

Genome-Wide Differences in Gene Expression and Alternative Splicing in Developing Embryo and Endosperm, and Between F1 Hybrids and Their Parental Pure Lines in Sorghum

Meishan Zhang

Jilin Agricultural University <https://orcid.org/0000-0002-0141-2723>

Ning Li

Northeast Normal University

Weiguang Yang

Jilin Agricultural University

Bao Liu (✉ baoliu@nenu.edu.cn)

Northeast Normal University

Research Article

Keywords: gene expression, alternative splicing, imprinting, development, heterosis, Sorghum bicolor

Posted Date: June 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-556502/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Plant Molecular Biology on November 30th, 2021. See the published version at <https://doi.org/10.1007/s11103-021-01196-y>.

Abstract

Differential regulation of gene expression and alternative splicing (AS) are major molecular mechanisms dictating plant growth and development, as well as underpinning heterosis in F1 hybrids. Here, using deep RNA-sequencing we analyzed differences in genome-wide gene expression and AS between developing embryo and endosperm, and between F1 hybrids and their pure-line parents in sorghum. We uncover dramatic differences in both gene expression and AS between embryo and endosperm with respect to gene features and functions, which are consistent with the fundamentally different biological roles of the two tissues. Accordingly, F1 hybrids showed substantial and multifaceted differences in gene expression and AS compared with their pure-line parents, again with clear tissue specificities including extents of difference, genes involved and functional enrichments. Our results provide useful transcriptome resources as well as novel insights for further elucidation of seed yield heterosis in sorghum and related crops.

Key Message

Developing embryo and endosperm of sorghum show substantial and multifaceted differences in gene expression and alternative splicing, which are potentially relevant to heterosis.

Introduction

Sorghum (*Sorghum bicolor* L.) ranks fifth in global cereal production, and its seeds provide important resources for food, feed and biofuel worldwide (Mace et al. 2013). Owing to its relatively small genome (730 Mb), sorghum has also been used as a model species for genetic and genomic studies in sugarcane and other C4 grasses (Paterson et al. 2009). In cereal crops, endosperm and embryo are major components of seeds and hence are direct determinants of yield. As in all angiosperms, sorghum seed development is initiated following double fertilization, which gives rise to diploid embryo and triploid endosperm, respectively, with the two tissues exhibiting dramatic differences in many respects. The main function of embryo is to perpetuate genetic information to the next generation, while that of cereal endosperm is mainly to nourish the developing embryo and germinating seedling (Lu et al. 2013; Xin et al. 2013). In flowering plants, developing endosperm is the primary tissue of imprinted gene expression, i.e., unequal expression of maternal and paternal alleles depending on parent of origin (Gehring and Satyaki 2017; Zhang et al. 2011). In a previous study, we identified 101 sorghum imprinted genes (≥ 5 -fold allelic expression difference) including 85 maternally expressed imprinted genes (MEGs) and 16 paternally expressed imprinted genes (PEGs) from 14 DAP (days after pollination) sorghum endosperm based on RNA-sequencing (Zhang et al. 2016).

Hybridization is a prominent process in the evolution, domestication and genetic improvement of sexually reproducing crop species. Heterosis or hybrid vigor results from genome-wide interactions between paternal and maternal alleles and *de novo* regulatory changes in the F1 hybrids (Birchler et al. 2010; Huang et al. 2016). Crop growth and development are regulated by complex global gene expression

networks. Gene expression patterns in F1 hybrids can be divided into two broad categories, i.e., additive and non-additive expression, compared with that of the parental average. Nonadditive gene expression is defined as expression of a gene in a hybrid, which is significantly deviated from the average value of the parental lines (mid-parent value), and which has been shown as significantly associated with heterosis (Paschold et al. 2012). The proportions of genes showing nonadditive expression differ greatly across species, tissues and organs, experimental procedures, and surveyed gene sets (Springer and Stupar 2007). In recent years, several studies have used RNA-seq to analyze globally differentially expressed genes (DEGs) between a hybrid and its parents, provided new insights into the genetic mechanisms of heterosis (Chen et al. 2018; Paschold et al. 2012; Zhang et al. 2017a). For example, RNA-seq analysis of the primary root of maize inbred lines B73 and Mo17 and their reciprocal F1 hybrids revealed that 42%~55% of all expressed genes were differentially expressed between at least one of the parents and one of the hybrids, and about 10% of expressed genes were nonadditively expressed in both hybrids (Paschold et al. 2012). A transcriptome analysis between super-hybrid rice and its parents of young panicles identified a total of 5,910 DEGs, which were significantly enriched in carotenoid biosynthesis and hormone signal transduction pathways (Chen et al. 2018).

Alternative splicing (AS) is another common mechanism to cause changes in gene expression as well as transcript diversity. AS is a process occurring in all higher eukaryotes whereby the same precursor mRNA generates multiple transcripts with variable quantities. AS is a major mechanism to achieve proteomic diversity with defined gene numbers (Ast 2004; Nilsen and Graveley 2010; Song et al. 2019). In human, more than 95% of multiexonic genes are alternatively spliced (Syed et al. 2012). Nevertheless, according to recent studies, the occurrence of AS varies tremendously among species, ranging from 37–63% in intron-containing genes. At least some of these variations can be attributed to differences in depth of the RNA-seq data, technology platforms, and statistical analyses (Mei et al. 2017; Satyawati et al. 2017; Shen et al. 2014; Zhang et al. 2019). Meanwhile, extensive tissue-specific, developmental stage-specific splicing patterns, or altered AS behaviors, induced by environmental stresses also contribute to the diversity of AS (Mei et al. 2017; Thatcher et al. 2014; Wang et al. 2018). A previous study based on transcriptomic data indicated that 18,741 (45%) and 13,327 (38.5%) of expressed genes were alternatively spliced in maize and sorghum, respectively, but the proportions of different splicing events varied among tissues (Wang et al. 2018). In maize, 50.7% of multiexonic genes were alternatively spliced and 7.0% of AS events were differentially regulated between 9 DAP embryo and endosperm, and a higher frequency of tissue-regulated events was detected in embryo than in endosperm (Lu et al. 2013). A pairwise analysis of different maize tissues showed that embryo had the highest number of AS events when compared with the rest 14 tissues (Thatcher et al. 2014). Intron retention (IR) is the most frequent AS event in plants, whereas exon skipping (ES) accounts for the most common event in mammalian species (Ast 2004; Nilsen and Graveley 2010; Syed et al. 2012). Some alternatively spliced transcripts play a significant role in development, stress responses and tissue-specific physiology (Lin and Zhu 2021; Liu et al. 2018; Mei et al. 2017). In contrast, most splice variants may not be functional, and degraded by RNA surveillance mechanisms such as the nonsense-mediated mRNA decay (NMD) pathway (Chaudhary

et al. 2019; Chen et al. 2019; Satyawati et al. 2017). In some cases, degradation of AS isoforms through the NMD pathway serves to regulate the concentration of transcripts in the cells (Drechsel et al. 2013).

Although extensive studies have been conducted to identify AS events in different plant species, tissues and environmental conditions, little attention has been paid to AS events in crop plant imprinted genes at genome-wide scale. In a previous study, we compared the variant splicing differences in maize for several imprinted genes between embryo and endosperm, and found that more IR events occurred in embryo than in endosperm (Zhang et al. 2018). In this study, based on genome-wide comparative transcriptome analyses, we aimed to identify differentially expressed genes (DEGs) in developing embryo and endosperm tissues of a pair of reciprocal sorghum F1 hybrids and their parental pure lines. We also performed AS analysis, including the identification and characterization of AS events, differential splicing between tissues and among genotypes, and AS features of sorghum imprinted genes. Experimental validation as well as gene ontology enrichment analyses were also conducted.

Materials And Methods

Plant tissue collection

The sorghum (*Sorghum bicolor* L.) parental pure lines, BTx623 (B) and Y49 (Y) were grown in the field at Jilin Agricultural University experimental farms near Changchun. A pair of reciprocal inter-strain F1 hybrids, BY (BTx623 as female pollinated with Y49) and YB (Y49 as female pollinated with BTx623) were produced by manual pollination. Embryo and endosperm samples were dissected manually from the 14 days after pollination (DAP) immature seeds. Embryos were washed three times with RNA-free water to remove possible endosperm contamination on the surface of embryo. Materials were quickly frozen in liquid nitrogen and store in -70°C until RNA extraction. Each sample was collected from at least five different plants.

RNA isolation

Embryo and endosperm samples from 4 genotypes in biological triplicate were used for RNA isolation. The total RNA was isolated using Plant RNA Reagent (Invitrogen, #12322-012 <http://www.lifetechnologies.com>) and subsequently purified with QIAGEN RNeasy Plus Mini kit (#74134) according to the manufacturer's instructions.

