ORIGINAL ARTICLE

Transcriptome analysis of alternative splicing in different moso bamboo tissues

Long Li[1,2,3] · Qianqian Shi[3,4] · Dan Hou[2] · Zhanchao Cheng[2] · Juan Li[2] · Yanjun Ma[2] · Xueping Li[2] · Shaohua Mu[2] · Jian Gao[2]

Abstract

Alternative splicing (AS) is a key regulatory mechanism associated with proteome and transcriptome diversity. However, the prevalence of AS in the moso bamboo genome is still unclear. Transcriptome sequencing of four different tissues indicated that 36.17% of the genes in the moso bamboo genome undergo AS. The predominant type of AS found in moso bamboo was intron retention (38.70%), followed by alternative 5′ (31.86%) and 3′ (16.68%) splice sites and exon skipping (11.46%). The number of AS events in every gene was relevant to gene characteristics and gene expression. We also observed that the alternative 5′ and 3′ splice sites were significantly enriched in the fourth nucleotide downstream or upstream of the dominant splice sites. Furthermore, the frequency of AS types as well as the quantity of AS events significantly varied among different tissues. The findings of the present study provide a comprehensive view of AS events in moso bamboo.

Keywords Alternative splicing · Gene characteristics · Tissues

Abbreviations

AA	Alternative accept site
AD	Alternative donor site
ARF	Auxin response factor
AS	Alternative splicing
CDS	Sequence coding for amino acids in protein
DEG	Differentially expressed genes
FPKM	Fragments per kilobase of transcript per million mapped reads
GO	Gene ontology
ORF	Open reading frame
RPKM	Reads per kilobase of transcript per million mapped reads
SE	Skipped exon
SS	AS site selection
UTR	Untranslated regions

✉ Jian Gao
gaojian@icbr.ac.cn

1 College of Forestry, Northwest A&F University, Yangling 712100, Shaanxi, People's Republic of China

2 International Center for Bamboo and Rattan, Key Laboratory of Bamboo and Rattan Science and Technology, State Forestry Administration, Beijing 100102, People's Republic of China

3 Key Laboratory of Silviculture on the Loess Plateau State Forestry Administration, Yangling 712100, Shaanxi, People's Republic of China

4 College of Landscape Architecture and Art, Northwest A&F University, Yangling 712100, Shaanxi, People's Republic of China

Introduction

Alternative splicing (AS) of pre-mRNA is a common, crucial regulatory process in eukaryotes, generating various transcripts from a single gene locus (Emilio et al. 2015). Gilbert (1978) was the first to suggest that a template can generate different isoforms via AS. The previous studies have indicated that almost all human genes undergo AS (Wang et al. 2008). Recently, it has been revealed that AS occurs in 60% of intron-containing genes in *Arabidopsis thaliana* (Filichkin et al. 2010; Marquez et al. 2012), 52% of these genes in *Glycine max* (Shen et al. 2014), 40% in *Zea mays* (Thatcher et al. 2014), and 33% in *Oryza sativa*

(Zhang 2010). The real percentages might even be higher than reported, as increasing numbers of low-abundance splicing variants are being revealed with the development of sequencing technology. For *Arabidopsis*, the reported percentage of AS genes increased dramatically within a decade, from 1.2% in 2003 (Zhu et al. 2003) to 11.6% in 2004 (Iida et al. 2004), more than 30% in 2006 (Campbell et al. 2006), 42% in 2010 (Filichkin et al. 2010), and 61% in 2012 (Marquez et al. 2012). It has been reported that several AS events are conserved among various species (Birzele et al., 2008). For example, compared with those observed for a distant *SCL33* orthologue from *Arabidopsis*, the AS events observed for *Brachypodium distachyon SCL33* are strikingly conserved (Kranthi and Karen-Beth 2015). However, a large number of AS events are specific to individual species.

In general, AS events can occur via four major mechanisms: alternative 5′ splice sites (5′ SSs), exon skipping, intron retention, and alternative 3′ splice sites (3′ SSs) (Huang et al. 2015). In animals, exon skipping is the dominant AS event, while the frequency of intron retention is low; for instance, one-third of all AS events in humans are a result of exon skipping, whereas only 0.01% is due to intron retention (Wang et al. 2008). By contrast, intron retention is the main AS event, whereas exon skipping occurs with very low frequency in plants, such as in *Arabidopsis* (Filichkin and Mockler 2012), *Brachypodium* (Kranthi and Karen-Beth 2015), and *G. max* (Shen et al. 2014). AS events are involved in a wide array of biological functions, including plant development, biological rhythms, biotic and abiotic stress responses, and flowering in particular (Howard et al. 2013; Reddy et al. 2013; Staiger and Brown 2013). In plants, a number of AS events show tissue specificity (Yoshimura et al. 2002). Within a given tissue of soybean, the number of AS events is higher at younger developmental stages than that at older developmental stages (Shen et al. 2014). In addition, abiotic stress can have an impact on the AS of pre-mRNAs (Ali et al. 2001; Palusa et al. 2007) and cause alternatively spliced genes to be over-expressed (Filichkin et al. 2010). The serine/arginine-rich (SR) family plays crucial roles in the pre-mRNA splicing regulation and is dramatically altered under various abiotic stresses (Palusa et al. 2007; Rauch et al. 2014). In addition, environmental signals control the expression levels, AS patterns, phosphorylation status, and subcellular distribution of SR proteins (Kalyna and Barta 2004; Reddy et al. 2013), indicating that altered expression levels of splice variants in response to various abiotic or biotic stresses may play a role in the improvement of plant stress resistance. In addition, AS affects the activity, localization, stability, and interaction capacity of transcripts (Graveley 2009). However, the AS characteristics and AS ratios in different moso bamboo tissues have not been explored thus far.

Moso bamboo is a large family of woody bamboos with the highest economic and ecological value in China (Gao et al. 2014). Its bamboo shoots grow as high as 1 m within 1 day and grow to a final height of 5–20 m at 1.5–2 months (Peng et al. 2010). Due to the striking growth rate, it is the foremost non-timber forestry resource in southeast China. Interestingly, moso bamboo often simultaneously flowers and dies after flowering (Peng et al. 2013a, b). However, moso bamboo rarely flowers in natural settings. Transcriptomic studies in moso bamboo shoots (Peng et al. 2013a) have revealed that hormone signalling-associated genes play a vital role in the fast-growing shoots. RNA sequencing suggested that the flowering of bamboo may be influenced by drought stress (Gao et al. 2014). Although bioinformatics analyses have been carried out in moso bamboo, the genome-wide study of the dynamic variation of AS in different tissues and different developmental stages is still relatively scarce. Here, we performed transcriptome analyses of different tissues in moso bamboo based on the Illumina HiSeq™ 2000 platform and described the features of AS in different tissues. We believe that our work will provide useful resources for future genomic research on moso bamboo.

