*CORRESPONDENCE
Hang He
✉ hehang@pku.edu.cn
Xing Wang Deng
✉ deng@pku.edu.cn

Gene expression variations and allele-specific expression of two rice and their hybrid in caryopses at single-nucleus resolution

Han Zhou[1,2], Xing Wang Deng[1,2]* and Hang He[1,2]*

[1]State Key Laboratory of Protein and Plant Gene Research, School of Advanced Agriculture Sciences and School of Life Sciences, Peking University, Beijing, China, [2]Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Peking University Institute of Advanced Agricultural Sciences, Weifang, Shandong, China

Seeds are an indispensable part of the flowering plant life cycle and a critical determinant of agricultural production. Distinct differences in the anatomy and morphology of seeds separate monocots and dicots. Although some progress has been made with respect to understanding seed development in *Arabidopsis*, the transcriptomic features of monocotyledon seeds at the cellular level are much less understood. Since most important cereal crops, such as rice, maize, and wheat, are monocots, it is essential to study transcriptional differentiation and heterogeneity during seed development at a finer scale. Here, we present single-nucleus RNA sequencing (snRNA-seq) results of over three thousand nuclei from caryopses of the rice cultivars Nipponbare and 9311 and their intersubspecies F_1 hybrid. A transcriptomics atlas that covers most of the cell types present during the early developmental stage of rice caryopses was successfully constructed. Additionally, novel specific marker genes were identified for each nuclear cluster in the rice caryopsis. Moreover, with a focus on rice endosperm, the differentiation trajectory of endosperm subclusters was reconstructed to reveal the developmental process. Allele-specific expression (ASE) profiling in endosperm revealed 345 genes with ASE (ASEGs). Further pairwise comparisons of the differentially expressed genes (DEGs) in each endosperm cluster among the three rice samples demonstrated transcriptional divergence. Our research reveals differentiation in rice caryopsis from the single-nucleus perspective and provides valuable resources to facilitate clarification of the molecular mechanism underlying caryopsis development in rice and other monocots.

KEYWORDS

snRNA-seq, endosperm, rice, caryopsis, development

1 Introduction

Seeds of cereal crops greatly contribute to human food, animal feed, biofuels and raw material for manufactured goods. Cereal caryopses mainly consist of three components with distinct genetic origins: filial diploid tissues of the embryo, filial triploid tissues of the endosperm and maternal diploid tissues such as the pericarp, testa, integuments, and nucellus (Lafon-Placette and Köhler, 2014; Liu et al., 2022). In rice, the embryo initiates egg cell fertilization with a sperm nucleus to form a zygote. From the first zygotic division, it takes 10 stages for the single cell (fertilized egg) to develop into the mature embryo, which is marked by landmark events including embryonic shoot apical meristem (SAM) formation, leaf primordia emergence, leaf primordia formation and organ enlargement (Xue et al., 2012). The development of rice endosperms starts from coenocytic nuclear divisions of the triploid central nucleus in the embryo sac after fertilization and continues with cellularization to form cell walls between individual nuclei. The cellular endosperm further differentiates into various tissue types, including aleurone and starchy endosperm, and rapidly accumulates storage products (Liu et al., 2022). The maternal tissues are fused tightly together, encasing the embryo and endosperm. In these maternal tissues, cell division, cell expansion, tissue differentiation and degeneration play important roles in sugar loading and grain filling (Wu et al., 2016; Ingram, 2017).

To more accurately dissect the spatial-temporal partitions of rice caryopses, approaches such as laser capture microdissection (Ram et al., 2021), mechanical squeezing (Luo et al., 2011), glass micropipette aspiration (Kuang et al., 2015), and cell type-specific expression profiling (Birnbaum et al., 2005) have been applied, facilitating the analysis of tissue-specific expression profiles. As a result, a series of genes with tissue-specific expression, transcription factors, and imprinted genes have been systematically identified. Despite such extensive studies, a comprehensive analysis of the identity of every cell in the caryopsis is lacking, limiting our ability to prospectively characterize functions and delineate the differentiation trajectory of cells in rice caryopses. Quantitation of gene expression patterns at the single-cell level is crucial for understanding the development and differentiation process.

Recent advances in single-cell RNA sequencing (scRNA-seq) have revolutionized transcriptomics studies by allowing the detection of transcriptional differentiation at the single-cell level in thousands of cells simultaneously (Grün and van Oudenaarden, 2015). The molecular heterogeneity and developmental progression of tissues are also accurately reflected by scRNA-seq. To reveal dynamic gene expression regulation in plants, scRNA-seq has been successfully applied in protoplasts from diverse cell types, such as roots, shoot apices, leaves, and gametophytes (Shaw et al., 2021). However, this approach is problematic since gene expression noise is created during the protoplasting procedure, and protoplast isolation is difficult or even impossible in certain tissues. To address these difficulties, single-nucleus RNA sequencing (snRNA-seq) has served as an alternative methodology (Habib et al., 2016). Nucleus isolation-based methods have been proven

to be applicable to a wide range of plant materials from *Arabidopsis*, including seedlings, roots, inflorescences, embryos, developing endosperm, and seeds, which greatly extend the application prospects of snRNA-seq (Long et al., 2021; Picard et al., 2021).

To date, snRNA-seq of seeds, endosperm, and embryos has mainly been conducted in the dicot *Arabidopsis*, demonstrating distinct transcriptional regulatory mechanisms across emerging cell types. Spatiotemporally heterogeneous gene imprinting, which can potentially explain heterosis at the seed stage, has also been identified (Picard et al., 2021). The seeds in monocots are very different in anatomy and morphology from those in dicots, such as the number and structure of cotyledons and the size of endosperm (Agarwal et al., 2011). In rice, morphological differentiation of the embryo takes place at 4 days after pollination (DAP), and cellularization of the endosperm completes at 5 DAP. 5 DAP is an ideal stage for underlying cell differentiation in rice caryopses. The endosperm determines the quality and yield of rice production and supplies an important food source for humans. Previous research has focused on the molecular mechanism underlying the regulation of rice endosperm development (Vijayakumar and Gowda, 2012; Nie et al., 2013; Zhou et al., 2013; Kuang et al., 2015; Inukai, 2017). In developing rice endosperm, starch mainly accumulates in the central endosperm cells. Other nutrients, such as protein and lipids, tend to be distributed in the aleurone cells on the peripheral endosperm layers (Zhou et al., 2013). However, the molecular mechanisms underlying cell type differentiation and cell fate transition in rice endosperm remain elusive. Although bulk transcriptomics of seeds from various developing stages have been well documented by microarray or RNA-seq, the transcriptome at the early developmental stage at the single-cell level remains largely unknown. Since the most important cereal crops, such as maize, rice, and wheat, are monocots, it is essential to uncover the molecular basis of monocot seed cell types and their differentiation trajectories during seed development at a finer scale.

Allele-specific expression (ASE) is defined as the characteristic of preferentially expressing a particular allele of two alleles in a hybrid (Gaur et al., 2013). One possible mechanism underlying ASE may be parent-of-origin-specific differential expression, which is caused by physiological effects or the imprinting of parental alleles. Under normal circumstances, hybrids evenly express complete sets of chromosomes per nucleus, combining the genetic information from both parents (Tucci et al., 2019); however, epigenetic gene regulatory process violates this rule, resulting in the differential expression of maternally and paternally inherited alleles (Bartolomei et al., 2020). In other words, because of ASE, any given gene could be either a maternally expressed gene (MEG) or a paternally expressed gene (PEG) depending on the preferentially expressed allele. In flowering plants, endosperm has been identified as the major tissue of ASE, while a few ASE genes have been found in the embryo (Yang et al., 2020; Rodrigues et al., 2021). The primary factor controlling the ASE status of many genes is the DNA methylation asymmetries between parental genomes (Picard et al., 2021). Moreover, considering that rice endosperm is a

tissue with relatively high heterogeneity, nuclei carrying amplified MEGs are likely intermingled with those carrying amplified PEGs. This heterogeneity among endosperm cell/nucleus types may also account for the varied ASE status in rice endosperm.

