAs based on their expression profiles across 18 samples. A total of 421 candidate lncRNAs were identified, comprising 218 (51.8%) intergenic transcripts, 176 (41.8%) intronic transcripts and 27 (6.4%) antisense transcripts (Fig. 1a).

To validate the reliability of RNAseq data for lncRNA expression profiles, we randomly selected four lncRNAs (*LINC.CUFF.1743*, *INTRONIC.CUFF.35400*, *INTRONIC.CUFF.34987* and *INTRONIC.CUFF.32583*) in three seed development stages in two environments and evaluated their expression by qRT-PCR. Notably, the four lncRNA expression trends were consistent with the expression levels calculated from the deep sequencing data in the two environments (Fig. 1b), indicating that overall, the RNAseq data sets were reliable. The primer sequences used are listed in Table S1.

Comparative Analysis of Features of mRNAs and lncRNAs

To comprehensively understand the features of lncRNAs, we analyzed their gene structure and expression in comparison to mRNAs. There were 71.73% lncRNAs shorter than 1000 bp and 75.82% mRNAs longer than 1000 bp (Fig. 2a and Table S2), indicating that the length of lncRNAs was generally shorter than that of mRNAs. Meanwhile, 83.85% (353 of 421) of lncRNAs were single-exon transcripts, while some mRNAs contained more than 25 exons, and a large fraction of mRNAs were multiple-exon transcripts (Fig. 2b). We also found that the expression of lncRNAs was much lower than that of mRNAs (Fig. 2c), which was consistent with previous research (Zhao *et al*. 2020). We compared the length of open reading frames (ORFs) between lncRNAs and mRNAs and found that all of the lncRNAs were shorter than 100 aa, while the ORFs of mRNAs were much larger than 100 aa (Fig. 2d).
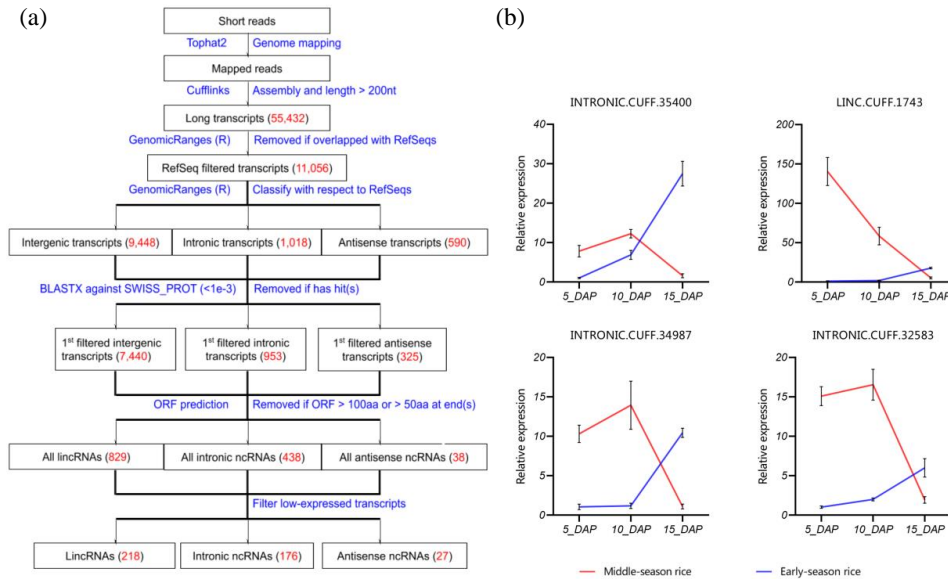
Fig. 1: The identification pipeline and validation of lncRNAs (a) LncRNAs identification pipeline and corresponding numbers of transcripts (in red) for each step (b) Validation of lncRNAs relative expression levels under early and middle-season rice at three intervals. 5_DPA, 10_DPA and 15_DPA respectively represent panicle at 5, 10 and 15 days post-anthesis, Blue and red bars stand for early and middle-season rice, respectively. GADPH was used as an endogenous control
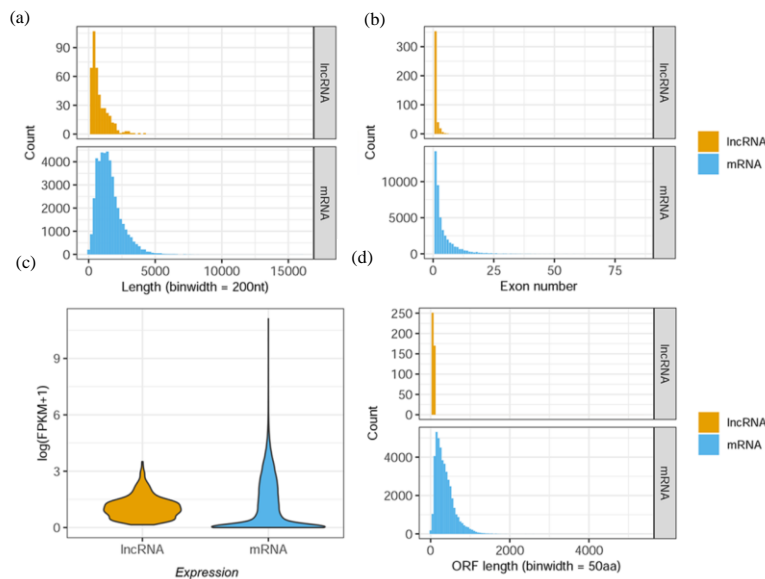