Materials and methods

Plant materials

Samples of moso bamboo shoots (mixed by different growth stages) and mature culms were collected in Huoshan County (E115u529–116u329; N31u039–31u339), Anhui Province, China, from spring to autumn 2014. Mixed flowering samples from different floral stages and leaf samples were harvested in Guilin City (E110u179–110u479; N25u049–25u489), Guangxi Zhuang Autonomous Region, China, from April to August, 2014. All samples were harvested from sites suitable for moso bamboo without human intervention or insect pests (Gao et al. 2014; Peng et al. 2013a).

cDNA libraries construction and transcriptome sequencing

The TRIzol reagent was used to isolate total RNA from each sample (Invitrogen, California, USA). After examining the purity and integrity of all RNAs, strand-specific transcriptome sequencing libraries of four different samples were constructed using an Illumina kit (Illumina, San Diego, CA, USA). Thereafter, these four cDNA libraries were sequenced on the HiSeq™ 2500 platform from Illumina and generated 101-nt paired-end reads at the Beijing Genomics Institute (Shenzhen, China). The detailed methods were described in the previous studies (Gao et al. 2014; Peng et al. 2013a).

Read alignment and gene expression analysis

After removing adapter sequences and low-quality reads, the remaining clean reads were assembled according to moso bamboo genome database (http://www.bamboogdb.org/) (Peng et al. 2013b) using SOAP aligner/SOAP2 (Li et al. 2008), allowing up to five mismatches. The expression levels of genes and isoforms were calculated using Cufflinks software and were then normalized to transcripts per million (reads per kilobases per million reads, RPKM) values (Trapnell et al. 2010).

AS event identification, statistical analysis, and gene ontology (GO) annotation

The repertoire of AS events was established using ASTA-LAVISTA. Four main types of AS events, including intron retention, exon skipping, 3′ SSs, and 5′ SSs, were analysed with the methods described by Sammeth et al. (2008). Exon lengths and exon numbers were calculated based on the gtf files obtained from the moso Bamboo Genome Database as follows: mRNA length = $exon_1$ length + $exon_2$ length + ⋯ + $exon_n$ length, assuming that one mRNA was formed by n exons. Weblogo was used to visualize the nucleic acid sequences from the 5′ SSs and 3′ SSs. In addition, genes with AS events were automatically annotated and categorized using the blast2go program (Conesa et al. 2005). AS events were classified into three categories based on their positions at different mRNA features, including 5′ untranslated regions (UTRs), open reading frames (ORFs), and 3′ UTRs. The ORFs within AS events could cause frame shifts if the altered fragments are not filtered in multiples of three nucleotides. mRNA sequences were assembled according to moso bamboo genome database and gtf files using Perl code, which was provided by the Biomarker company (Biomarker, Beijing, China), and then were used to compute GC content. After the isoforms being translated into amino acids according to the longest ORF using Getorf software, the conserved domains of transcript isoforms were identified by an online conserved domain program service of the National Center for Biotechnology Information (NCBI). The expression levels of the isoforms in different tissues were hierarchically clustered using Cluster 3.0 software based on Euclidean distance with complete linkage.

Reverse transcription PCR (RT-PCR) validation

Eight genes exhibiting 11 AS events were randomly selected and used for further experimental verification. Semi-quantitative RT-PCR was performed with specific primers that were designed using Oligo 7 software (Supplementary material: Table S1). Total RNA from four different tissues was extracted according to the method described above using TRIzol and was treated with DNAase I. cDNA was synthesized with random primers using an M-MLVRT kit (Promega, Madison, WI, USA). The thermal conditions for the PCR procedures were set as follows: 94 °C, 5 min; 32 cycles: 94 °C for 30 s, 60 °C for 50 s, and 72 °C for 60 s; and then 72 °C, 10 min. The purified PCR products were transformed into pMD18-T vector (Takara, Dalian, China), which was then transferred into the *Escherichia coli* DH5α bacterial strain (Takara, Dalian, China). After blue–white selection, the positive clones were further sequenced using ABI 3100 DNA sequencer (Applied Biosystems, Foster, CA, USA).

Results

Quality analyses of transcriptome sequencing data

To identify AS events at the whole-genome-wide level and evaluate potential factors influencing AS in moso bamboo, transcriptomic analyses of culms, flowers, leaves, and shoots was performed using Illumina HiSeq 2500 sequencing technology. We obtained a total of 269.32 million reads, among which 52,805,612 (81.18%), 53,035,796 (78.43%), 52,489,437 (76.85%), and 55,520,086 (81.13%) clean reads from the culm, flower, leaf, and shoot libraries, respectively, were matched to the referenced moso bamboo genome (Table 1).

Characteristics of different AS types

In total, 29,731 AS events were identified. In moso bamboo genome, approximately 36.17% (11,344 genes) of all annotated genes were predicted to undergo AS events (Fig. 1a). Among the 11,344 AS genes, 5389 genes exhibited only one AS event. In contrast, only 117 genes exhibited more than 11 AS events. In general, the number of genes displaying fewer AS events was much higher than the number displaying more AS events (Fig. 1b). Furthermore, we calculated the AS distribution in each moso bamboo tissue and observed no significant differences among different tissues (Supplementary material: Figure S1).

The AS events could be classified into four main categories: intron retention, skipped exons, 5′ SSs, and 3′ SSs. We found that the most common event among the four tissues was intron retention, accounting for 38.70% of the AS events, followed by 3′ SSs (31.86%), 5′ SSs (16.68%), and skipped exons (11.46%). Nearly 64.7% of the AS events in moso bamboo affected coding regions. Another 35% occurred within non-coding regions (19% in 5′ UTRs and 16% in 3′ UTRs) (Fig. 1c). In addition, we found that up to 64.7% of AS events were located in ORFs (42.1% of the total AS events). Furthermore, approximately 35.3% (28.2% of total AS events) of the AS locating in ORFs maintained the