In addition, ASE has been suggested to play a key role in the differing levels of heterosis between hybrids and parental lines (Shao et al., 2019; Botet and Keurentjes, 2020). Heterosis refers to the superior phenotypes of hybrids relative to their inbred parents. These superior phenotypes are mostly observed with respect to agronomic traits, including the growth rate, reproductive capacity, yield, and resistance. Utilization of heterosis has greatly promoted agricultural production for over a century (Lippman and Zamir, 2007). There are three classic hypotheses for the genetic mechanisms of heterosis: the dominance and overdominance hypotheses, which are based on allelic interactions, and the epistasis hypothesis, which is based on nonallelic interactions. Other hypotheses, such as the active gene effect, gene network system, and genomic dosage effects, have also been suggested to contribute to heterosis (Huang et al., 2016); however, the genetic and molecular mechanisms underlying heterosis remain elusive and controversial. Previous studies have shown that phenotypic novelty in hybrids is regulated by differential gene expression (Shao et al., 2019). ASE has been proposed to play a role in heterosis in many plants, including *Arabidopsis* (Todesco et al., 2010; Ng et al., 2014; Picard et al., 2021), rice (He et al., 2010; Chodavarapu et al., 2012; Shao et al., 2019), and maize (Springer and Stupar, 2007; Paschold et al., 2012; Paschold et al., 2014).

In rice, there is more genetic variation in intersubspecies hybrids than in intrasubspecies hybrids. The heterosis of intersubspecies is usually stronger than that of intrasubspecies (Zhang, 2020). Based on this information, we chose two completely sequenced cultivars, Nipponbare (*japonica*) and 9311 (*indica*), as parents (Goff et al., 2002; Yu et al., 2002). The F_1 hybrid was obtained from a cross between female Nipponbare and male 9311. A previous study showed that the hybrid exhibited more days to heading, increased plant height, narrower plant type, increased number of tillers, slightly shorter main panicle length, decreased number of kernels per main panicle, decreased total kernel weight per main panicle, intermediate 1000 grain weight and decreased total grain yield per plant compared to the inbred parents (Venu et al., 2014). Transcriptomic analysis of hybrids and their parents may provide comprehensive insight into gene expression variations and their correlation with heterosis.

Here, we present snRNA-seq results of over three thousand nuclei from rice caryopses of Nipponbare, 9311, and their F_1 hybrid. A transcriptomics atlas covering most cell types present during the early developmental stage of rice caryopses was constructed, and novel nuclei type-specific marker genes were defined for each cluster. With a focus on rice endosperm, the differentiation trajectory of endosperm subclusters was reconstructed at the single-nuclei level. By profiling ASE in endosperm, 345 genes with ASE (ASEGs) were identified. Further comparisons of differentially expressed genes (DEGs) in each endosperm subcluster among the three samples also demonstrated transcriptional divergence during early endosperm development.

2 Materials and methods

2.1 Plant materials

The *japonica* cultivar Nipponbare was used as the maternal parent, and the *indica* cultivar 9311 was used as the paternal parent to develop the intersubspecies F_1 hybrid. All plants were planted in the fields of the Chinese Academy of Agricultural Sciences (Beijing, China) during the normal growth season (May to September) in 2021. The sowing time of these two varieties was calculated by their heading date. To ensure the coordinated flowering process, each rice variety had three sowing dates, separated by 5 days. Plants were emasculated in the morning and pollinated at noon. The young caryopses were harvested 5 days after pollination (DAP); however, the development rate was different between the hybrid and its parents. Nipponbare had already stored a great deal of starch at 5 DAP, whereas the hybrid had just begun to accumulate starch at that time. A portion of the rice caryopses was used to extract DNA for hybrid identification by DNA markers. DNA extraction was performed according to the CTAB (hexadecyl trimethyl ammonium bromide) protocol. The polymorphic InDel marker RID2-1 (an InDel marker in rice) was used. The amplified products were separated by 1.5% agarose gel electrophoresis (Yang et al., 2020).

2.2 Nuclear extraction from seeds

The rice endosperm at 5 DAP is a syncytium with partially established cellularization. To overcome the difficulties in constructing protoplasts, we extracted nuclei from rice seeds. The fresh rice seeds were homogenized in Honda buffer (2.5% Ficoll^{TM} 400, 5% dextran T40, 0.4 M sucrose, 25 mM Tris-HCl, pH 7.4, 10 mM MgCI$_2$, 10 mM β-mercaptoethanol, and a proteinase inhibitor cocktail) using a mortar and pestle and subsequently passed through a 10-μm (pore-size) filter. Triton^{TM} X-100 was added at a final concentration of 0.5%, and the mixture was incubated on ice for 15 min and centrifuged at 1,500 × g for 5 min at 4°C. The pellet was washed with Honda buffer containing 0.1% Triton^{TM} X-100, gently resuspended in 1 mL Honda buffer, and transferred to a microcentrifuge tube. A 200-μL aliquot of 2.5 M sucrose was placed into a chilled microcentrifuge tube, and 400 μL 60% Percoll$^®$ solution in Honda buffer was carefully overlaid. The nucleus-enriched preparation was slowly released onto the side of the tube above the 60% Percoll$^®$ layer, after which the tube was centrifuged at 1,000 × g for 20 min at 4°C. The 60% Percoll$^®$ layer containing most of the nuclei was carefully removed using a Pasteur pipette, slowly diluted in 1 mL 10× buffer (1% BSA, 0.1% Tween$^®$20, 10 mM Tris-HCl, 10 mM NaCl, and 3 mM MgCl$_2$), and centrifuged at 800 × g for 10 min to pellet the nuclei. The nuclear pellet was resuspended in 10× buffer to achieve a target concentration of 700–1,200 nuclei/μL.

The caryopses were ground and homogenized in buffer to release nuclei. Then, the impurities were filtered out, and only the nuclei were kept in the precipitate. Since the large amount of starch granules in rice caryopses severely influences nuclear isolation, sucrose density gradient centrifugation was applied to refine the nuclei prior to loading into the 10× Genomics Chromium

Controller (Annunziata et al., 2021). The isolated nuclei were mixed with 10× Genomics single nucleus reaction reagents, which included gel bead barcodes. Subsequently, snRNA-seq libraries were generated and sequenced on the high-throughput Illumina platform. Sequencing data were prefiltered at both the cell and gene levels.

2.3 snRNA sequencing

The nuclear suspensions were loaded onto a 10X Genomics GemCode Single-cell instrument that generates single-cell Gel Bead-In-Emulsions (GEMs). Libraries were generated and sequenced from the cDNAs with Chromium Next GEM Single Cell 3' Reagent Kits v3.1. Silane magnetic beads were used to remove leftover biochemical reagents and primers from the post GEM reaction mixture. Full-length, barcoded cDNAs were then amplified by PCR to generate sufficient mass for library construction. R1 (read 1 primer sequence) was added to the molecules during GEM incubation. P5, P7, a sample index, and R2 (read 2 primer sequence) were added during library construction via end repair, A-tailing, adaptor ligation, and PCR. The final libraries contained the P5 and P7 primers used in Illumina bridge amplification. The Single Cell 3' Protocol was followed to produce Illumina-ready sequencing libraries. A Single Cell 3' Library comprised standard Illumina paired-end constructs that begin and end with P5 and P7. The Single Cell 3' 16 bp 10x Barcode and 10 bp UMI were encoded in Read 1, while Read 2 was used to sequence the cDNA fragment. Sample index sequences were incorporated as the i7 index read. Read 1 and Read 2 are standard Illumina® sequencing primer sites used in paired-end sequencing on the Illumina HiSeq2000 platform.

2.4 snRNA-seq data processing

snRNA-seq data were aligned to the Nipponbare reference genome (IRGSP-1.0) and counted using Cell Ranger pipelines (version 6.0.1) (Zheng et al., 2017). The raw count matrix data were imported into R using Seurat (version 4.1.1) (Hao et al., 2021) for further analysis. Cells with fewer than 3000 genes and genes detected in fewer than 5 cells were removed. Each individual dataset was scaled and normalized using the ScaleData and NormalizeData functions. The top 2000 highly variable genes were used for principal component analysis (PCA) dimensionality reduction. The first 15 principal components (PCs) were selected according to the PCA elbow plot and used for clustering with a resolution parameter of 0.4. The clusters were visualized and explored by t-SNE and UMAP. Marker genes (cluster-enriched genes) were identified using the function FindAllMarkers. Differential expression analysis was performed based on the Wilcoxon rank-sum test. The cluster-enriched genes were determined using the following parameters: greater than a 1.5-fold difference (logfc.threshold = 0.58) between the two groups of cells and test genes with a minimum fraction of at least 0.1. The cluster-specific

marker genes were determined using the following parameters: a log2-fold change greater than 0.25 and a proportion of marker genes expressed in cells among all other clusters (PCT2) of less than 10%. Spearman's correlation coefficients among the seven identified rice seed clusters and tissue type-specific transcriptomes from two previously published bulk RNA-seq datasets were calculated using the rcorr function in the Hmisc R package.