Fig. 2: Comparative analysis of features of mRNAs and lncRNAs (a) Length distribution of lncRNAs compared to protein-coding RNAs (mRNAs). (b) Exon numbers of lnRNAs compared to mRNAs. (c) Expression levels of lncRNAs compared to mRNAs.Yellow and blue respectively represent lncRNAs and mRNAs (d) ORF length of lncRNAs compared to mRNAs

Expression Profiles of lncRNAs During Seed Development

LncRNAs play critical roles in regulating coding gene expression (Wang *et al.* 2018). A multidimensional scaling (MDS) plot based on the gene expression profiles of eighteen samples showed the clustering of global expression of lncRNAs at 5, 10 and 15 DPA in two environments (early and middle-season rice). The three stages mainly cover the cellularization and maturation of endosperm in the early and middle-season rice. The results showed a relatively high repeatability of the experiment in terms of data analysis (Fig. 3a). The FPKMs of all lncRNAs detected in the eighteen samples were analyzed, and the results showed that the expression profiles of lncRNAs were different during the three seed development stages (Fig. 3b).
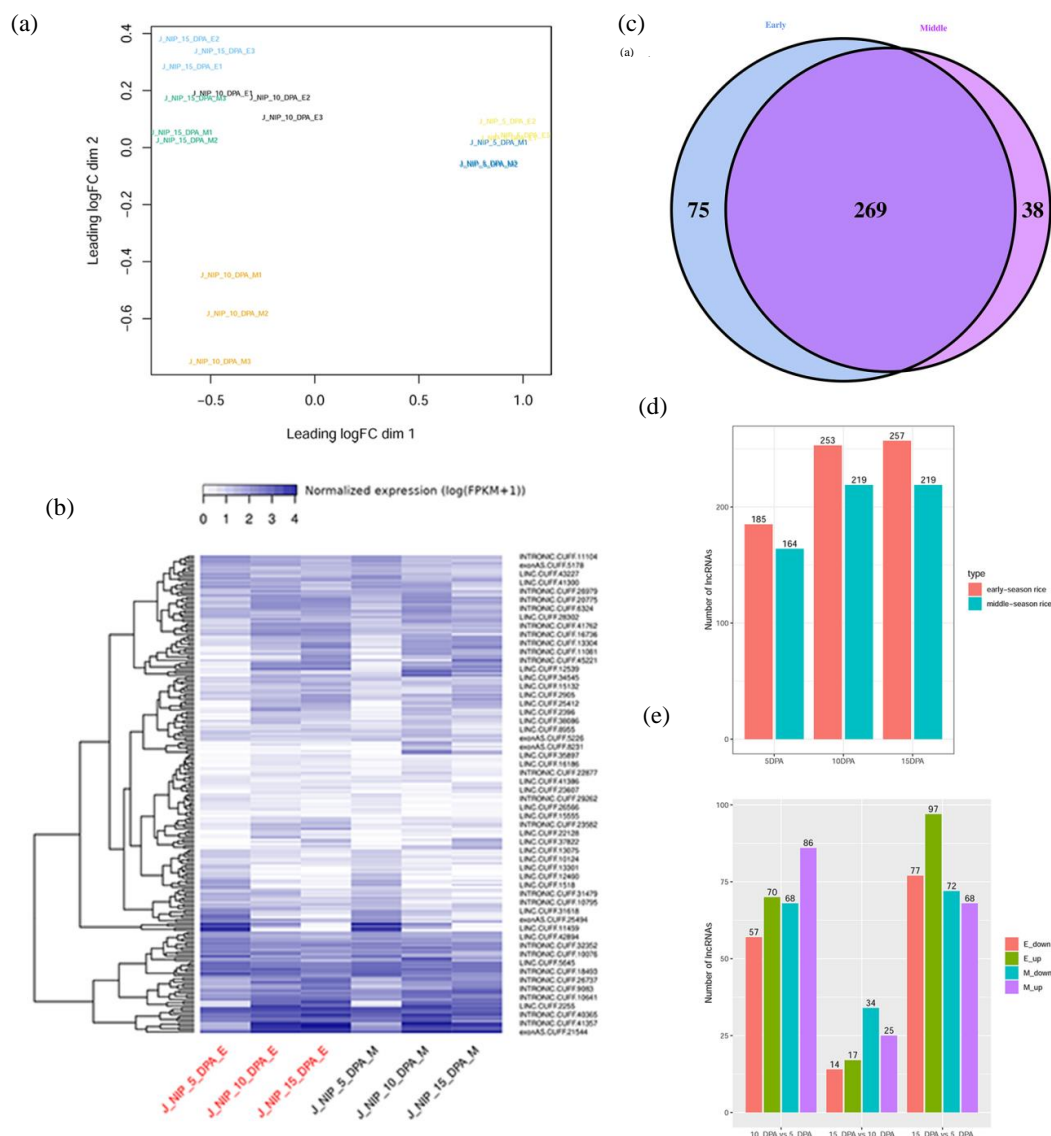