Table 1 Alignment statistics results

	Culm		Flower		Leaf		Shoot	
	Number	Percentage	Number	Percentage	Number	Percentage	Number	Percentage
Total reads	65,045,670	100.00	67,617,866	100.00	68,305,136	100.00	68,431,884	100.00
Total base pairs	5,854,110,300	100.00	6,085,607,940	100.00	6,147,462,240	100.00	6.16E+09	100.00
Total mapped reads	52,805,612	81.18	53,035,796	78.43	52,489,347	76.85	55,520,086	81.13
Perfect match	41,749,711	64.19	39,386,902	58.25	39,026,185	57.14	44,359,624	64.82
≤5 bp mismatch	11,055,901	17.00	13,648,894	20.19	13,463,162	19.71	11,160,462	16.31
Unique match	47,690,902	73.32	47,709,687	70.56	47,847,557	70.05	51,402,531	75.11
Multi-position match	5,114,710	7.86	5,326,109	7.88	4,641,790	6.80	4,117,555	6.02
Total unmapped reads	12,240,058	18.82	14,582,070	21.57	15,815,789	23.15	12,911,798	18.87
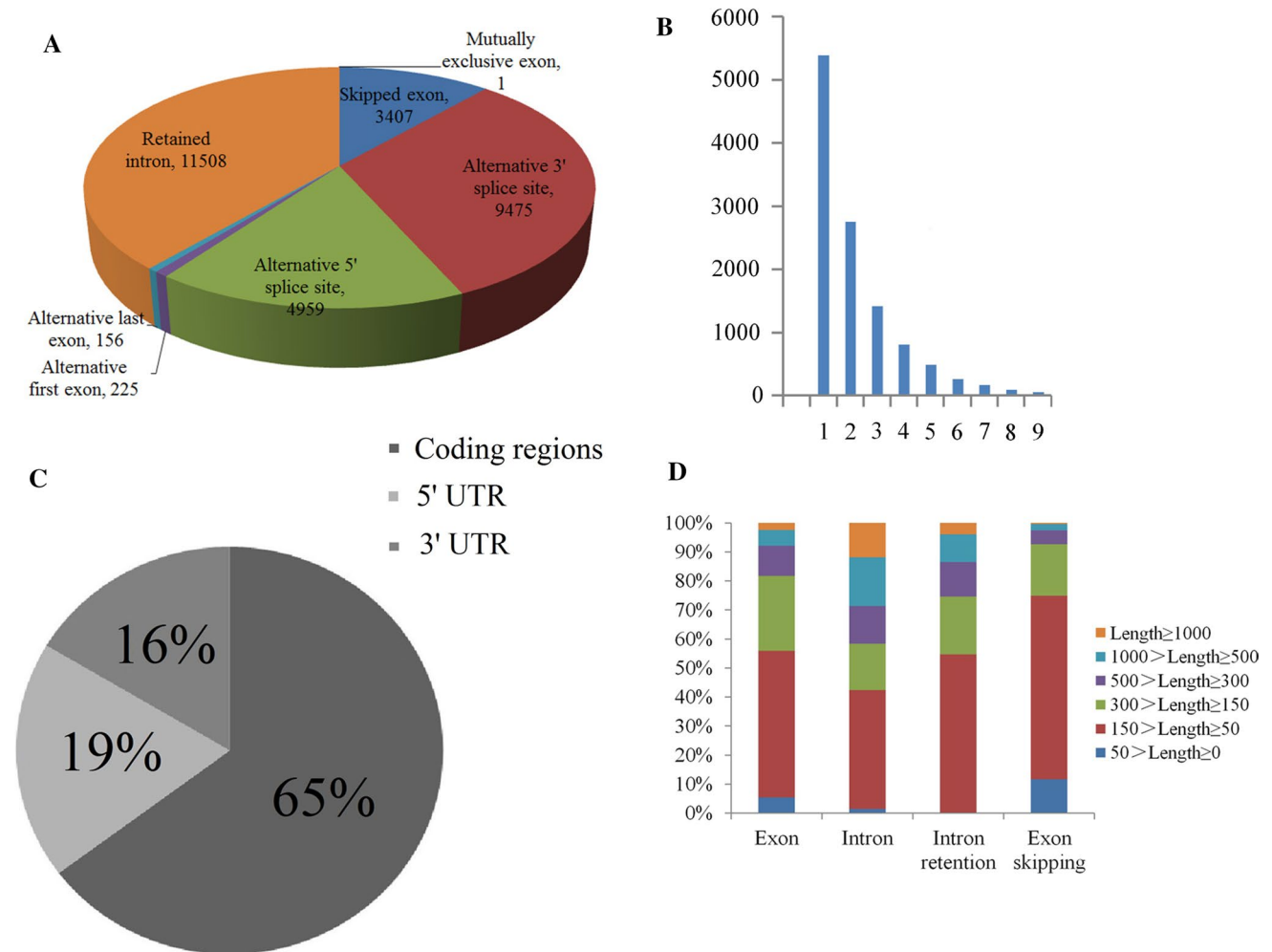
Fig. 1 **a** Percentage of different AS types. **b** Distribution of different numbers of AS events per gene. **c** Localization of AS events in mRNAs. **d** Length distribution of exons, introns, retained introns, and skipped exons

original reading frame, because the change in length was a multiple of three nucleotides (Fig. 1d).

The size distribution of retained introns was considerably smaller than the size of all introns (average sizes of 231.01 and 480.93 bp, respectively). Similarly, the size distribution of skipped exons was considerably smaller than that of all exons (average sizes of 123.97 and 269.86 bp, respectively) (Fig. 1d).

The positions of splice sites are another characteristic of an AS event. In this study, AS site selection (alternative 5′ SS and 3′ SS) was found to be significantly enriched four nucleotide downstream or upstream of the dominant splice sites (Fig. 2a, b). Sequence analysis of 5′ SSs and 3′ SSs indicated that activated splice sites also involved AG and GU dinucleotides (Fig. 2c, d).

## Analysis of the relationship between gene characteristics and AS frequency

To determine putative factors influencing AS, we analysed the relationship between the AS number and gene characteristics such as GC content, intron length, exon length, exon number, and mRNA length (Fig. 3a, b). Genes that presented equal to or greater than 11 AS events exhibited the longest mRNA lengths (median length: 3650 bp) and greatest exon numbers (median exon number: 12). In terms of exon length and GC content, genes that could generate more isoforms presented lower GC contents and shorter exon lengths (Fig. 3c, d). Non-AS genes display much greater GC

contents (median content: 60%) and exon lengths (median length: 146 bp) than AS genes. Among all four tissues, AS genes showed higher expression levels than non-AS genes (Fig. 4).

Comparison of AS events in different tissues

We next compared the differences in AS between different tissues. AS events occurred more frequently in leaves than in flowers, shoots, and culms (Fig. 4). In leaves, we identified a total of 15,575 AS events, including 2319 5′ SSs, 4412 3′ SSs, 6570 intron retention, and 2113 exon skipping events. In contrast, only 11,117 AS events, including 1348 5′ SSs, 2860 3′ SSs, 1953 skipped exons, and 4831 intron retention events, were identified in culms (Fig. 5).