RNA-seq data generation and read alignment to the reference sorghum genome

RNA-seq data of endosperm was downloaded from SRA, the accession number was PRJNA302593, which we uploaded 2016. For the embryo, the Agilent 2100 BioAnalyzer (Agilent Technologies), and Nanodrop were used to measure quality and quantity of each RNA sample for library construction. The library construction, cluster generation and HiSeq 2500 sequencing were conducted in accordance with the standard protocols. All samples were treated and sequenced as parallel experiments. Adapters and low-

quality reads were removed from raw sequence by using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), leaving a total of 238.67 million 150 bp paired-end clean reads (71.6 Gb clean data). Reads were mapped to the sorghum reference genome (<https://phytozome.jgi.doe.gov/pz/portal.html>) by the HISAT2 under default parameter. The raw data information and mapping efficiency are given in Table S1.

Gene expression estimation and putative transcript assembly

We used StringTie (Pertea et al. 2016) to estimate the GTF file for each sample, and then merged all GTFs into one final GTF with StringTie Merge. Transcripts were normalized using the transcript per million base pairs (TPM). HTSeq (Anders et al. 2015) was used to count reads per gene, and DESeq2 (Love et al. 2014) was used to identify the differentially expressed genes (DEGs) and transcripts.

Identification of AS events

We identified the AS events using ASTALAVISTA algorithm (Foissac and Sammeth 2007). The different types of AS events were analyzed as previously described (Trincado et al. 2018).

Reverse transcription and experimental validation with Semi-quantitative RT-PCR

About 800 ng total RNA from each sample was used for cDNA synthesis using a Reverse Transcriptase kit (TransGen Biotech) following the manufacturer's instructions. The PCR primers sets that flank the differentially spliced region are used to amplify splice variants from cDNA prepared from different samples, with three biological replicates. The primer information is listed in Supplementary Information, Table S3. PCR was performed using rTaq DNA polymerase (TAKARA) in 20 μ l reaction volume. The PCR profile was: initial denaturation at 94°C for 2 min followed by 25–32 cycles (according to the primer pair) at 94°C for 40 s, 55–60°C (according to the primer pair) for 40 s, 72°C for 40–60 s (according to the PCR products length), and a final extension at 72°C for 8 min. The amplified products were visualized in a 1.8 % agarose gel stained with ethidium bromide. The sorghum actin gene (Genbank accession X79378) was used as an internal reference gene.

Gene Ontology (GO) analysis

GO term enrichment analysis was performed using the hypergeometric statistic (R function phyper). GO categories with p value < 0.05 were considered as enriched.

Results

We performed high-throughput RNA-seq using the Illumina HiSeq 2500 platform for embryo and endosperm tissues dissected from sorghum seeds at 14 DAP (days after pollination) of four genotypes (two parental pure lines BTx623 [B] and Y49 [Y], and their reciprocal F1 hybrids, BY and YB). We obtained a total of 487.43 million 150 bp paired-end clean reads which were comprised of 77.9 Gb of high-quality

raw data (Table S1). The sequence reads were aligned to the BTx623 v3 reference genome (<https://phytozome.jgi.doe.gov/pz/portal.html>) by HISAT2. An overview of the distribution of the aligned reads along each of the 10 sorghum chromosomes showed that, as expected, most of the reads were located in the distal chromosome arm regions, while transposable elements (TEs) were mainly aggregated to gene-poor, pericentromeric regions (Fig. 1). Moreover, approximately 80% of the mapped reads were in exonic regions and only a small number of the reads were mapped to non-genic regions (Table S1). The aligned reads located in exonic and intronic regions were used to detect gene expression and alternative splicing (AS) for each gene in each sample (tissue and genotype). We assembled RNA-seq alignments into potential transcripts for each sample with StringTie. Gene and transcript expression levels were estimated with Ballgown and measured as transcript per million base pairs (TPM). The cutoff value for determining the expressed genes or transcripts was TPM > 0.

We found that although most genes were commonly expressed in both tissues of all genotypes, transcript isoform abundance seems higher in embryo than in endosperm, consistent with results in a maize transcriptome analysis (Lu et al. 2013). We detected a total of 46,035 genes expressed in the developing sorghum seed transcriptome, with 43,676 and 33,763 genes expressed in embryo and endosperm, respectively, which corresponded to 80.3–64.1% of the annotated genes, producing 81,800 and 66,632 transcripts, with 31,404 (63.2%) genes being commonly expressed in both tissues (Table S2). In addition, we detected expression of 15,690 and 6,637 unannotated genes, which produced 44,710 and 32,054 transcripts in sorghum embryo and endosperm, respectively. Of note, we found that the total numbers of expressed genes in F1 hybrids were higher than those in the parental lines in both tissues (Table S2).

Differentially expressed genes (DEGs) between embryo and endosperm, and between hybrids and parents

Differential expression analyses between the tissues and among the genotypes were calculated by DESeq2. Tissue-specific or tissue-preferential expressed genes were categorized based on the following criteria: (i) false discovery rate (FDR) was less than 0.01; (ii) genes with > 10 reads in one tissue and ≤ 1 read in another are defined as tissue-specific; (iii) genes showing > 2-fold differences in expression level are defined as tissue-preferential. According to these criteria, we identified 1,994 (4.6%) embryo-specific genes, 14,073 (32.2%) embryo-preferential genes, 2,335 (6.9%) endosperm-specific genes and 4,078 (12.1%) endosperm-preferential genes, respectively (Fig. 2a). Clearly, a greater proportion of the genes were specifically or exclusively expressed in endosperm than in embryo, whereas the proportion of preferentially expressed genes was 3.5 times more in embryo than in endosperm (Fig. 2a). However, for imprinted genes (defined as ≥ 5 -fold differences in read counts between parental alleles), the great majority is endosperm-specific or -preferential, with embryo-preferential genes accounting < 10% of the total number of imprinted genes, and none imprinted gene showed embryo-specific expression.

Next, we analyzed distribution of the four types of DEGs (embryo-specific genes, embryo-preferential genes, endosperm-specific genes and endosperm-preferential genes) in the different genotypes, including a pair of reciprocal F1 hybrids (BY and YB) and their two parental pure lines (B and Y) (Fig. 2b, c). We

found that tissue-specific expressed genes were less shared by the genotypes, i.e., they are more prone to showing genotype-specific expression. Approximately 18% and 26% of the embryo- and endosperm-specific genes, respectively, were common across all four genotypes. Additionally, ~ 52% and ~ 34% of the embryo- and endosperm-preferential expressing genes, respectively, were shared by all four genotypes (Fig. 2b). Notably, we found that the numbers of embryo-specific expressing genes in hybrids were significantly higher than those in their parents, whereas the other three types of between-tissue DEGs showed similar gene expression between hybrids and parental lines (Fig. 2b). Finally, there were more genes expressed at a higher level in embryo than in endosperm (Fig. 2c), which is consistent with a previous study in maize (Lu et al. 2013).

Additive and nonadditive gene expression patterns in embryo and endosperm of F1 hybrids

To determine the relative expression levels of genes in the sorghum reciprocal F1 hybrids (BY and YB) as compared to their parental pure lines (B and Y), and to identify genes displaying expression levels that deviate from additivity in the hybrids, we analyzed the RNA-seq data and identified nonadditively expressed genes in the hybrids. Here, nonadditive gene expression is defined as expression level in a hybrid that is significantly deviated from the averaged parental value, i.e., mid-parent value (MPV) ($FDR < 0.05$). In embryo, no significant differences between the reciprocal hybrids with respect of the proportions of genes showing additive vs. nonadditive expression were detected; both hybrids showed a predominant proportion of additive expression (99.6%) and very small proportion of nonadditive expression (0.39%) (Fig. 3). In contrast, in endosperm, the reciprocal F1 hybrids showed clear difference in the proportions of genes showing additive vs. nonadditive expression. Moreover, the proportions of nonadditively expressed genes in endosperm were markedly greater than those in embryo, with 10.3% and 7.6% genes showing nonadditive expression in endosperm of the reciprocal hybrids, BY and YB, respectively (Fig. 3). This suggests the influence of cytoplasm and other maternal effects were significantly exacerbated by nuclear genome dosage as the triploid endosperm has a 2:1 maternal vs. paternal nuclear genome contribution.

Differential alternative splicing (AS) between tissues and among genotypes

The RNA-seq data also enable global analysis of dynamic changes in alternative splicing (AS) patterns. We investigated spliced events in types of genes, between tissues and among the four genotypes using the algorithm ASTALAVISTA (Foissac and Sammeth 2007). Overall, of the 46,035 total expressed genes, 27,048 (58.8%) were multiexonic genes, and 41.9% of these genes were alternatively spliced, resulting in 31,012 AS events in total (Table 1). We found the overall AS frequency is slightly higher than that reported previously for sorghum (Wang et al, Genome Res 2018; Min et al. BMC Genomics, 2015), but similar to those reported in maize (Thatcher et al. 2014) and rice (Lu et al. 2010), and significantly lower than other plant species, especially Arabidopsis and soybean, in which AS is reported to occur in > 60% of multiexonic genes (Marquez et al. 2012; Shen et al. 2014). Specifically, 38.6% and 37.4% of expressed multiexonic genes were alternatively spliced and showing 29,600 and 25,427 AS events in sorghum

embryo and endosperm, respectively (Table 1). Among these, a total of 22,859 AS events occurred in both tissues but embryo-specific AS events (6,741) were 2.6 times more than endosperm-specific AS events (2,568) (Fig. S1). Then, we analyzed AS events, in separation, for the four types of the DEGs and of imprinted genes. We found that 30.4% and 56.4% of embryo-specific and –preferential genes were alternatively spliced respectively, however, the ratios were much lower in both endosperm-specific and –preferential genes than in embryo DEGs, showing only 17.8% and 29.4%, respectively (Table 1). Additionally, the frequency of AS in imprinted genes was 46.3%, which was higher than those of total expressed genes and endosperm DEGs (Table 1).