Fig. 3: LncRNAs differential expression analysis (a) MDS plot showing the cluster of global expression of lncRNAs among eighteen samples. Each color represents a kind of sample; each sample has three biological replicates. (b) Heatmap representation for the expression profiles of lncRNAs in 5_DPA, 10_DPA and 15_DPA at early and middle-season rice, all expression levels are normalized by FPKMs. (c) Venn diagram compares of the lncRNAs in early and middle-season rice. (d) Bar graph showing the number of lncRNAs in 5_DPA, 10_DPA and 15_DPA at early and middle-season rice, Red and blue respectively represent early and middle-season rice. (e) Bar graph showing the number of up-regulated and down-regulated genes in groups 10_DPA vs 5_DPA, 15_DPA vs 10_DPA and 15_DPA vs 5_DPA at early and middle-season rice. Red and blue respectively represent down-regulated genes in early and middle-season rice, Green and purple respectively represent up-regulated genes in early and middle-season rice

We removed the lncRNAs with low expression according to FPKM (Chen *et al*. 2018), and a total of 382 lncRNAs were obtained for further research, which included 344 and 309 lncRNAs from early and middle-season rice, respectively; 70.42% (269 of 382) of these lncRNAs were expressed in both environments (Fig. 3c). In addition, 72.77% (147 of 202), 60.54% (178 of 294), and 64.71% (187 of 289) of lncRNAs were expressed in both the early and middle-season rice, at 5 DAP, 10 DAP and 15 DAP,

respectively (Fig. 3d).