2.5 Integrative analysis of the Nipponbare and hybrid snRNA-seq datasets

Normalized snRNA-seq data from Nipponbare and the hybrid were combined into one dataset. Variable genes were selected using the SCtransform function with default parameters. The FindintegrationAnchors function command was used with default parameters to discover integration anchors across all samples. The IntegrateData function was run on the anchor set with default additional arguments. ScaleData and RunPCA were performed on the integrated assay to compute 12 PCs. UMAP dimensionality reduction was carried out, and a shared nearest neighbor (SNN) graph was constructed using dimensions 1:12 as input features and default PCA reduction. Clustering analysis was performed on the integrated assay at a resolution of 1.

2.6 Pseudotime analysis

Pseudotime analysis of rice endosperm differentiation and the determination of cell fate were performed using the Monocle2 (version 2.24.0) R package (Qiu et al., 2017). Dimensionality reduction was performed using the reduceDimension function (reduction_method = 'DDRTress'), and the cells along the trajectory were ordered using the orderCells function.

2.7 Gene ontology analysis

Gene Ontology (GO) is an international standardized gene functional classification system that offers a dynamically updated and controlled vocabulary and a strictly defined concept to comprehensively describe the properties of genes and their products in any organism. GO has three ontologies: molecular function, cellular component, and biological process. The basic unit of GO is the GO term. Each GO term belongs to a type of ontology (Ashburner et al., 2000).

GO enrichment analysis provides all GO terms that are significantly enriched in DEGs compared to the genome background and filters the DEGs that correspond to biological functions. First, all peak-related genes were mapped to GO terms in agriGO v.2.0 with default settings (Tian et al., 2017). Gene numbers were calculated for every term, and significantly enriched GO terms in DEGs compared to the genome background were defined by a hypergeometric test. The formula for calculating the P value is:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where N is the number of all genes with GO annotation; n is the number of DEGs in N; M is the number of all genes that are annotated to certain GO terms; and m is the number of DEGs in M. The calculated p values were determined through FDR correction, taking FDR ≤ 0.05 as a threshold. GO terms meeting this condition were defined as significantly enriched GO terms in DEGs. This analysis allowed us to recognize the main biological functions that DEGs perform.

2.8 ASE gene identification

Expressed genes with 1,088,928 SNPs between Nipponbare and 9311, previously identified by RiceVarMap2 (Zhao et al., 2021), were used to calculate the maternal, paternal, and total gene expression levels in each rice endosperm nucleus. The allele-specific read count matrixes for rice endosperm were constructed based on these expressed genes. ASE was then assessed using an existing method that accounts for maternal and paternal dosage in endosperm for each gene (single_cell_ASE_analysis.R) (Picard et al., 2021).

2.9 Differentially expressed gene identification and classification

DEGs among Nipponbare, 9311, and their F_1 hybrid were identified using the edgeR package (Robinson et al., 2010). A fold change greater than 1.5 was used as the filtering criterion. DEGs were classified into five major expression patterns based on their expression levels. Moreover, each gene was classified based on its expression level in the F_1 hybrid versus that in both parental lines. The five patterns were OHP (over-higher parents, $F_1 > LP \geq 3HP$), BLP (below-lower parents, $F_1 < LP/3 \leq HP$), OMP (over-mid parents, $HP < F_1 < 3HP$), BMP (below-mid parents, $F_1 > LP/3$ & $F_1 < LP$), and MPV (similar to mid parents, $F_1 > LP$ & $F_1 < 3HP$). Additive, dominant, and overdominant expressed genes were identified by comparing their expression levels in the hybrid to those in the parents (Li et al., 2016; Liu et al., 2021).

3 Results

3.1 snRNA-seq and identification of nucleus types in rice caryopses

The rice caryopsis, comprised of the embryo, endosperm, pericarp, and seed coat possessing different ploidy levels, is technically challenging to separate, and profiling of the transcriptome of each tissue type is difficult. To profile transcriptional complexity during the early development of rice caryopses, caryopses at 5 DAP were collected from the *japonica* cultivar Nipponbare, the *indica* cultivar 9311, and their intrasubspecies hybrid obtained from a cross between female Nipponbare and male 9311. A single-nucleus RNA sequencing (snRNA-seq) method was employed to generate a transcriptional atlas of rice caryopses. In total, 3,617 individual nuclei were profiled from these three rice samples, and 20,156, 27,213, and 21,817 expressed genes were detected in Nipponbare, 9311, and their F_1 hybrid, respectively, using the IRGSP1.0 reference genome (Figure 1A; Table S1) (Kawahara et al., 2013).

The anatomical structures of the three rice accessions could all be roughly divided into the same three parts (embryo, endosperm and maternal tissues). To this end, snRNA-seq data of Nipponbare, 9311, and their hybrid were normalized to correct for experimental variation. Then, for each nucleus, the genes with the strongest biological signal relative to the technical noise were separately identified based on their expression levels. To reduce the dimensionality of the strongest gene expression features, principal component analysis (PCA) and graph-based clustering analysis [Uniform Manifold Approximation and Projection (UMAP) and *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE)] were performed on the qualifying 3,345 single-nucleus transcriptomes using Seurat (Version 4.1.0) (Figures 1B, S1). Clusters of nuclei were identified by a shared nearest neighbor modularity optimization-based clustering algorithm. Nine distinct clusters, each containing between 52 and 1,689 nuclei, were identified in the caryopsis.

We noted that three meta-clusters exhibited relatively distinct transcriptomics profiles, and the nuclei clusters in each meta-cluster were grouped on the UMAP and *t*-SNE plots (Figures 1B, S1). A series of enriched and specific genes for each cluster was identified (Table S2). To attribute nuclear identities to these clusters, the expression levels of marker genes whose patterns have been well studied in the endosperm, embryo, pericarp, and seed coat were evaluated for each cluster (Figure 1C). The average expression levels of previously reported tissue-specific genes helped us to identify nuclei clusters (Zhou et al., 2013; Wu et al., 2020; Ram et al., 2021) (Table S3). The results suggested that clusters 2 and 7 likely represented the pericarp and seed coat, clusters 3 and 5 could be represented the embryo, and the remaining clusters (including clusters 0, 1, 4, 6, and 8) were likely to be the endosperm. The representative cluster marker genes included *OsLEA3-2* (Duan and Cai, 2012), *OsVPE3* (Lu et al., 2016), and *OsLEA3* (Xiao et al., 2007), which are known to be expressed in embryos; *OsFd1* (He et al., 2020) and *OsMADS29* (Yin and Xue, 2012), which are predominantly expressed in maternal tissues (seed coat and pericarp); and *OsAPS2* (Akihiro et al., 2005), *OsNAC23* (Li et al., 2022), and *OsSWEET15* (Yang et al., 2018), which are important for starch synthesis in the endosperm. These known marker genes were highly and specifically enriched in their corresponding clusters (Figure 1D).

Moreover, Gene Ontology (GO) enrichment analysis of the marker genes in each meta-cluster was performed to annotate their putative functions (Table S4). For example, genes related to photosynthesis, reproduction, and response to chemical stimuli were enriched in the pericarp like meta-cluster (clusters 2 and 7); genes related to carbohydrate metabolism, pyrophosphatase
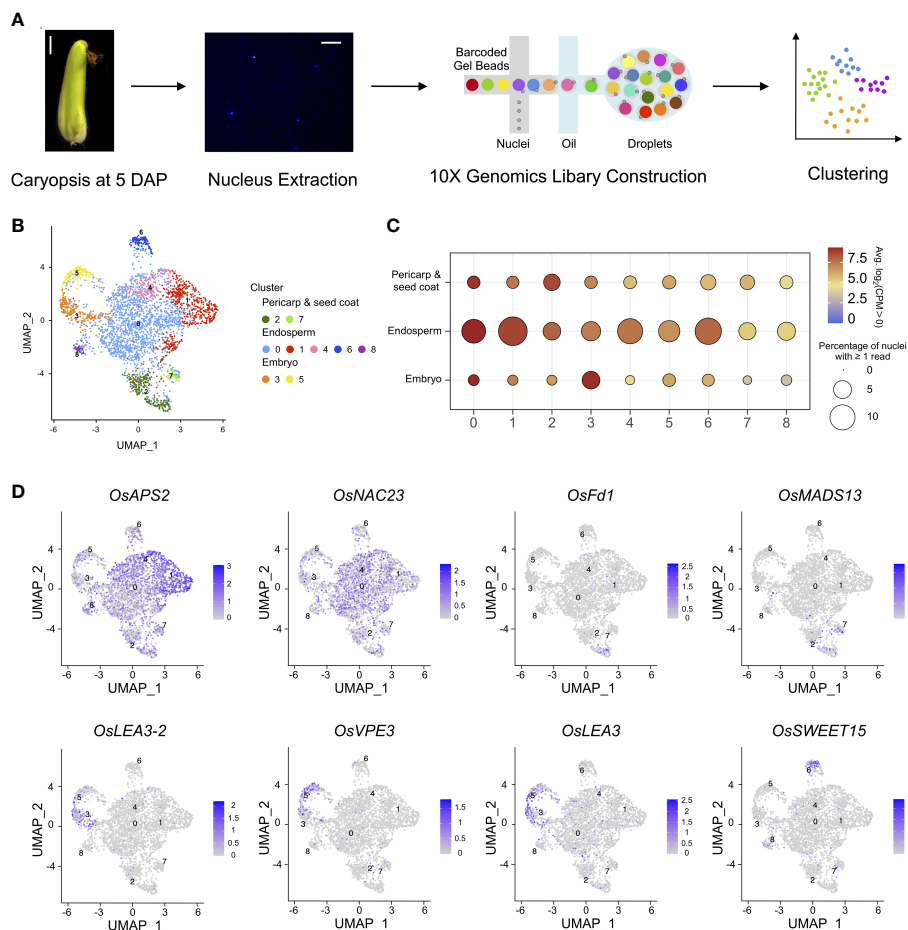
FIGURE 1
Nuclear heterogeneity in the early developmental stage of rice caryopses. **(A)** Overview of the experimental design. The first panel shows a microscopic image of a Nipponbare caryopsis at 5 DAP (scale bar = 2 mm), and the second panel shows an image of rice nuclei (scale bar = 25 μm). **(B)** Uniform manifold approximation and projection (UMAP) plot of unbiased clustering of the 3,617 nuclei in rice caryopsis. **(C)** Average expression levels of marker genes in the nine identified nuclei clusters in the pericarp and seed coat, endosperm, and embryo. **(D)** UMAP visualization of meta-cluster-specific genes in rice caryopses. The colors represent the normalized gene expression levels, log_{10} (CPM+1), on the UMAP graph.