Table 1
AS events in sorghum embryo and endosperm

	Expressed genes	Expressed multiexonic genes	AS genes (% of multiexonic genes)	AS events of multiexonic genes
Total	46,035	27,048	11,332 (41.9)	31,012
Embryo	43,676	25,745	9,949 (38.6)	29,600
Endosperm	33,763	24,859	9,299 (37.4)	25,427
Embryo-specific	1,994	1,466	446 (30.4)	589
Embryo-preferential	14,073	12,681	7,153 (56.4)	20,503
Endosperm-specific	2,335	1,561	278 (17.8)	374
Endosperm-preferential	4,078	2,942	865 (29.4)	1,526
Imprinted genes	101	80	37 (46.3)	70

Next, we projected these comparisons with emphasis on genotype differences in both tissues. Results showed that (i) 16,920 and 10,847 events were conserved among all four genotypes in embryo and endosperm, respectively; (ii) compared with parents, 3,093 and 3,522 AS events were unique to embryo and endosperm of at least one hybrid, and 624 and 537 of both hybrids; (iii) compared with parents, 3,093 and 3,522 AS events were lost from embryo and endosperm from at least one hybrid, and 444 and 552 from both hybrids (Fig. 4a).

We further investigated the different AS types in tissue-specific or tissue-preferential expressing genes, as well as in imprinted genes. The seven AS modes were illustrated (Fig. 4b). Overall, the four basic AS types accounted for ~ 78% of the total AS events in both tissues, and the distribution of different types of AS events was also similar between embryo and endosperm. Specifically, as described in other plant species, we found that intron retention (IR) accounted for the most abundant AS type in both tissues (~ 25%), though it was only slightly higher than the second type, alternative 3' splicing site (A3SS). A similar trend

was observed in 20 DAP sorghum BTx623 embryo and endosperm (Wang et al. 2018). Exon skipping (ES) events varied in different plant species from 3% in Arabidopsis (Marquez et al. 2012) to 25% in rice (Zhang et al. 2010) and 33% in maize (Thatcher et al. 2014). The number of ES events in our data fell in the middle, accounting for ~ 17% of the total AS events. Alternative 5' splice site (A5SS) was the least frequent mode of the four main AS types (~ 13%) in both tissues (Fig. 4c). We also calculated the rare AS types. For example, both the skipping two exons, "ES1 + ES2" and "A5SS or A3SS" counted *ca.* ~2.5% of the total AS events. The mutually exclusive exon (MXE) AS type, which has been detected in rice (Lu et al. 2010; Zhang et al. 2010), but not in Arabidopsis (Marquez et al. 2012), accounted for ~ 0.6% of the AS in sorghum, ranked the least AS type in our data (Fig. 4c). This is consistent with the results in maize (Lu et al. 2013). For those complex AS models, we put them all together to the type of "others", which likely resulted from different combinations of the four basic types or arose from different mechanisms (Syed et al. 2012). These complex models accounted for approximately 17% of the total AS events in embryo and endosperm of sorghum (Fig. 4c).

For the DEGs, no significant differences in the proportion of AS patterns were observed in tissue-preferentially expressed genes of both embryo and endosperm compare to all genes from the sorghum genome. However, the tissue-specific genes in both tissues showed obviously different splice patterns. Specifically, (i) in embryo-specific genes, IR was increased 16%, whereas in endosperm-specific genes, IR was decreased 28%, relative to those of the total expressed genes; (ii) the exon skipping type did not change in embryo-specific expressing genes but was increased by 27% in endosperm-specific expressing genes; (iii) in both embryo- and endosperm-specific genes, the "others" AS type was decreased compared to the total expressed genes (Fig. 4c).

Interestingly, we found that all the four common types of AS, especially IR, were decreased in sorghum imprinted genes, while those of the rare types, like "skip 2 exons", "A5SS or A3SS", MXE and other complex types were increased, led to nearly 30% of the AS events were accounted by the "others" type in imprinted genes (Fig. 4c). This suggests that compared with embryo, endosperm likely has unique AS splicing mechanisms.

We further analyzed the sequence length covered by each AS type in sorghum embryo and endosperm. We found that the four major types of AS events showed different sequence lengths. Specifically, (i) the retained intron length increased from 60 to 100 bp; (Fig. 5); (ii) both of A3SS and A5SS showed similar features of sequence length, with the most frequent length < 10 bp; (iii) the skipped exon lengths ranged from 1 to 200 bp and peaked at 70–80 bp, but this peak was not as sharp as the other types of AS events (Fig. 5). These features of AS are consistent with previous findings from other species (Campbell et al. 2006; Shen et al. 2014).

Generation of mature mRNAs from genes requires precise removal of introns from precursor mRNAs (pre-mRNAs) and joining of exons. Four loosely conserved core sequence elements in introns, the 5'SS with a conserved GT, the 3'SS with a conserved AG, therefore, usually splice at a 5'-GT-AG-3' exon-intron junction (Reddy et al. 2013). We compared splice motif preference among the different AS types and between

tissues. We found that the splicing junction site of 5'-GT-AG-3' pair represented the highest proportion of all splicing sites, followed by the 5'-GC-AG-3' pair. Specifically, of the IR and A5SS events, splice junctions (SJs) were comprised of *ca.* 96% GT-AG and 4% GC-AG, whereas in A3SS and ES, SJs were comprised of 98% and 99% GT-AG, 2% and 1% GC-AG, respectively. (Fig. S2). This result was in line with previous studies in other plant species (Li et al. 2014; Marquez et al. 2012; Shen et al. 2014). Further Chi-square test indicated that there was no significant difference in the frequencies of the types of splicing sites between embryo and endosperm.

Experimental validation of differential expression and AS events

To validate the DEGs and AS events detected in our transcriptomic data, we selected 20 AS genes (17 of which were also DEGs between tissues) and interrogated the accuracy via RT-PCR (primers were listed in Table S3). Results showed that most of the DEGs and AS events were very well supported by this independent experiment, with 82% of the DEGs and 80% of the AS events were supported in both tissues (Fig. 6; Table S3).

Gene ontology (GO) enrichment analyses of DEGs and AS genes

To assess the possible functional relevance of the DEGs, we conducted Gene Ontology (GO) enrichment analysis by using a hypergeometric statistic (R function *phyper*). The significantly enriched (p value < 0.05) top GO terms in each category of the DEGs were presented in Fig. 7 and all enriched GO terms were listed in Table S4-S6. We found that more GO terms were enriched in embryo than in endosperm when controlling FDR at the 0.01 level. Those genes that exhibited embryo-specific expression were involved in all three main GO categories: biological processes, molecular function, and cellular component. For example, the top significantly enriched GO subcategories in the biological process category were “cellular glucan metabolic process”, “regulation of transcription” and “protein amino acid phosphorylation”. In the molecular function category, “ADP binding”, “protein dimerization activity”, “heme binding”, and “transporter activity” were significantly enriched. We also found that the embryo-specific genes were enriched in a few cellular component categories including “cell wall”, “apoplast” and “membrane” (Fig. 7 and Table S4). In contrast, there were limited functional terms involved in endosperm-specific genes when controlling FDR at 0.01. “Biological processes” of endosperm-specific genes were mainly enriched in “regulation of transcription (DNA-dependent)” and “proteolysis”. In molecular function category, “transcription factor activity”, “enzyme inhibitor activity” and “aspartic-type endopeptidase activity” were significantly enriched (Fig. 7 and Table S4). GO enrichment was not detected for endosperm-DEGs in the “cellular component” category.

Analysis of GO enrichment for nonadditive expressing genes, using all sorghum genes as background, revealed that they were significantly overrepresented by some specific GO terms (Fig. 7). All the enriched GO terms were listed in Table S5. Importantly, GO terms “defense response”, “response to biotic stimulus”,

and “oxidoreductase activity” were significantly enriched in both tissues, indicating that nonadditively expressed genes in hybrids were possibly related to increased resistance or resilience to biotic and abiotic stresses of the hybrids. In addition, in embryo, the significantly enriched top GO terms were “transferase activity (transferring hexosyl groups)” and “ribonuclease T2 activity”, while the nonadditively expressed genes in endosperm were highly enriched in “carbohydrate metabolic process”, “proteolysis”, “oxidation reduction”, “serine-type carboxypeptidase activity”, “extracellular region”, and so on (Fig. 7 and Table S5).