To identify lncRNAs potentially involved in the regulation of seed development, we analyzed the differentially expressed lncRNAs for seed developmental stages in early and middle-season rice. We found that there were 70, 98, and 17 significantly upregulated lncRNAs and 57, 74, and 14 downregulated lncRNAs in 10_DPA vs 5_DPA, 15_DPA vs 5_DPA_ and 15_DPA vs 10 DPA, respectively, in early-season rice, and there were 86, 25, and
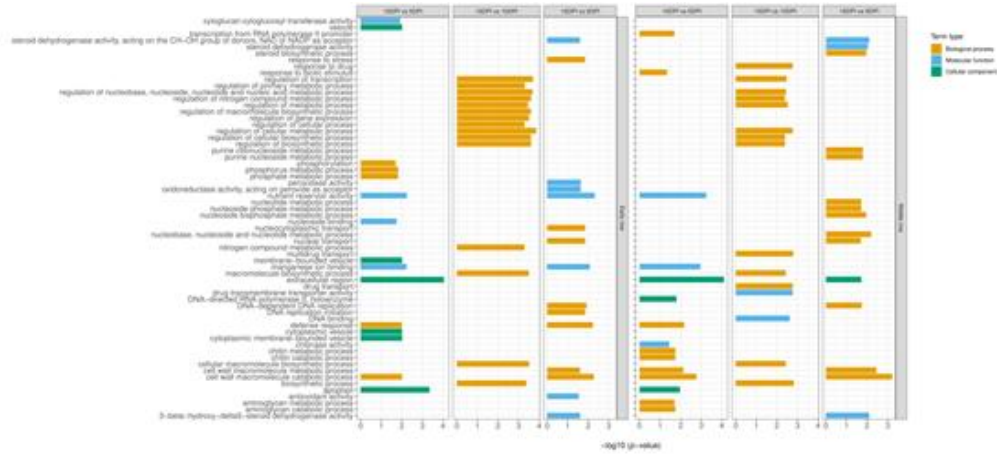
Fig. 4: Overview of Gene Ontology analysis of all DElncs in groups 10_DPA vs 5_DPA, 15_DPA vs 10_DPA and 15_DPA vs 5_DPA in early and middle-season rice. The x-axis represents the negative log of the P-value, and y-axis stands for GO terms
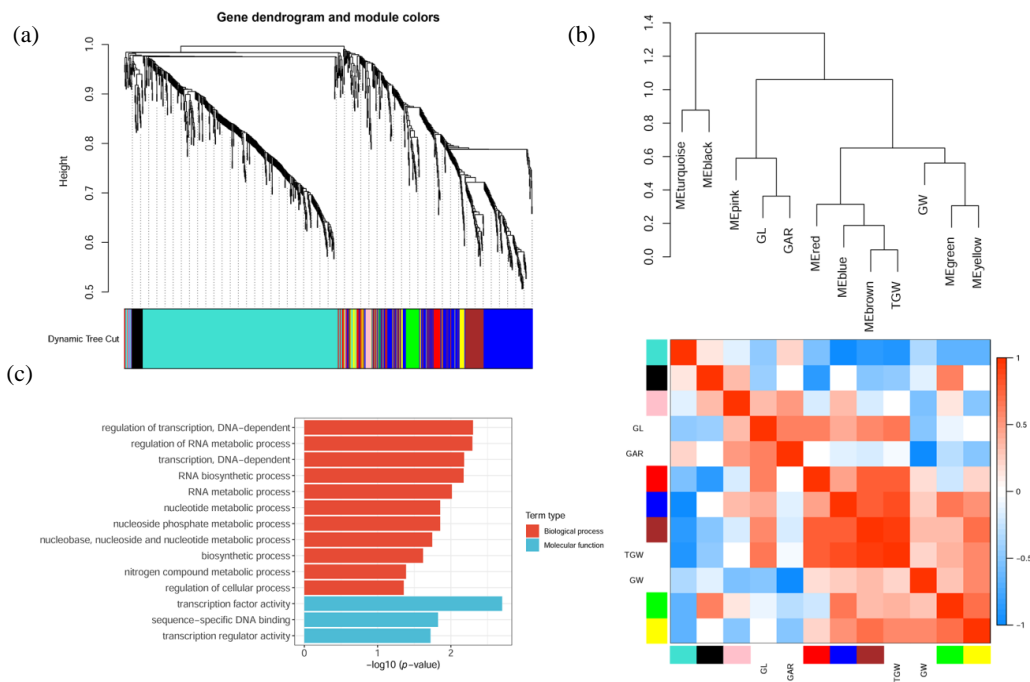