activity, macromolecule biosynthesis, and protein modification were enriched in the endosperm like meta-cluster (clusters 0, 1, 4, 6, and 8); and genes related to transcription factor activity and RNA biosynthesis were enriched in the embryo like meta-cluster (clusters 3 and 5). GO analysis of each subcluster further demonstrated more detailed functional categories for each nucleus type. In the pericarp like cluster, cluster 2 included specific gene signatures for photosynthesis and signal transmission, while genes highly expressed in cluster 7 are involved in transcription and response to stress. Similarly, for endosperm like clusters, genes involved in lipid metabolism were enriched in clusters 6 and 8, those related to glucan biosynthesis and metabolism and carbohydrate biosynthesis were overrepresented in cluster 1, and those related to nutrient reservoir activity were enriched in cluster 4.

The nucleus type in each cluster was further validated by correlation coefficiency analysis between the cluster-enriched genes and published seed cell-type expression profiles (Figure S2) (Wu et al., 2020; Ram et al., 2021). Furthermore, conserved and divergently expressed genes for each homologous cell type were compared among the three rice samples. Venn diagram analysis demonstrated that

greater than 70% of hybrid-expressed genes were also expressed in Nipponbare or 9311 in all three meta-clusters. For hybrid-specific genes, the highest percentage was found in the endosperm, which is in accordance with the developmental process of endosperm being the most divergent, while other cell types were more similar between the hybrid and its parents (Figure S3).

Taken together, these results show that the nine nuclei clusters yielded two clusters for the embryo like, two clusters for the seed coat and pericarp like, and five clusters for the endosperm like, and reveal a high degree of nuclear transcriptional heterogeneity in rice caryopses.

3.2 Construction of an endosperm developmental trajectory

To obtain a better understanding of the regulation of rice endosperm development, the corresponding endosperm cells from different stages, including the early stage, differentiated stage, and terminal stage, should be identified. Taking advantage of our

snRNA-seq data from rice caryopses, the endosperm nuclear population, consisting of the aforementioned clusters 0, 1, 4, 6, and 8 identified by Seurat, was reclustered. The second subnuclear clustering revealed 8 subclusters from the endosperm, named E0–E7 (Figure 2A). The nuclear identities of these endosperm subclusters were determined by examining the enriched expression of marker genes (Table S5). The specific enrichment of glucose and starch genes in endosperm subclusters E2, E4, and E5 and translation- and immunity-related genes in endosperm

subcluster E3 showed that these two groups of subclusters are hotspot regions for starch and protein biosynthesis, revealing their differentiation in cell fate and storage product preservation during the developmental trajectory. For the remaining endosperm subclusters, genes playing a role in the stress response were enriched in E1, those involved in transducer activity and signal transduction were enriched in E6, and those associated with ribonucleotide metabolism were enriched in E7, indicating potential functions in development and regulation rather than only in storage (Table S5).
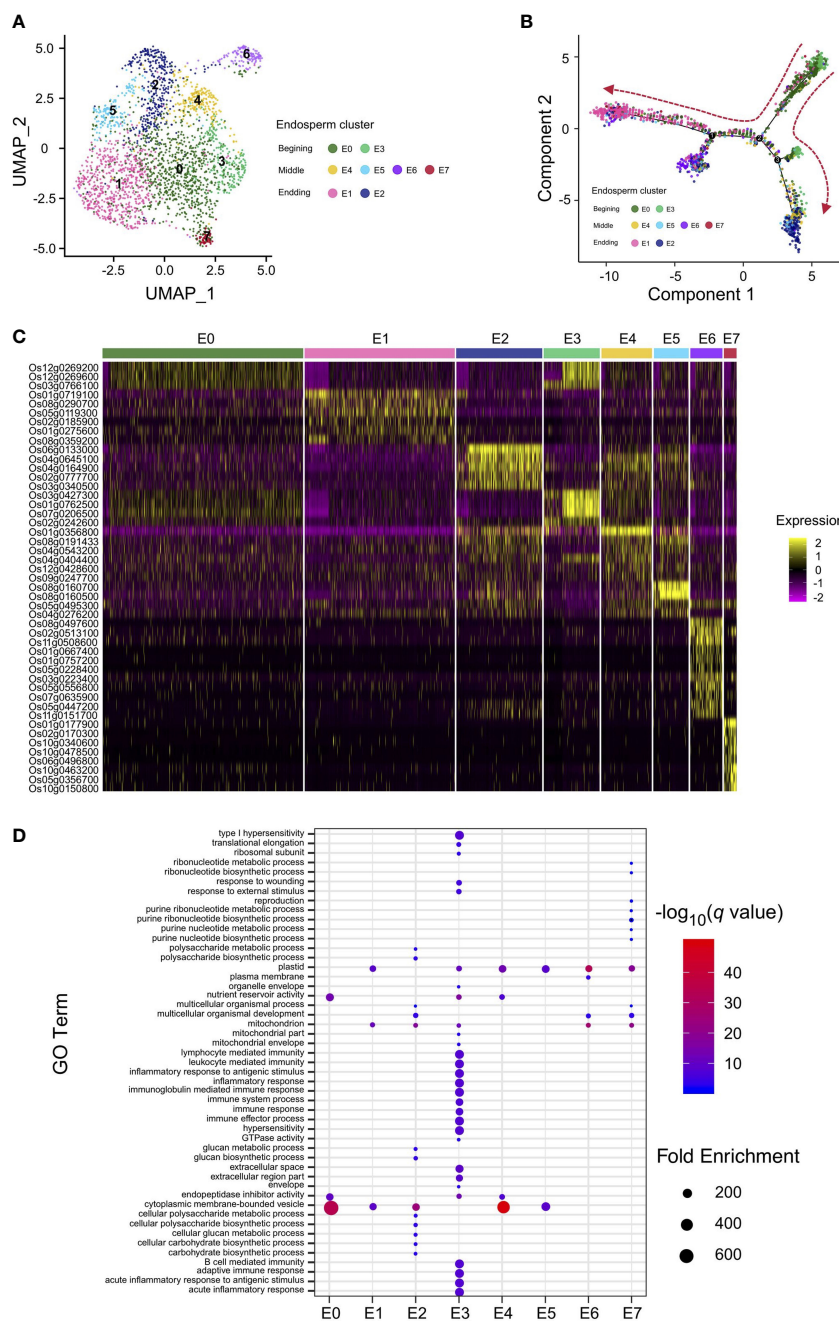

FIGURE 2

Nucleus type identification and differentiation trajectory reconstruction in rice endosperm. **(A)** UMAP plot of unbiased clustering of the 2,600 rice endosperm nuclei. **(B)** Pseudotime analysis using endosperm nuclei to show the developmental trajectory of rice endosperm. Colored dots indicate nuclei from clusters in **(A)**. **(C)** Heatmap of the scaled expression of markers for each endosperm cluster in all individual nuclei. Nuclei are arranged by cluster. **(D)** Enriched Gene Ontology (GO) terms identified for each cluster according to the cluster-specific genes in each cluster.