We also conducted GO enrichment analysis for AS events in the different tissues and genotypes. The numbers of genes categorized to the GO terms were listed in Table S6, and the significant top GO terms in each category were presented in Fig. 7. Overall, fewer AS genes were involved in functional terms compared to the DEGs when controlling FDR at the 0.01 level. Nevertheless, the AS genes of the “IR” type were highly enriched in protein metabolism process in both tissues. Importantly, the GO term “nutrient reservoir” (GO:0045735) was found to be significantly enriched in the genes of the “A3SS” and “A5SS” AS types in endosperm but not in embryo of all four genotypes, indicating functional difference of genes underwent specific types AS between the tissues. In addition, genes spliced by “A3SS”, “ES” and “A5SS” patterns of both tissues showed highly enriched in GO terms “nucleic acid binding” or “RNA binding” (FDR corrected P values, 0.01) (Fig. 7 and Table S6).

Discussion

Our RNA-seq analyses demonstrated that more than one third of genes in 14 DAP sorghum seeds showed tissue differential gene expression, with transcript abundance being higher in embryo than in endosperm, and more genes were preferentially expressed in embryo than in endosperm, whereas endosperm DEGs tended to show a higher proportion of specific expression. We also found the great majority of sorghum imprinted genes exhibiting endosperm-specific or -preferential expression. Further comparative transcriptome analyses of the F1 hybrids and parental lines indicated that the genes exhibiting nonadditive expression was significantly higher in endosperm than in embryo. These results suggest there exist distinct regulatory systems of gene expression between embryo and endosperm during seed development. A reasonable explanation of the multifaceted genome-wide gene expression differences between embryo and endosperm is that they are of different ploidy levels and exist in distinct transcriptional and chromatin states immediately after double fertilization. First, embryo is diploid, with one paternal genome and one maternal genome, while endosperm is triploid, with two maternal genomes and one paternal genome. Second, while the major task of embryo is to ensure totipotency and perpetuation of genetic information to the next generation, that of endosperm is to nourish embryo development and the germinating seedling, but genetically inconsequential to the next generation (Gehring and Satyaki 2017). Conceivably, such distinct functional partitioning entails substantially variable gene regulation circuits and hence their direct molecular readouts.

Numerous observations have suggested that, like animals, there are many alternative splicing (AS) events in plants that are regulated in a tissue-specific manner. Multicellular organisms may generate different splice forms of the same gene in different tissues, or even within different cells in the same tissue

(McGuire et al. 2008). Our study demonstrated that the number of AS events and the frequency of AS types varied in embryo- or endosperm-specific genes and imprinted genes. The frequency of AS events was higher in embryo than in endosperm, a similar trend was found in soybean (Shen et al. 2014). These tissue-specific AS events may function in a coordinated manner in specific pathways or interaction networks as groups of genes that are coregulated at the transcriptional level (Shen et al. 2014). Moreover, in this study we found that sorghum imprinted genes showed different AS patterns from the rest “normal” genes, implying a role of AS in regulating imprinted gene expression.

The most common AS type in plants is intron retention (IR) whose rates vary among plant species and individuals of the same species, and even differ in the same genotype, ranging from 30–62% (Lu et al. 2013; Mei et al. 2017; Thatcher et al. 2014). One possible reason is that different statistical approach or classification methods were used. In some studies, IR only counts the mode that retains one intron, and those from different combinations of the four basic types are categorized into the “Others” AS type (illustrated in Fig. 4b); other studies not only count the mode that retain one intron, but also all of the IR-containing complicated cases, for instance, “IR1 + IR2”, “IR1 or IR2”, “A3SS and IR”, and “A3SS or IR” are all included in IR (Marquez et al. 2012; Zhang et al. 2017b). In the later categorizations, the percentage of IR events can reach to 40%-47% of all AS events (Marquez et al. 2012; Zhang et al. 2017b). If count the simple IR only, the IR frequency will be down to about 26% in the studies of Arabidopsis (Marquez et al. 2012; Zhang et al. 2017b). Our analysis also indicated that IR was the most variable AS type in sorghum. A previous study in maize suggested that IR may act to fine-tune gene expression across seed development (Mei et al. 2017). However, some IR events were recently shown to be more likely to represent partially spliced transcripts due to their low abundance (Syed et al. 2012). Although the biochemical mechanisms of AS are complex and in large part remain poorly understood, AS is an important mechanism whereby plants modulate gene expression in development (Syed et al. 2012).

Gene ontology (GO) analysis of the DEGs between the two tissues revealed that genes associated with different GO terms are potentially related to the functions of a given tissue during seed development. There were 1,994 and 2,335 genes specifically expressed in embryo and endosperm, respectively; however, the greater number of GO terms enriched in embryo-specific genes than in endosperm-specific genes suggests embryo may perform more specialized functions. In contrast, nonadditive gene expression was more prevalent in endosperm than in embryo; accordingly, more enriched GO terms are involved in nonadditive genes of endosperm than in those of embryo. This may suggest that compared with embryo, gene regulation in endosperm is more permissive (i.e., regulated with less stringency), consistent with inconsequential of endosperm to the perpetuation of genetic information. Moreover, analysis of GO enrichment on AS genes indicated that GO terms with protein metabolism and nucleic acid binding was enriched in embryo and the molecular function term “nutrient reservoir activity” was enriched in endosperm, consistent with the role of endosperm in nutrient storage, and again indicating fundamental functional difference between the two tissues (Merkin et al. 2012; Wang et al. 2018).

In sum, the comprehensive gene expression data, and the elucidation of differential gene expression (including tissue-specific expression) and differential AS patterns between tissues and between F1

hybrids and parents in sorghum provide foundational information for further functional genetic and genomic research on sorghum and other related cereal crops.

Declarations

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No. 31370341) and the Natural Science Foundation of Jilin Province, China (No. 20150101086JC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

M.S.Z., B.L. and W.G.Y. designed the research. Material preparation, data analysis and experimental validation were performed by N.L. and M.S.Z. M.S.Z., B.L. and N.L. wrote the manuscript. All authors read and approved the manuscript.

Conflict of interest: The authors declare that they do not have any conflict of interest.

Ethical statement: We declare that this manuscript complies with all ethical rules required by this journal.

Data availability

RNA-seq data of endosperm can download from SRA, the accession number is PRJNA30259. For the embryo, transcriptome data generated in this study have been deposited to NCBI under the accession number of PRJNA732112, and are available at: <https://dataview.ncbi.nlm.nih.gov/object/PRJNA732112?reviewer=j64as52ku12oajl4d2la2ain2h>

References

1. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169
2. Ast G (2004) How did alternative splicing evolve? *Nat Rev Genet* 5:773–782
3. Birchler JA, Yao H, Chudalayandi S, Vaiman D, Veitia RA (2010) Heterosis. *Plant Cell* 22:2105–2112
4. Campbell MA, Haas BJ, Hamilton JP, Mount SM, Buell CR (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genom* 7:327
5. Chaudhary S, Jabre I, Reddy ASN, Staiger D, Syed NH (2019) Perspective on Alternative Splicing and Proteome Complexity in Plants. *Trends Plant Sci* 24:496–506
6. Chen L, Bian J, Shi S, Yu J, Khanzada H, Wassan GM, Zhu C, Luo X, Tong S, Yang X, Peng X, Yong S, Yu Q, He X, Fu J, Chen X, Hu L, Ouyang L, He H (2018) Genetic analysis for the grain number heterosis of a super-hybrid rice WFYT025 combination using RNA-Seq. *Rice (N Y)* 11:37

7. Chen MX, Zhu FY, Wang FZ, Ye NH, Gao B, Chen X, Zhao SS, Fan T, Cao YY, Liu TY, Su ZZ, Xie LJ, Hu QJ, Wu HJ, Xiao S, Zhang J, Liu YG (2019) Alternative splicing and translation play important roles in hypoxic germination in rice. *J Exp Bot* 70:817–833
8. Drechsel G, Kahles A, Kesarwani AK, Stauffer E, Behr J, Drewe P, Ratsch G, Wachter A (2013) Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. *Plant Cell* 25:3726–3742
9. Foissac S, Sammeth M (2007) ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic acids research* 35:W297–W299
10. Gehring M, Satyaki PR (2017) Endosperm and Imprinting, Inextricably Linked. *Plant Physiol* 173:143–154
11. Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, Zhao Y, Li W, Cheng B, Xia J, Chen N, Huang T, Zhang L, Fan D, Chen J, Zhou C, Lu Y, Weng Q, Han B (2016) Genomic architecture of heterosis for yield traits in rice. *Nature* 537:629–633
12. Li Q, Xiao G, Zhu YX (2014) Single-nucleotide resolution mapping of the *Gossypium raimondii* transcriptome reveals a new mechanism for alternative splicing of introns. *Mol Plant* 7:829–840
13. Lin J, Zhu Z (2021) Plant responses to high temperature: a view from pre-mRNA alternative splicing. *Plant Mol Biol* 105:575–583
14. Liu Z, Qin J, Tian X, Xu S, Wang Y, Li H, Wang X, Peng H, Yao Y, Hu Z, Ni Z, Xin M, Sun Q (2018) Global profiling of alternative splicing landscape responsive to drought, heat and their combination in wheat (*Triticum aestivum* L.). *Plant Biotechnol J* 16:714–726
15. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550
16. Lu T, Lu G, Fan D, Zhu C, Li W, Zhao Q, Feng Q, Zhao Y, Guo Y, Li W, Huang X, Han B (2010) Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res* 20:1238–1249
17. Lu X, Chen D, Shu D, Zhang Z, Wang W, Klukas C, Chen LL, Fan Y, Chen M, Zhang C (2013) The differential transcription network between embryo and endosperm in the early developing maize seed. *Plant Physiol* 162:440–455
18. Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frere C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat Commun* 4:2320
19. Marquez Y, Brown JW, Simpson C, Barta A, Kalyna M (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res* 22:1184–1195
20. McGuire AM, Pearson MD, Neafsey DE, Galagan JE (2008) Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol* 9:R50
21. Mei W, Liu S, Schnable JC, Yeh CT, Springer NM, Schnable PS, Barbazuk WB (2017) A Comprehensive Analysis of Alternative Splicing in Paleopolyploid Maize. *Front Plant Sci* 8:694