Fig. 5: Co-expression networks for lncRNAs and cis-target genes during seed development. (a) Hierarchical clustering dendrograms (Dendrogram) showing the cluster of transcripts and co-expression modules, the color below the dendrogram demonstrates the module assignment determined by the dynamic tree cut. (b) The dendrograms shows the relation of modules with grain yield and the heatmap shows the eigengene adjacency, GL represents grain length, GW represents grain width, TGW represents thousand grain weight, GAR represents grain aspect ratio. (c) Gene ontology (GO) enrichment analysis for co-expressed genes in brown module

68 significantly upregulated lncRNAs and 68, 34, and 72 downregulated lncRNAs in 10_DPA vs 5_DPA, 15_DPA vs 5_DPA_ and 15_DPA vs 10 DPA, respectively, in middle-season rice (Fig. 3e).

Enrichment Analysis of cis target Genes During Seed Development

To further explore the potential functions of DElncs involved in seed development, it is well known that the corresponding neighboring genes of lncRNAs are likely to be potential target genes (Wang *et al.* 2011). Therefore, the potential neighboring target genes spaced 100 kb upstream and downstream of these DElncs (Chen *et al.* 2018), and there were 1389, 306 and 1675 matched lncRNA-mRNA pairs in 10_DPA vs 5_DPA, 15_DPA vs 10_DPA and 15_DPA vs 5_DPA in early-season rice, respectively. In addition, there were 892, 146, and 1099 matched lncRNA-

mRNA pairs in middle-season rice, respectively (Table S3).

These neighboring potential target genes were analyzed with GO enrichment analysis to predict their function (Fig. 4 and Table S4). The results showed that potential target genes had different functions in regulating many biological processes at the three different seed developmental stages not only in the early-season rice but also in the middle-season rice. Simultaneously, in the early-season rice, the top three most significantly enriched biological processes were extracellular region, apoplast and nutrient reservoir activity in 10_DPA vs 5_DPA; regulation of cellular metabolic process, regulation of cellular biosynthetic process and cellular macromolecule biosynthetic process in 15_DPA vs 10_DPA; and cell wall macromolecule catabolic process, nutrient reservoir activity and defense response in 15_DPA vs 5_DPA.

On the other hand, in the middle-season rice, the top three most significantly enriched biological processes were extracellular region, cell wall macromolecule catabolic process and nutrient reservoir activity in 10_DPA vs 5_DPA; biosynthetic process, multidrug transport and regulation of cellular metabolic process in 15_DPA vs 10_DPA; and cell wall macromolecule catabolic process, cell wall macromolecule metabolic process and nucleobase, nucleoside and nucleotide metabolic process in 15_DPA vs 5_DPA.

Weighted Gene Coexpression Network Analysis

To elaborate the potential correlated pairs of lncRNAs and genes during seed development stages, we performed weighted gene coexpression network analysis (WGCNA) based on paired-end RNA-seq data. A total of 382 DE lncRNAs and their neighboring genes were assembled into 8 modules/subnetworks by hierarchical clustering and dynamic branch cutting (Fig. 5a and Table S5). Each module was defined with a unique color as an identifier, and gray modules represent the set of genes that were not assigned to any modules (Fig. S1). We investigated whether any module was correlated with grain yield and tested the relationships between each module and grain yield traits. We found that the most relevant module, brown (r = 0.95, P =4.3e-10), had the strongest association with thousand grain weight (TGW) (Fig. 5b and Table S6). All target genes of the lncRNAs in the brown module were subjected to Gene Ontology (GO) analysis for further elucidation of the functional properties. The top three most significant pathways in the brown module were transcription factor activity, regulation of transcription, DNA−dependent and regulation of RNA metabolic process (Fig. 5c and Table S7).