The analysis of AS events shared among different tissues indicated that most AS events did not occur in all samples. Among the four tissues, leaves presented the most specific AS events, including 272 skipped exons, 1300 5′ SSs, 1702 3′ SSs, and 2536 intron retention events, while 750 skipped
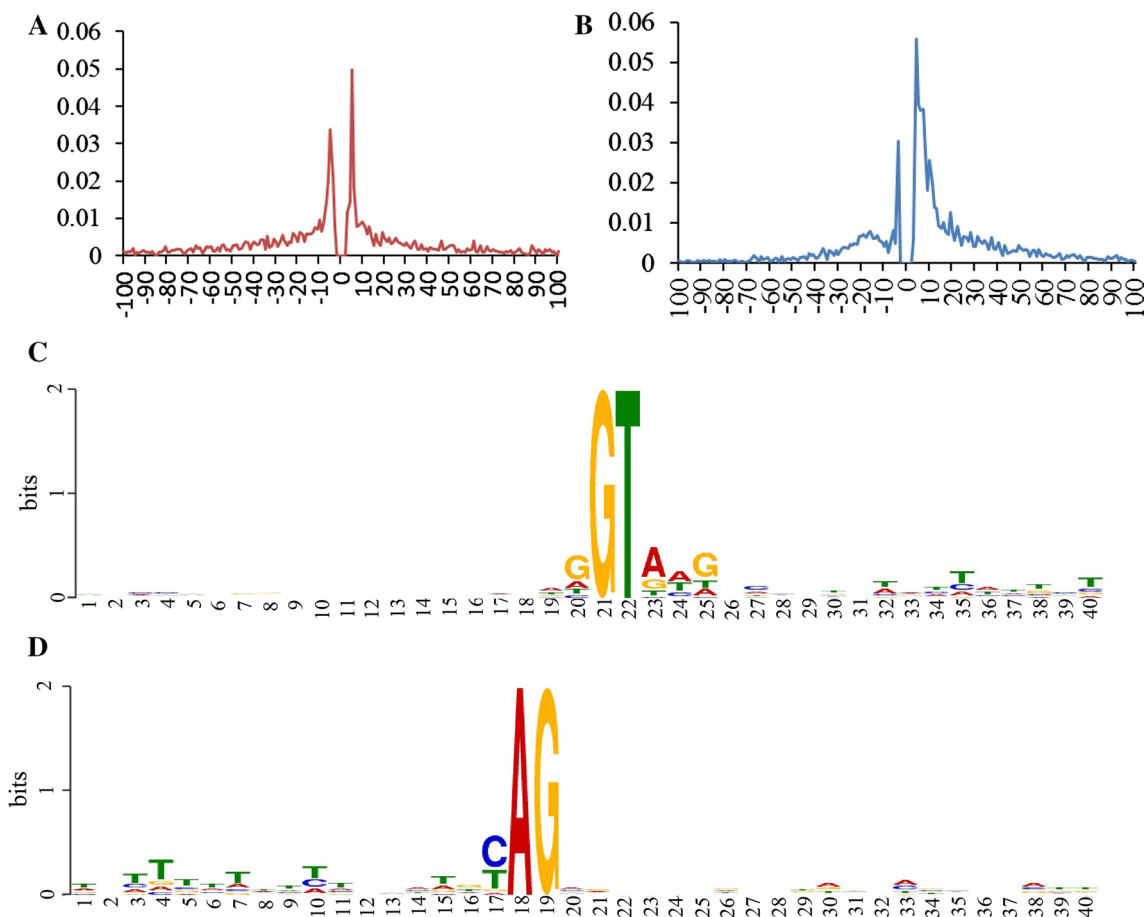

Fig. 2 Distribution of the activated 5′ SSs (**a**) and the activated 3′ SSs (**b**) around the dominant splice sites. Weblogo was used to visualize the nucleic acid sequences around the 5′ SSs (**c**) and 3′ SSs (**d**)

Fig. 3 **a** Comparison of mRNA length in different numbers of AS events in each gene. **b** Comparison of exon numbers in different numbers of AS events in each gene. **c** Comparison of GC content in different numbers of AS events in each gene. **d** Comparison of exon length in different numbers of AS events in each gene. The *x*-axis represents different numbers of AS events in each gene, and the *y*-axes in **a**–**d** represent mRNA length, exon number, GC content, and exon length, respectively
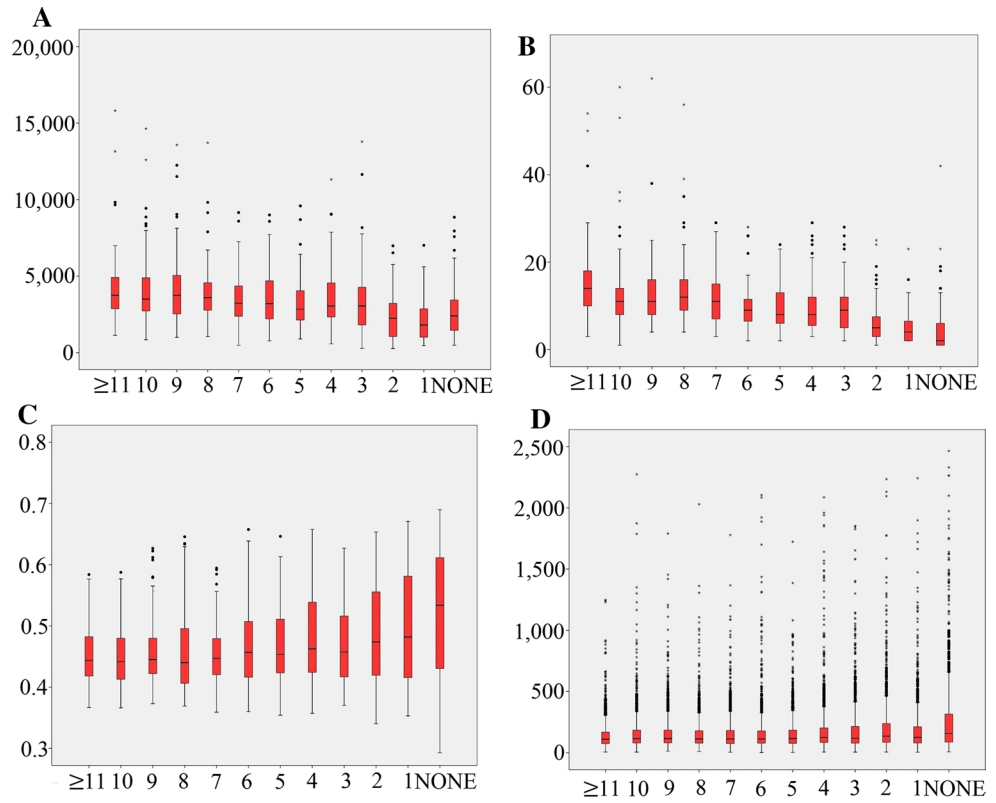