22. Merkin J, Russell C, Chen P, Burge CB (2012) Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338:1593–1599
23. Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463:457–463
24. Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, Hochholdinger F (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. *Genome Res* 22:2445–2454
25. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberler G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otiillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457:551–556
26. Perteau M, Kim D, Perteau GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* 11:1650–1667
27. Reddy AS, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative splicing landscape in plants. *Plant Cell* 25:3657–3683
28. Satyawati D, Kim MY, Lee SH (2017) Stochastic alternative splicing is prevalent in mungbean (*Vigna radiata*). *Plant Biotechnol J* 15:174–182
29. Shen Y, Zhou Z, Wang Z, Li W, Fang C, Wu M, Ma Y, Liu T, Kong LA, Peng DL, Tian Z (2014) Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26:996–1008
30. Song QA, Catlin NS, Brad Barbazuk W, Li S (2019) Computational analysis of alternative splicing in plant genomes. *Gene* 685:186–195
31. Springer NM, Stupar RM (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell* 19:2391–2402
32. Syed NH, Kalyna M, Marquez Y, Barta A, Brown JW (2012) Alternative splicing in plants—coming of age. *Trends Plant Sci* 17:616–623
33. Thatcher SR, Zhou W, Leonard A, Wang BB, Beatty M, Zastrow-Hayes G, Zhao X, Baumgarten A, Li B (2014) Genome-wide analysis of alternative splicing in *Zea mays*: landscape and genetic regulation. *Plant Cell* 26:3472–3487
34. Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E (2018) SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* 19:40
35. Wang B, Regulski M, Tseng E, Olson A, Goodwin S, McCombie WR, Ware D (2018) A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing. *Genome Res* 28:921–932

36. Xin M, Yang R, Li G, Chen H, Laurie J, Ma C, Wang D, Yao Y, Larkins BA, Sun Q, Yadegari R, Wang X, Ni Z (2013) Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. *Plant Cell* 25:3212–3227
37. Zhang C, Lin C, Fu F, Zhong X, Peng B, Yan H, Zhang J, Zhang W, Wang P, Ding X, Zhang W, Zhao L (2017a) Comparative transcriptome analysis of flower heterosis in two soybean F1 hybrids by RNA-seq. *PLoS One* 12:e0181061
38. Zhang G, Guo G, Hu X, Zhang Y, Li Q, Li R, Zhuang R, Lu Z, He Z, Fang X, Chen L, Tian W, Tao Y, Kristiansen K, Zhang X, Li S, Yang H, Wang J, Wang J (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res* 20:646–654
39. Zhang M, Li N, He W, Zhang H, Yang W, Liu B (2016) Genome-wide screen of genes imprinted in sorghum endosperm, and the roles of allelic differential cytosine methylation. *Plant J* 85:424–436
40. Zhang M, Lv R, Yang W, Fu T, Liu B (2018) Imprinted gene expression in maize starchy endosperm and aleurone tissues of reciprocal F1 hybrids at a defined developmental stage. *Genes Genomics* 40:99–107
41. Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y, Song W, Lai J (2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci U S A* 108:20042–20047
42. Zhang R, Calixto CPG, Marquez Y, Venhuizen P, Tzioutziou NA, Guo W, Spensley M, Entizne JC, Lewandowska D, Ten Have S, Frei Dit Frey N, Hirt H, James AB, Nimmo HG, Barta A, Kalyna M, Brown JWS (2017b) A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Res* 45:5061–5073
43. Zhang Z, Fu T, Liu Z, Wang X, Xun H, Li G, Ding B, Dong Y, Lin X, Sanguinet KA, Liu B, Wu Y, Gong L (2019) Extensive changes in gene expression and alternative splicing due to homoeologous exchange in rice segmental allopolyploids. *Theor Appl Genet* 132:2295–2308

Figures

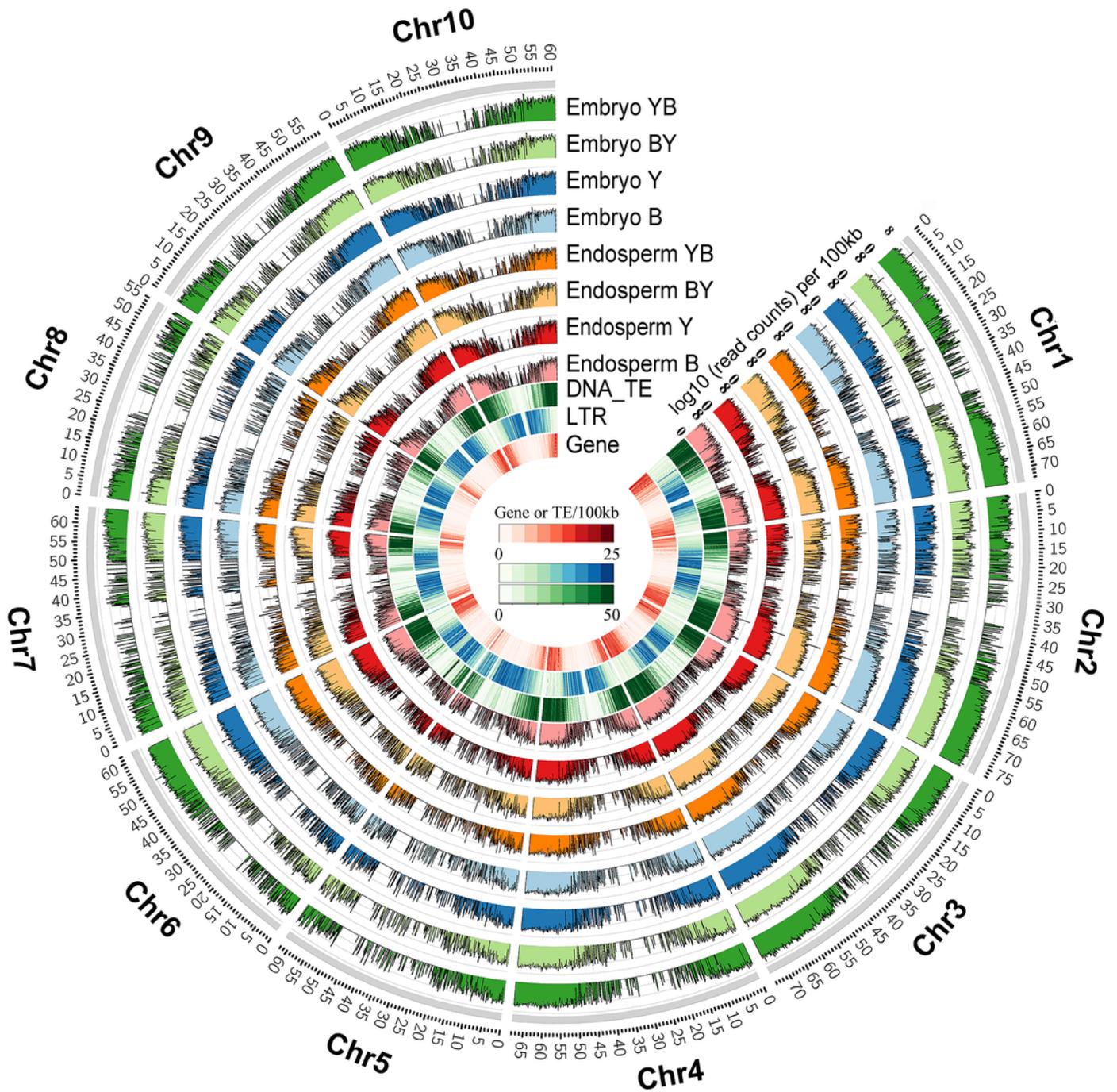


Figure 1

Genome expression profiles in sorghum embryo and endosperm of a pair of reciprocal F1 hybrids (BY and YB) and their parental pure lines (B and Y), based on deep RNA-seq. Global distribution of aligned RNA-seq reads and genomic features along 10 sorghum chromosomes that were drawn using the Circos program. From inner to outer circles, heat-maps indicate the density of expressed genes (red panel), LTR retrotransposons (blue panel) and DNA transposons (green panel). Histograms represent the log₁₀-transformed read counts per 100 kb non-overlapping windows on the indicated chromosomes for each genotype. Surrounding circle represents the 10 sorghum chromosomes and physical positions.