Discussion

Over the past decades, high-throughput sequencing technologies have emerged for both plants and mammals,

and many lncRNAs have been identified and analyzed (Batista and Chang 2013; Zhao *et al*. 2015; Deng *et al*. 2018). High-throughput sequencing has the advantages of low cost, advanced technology and the ability to perform large-scale parallel deep sequencing; therefore, transcriptome sequencing is widely used to analyze lncRNA function (Lu *et al*. 2016). Many studies have demonstrated that lncRNAs have ubiquitous biological functions in almost every aspect of biological processes and are involved in regulating gene expression (Guttman *et al*. 2011; Ariel *et al*. 2014; Berry and Dean 2015). However, the regulatory mechanisms of lncRNAs related to rice seed development are not well characterized. In this study, the genes expressed in three key growth stages (5, 10, 15 days post anthesis) in two environments were analyzed in eighteen samples by paired-end transcriptome sequencing. LncRNAs have a relatively shorter length and lower exon number than mRNAs (Yu *et al*. 2020). Our data showed that lncRNAs were significantly shorter in length and expressed at much lower levels than mRNAs; moreover, 83.6% of lncRNAs were single-exon transcripts, while most mRNAs had multiple-exon transcripts. These results are similar to previous reports on lncRNA characteristics in rice (Zhao *et al*. 2020).

In total, we screened 382 lncRNAs that were differentially expressed throughout seed development among three stages for two different environments. For these lncRNAs, 344 and 307 lncRNAs were detected in early and middle-season rice, respectively, and 70.42% (269 of 382) of the lncRNAs were found in both environments. Furthermore, we found that 72.77% (147 of 202), 60.54% (178 of 294), and 64.70% (187 of 289) of lncRNAs were expressed in both environments at 5 DAP, 10 DAP and 15 DAP, respectively. In addition, the metabolic pathways in early-season rice were the same as those in middle-season rice; the most significant GO pathways in both environments were extracellular region, regulation of cellular metabolic process and cell wall macromolecule catabolic process in 10_DPA vs 5_DPA, 15_DPA vs 10_DPA and 15_DPA vs 5_DPA, respectively. The results indicated that the environment had little effect on the expression of lncRNAs during the rice seed development process.

On the other hand, the expression profile and metabolic pathways were different in the three stages. For early-season rice, there were 127, 172, and 31 DElncs in 10_DPA vs 5_DPA, 15_DPA vs 5_DPA and 15_DPA vs 10_DPA, respectively. In contrast, in middle-season rice, there were 154, 140, and 59 DElncs in 10_DPA vs 5_DPA, 15_DPA vs 5_DPA and 15_DPA vs 10_DPA, respectively. The results indicated that the number and expression level of lncRNAs at 5 DAP were significantly different from those at 10 DAP and 15 DAP. In other words, the gene mechanism regulating rice grain filling in the early stage might be different from that of the middle and late stages. Furthermore, the GO pathway analysis of cis target genes of

DElncs in 10_DPA vs 5_DPA and 15_DPA vs 5_DPA revealed that the significant GO pathways were mainly extracellular region, nutrient reservoir activity and cell wall macromolecule catabolic process, which are involved in regulating the production of seed nutrients and macromolecules in the grain filling stage.

Conclusion

This study provided the first systematic analysis of lncRNA dynamic regulatory profiles among three seed development stages in early and middle-season rice. The results showed that a large proportion of lncRNAs were only slightly affected by the environment during the grain development stage. Meanwhile, the number and expression level of lncRNAs at 5 DAP were significantly different from those at 10 DAP and 15 DAP. The enriched GO pathways were mainly involved in regulating seed nutrients and macromolecules in the grain filling stage. Our study overall characterized lncRNA expression and molecular pathways related to the differences in three seed developmental stages in early and middle-season rice and will provide a valuable resource for future high-yield breeding.

Acknowledgments

Author Contributions

Haohua He, Jianmin Bian and Jingai Tan designed the research work, annotated the data and drafted the manuscript. Jingai Tan performed the experiment, Jianfeng Yu took the samples, Peng Wang took validation of lncRNAs. Haodong Deng, Guangliang Wu, Xin Luo, Shan Tong, Xiangyu Zhang, Yanning Wang, Qin Cheng and Caijing Li interpreted the data. All authors read and approved the final manuscript.