Interestingly, UMAP analysis revealed that the majority of nuclei from the hybrid endosperm were present in cluster E1, while the proportions of nuclei in clusters E0 and E4 were obviously lower than those from the Nipponbare and 9311 endosperms. The remaining five subclusters were proportionally distributed across all three rice accessions (Figure S4). To further clarify the differentiation process in rice endosperm, a developmental trajectory over pseudotime was constructed (Figures 2B, S5). The expression profiles of the unique marker genes for the eight endosperm subclusters were also compared (Figure 2C). At the beginning of the developmental trajectories, nuclei from endosperm subclusters E0 and E3 were specifically distributed, and their expression profiles were enriched in translation, enzyme regulation and endopeptidase inhibitor activity; therefore, they were classified as nuclei where active transcription occurs at the very early stages of endosperm development. Concerning the branch on the right side of the differentiated map, early-stage endosperm subclusters E0 and E3 gradually transitioned into E4 and E2 at the end of the branch, displaying specific expression patterns of glucose and starch synthesis marker genes; therefore, subclusters E2 and E4 were classified as starchy endosperm. Nuclei from subcluster E1, E6 and E7 were mainly grouped at the left of the branch, with marker gene signatures enriched in nitrogen compound metabolic process (Figure 2D; Tables S5, S6), suggesting an active role in the nitrogen compound metabolism. Besides, endosperm transfer cell-containing region-specific gene *AL2* was specifically expressed in E7 (Kuwano et al., 2011). As the aleurone layer adjacent to the vascular bundle would further develop into transfer cells (Thompson et al., 2001), the subcluster E1 were likely to be classified as aleurone layer and E7 to be classified as transfer cells.

3.3 ASE of genes from Nipponbare × 9311 in endosperm subclusters

Analysis of expression patterns in the endosperm of hybrids showed that among genes with ASE, the vast majority overexpressed the maternal allele. Maternal traits in plants are predominantly related to seed formation due to the maternal genotype of the endosperm (Botet and Keurentjes, 2020). Therefore, we investigated ASE of genes in endosperm nuclei from the hybrid. We took advantage of the allele-specific nature of our snRNA-seq data to examine ASE across the defined endosperm nuclear subclusters. To achieve reliable allele-specific estimation with low per-cell coverage data, only the 1,088,928 previously reported SNPs between Nipponbare and 9311 from the RiceVarMap2 website were used as a reference for ASE calling. The allele from Nipponbare was deemed the maternal allele, and the allele from 9311 was deemed the paternal allele.

A method for the assessment of ASE that accounts for maternal and paternal dosage in endosperm was employed. Of the 11,426 genes expressed in hybrid endosperm nuclei, ASE was assessed for 2,056 genes. Maternal bias was detected for 189 genes (MEGs), and paternal bias was detected for 156 genes (PEGs). MEGs and PEGS were defined as having strong, intermediate, or weak potential based on the extent of parental bias (Figure 3). For MEGs, GO categories were enriched in translation, protein metabolism, protein modification, carbohydrate metabolism, and biosynthesis regulation, while PEGs were enriched in regulatory processes such as the regulation of transcription and RNA metabolism (Table S8), which is in accordance with previously reported GO enrichment preferences in rice, maize, and *Arabidopsis* from studies using bulk RNA-seq data (Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2011; Chen et al., 2018).

To determine whether the ASEGs were preferentially expressed in specific nucleus types within rice seeds, total and allelic expression patterns were examined across the six newly identified endosperm clusters, two pericarp and seed coat clusters, and one embryo cluster. The heatmap of the total expression of MEGs and PEGs demonstrated that different subclusters of ASEGs were enriched in distinct endosperm nucleus types, with only a few exceptions of common highly expressed genes in several subclusters (Figure 3). The highest expression level for both MEGs and PEGs was observed in endosperm subclusters E1 and E2, deemed endosperm nuclei from the late developmental stage. In agreement with this result, endosperm subclusters E0 and E3 showed the lowest expression levels of ASEGs at the beginning of the pseudotime trajectory. Moreover, allele-specific expression values for ASEGs revealed that the enriched expression of MEGs in the specific endosperm clusters was stimulated by increased expression of the maternal allele, whereas expression of the paternal allele remained low and largely unchanged across all endosperm clusters. Expression patterns were the same for PEGs. Essentially, the ASE level is also heterogeneous among endosperm cell/nucleus types.

3.4 Gene expression differences in rice endosperm clusters among the hybrid and its parents

To explore the differences in gene expression during the early development of rice endosperm at a finer scale, the transcriptome was profiled across the eight endosperm subclusters, and the gene expression levels from the pairwise comparisons among Nipponbare, 9311, and their hybrid were evaluated. A total of 1,468 genes were identified as differentially expressed in at least one pairwise comparison when treating the eight endosperm clusters as a whole, and 1,148–2,503 DEGs were separately detected in single clusters (from E0 to E7, except for E4, owing to only one nucleus being defined as E4 in the hybrid). The late-stage endosperm nuclei clusters E1 and E2 tended to contain more DEGs than the early-stage clusters E0 and E3 (Figure 4A). During the process of endosperm development, the determination of nucleus type was accompanied by the identification of a larger set of DEGs. DEGs were classified into five major types based on expression patterns in the hybrid and its parental lines, namely, below-lower parents (BLP), below-mid parents (BMP), similar to mid parents (MPV), over-mid parents (OMP), and over-higher parents (OHP), from low to high relative transcript levels. The results revealed that the expression levels in the hybrid were restricted within the scope of its
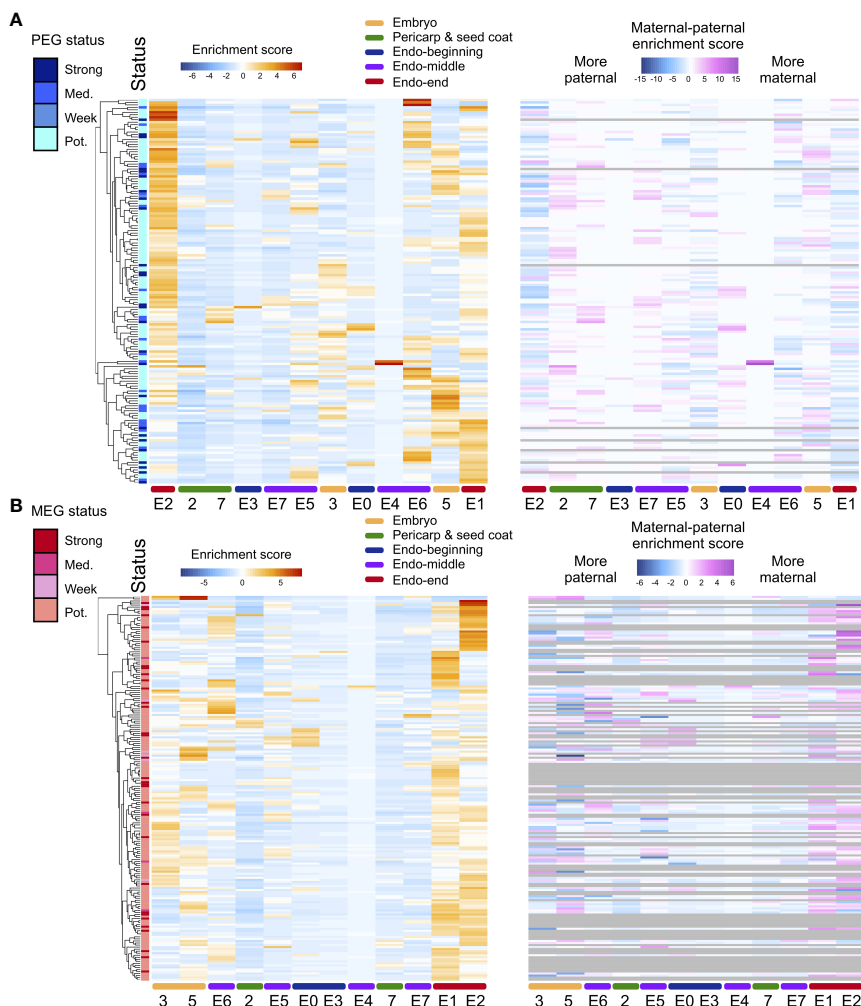
FIGURE 3