Genotypes B (BTx623) and Y (Y49) are a pair of sorghum pure lines, while BY (BTx623 × Y49) and YB (Y49 × BTx623) are their reciprocal F1 hybrids.

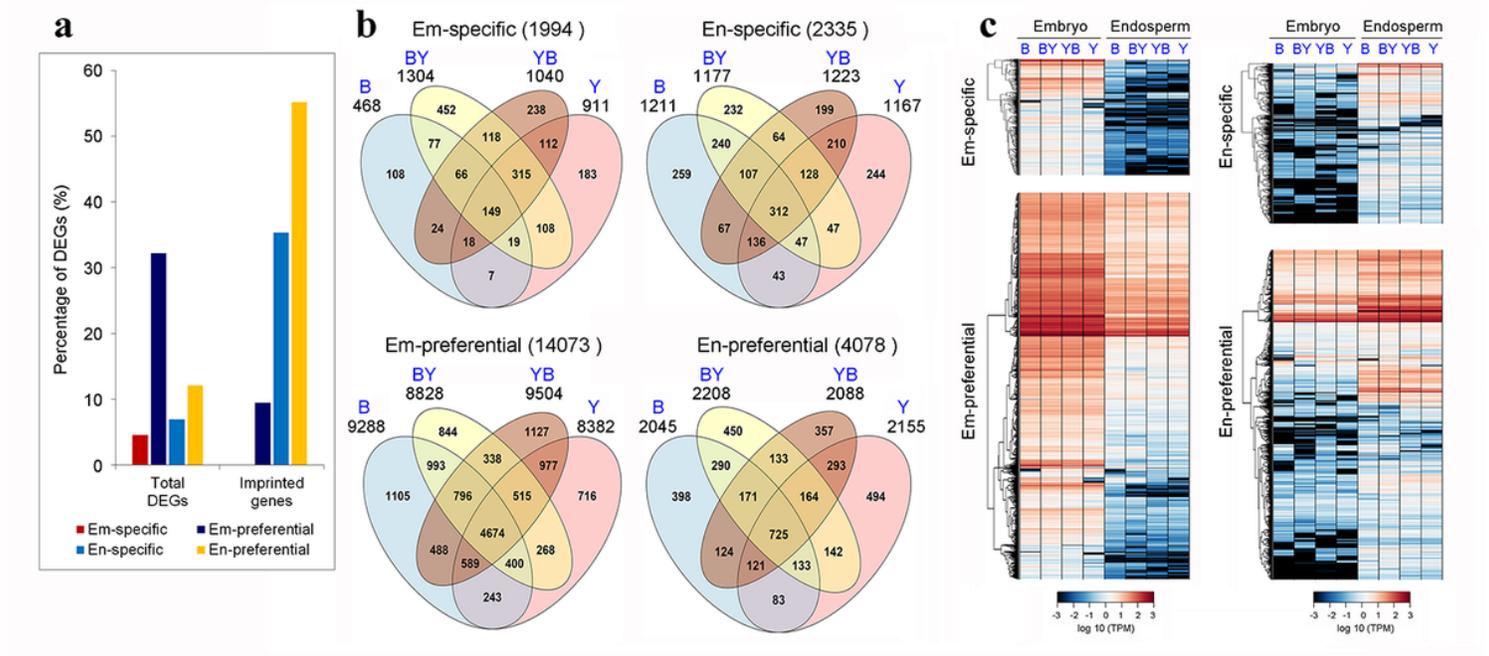


Figure 2

Differentially expressed genes (DEGs) between embryo and endosperm in the four sorghum genotypes. a. Summary of the identified differentially expressed genes or DEGs (embryo-specific, embryo-preferential, endosperm-specific and endosperm-preferential) in all genes and imprinted genes, respectively. b. Venn diagrams show common and unique DEGs among the four genotypes. The numbers below the genotypes denote total number of DEGs detected in the corresponding patterns. c. Heatmaps showing expression patterns of the DEGs in the four genotypes. The DEG types are located on the left and the tissues and genotypes are indicated at the top of each column. The color bar represents the log₁₀-transformed expression value. Genotypes B (BTx623) and Y (Y49) are sorghum parental pure lines, and BY (BTx623 × Y49) and YB (Y49 × BTx623) are their reciprocal F1 hybrids.

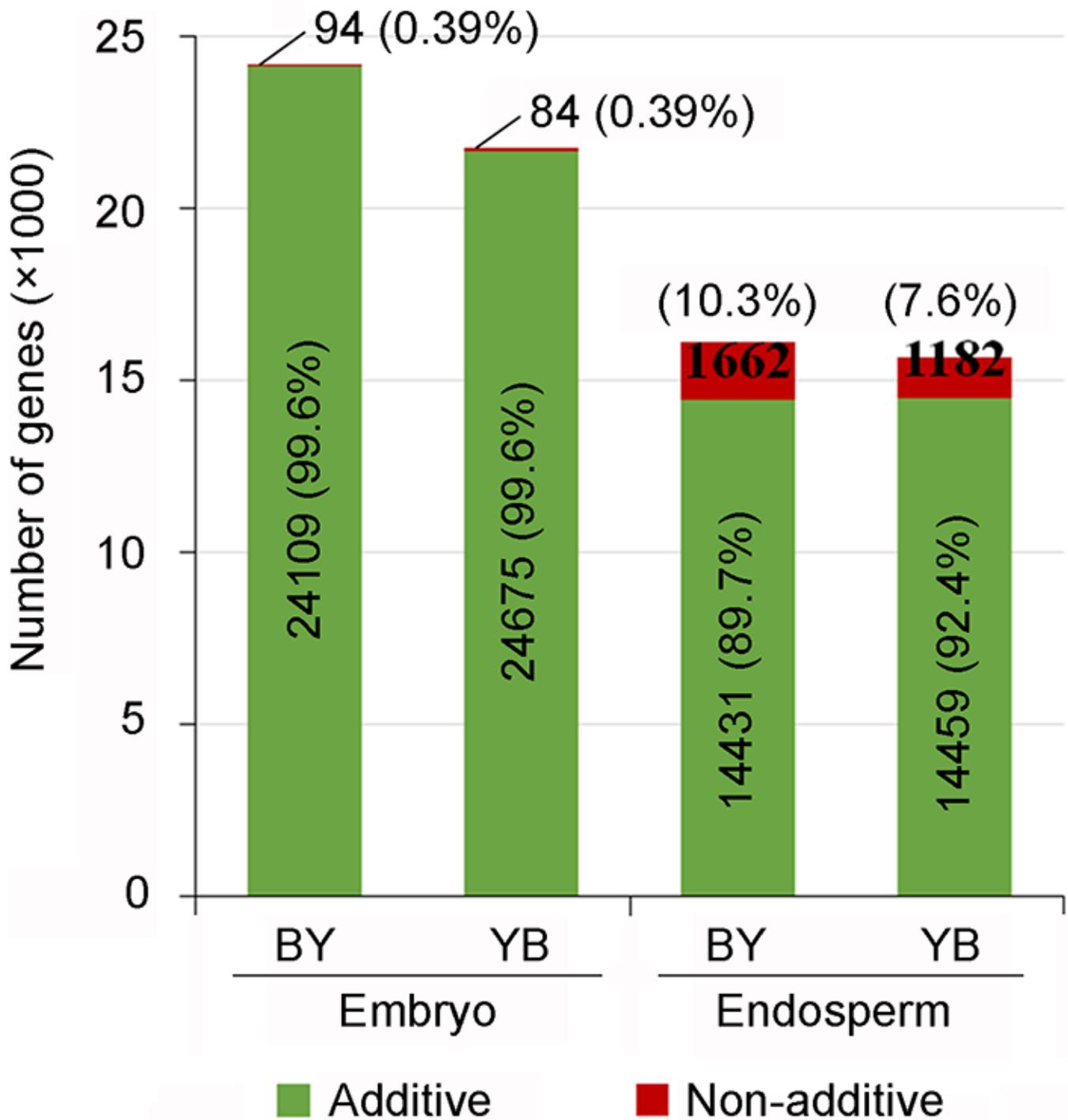


Figure 3

Additive and nonadditive gene expression between F1 hybrids and parental lines in sorghum embryo and endosperm. The numbers above individual columns present numbers and proportions of genes detected in the corresponding tissues and genotypes relative to total expressed genes. The statistical significance was determined using p-value (false discovery rate, FDR < 0.05). BY (BTx623 x Y49) and YB (Y49 x BTx623) are reciprocal F1 hybrids, while B (BTx623) and Y (Y49) are their parental pure lines.