Conflict of Interest

It is hereby declared that the authors have no competing interest

Data Availability Declaration

It is declared that data relevant to this article are available with the corresponding authors and will be made available on demand

References

Ariel F, T Jegu, D Latrasse, N Romero-Barrios, A Christ, M Benhamed, M Crespi (2014) Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell* 55:383–396

Bairoch A, R Apweiler (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl Acids Res* 28:45–48

Bardou F, F Ariel, CG Simpson, N Romero-Barrios, P Laporte, S Balzergue, JW Brown, M Crespi (2014) Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Dev Cell* 30:166–176

Batista PJ, HY Chang (2013) Long Noncoding RNAs: Cellular Address Codes in Development and Disease. *Cell* 152:1298–1307

Berry S, C Dean (2015) Environmental perception and epigenetic memory: Mechanistic insight through FLC. *Plant J* 83:133–148

Bolger AM, M Lohse, B Usadel (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120

Bonasio R, R Shiekhattar (2014) Regulation of transcription by long noncoding RNAs. *Ann Rev Genet* 48:433–455

Cech TR, JA Steitz (2014) The noncoding RNA Revolution-trashing old rules to forge new ones. *Cell* 157:77–94

Chen L, SL Shi, NF Jiang, H Khanzada, GM Wassan, CL Zhu, XS Peng, J Xu, YJ Chen, QY Yu, XP He, JR Fu, XR Chen, LF Hu, LJ Ouyang, XT Sun, HH He, JM Bian (2018) Genome-wide analysis of long non-coding RNAs affecting roots development at an early stage in the rice response to cadmium stress. *BMC Genomics* 19:460

Deng P, S Liu, X Nie, WN Song, L Wu (2018) Conservation analysis of long non-coding RNAs in plants. *Sci China Life Sci* 61:190–198

Derrien T, R Johnson, G Bussotti, A Tanzer, S Djebali, H Tilgner, G Guernec, D Martin, A Merkel, DG Knowles, J Lagarde, L Veeravalli, XA Ruan, YJ Ruan, T Lassmann, P Carninci, JB Brown, L Lipovich, JM Gonzalez, M Thomas, CA Davis, R Gingeras, TR Gingeras, TJ Hubbard, C Notredame, J Harrow, R Guigó (2012) The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genom Res* 22:1775–1789

Du Z, X Zhou, Y Ling, Z Zhang, Z Su (2010) agriGO: A GO analysis toolkit for the agricultural community. *Nucl Acids Res* 38:W64–W70

Fang J, FT Zhang, HG Wang, W Wang, F Zhao, ZJ Li, CH Sun, FM Chen, F Xu, SQ Chang, L Wu, QY Bu, PG Wang, JK Xie, F Chen, XH Huang, YJ Zhang, XG Zhu, B Han, XJ Deng, CC Chu (2019) Ef-cd locus shortens rice maturity duration without yield penalty. *Proc Natl Acad Sci USA* 116:18717–18722

Finnie C, S Melchior, P Roepstorff, B Svensson (2002) Proteome analysis of grain filling and seed maturation in barley. *Plant Physiol* 129:1308–1319

Guttman M, J Donaghey, BW Carey, M Garber, JK Grenier, G Munson, G Young, AB Lucas, R Ach, L Bruhn, X Yang, I Amit, A Meissner, A Regev, JL Rinn, DE Root, ES Lander (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477:295–300

Hu ZJ, SJ Lu, MJ Wang, HH He, L Sun, HR Wang, XH Liu, L Jiang, JL Sun, XY Xin, W Kong, CC Chu, HW Xue, JS Yang, XJ Luo, JX Liu (2018) A Novel QTL qTGW3 Encodes the GSK3/SHAGGY-like kinase OsGSK5/OsSK41 that interacts with OsARF4 to negatively regulate grain size and weight in rice. *Mol Plant* 11:736–749

Kawahara Y, M de la Bastide, JP Hamilton, H Kanamori, WR McCombie, S Ouyang, DC Schwartz, T Tanaka, JZ Wu, SG Zhou, KL Childs, RM Davidson, HN Lin, L Quesada-Ocampo, B Vaillancourt, H Sakai, SS Lee, J Kim, H Numa, T Itoh, CR Buell, T Matsumoto (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6; Article 4

Khemka N, VK Singh, R Garg, M Jain (2016) Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development. *Sci Rep* 6; Article 33297