Fig. 4 Expression profiles of AS occurrence-related genes and non-AS genes in different tissues. The *y*-axis represents RPKM. **a** Culms, **b** flowers, **c** leaves, and **d** shoots
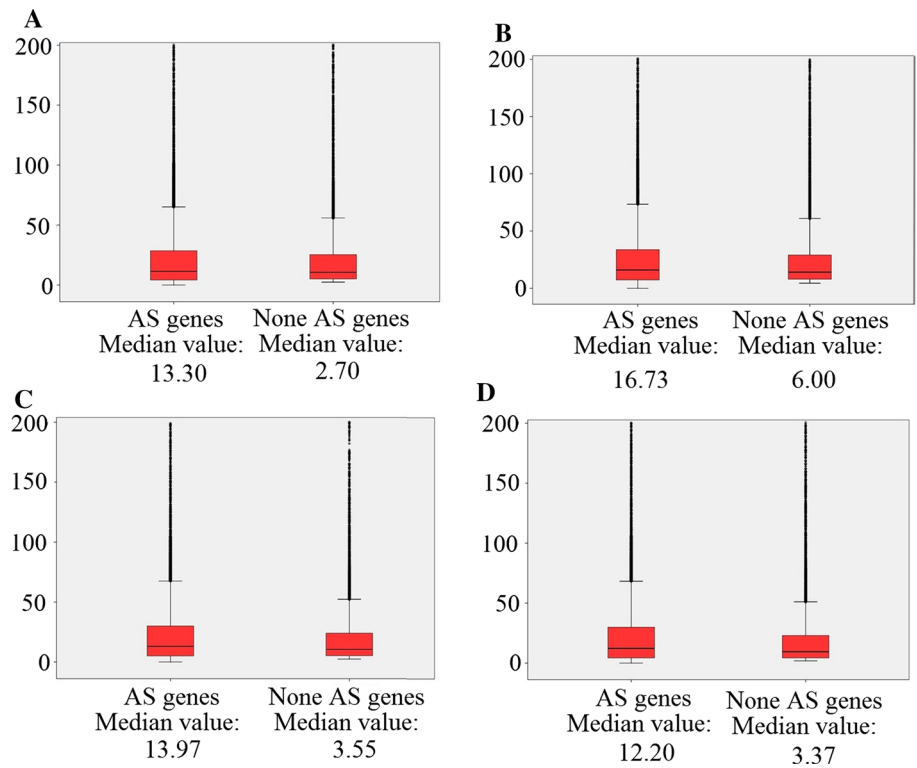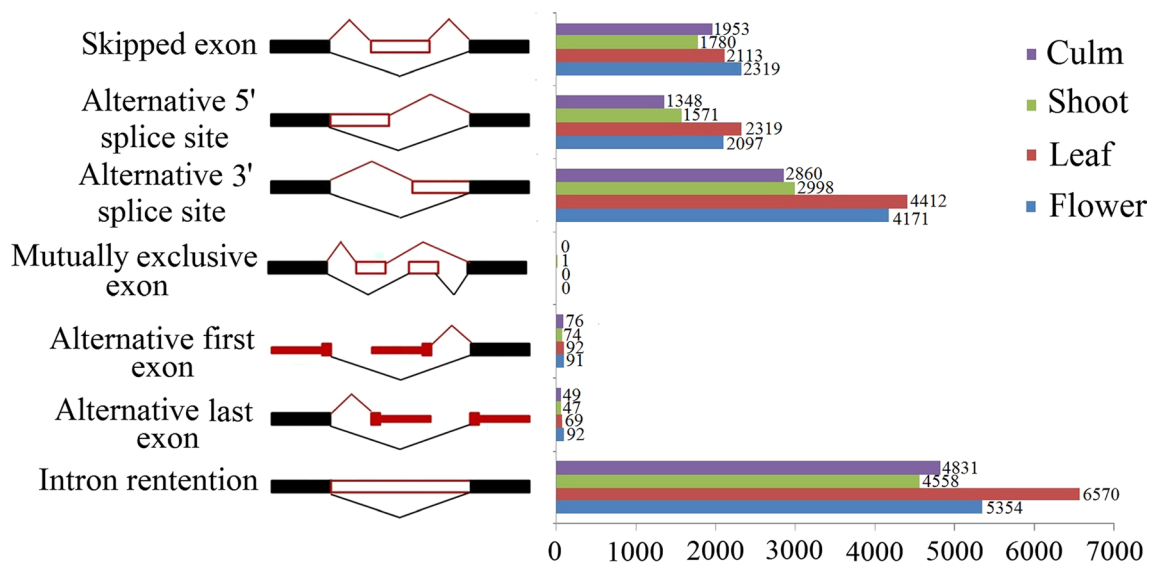
Fig. 5 Numbers of different types of AS event in culm, shoot, leave, and flower