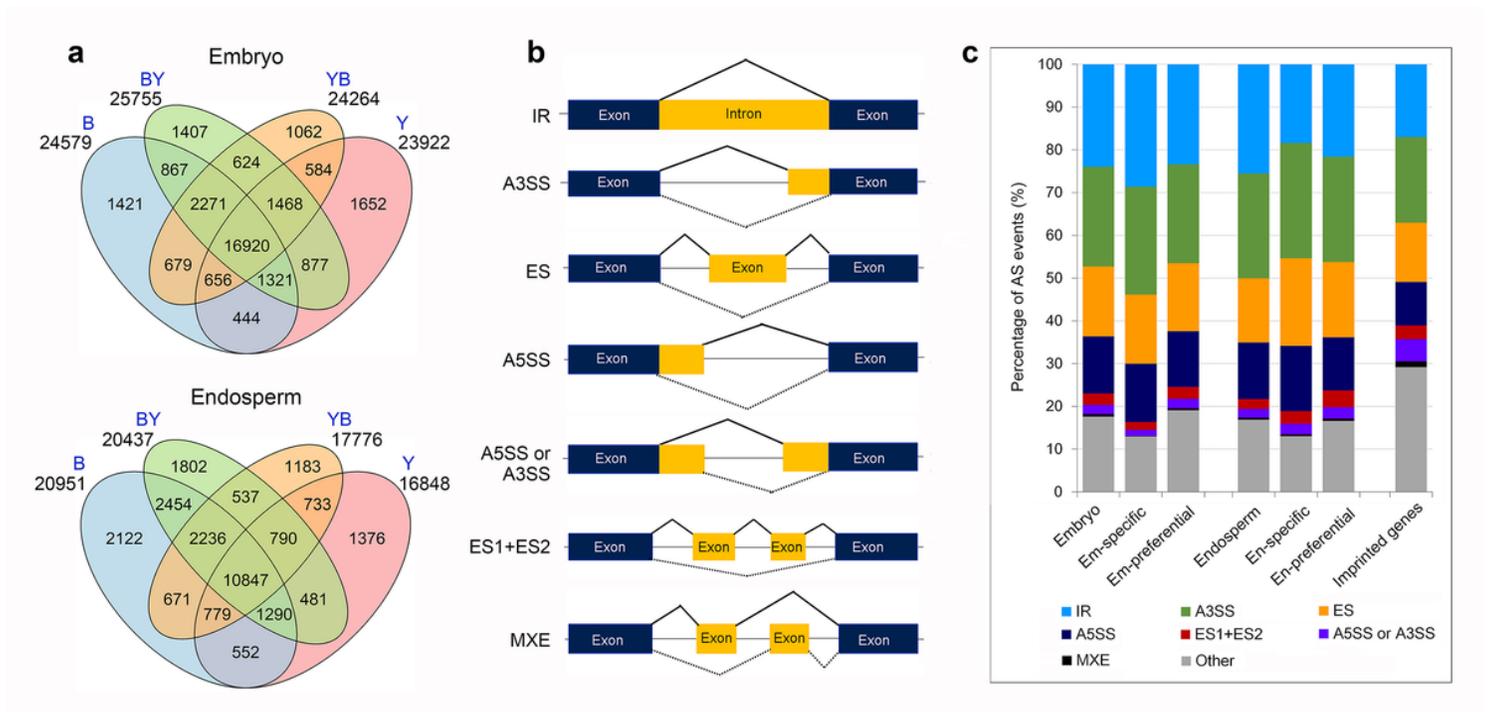


Figure 4

Summary of alternative splicing (AS) events in sorghum embryo and endosperm. a. Venn diagrams show AS events among four genotypes in embryo and endosperm, respectively. b. Visualization of seven AS modes. The exon-intron structure of the AS events are illustrated. IR: intron retention; A3SS: alternative 3' splicing site; ES: exon skipping; A5SS: alternative 5' splicing site; ES1 + ES2: skipping two exons; MXE: mutually exclusive exon. c. Distribution of various splicing patterns in DEGs of embryo and endosperm, as well as in sorghum imprinted genes. BY (BTx623 × Y49) and YB (Y49 × BTx623) are reciprocal F1 hybrids, while B (BTx623) and Y (Y49) are their parental pure lines.

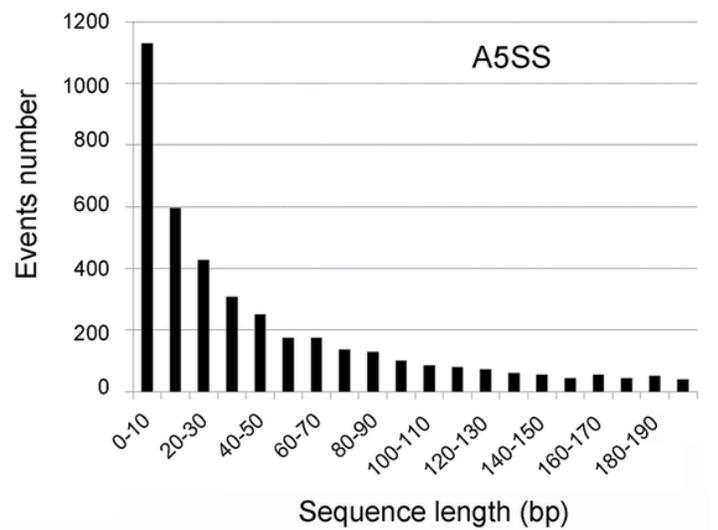
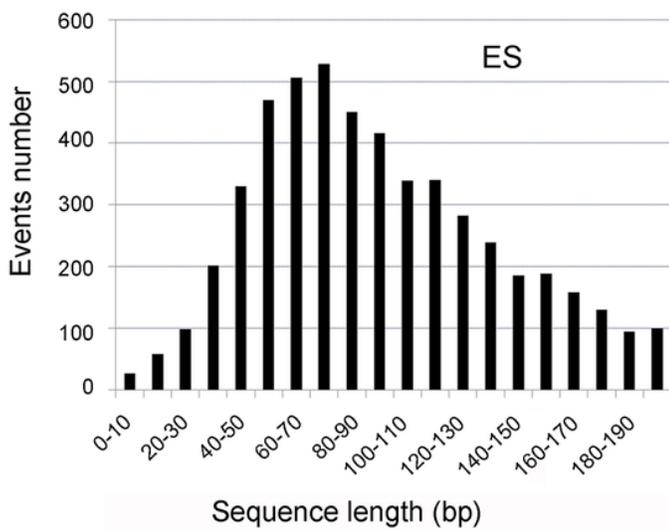
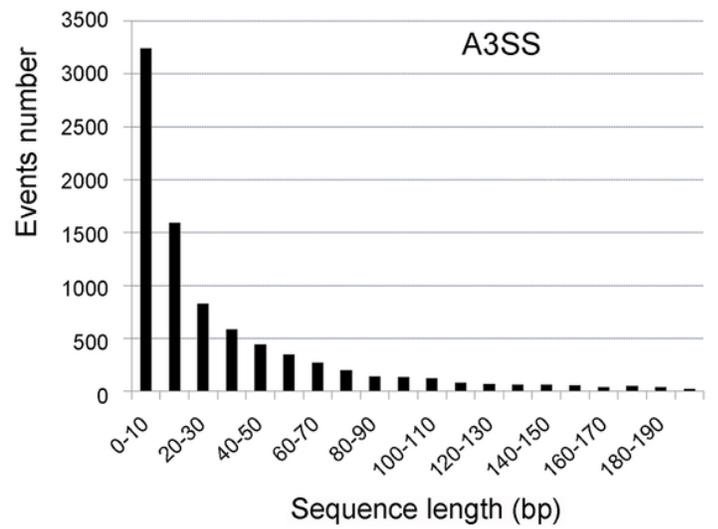
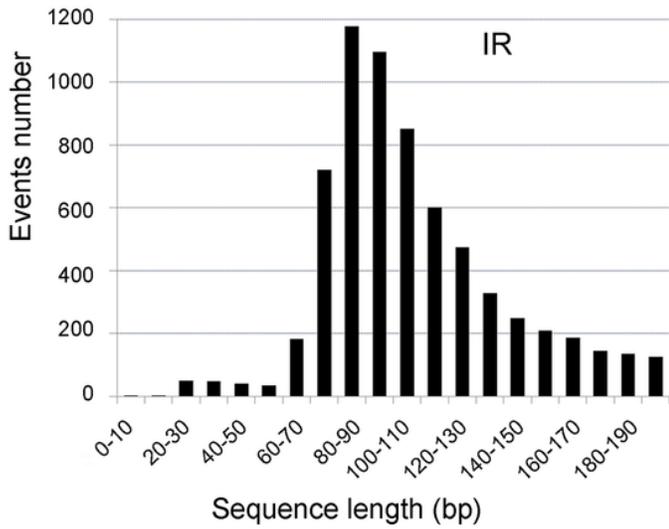


Figure 5

Sequence length distribution of the four main AS types. IR: intron retention; A3SS: alternative 3' splicing site; ES: exon skipping; A5SS: alternative 5' splicing site.