Kim D, G Pertea, C Trapnell, H Pimentel, R Kelley, SL Salzberg (2013) TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14; Article R36

Kindgren P, R Ard, M Ivanov, S Marquardt (2018). Transcriptional read-through of the long non-coding RNA SVALKA governs plant cold acclimation. *Nat Commun* 9; Article 4561

Langfelder P, S Horvath (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform* 9; Article 559

Liu JF, J Chen, XM Zheng, FQ Wu, QB Lin, YQ Heng, P Tian, ZJ Cheng, XW Yu, KN Zhou, X Zhang, XP Guo, JL Wang, HY Wang, JM Wan (2017) GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat Plants* 3; Article 17043

Lu XK, XG Chen, M Mu, JJ Wang, XG Wang, DL Wang, ZJ Yin, WL Fan, S Wang, LX Guo, WW Ye (2016) Genome-wide analysis of long noncoding RNAs and their responses to drought stress in cotton (*Gossypium hirsutum* L.). *PLoS One* 11:e0156723

Mao HL, SY Sun, JL Yao, CR Wang, SB Yu, CG Xu, XH Li, QF Zhang (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc Natl Acad Sci USA* 107:19579–19584

Rinn JL, HY Chang (2012) Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81:145–166

Robinson MD, DJ McCarthy, GK Smyth (2010). edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140

Shannon P, A Markiel, O Ozier, NS Baliga, JT Wang, D Ramage, N Amin, B Schwikowski, T Ideker (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504

Song XJ, W Huang, M Shi, MZ Zhu, HX Lin (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630

Tong HN, LC Liu, Y Jin, L Du, YH Yin, Q Qian, LH Zhu, CC Chu (2012) DWARF AND LOW-TILLERING acts as a direct downstream target of a GSK3/SHAGGY-like kinase to mediate brassinosteroid responses in rice. *Plant Cell* 24:2562–2577

Trapnell C, BA Williams, G Pertea, A Mortazavi, G, MJ van Baren, SL Salzberg, BJ Wold, L Pachter (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515

Wang D, ZP Qu, L Yang, QZ Zhang, ZH Liu, T Do, DL Adelson, ZY Wang, I Searle, JK Zhu (2017) Transposable elements (TEs) contribute to stress-related long intergenic noncoding RNAs in plants. *Plant J* 90:133–146

Wang KC, YW Yang, B Liu, A Sanyal, R Corces-Zimmerman, Y Chen (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472:120–124

Wang Y, XJ Luo, F Sun, JH Hu, XJ Zha, W Su, JS Yang (2018) Overexpressing lncRNA LAIR increases grain yield and regulates neighbouring gene cluster expression in rice. *Nat. Commun* 9; Article 3516

Yu Y, YF Zhou, YZ Feng, H He, JP Lian, Y Yang, Wei, MQ Lei, YC Zhang, YQ Chen (2020) Transcriptional landscape of pathogen-responsive lncRNAs in rice unveils the role of ALEX1 in jasmonate pathway and disease resistance. *Plant Biotechnol J* 18:679–690

Zhan JP, D Thakare, C Ma, A Lloyd, NM Nixon, AM Arakaki, WJ Burnett, KO Logan, DF Wang, XF Wang, GN Drews, R Yadegari (2015) RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* 27:513–531

Zhao J, AA Ajadi, YF Wang, XH Tong, HM Wang, LQ Tang, ZY Li, YZ Shu, J Zhang (2020) Genome-wide identification of lncRNAs during rice seed development. *Genes* 11; Article 243

Zhao XY, JG Li, B Lian, HQ Gu, Y Li, YJ Qi (2018) Global identification of Arabidopsis lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. *Nat Commun* 9; Article 5056

Zhao Z, J Bai, AW Wu, Y Wang, JW Zhang, ZS Wang, YS Li, J Xu, X Li (2015) *Co-LncRNA: Investigating the lncRNA combinatorial effects in GO annotations and KEGG pathways based on human RNA-Seq data*. Database 2015

Zou CS, QL Wang, CR Lu, WC Yang, YP Zhang, HL Cheng, XX Feng, MA Prosper, GL Song (2016) Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). *Sci Chin Life Sci* 59:164–171