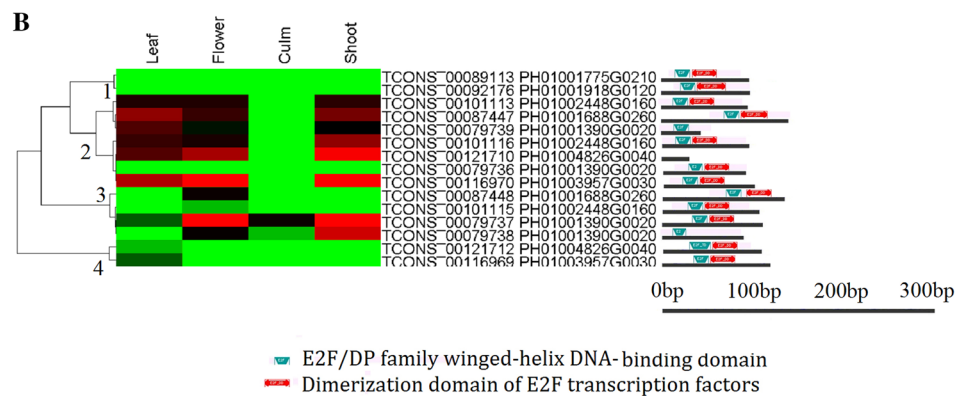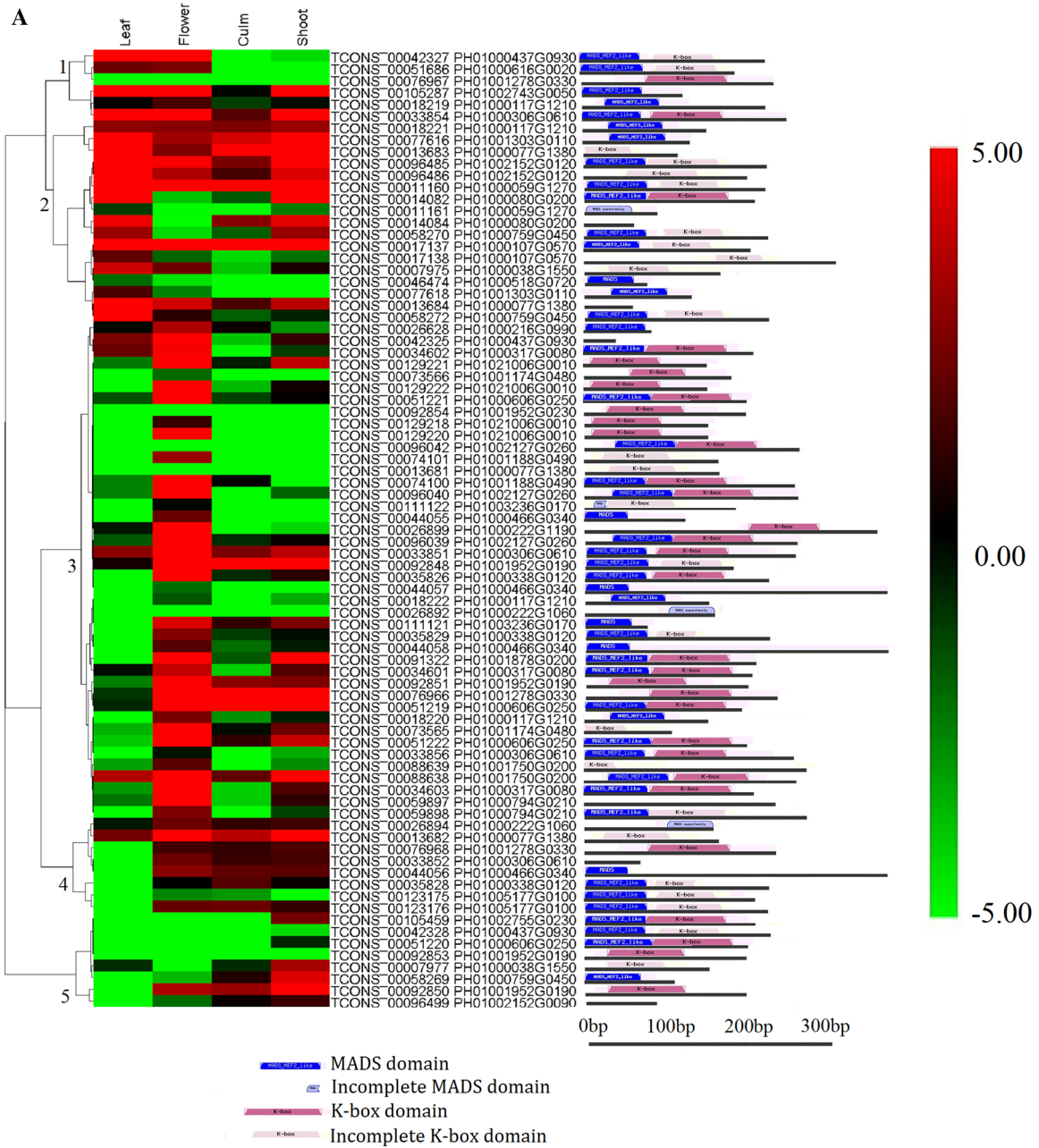
Fig. 6 Venn analyses of various types of AS events in different tissues. **a** Skipped exon, **b** alternative 5′ site, **c** alternative 3′ site, and **d** intron retention
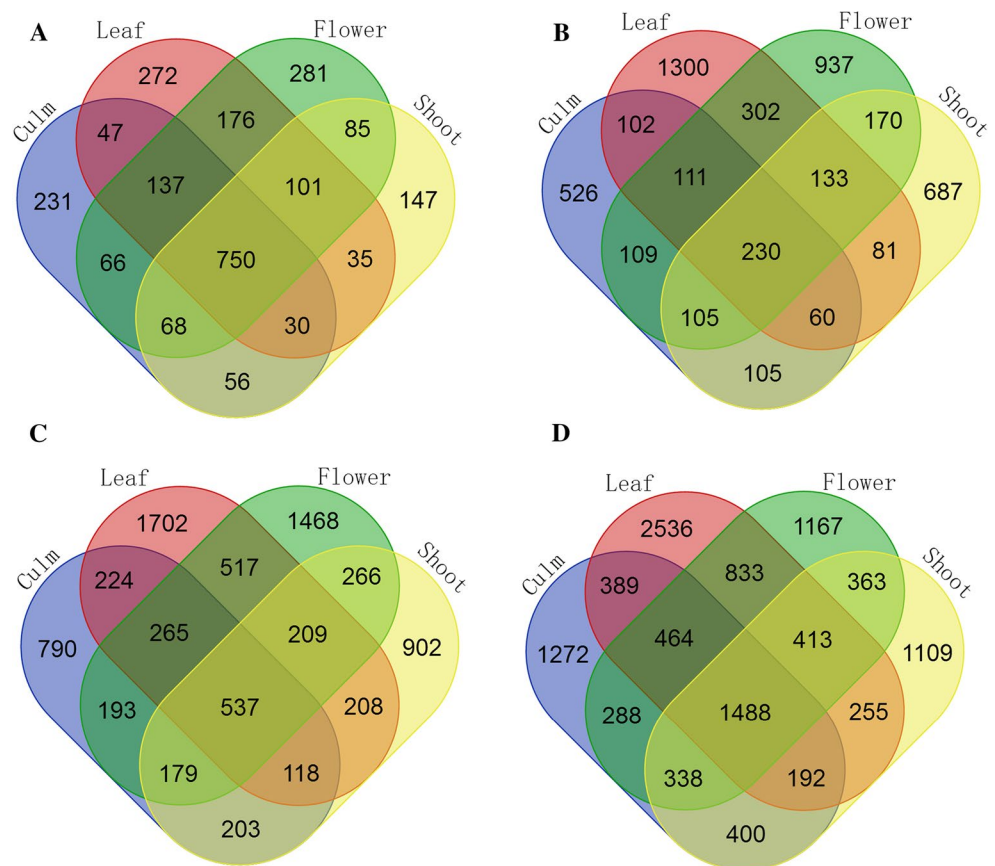
exons, 230 5′ SSs, 537 3′ SSs, and 1488 intron retention events could be detected in all samples in this study (Fig. 6).