ASE heterogeneity in rice seeds. **(A)** Heatmap of the total expression enrichment score and the difference between the allele-specific maternal and paternal expression enrichment scores for all PEGs in two pericarp and seed coat clusters, one embryo cluster, and six endosperm clusters in the hybrid. **(B)** Heatmap of the total expression ES and ES (maternal) to ES (paternal) for all MEGs. Numbers and characters on the bottom of the figure represent the embryo and pericarp and seed coat, nuclei clusters and endosperm subclusters. Numbers (2, 3, 5 and 7) represent the rice caryopsis clusters except for the endosperm clusters. E0-E7 represent the endosperm subclusters from 0 to 7.
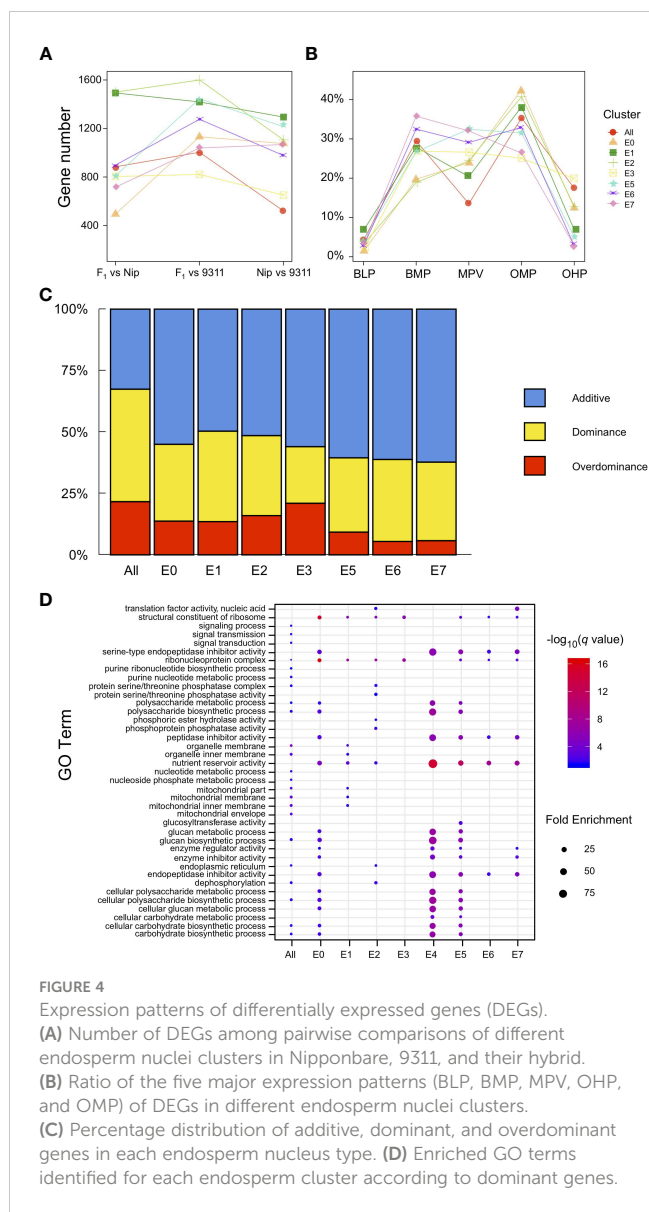
parents for most of the DEGs. Nevertheless, there were a number of DEGs in the hybrid, ranging from 1.3% to 19.9%, that significantly differed from the lower or higher parent (Figure 4B). Of the 1,148–2,503 DEGs with SNPs between Nipponbare and 9311, 3.26–4.25% showed preferential allelic expression of only one maternal or paternal allele.

According to transcriptional changes identified in the hybrid (compared with its parents), DEGs were further classified into additive, dominant, and overdominant. The additive genes were established from early-stage endosperm nuclei E0 and E3 with a high level (55.0–55.9%), reaching a peak in the intermediate stage (60.5–62.2%). Notably, the dominance effects increased with endosperm development, while the overdominance effects varied among early nucleus types and reconciled gradually (Figure 4C). GO analysis of the dominant genes in different endosperm clusters demonstrated enrichment in the biosynthesis and metabolism of a variety of substances, such as glucan, nucleotides, carbohydrates, and polysaccharides, and these genes also play a role in nutrient reservoir

activity (Figure 4D). In contrast, overdominant genes tended to be involved in gene regulation or interactions with other genes (Table S9). For example, *RPBF* acts as a transcriptional activator in the regulation of free lysine content and protein accumulation in rice grain (Paponov et al., 2009), and *FLO10* functions by affecting the trans-splicing of mitochondrial *nad1* intron 1, in addition to playing an important role in the maintenance of mitochondrial function and endosperm development (Wu et al., 2019). These results indicate that during rice endosperm development, dominance effects play an important role in substance accumulation, while overdominance effects orchestrate the varied endosperm clusters as a whole.

4 Discussion

Rice caryopses display relatively high heterogeneity, with obvious differences in tissue type, ploidy, and source of the genetic material within cells (Itoh et al., 2005). Previous studies

Expression patterns of differentially expressed genes (DEGs).
(A) Number of DEGs among pairwise comparisons of different
endosperm nuclei clusters in Nipponbare, 9311, and their hybrid.
(B) Ratio of the five major expression patterns (BLP, BMP, MPV, OHP,
and OMP) of DEGs in different endosperm nuclei clusters.
(C) Percentage distribution of additive, dominant, and overdominant
genes in each endosperm nucleus type. (D) Enriched GO terms
identified for each endosperm cluster according to dominant genes.

captured during the early stage of rice caryopsis development and
assigned to their respective tissue types, including maternal tissue
(such as the pericarp and seed coat) and filial tissue (such as the
embryo and endosperm), based on well-known marker genes from
the literature and bulk RNA-seq results from bioinformatics
correlation analysis.

Rice endosperm is one of the most important sources of human
food. To date, many molecular studies have identified the important
genes required for endosperm development, including the syncytium-
cellularization transition, endosperm cell differentiation and storage
compound accumulation (Zhang et al., 2021; Liu et al., 2022). In rice
interspecies hybrids, the rate of syncytial nuclear division in the
endosperm is different from that of their parent lines (Ishikawa et al.,
2011). A similar phenomenon was found in our intersubspecies
hybrids. Cellularization was not completed in the F_1 caryopses at 5
DAP. Nuclei extraction rather than protoplast isolation methods were
more appropriate to construct scRNA-seq libraries (Wu et al., 2016;
Picard et al., 2021). Here, the developmental trajectory of rice
endosperm over pseudotime was successfully reconstructed
(Figure 2). Interestingly, genes required for sugar transportation in
grain filling (*OsSWEET11* and *OsSWEET15*) were preferentially
expressed in the left branch subclusters (E6 and E7) of the
endosperm developmental trajectory (Figure 2B; Table S5). The
highest mRNA levels were detected in regions of the nucellus and
the aleurone in a previous study. Considering that the left branch
subclusters (E2 and E4) specifically expressed glucose and starch
synthesis-related genes, we hypothesized that our snRNA-seq
analysis reproduced endosperm cell differentiation in storage product
accumulation. In addition, compared with its parents Nipponbare and
9311, a greater number of endosperm nuclei were differentiated to E1
than to E2 in the F_1 hybrid, and these clusters were concentrated on
one of the branches of the developmental trajectory. We propose that
the incompatibility caused by interspecies hybridization may influence
endosperm development in such a way that it reduces normal nutrient
storage. Accordingly, the percentage of nuclei in substance biosynthesis
and metabolism-related clusters (E2, E3, and E4) was much lower in
the hybrid than in Nipponbare and 9311, and these clusters may
critically influence grain filling, cause inadequate energy supply for seed
germination, and ultimately lead to a decrease in seed fertility.

Using the intersubspecies hybrid seeds of two rice cultivars
(Nipponbare and 9311), we identified 345 genes as ASEGs in the
endosperm, displaying deviation in gene expression levels from the
predicted 2:1 maternal-to-paternal allele expression ratio (Figure 3).
Previous research suggests that the functions of imprinted genes,
which are important components of ASEGs, are enriched in the
synthesis and transport of nutrients. A bulk RNA-seq study in rice
endosperm revealed that most of these genes play a role in
carbohydrate biosynthesis and metabolism, nitrogen metabolism,
and transmembrane transport (Yang et al., 2020). Similar findings
were generated in the present study. For MEGs, GO categories were
enriched in protein metabolism, protein modification, carbohydrate
metabolism, and the regulation of biosynthesis, whereas PEGs were
enriched in regulatory processes such as the regulation of
transcription and RNA metabolism. These results indicate that

have devoted great effort to physically dissecting rice grains using
microdissection approaches. Nevertheless, due to significant cell
heterogeneity, conventional bulk RNA-seq analysis causes a certain
degree of bias by averaging differential gene expression levels
among cell populations or overlooking rare cell types (Xue et al.,
2012; Shen et al., 2018; Wu et al., 2020; Ram et al., 2021). Owing to
the advantages of single-cell techniques, snRNA-seq has been
widely applied to study the diversity of transcriptionally distinct
cell/nucleus types in many plant tissues, such as developing seeds in
Arabidopsis thaliana (Picard et al., 2021). In contrast to *Arabidopsis*,
knowledge of rice caryopsis development at a fine scale remains
limited. We performed the snRNA-seq in rice seeds for the first
time. In our study, a transcriptomics atlas of caryopses from the
japonica cultivar Nipponbare, the *indica* cultivar 9311, and an F_1
intersubspecies hybrid at 5 DAP was reconstructed at single-
nucleus resolution (Figure 1). A series of nuclear states were

ASEGs play an important role in rice endosperm energy metabolism and seed development. Heterogeneity was also found in the expression levels of ASEGs in the endosperm. The highest expression of ASEGs was detected in early developmental stage endosperm, while the lowest expression was detected in the highly differentiated starchy endosperm. These results suggest that the conflict between maternal and paternal alleles is gradually reconciled during seed development. Future cell/nucleus type-specific studies will extend the knowledge of the regulatory mechanism underlying imprinting.