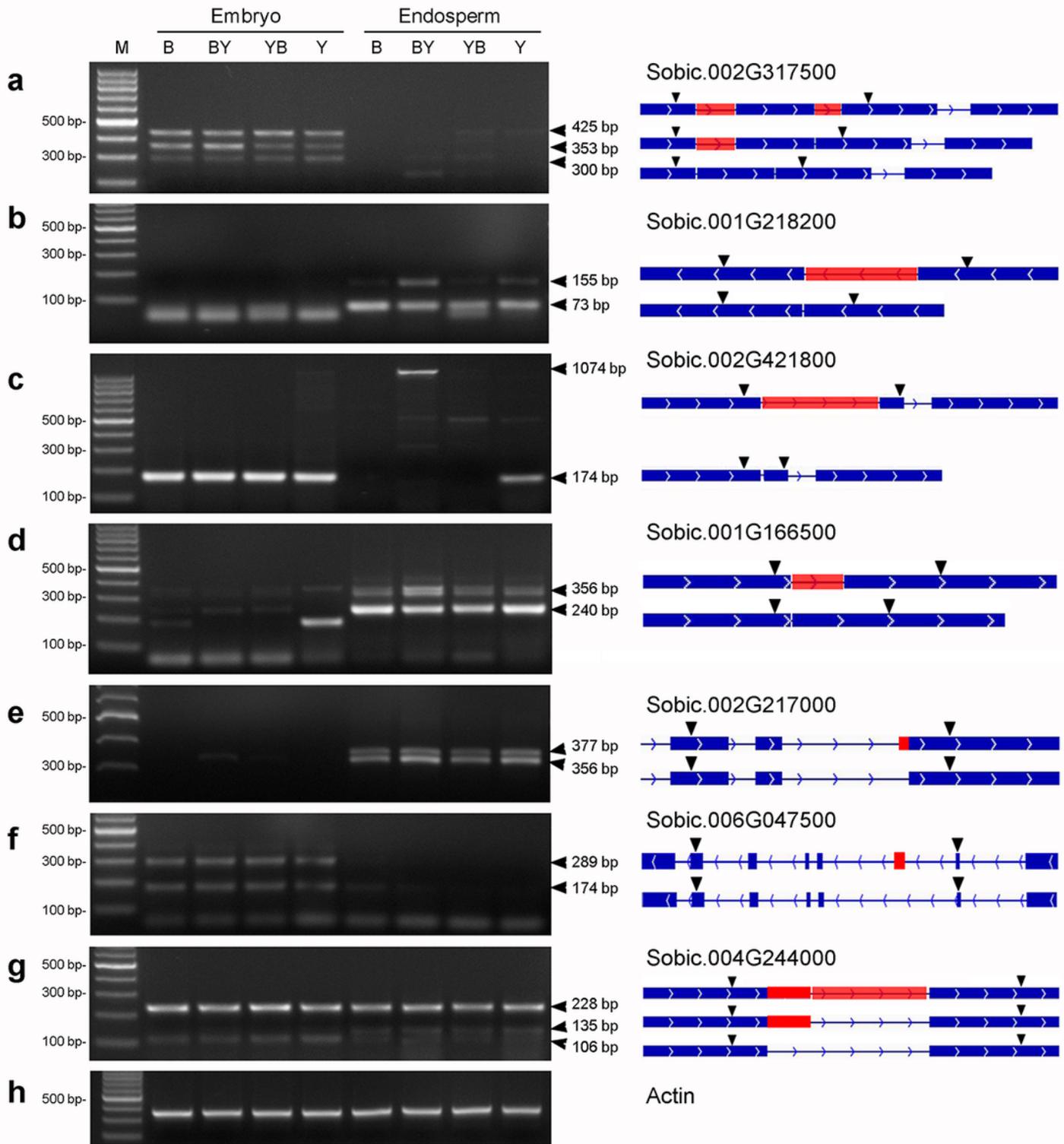


Figure 6

RT-PCR validation of DEGs and AS events in sorghum embryo and endosperm. In the agarose gel images, arrows indicate the alternative transcripts. The length (bp) of PCR product was indicated. The gene IDs were shown above the gene model. In the gene model diagrams (not drawn to scale), blue bar indicates exon; red bar indicates the alternatively spliced regions; blue line indicates intron; the arrowhead on the gene structure diagrams indicates the position of PCR primers used for RT-PCR. a. Embryo-specific

expression and intron retention (IR); b. Endosperm-specific expression and IR; c. Embryo-preferential expression and IR; d. Endosperm-preferential expression and IR; e. Endosperm-specific expression and alternative 3' splicing site (A3SS). f. Embryo-specific expression and exon skipping (ES); g. IR and alternative 5' splicing site (A5SS); h. A sorghum actin gene used as control. BY (BTx623 × Y49) and YB (Y49 × BTx623) are reciprocal F1 hybrids, while B (BTx623) and Y (Y49) are their parental pure lines.

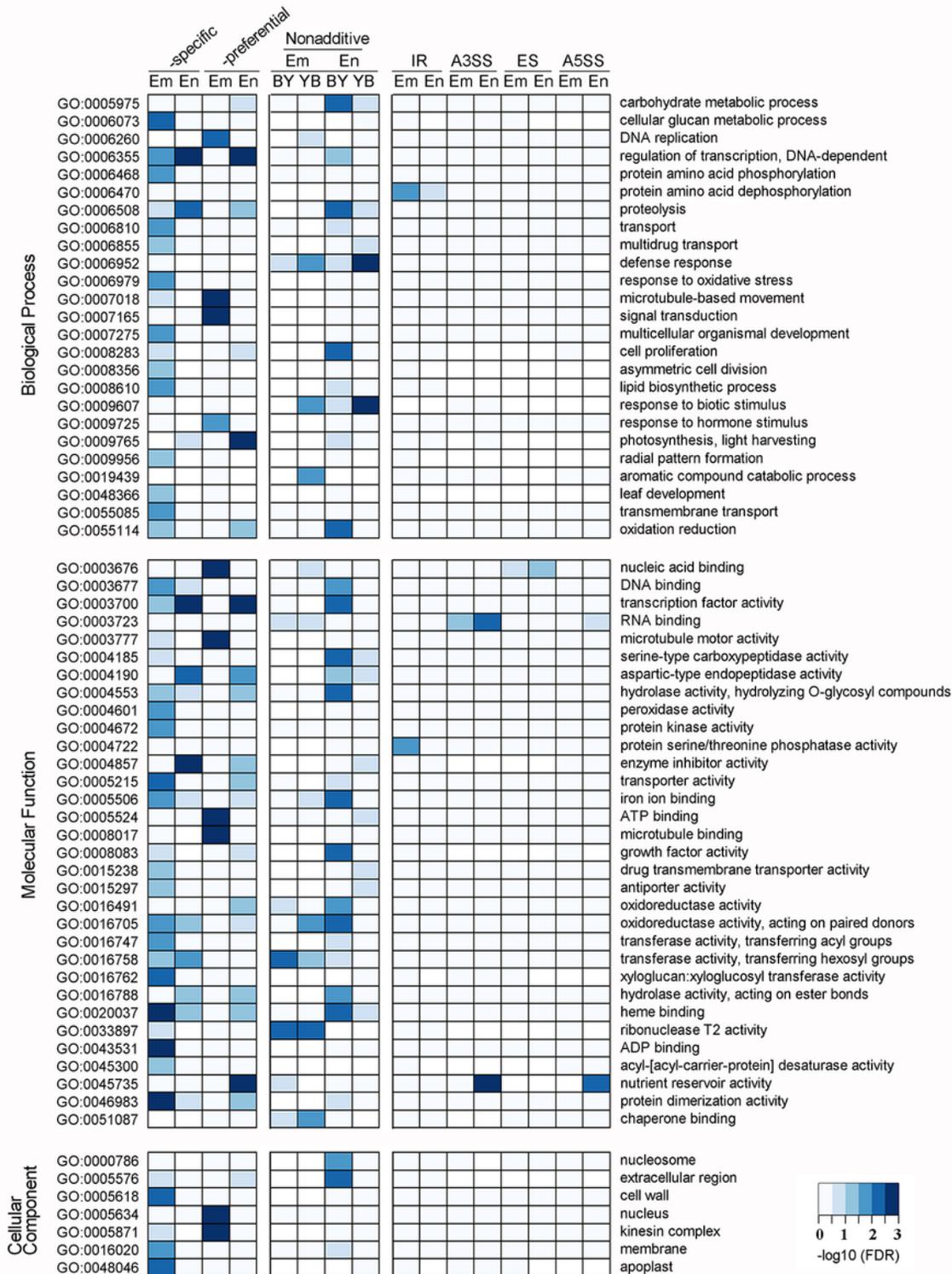


Figure 7

Log10 (FDR) fold change heatmap showing gene ontology (GO) terms of DEGs and AS genes in sorghum embryo and endosperm. GO terms with significant enrichment (q value < 0.05) in at least one comparison were shown. Color in heatmap represent the log10 (q value). Em and En are abbreviations for embryo and endosperm, respectively. IR: intron retention; A3SS: alternative 3' splicing site; ES: exon skipping; A5SS: alternative 5' splicing site.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationZhangetal.2.pdf](#)