Functional analysis of alternatively spliced genes

To analyse the functions of alternatively spliced genes, GO term enrichment was investigated using blast2go. A total of

A

MADS domain
Incomplete MADS domain
K-box domain
Incomplete K-box domain

B

E2F/DP family winged-helix DNA-binding domain
Dimerization domain of E2F transcription factors

◄**Fig. 7** Expression levels and conserved domain of isoforms of MADS (**a**) and E2F (**b**) families. Fold difference was designated the log2-transformed reads per kilobase per million reads (RPKM) value. Green rectangles, black rectangles, and red rectangles represent low, intermediate, and high expression levels, respectively. The schematic diagrams of all conserved domains are shown under the corresponding heatmap

11,344 AS occurrence-related genes were categorized into the three primary GO categories of cellular components, molecular functions, and biological processes (Table S1). Among cellular components, "chloroplast", "cytosol", and "organelle part" were highly represented. Among molecular functions, "catalytic activity" was the dominant GO term, followed by "binding" and "ion binding" (Table S1). Within the biological category, the groups with the most significant enriched AS genes were associated with "single-organism cellular process", "Single-organism metabolic process", and "metabolic process" categories (Supplementary material: Table S2).

Expression and conserved domain analysis of the MADS and E2F gene families in different tissues

MADS and E2F gene families play vital roles in shoot elongation and flower development (Li et al. 2018; Cheng et al. 2017). To evaluate the effects of AS events on the MADS and E2F gene families, domain modification and loss of these two important families were analysed. In total, 35 MADS-box genes were identified from moso bamboo. Conserved domain analysis indicated that five truncated MADS isoforms lost all functional domains, and 48 isoforms contained the integrated MADS-box domain (Fig. 7a). The expression profiles of MADS isoforms were clustered into five clusters, among which cluster 3 accounted for the largest proportion of isoforms and showed high abundance in flowers. Expression analysis indicated that the many of the different isoforms generated from a single gene exhibited distinct expression trends. For example, PH01000606G0250, TCONS_00042325, with integrated MADS-box and k-box domains, showed high abundance in flowers, while TCONS_00042327, whose k-box domain was incomplete, showed high abundance in leaves. However, some splicing variants showed similar expression patterns to the predominant variant. For example, TCONS_00059898 lost all functional domains and showed a similar change in expression to the predominant variant, TCONS_00059897, which contained the integrated MADS-box domain (Fig. 7a).

In the E2F family, with the exception of one isoform (TCONS_0012712), all other isoforms contained the integrated E2F/DP family winged-helix DNA-binding domain (Fig. 7b). Similar to the phenomenon observed in the MADS gene family, the different isoforms generated from a single gene could exhibit similar expression patterns or distinctly different expression patterns.

Experimental verification

To validate the identified AS events, semi-quantitative RT-PCR was conducted for the eight genes that exhibited 11 AS events. Nearly four-fifths of the AS events were detected through RT-PCR verification (Fig. 8). Some electrophoretic bands generated from 5′ SS, 3′ SS, intron retention, or exon skipping events could not be separated via agarose gel electrophoresis, because their AS sites were proximal to the dominant site. However, after gel purification, vector construction, and blue–white selection experiments, several of these AS sites could be confirmed via sequencing validation using an ABI 3100 DNA sequencer. For example, PH01000604G0860 generated two splicing variants (designated a and b), due to 5′ SS and intron retention. After agarose gel electrophoresis, spicing variant b could be separated from PH01000604G0860 (Fig. 8c). In contrast, splicing variant a could not be separated from PH01000604G0860, because splicing variant a was only 12 bp longer than PH01000604G0860. Using the ABI 3100 automated sequencer, both splicing variant a and PH01000604G0860 were confirmed through sequencing validation. In addition, some weakly detected AS events generated from low-abundance genes, such as PH01000346G0780, could not be detected via PCR verification (Fig. 8e).

Discussion

Since the first report on AS by Gilbert (1978), increasing amounts of evidence have indicated that AS exists not only in animals but also in plants. Among the intron-containing genes of plants, nearly 60% undergo AS (Reddy et al. 2013); for example, 35–60% of genes and 33% of genes undergo AS events in *Arabidopsis* and rice, respectively (Marquez et al. 2012; Filichkin et al. 2010). Here, 36.17% of moso bamboo genes were found to be alternatively spliced, with each AS gene undergoing 2.6 AS events on average.

Nearly 35% of the AS events observed in moso bamboo occurred in the 5′ and 3′ UTRs. This percentage is much higher than that reported in *Arabidopsis* (28.5%) (*Reddy* 2007) but very similar to that in rice (34%) (*Matthew* et al. 2006) and *Volvox carteri* (33%) (Arash et al. 2014). AS of UTRs probably plays a vital role in regulating and producing of mRNA diversity. AS of UTRs changes the mRNA secondary structure and in turn affects RNA processing, mRNA stability, and translation of the mRNA (Kudla et al. 2009). Different splice site selection may lead to differences in a series of nucleotides among different isoforms. However, the original ORF could be maintained when
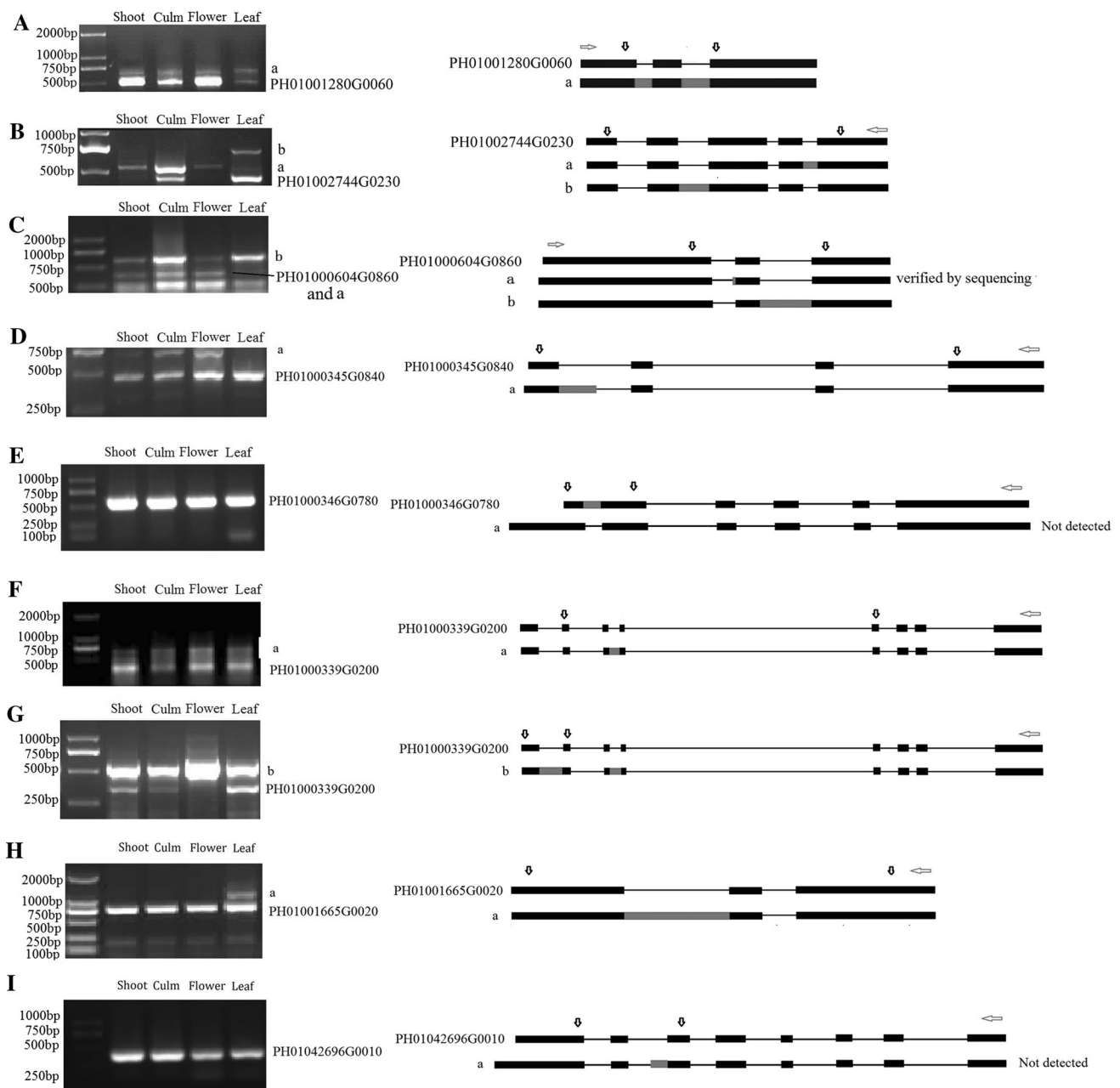
Fig. 8 Verification of AS events via RNA-seq and RT-PCR. **a** PH01001280G0060, **b** PH01002744G0230, **c** PH01000604G0860, **d** PH01000345G0840, **e** PH01000346G0780, **f**, **g** PH01000339G0200, **h** PH01001665G0020, and **i** PH01042696G0010. The gene structure, which was obtained from the Bamboo Genome Database, is indicated on top, and all splice variants identified via transcriptome sequencing are shown under the gene. Grey boxes indicate splicing events in isoforms; black boxes represent constitutively spliced exons; and arrows indicate primers