The snRNA-seq data of rice caryopses are likely to include some errors, especially in the discrimination of closely related cell types. Matching the messenger RNA profile of a nucleus with its exact position within a caryopsis is still impossible in this way. Spatial transcriptomics technology can be applied to characterize gene expression in barcoded regions of individual tissue sections (Rao et al., 2021). For plants, spatial transcriptomics has been applied in a limited number of studies. There will be broad application prospects for the *in situ* imaging of mRNAs in future research to directly assay gene expression while preserving positional information. Another major limitation of our work is the lack of evidence from biological experiments. The existing expression data are insufficient to support a causal relationship between gene expression and biological phenomena and are not adequate to resolve the molecular mechanisms underlying heterosis in hybrid rice. Further experiments in which gene functions are validated and phenotypic data of yield-related traits are acquired would improve this research.

In summary, we generated a gene expression map of rice caryopses at single-nucleus resolution. Based on snRNA-seq data, nuclei types from different tissues were divided. We focused on the development, ASE and DEG in endosperm, and these findings would help to utilize inter-subspecies hybridization. The identification of cells in distinct tissue types of the rice caryopsis based on unique transcriptomes now offers researchers the opportunity to investigate, at unprecedented resolution, how the caryopsis develops, how endosperms differentiate, and how ASE affects transcription in hybrid plants.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

HH carried out project design and material preparation and conceived data interpretation. HZ contributed to the experiments, data analyses, and manuscript writing. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

We thank the Computing Platform of the CLS (Peking University) for provisioning computing resources. We thank Di Cui for planting and crossing rice and Xue Han for instructing nuclear extraction. No conflicts of interest are declared.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1171474/full#supplementary-material

References

Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* 33, 189–202. doi: 10.1002/bies.201000107

Akihiro, T., Mizuno, K., and Fujimura, T. (2005). Gene expression of ADP-glucose pyrophosphorylase and starch contents in rice cultured cells are cooperatively regulated by sucrose and ABA. *Plant Cell Physiol.* 46, 937–946. doi: 10.1093/pcp/pci101

Annunziata, R., Balestra, C., Marotta, P., Ruggiero, A., Manfellotto, F., Benvenuto, G., et al. (2021). An optimised method for intact nuclei isolation from diatoms. *Sci. Rep.* 11, 1681. doi: 10.1038/s41598-021-81238-z

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Gene Ontology Consortium. Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bartolomei, M. S., Oakey, R. J., and Wutz, A. (2020). Genomic imprinting: an epigenetic regulatory system. *PLoS Genet.* 16, e1008970. doi: 10.1371/journal.pgen.1008970

Birnbaum, K., Jung, J. W., Wang, J. Y., Lambert, G. M., Hirst, J. A., Galbraith, D. W., et al. (2005). Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat. Methods* 2, 615–619. doi: 10.1038/nmeth0805-615

Botet, R., and Keurentjes, J. J. B. (2020). The role of transcriptional regulation in hybrid vigor. *Front. Plant Sci.* 11, 410. doi: 10.3389/fpls.2020.00410

Chen, C., Li, T., Zhu, S., Liu, Z., Shi, Z., Zheng, X., et al. (2018). Characterization of imprinted genes in rice reveals conservation of regulation and imprinting with other plant species. *Plant Physiol.* 177, 1754–1771. doi: 10.1104/pp.17.01621

Chodavarapu, R. K., Feng, S., Ding, B., Simon, S. A., Lopez, D., Jia, Y., et al. (2012). Transcriptome and methylome interactions in rice hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12040–12045. doi: 10.1073/pnas.1209297109

Duan, J., and Cai, W. (2012). *OsLEA3-2*, an abiotic stress induced gene of rice plays a key role in salt and drought tolerance. *PLoS One* 7, e45117. doi: 10.1371/journal.pone.0045117

Gaur, U., Li, K., Mei, S., and Liu, G. (2013). Research progress in allele-specific expression and its regulatory mechanisms. *J. Appl. Genet.* 54, 271–283. doi: 10.1007/s13353-013-0148-y

Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS One* 6, e23687. doi: 10.1371/journal.pone.0023687

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* l. ssp. *japonica*). *Science* 296, 92–100. doi: 10.1126/science.1068275

Grün, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. *Cell* 163, 799–810. doi: 10.1016/j.cell.2015.10.039

Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J. J., et al. (2016). Div-seq: single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. *Science* 353, 925–928. doi: 10.1126/science.aad7038

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell* 184, 3573–3587.e3529. doi: 10.1016/j.cell.2021.04.048

He, L., Li, M., Qiu, Z., Chen, D., Zhang, G., Wang, X., et al. (2020). Primary leaf-type ferredoxin 1 participates in photosynthetic electron transport and carbon assimilation in rice. *Plant J.* 104, 44–58. doi: 10.1111/tpj.14904

He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L., et al. (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* 22, 17–33. doi: 10.1105/tpc.109.072041

Hsieh, T. F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., et al. (2011). Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1755–1762. doi: 10.1073/pnas.1019273108

Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., et al. (2016). Genomic architecture of heterosis for yield traits in rice. *Nature* 537, 629–633. doi: 10.1038/nature19760

Ingram, G. C. (2017). Dying to live: cell elimination as a developmental strategy in angiosperm seeds. *J. Exp. Bot.* 68, 785–796. doi: 10.1093/jxb/erw364

Inukai, T. (2017). Differential regulation of starch-synthetic gene expression in endosperm between *Indica* and *Japonica* rice cultivars. *Rice (N Y)* 10, 7. doi: 10.1186/s12284-017-0146-5

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., and Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* 65, 798–806. doi: 10.1111/j.1365-313X.2010.04466.x

Itoh, J., Nonomura, K., Ikeda, K., Yamaki, S., Inukai, Y., Yamagishi, H., et al. (2005). Rice plant development: from zygote to spikelet. *Plant Cell Physiol.* 46, 23–47. doi: 10.1093/pcp/pci501

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* 6, 4. doi: 10.1186/1939-8433-6-4

Kuang, Q., Yu, X., Peng, X., and Sun, M. X. (2015). The isolation of early nuclear endosperm of *Oryza sativa* to facilitate gene expression analysis and screening imprinted genes. *Plant Methods* 11, 49. doi: 10.1186/s13007-015-0092-4

Kuwano, M., Masumura, T., and Yoshida, K. T. (2011). A novel endosperm transfer cell-containing region-specific gene and its promoter in rice. *Plant Mol. Biol.* 76, 47–56. doi: 10.1007/s11103-011-9765-1

Lafon-Placette, C., and Köhler, C. (2014). Embryo and endosperm, partners in seed development. *Curr. Opin. Plant Biol.* 17, 64–69. doi: 10.1016/j.pbi.2013.11.008

Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6026–e6035. doi: 10.1073/pnas.1610115113

Li, Z., Wei, X., Tong, X., Zhao, J., Liu, X., Wang, H., et al. (2022). The OsNAC23-Tre6P-SnRK1a feed-forward loop regulates sugar homeostasis and grain yield in rice. *Mol. Plant* 15, 706–722. doi: 10.1016/j.molp.2022.01.016

Lippman, Z. B., and Zamir, D. (2007). Heterosis: revisiting the magic. *Trends Genet.* 23, 60–66. doi: 10.1016/j.tig.2006.12.006

Liu, W., He, G., and Deng, X. W. (2021). Biological pathway expression complementation contributes to biomass heterosis in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 16, 118. doi: 10.1073/pnas.2023278118

Liu, J., Wu, M. W., and Liu, C. M. (2022). Cereal endosperms: development and storage product accumulation. *Annu. Rev. Plant Biol.* 73, 255–291. doi: 10.1146/annurev-arplant-070221-024405

Long, Y., Liu, Z., Jia, J., Mo, W., Fang, L., Lu, D., et al. (2021). FlsnRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. *Genome Biol.* 22, 66. doi: 10.1186/s13059-021-02288-0

Lu, W., Deng, M., Guo, F., Wang, M., Zeng, Z., Han, N., et al. (2016). Suppression of *OsVPE3* enhances salt tolerance by attenuating vacuole rupture during programmed cell death and affects stomata development in rice. *Rice (N Y)* 9, 65. doi: 10.1186/s12284-016-0138-x

Luo, M., Taylor, J. M., Spriggs, A., Zhang, H., Wu, X., Russell, S., et al. (2011). A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet.* 7, e1002125. doi: 10.1371/journal.pgen.1002125

Ng, D. W., Miller, M., Yu, H. H., Huang, T. Y., Kim, E. D., Lu, J., et al. (2014). A role for CHH methylation in the parent-of-Origin effect on altered circadian rhythms and biomass heterosis in *Arabidopsis* intraspecific hybrids. *Plant Cell* 26, 2430–2440. doi: 10.1105/tpc.113.115980

Nie, D. M., Ouyang, Y. D., Wang, X., Zhou, W., Hu, C. G., and Yao, J. (2013). Genome-wide analysis of endosperm-specific genes in rice. *Gene* 530, 236–247. doi: 10.1016/j.gene.2013.07.088

Paponov, I. A., Teale, W., Lang, D., Paponov, M., Reski, R., Rensing, S. A., et al. (2009). The evolution of nuclear auxin signalling. *BMC Evol. Biol.* 9, 126. doi: 10.1186/1471-2148-9-126

Paschold, A., Jia, Y., Marcon, C., Lund, S., Larson, N. B., Yeh, C. T., et al. (2012). Complementation contributes to transcriptome complexity in maize (*Zea mays* l.) hybrids relative to their inbred parents. *Genome Res.* 22, 2445–2454. doi: 10.1101/gr.138461.112

Paschold, A., Larson, N. B., Marcon, C., Schnable, J. C., Yeh, C. T., Lanz, C., et al. (2014). Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. *Plant Cell* 26, 3939–3948. doi: 10.1105/tpc.114.130948

Picard, C. L., Povilus, R. A., Williams, B. P., and Gehring, M. (2021). Transcriptional and imprinting complexity in *Arabidopsis* seeds at single-nucleus resolution. *Nat. Plants* 7, 730–738. doi: 10.1038/s41477-021-00922-0

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., et al. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982. doi: 10.1038/nmeth.4402

Ram, H., Singh, A., Katoch, M., Kaur, R., Sardar, S., Palia, S., et al. (2021). Dissecting the nutrient partitioning mechanism in rice grain using spatially resolved gene expression profiling. *J. Exp. Bot.* 72, 2212–2230. doi: 10.1093/jxb/eraa536

Rao, A., Barkley, D., França, G. S., and Yanai, I. (2021). Exploring tissue architecture using spatial transcriptomics. *Nature* 596, 211–220. doi: 10.1038/s41586-021-03634-9

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rodrigues, J. A., Hsieh, P. H., Ruan, D., Nishimura, T., Sharma, M. K., Sharma, R., et al. (2021). Divergence among rice cultivars reveals roles for transposition and epimutation in ongoing evolution of genomic imprinting. *Proc. Natl. Acad. Sci. U.S.A.* 29, 118. doi: 10.1073/pnas.2104445118

Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., et al. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5653–5658. doi: 10.1073/pnas.1820513116

Shaw, R., Tian, X., and Xu, J. (2021). Single-cell transcriptome analysis in plants: advances and challenges. *Mol. Plant* 14, 115–126. doi: 10.1016/j.molp.2020.10.012

Shen, F., Wu, X., Shi, L., Zhang, H., Chen, Y., Qi, X., et al. (2018). Transcriptomic and metabolic flux analyses reveal shift of metabolic patterns during rice grain development. *BMC Syst. Biol.* 12, 47. doi: 10.1186/s12918-018-0574-x

Springer, N. M., and Stupar, R. M. (2007). Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell* 19, 2391–2402. doi: 10.1105/tpc.107.052258

Thompson, R. D., Hueros, G., Becker, H., and Maitz, M. (2001). Development and functions of seed transfer cells. *Plant Sci.* 160, 775–783. doi: 10.1016/S0168-9452(01)00345-4

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community 2017 Update. *Nucleic Acids Res.* 45, W122–w129. doi: 10.1093/nar/gkx382

Todesco, M., Balasubramanian, S., Hu, T. T., Traw, M. B., Horton, M., Epple, P., et al. (2010). Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465, 632–636. doi: 10.1038/nature09083

Tucci, V., Isles, A. R., Kelsey, G., and Ferguson-Smith, A. C. (2019). Genomic imprinting and physiological processes in mammals. *Cell* 176, 952–965. doi: 10.1016/j.cell.2019.01.043

Venu, R. C., Ma, J., Jia, Y., Liu, G., Jia, M. H., Nobuta, K., et al. (2014). Identification of candidate genes associated with positive and negative heterosis in rice. *PLoS One* 9, e95178. doi: 10.1371/journal.pone.0095178

Vijayakumar, K. R., and Gowda, L. R. (2012). Temporal expression profiling of lipase during germination and rice caryopsis development. *Plant Physiol. Biochem.* 57, 245–253. doi: 10.1016/j.plaphy.2012.05.028

Waters, A. J., Makarevitch, I., Eichten, S. R., Swanson-Wagner, R. A., Yeh, C. T., Xu, W., et al. (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* 23, 4221–4233. doi: 10.1105/tpc.111.092668

Wu, X., Liu, J., Li, D., and Liu, C. M. (2016). Rice caryopsis development II: dynamic changes in the endosperm. *J. Integr. Plant Biol.* 58, 786–798. doi: 10.1111/jipb.12488

Wu, T. Y., Müller, M., Gruissem, W., and Bhullar, N. K. (2020). Genome wide analysis of the transcriptional profiles in different regions of the developing rice grains. *Rice (N Y)* 13, 62. doi: 10.1186/s12284-020-00421-4

Wu, M., Ren, Y., Cai, M., Wang, Y., Zhu, S., Zhu, J., et al. (2019). Rice *FLOURY ENDOSPERM10* encodes a pentatricopeptide repeat protein that is essential for the trans-splicing of mitochondrial nad1 intron 1 and endosperm development. *New Phytol.* 223, 736–750. doi: 10.1111/nph.15814

Xiao, B., Huang, Y., Tang, N., and Xiong, L. (2007). Over-expression of a *LEA* gene in rice improves drought resistance under the field conditions. *Theor. Appl. Genet.* 115, 35–46. doi: 10.1007/s00122-007-0538-9

Xue, L. J., Zhang, J. J., and Xue, H. W. (2012). Genome-wide analysis of the complex transcriptional networks of rice developing seeds. *PloS One* 7, e31081. doi: 10.1371/journal.pone.0031081

Yang, J., Luo, D., Yang, B., Frommer, W. B., and Eom, J. S. (2018). SWEET11 and 15 as key players in seed filling in rice. *New Phytol.* 218, 604–615. doi: 10.1111/nph.15004

Yang, L., Xing, F., He, Q., Tahir Ul Qamar, M., Chen, L. L., and Xing, Y. (2020). Conserved imprinted genes between intra-subspecies and inter-subspecies are involved in energy metabolism and seed development in rice. *Int. J. Mol. Sci.* 24, 21. doi: 10.3390/ijms21249618

Yin, L. L., and Xue, H. W. (2012). The MADS29 transcription factor regulates the degradation of the nucellus and the nucellar projection during rice seed development. *Plant Cell* 24, 1049–1065. doi: 10.1105/tpc.111.094854

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* l. ssp. *indica*). *Science* 296, 79–92. doi: 10.1126/science.1068037

Zhang, G.-q. (2020). Prospects of utilization of inter-subspecific heterosis between *indica* and *japonica* rice. *J. Integr. Agric.* 19, 1–10. doi: 10.1016/S2095-3119(19)62843-1

Zhang, J., Niu, B., E, Z., and Chen, C. (2021). Towards understanding the genetic regulations of endosperm development in rice. *Chin. J. Rice Sci.* 35, 326–341. doi: 10.3390/genes11070781

Zhao, H., Li, J., Yang, L., Qin, G., Xia, C., Xu, X., et al. (2021). An inferred functional impact map of genetic variants in rice. *Mol. Plant* 14, 1584–1599. doi: 10.1016/j.molp.2021.06.025

Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049. doi: 10.1038/ncomms14049

Zhou, S. R., Yin, L. L., and Xue, H. W. (2013). Functional genomics based understanding of rice endosperm development. *Curr. Opin. Plant Biol.* 16, 236–246. doi: 10.1016/j.pbi.2013.03.001