the change in length caused by the selection of different splice sites is a multiple of three nucleotides (Xu et al. 2014). In this work, approximately 35.3% (28.2% of the total AS events) of the changes in length were multiples of 3 nucleotides, which could maintain the original reading frame. The other 64.7% of changes in ORF length due to AS events could cause frame shifts.

Some studies have suggested that several AS events are organ-specific (Yoshimura et al. 2002). Our analysis of AS events shared among different tissues demonstrated that only a small number of genes underwent AS events in all four examined tissues. The number of AS events in leaves was highest, followed by those in the shoots, flowers, and mature culms. AS events were infrequent at older growth

stages compared with younger developmental stages (Shen et al. 2014; Shawn et al. 2014). A similar observation revealed that the frequency of AS in vigorously growing shoots was much higher than in mature culms. These tissue-specific AS events may act on the specific pathways or networks in concert when genes are co-regulated (Blencowe 2006).

The average intron length of animals is significantly greater than that of plants. It has been reported that the average length of introns in the human genome is 5 kb (Iwata and Gotoh 2011). By comparison, the average intron lengths in moso bamboo and rice are only 481 and 470 bp, respectively. However, it should be noted that outliers with large introns have been found in some plants in the recent years. For example, in *Picea abies* and *Picea glauca*, the lengths of some introns may reach 10 kb (Nystedt et al. 2013; Birol et al. 2013). Thus, we conclude that plant intron lengths may currently be underestimated. Due to the different roles of intron processing (Hartmuth and Barta 1986), intron sizes are quite different in plants and animals, leading that AS types are different in plants and animals. Unlike animals, in which exon skipping accounts for the largest proportion of AS events (Modrek and Lee 2002; Sultan et al. 2008), intron retention is the most common AS type in plants (Filichkin et al. 2010; Marquez et al. 2012). Our statistical results were relatively similar to these findings. In moso bamboo, intron retention was found to be the predominant AS type (38.70%), whereas the percentage of exon skipping among all events was only 8% and alternative 3′ and 5′ SSs were observed at intermediate frequencies.

AS affects protein functions through changing conserved domain modification and binding properties, amino acid sequences and composition, and the stability of secondary structures (Stamm et al. 2005). Protein domains are considered to represent distinct functional units, and domain loss is, therefore, also involved in afunction of proteins (Marchler-Bauer 2015). Statistical analysis indicated that 20 isoforms of the MADS gene family lost their MADS domain, which influenced their DNA-binding ability, thus affecting their function in floral formation and transition. In addition, we found that, in both the MADS and E2F families, isoforms generated from a single gene might exhibit distinct expression patterns. The results illustrated that gene activation and AS were independently regulated.

Statistical analysis of the gene structures of AS genes and non-AS genes indicated that gene structure (e.g., exon numbers, exon length, and mRNA length) could also affect the frequency of AS. More AS events were found to occur in shorter exons, which agreed with the previous reports (Tomasz and Konstantin 2011; Panahi et al. 2015). The possible reason for this pattern might be that exon shuffling events and duplications of exons could greatly enhance the complexity of the transcriptome and genome.

Moreover, 5′ SSs and 3′ SSs mainly occurred at the fourth nucleotides downstream and upstream of the predominant splice sites (Fig. 2d), respectively. These characteristics of 5′ SSs and 3′ SSs were similar with those of *Arabidopsis* genome (Ding et al. 2014). These activated SSs were related with GU and AG dinucleotides, suggesting that this AS pattern is conserved in eukaryotes. Thus, the variation of gene structure, particularly in terms of exon length, exon numbers, GC contents, and the motifs of splicing sites, was found to have a great impact on the frequency of AS in moso bamboo.

Expression analysis of AS genes and non-AS genes in four different tissues indicated that the transcription levels of genes also affected the frequency of AS events. The AS genes exhibited much abundant expression profiles than non-AS genes. However, it should be considered that the status of a transcript regarding whether it is turned on or off depends on its detectability, which in turn depends on the coverage (Nicola et al. 2014). Hence, we believe that some low-abundance genes produced only as small number of isoforms, which were undetectable. These isoforms could be detectable with the development of sequencing technology in the future.

In the present study, 36.17% of moso bamboo genes were subjected to AS. Our comprehensive investigation elucidated the influences of varied gene structures and gene expression levels on AS event, of and provided insights for future functional analyses. In addition, our analysis indicated that splice site sequence could affect the corresponding frequencies of isoforms resulting from altered 5′ SSs and 3′ SSs. These results provide a comprehensive view of AS in moso bamboo, and additional association studies will be aimed at better understanding the relationship of AS with growth and development.

Author contributions statement LL and QS performed the bioinformatics analysis, and drafted and edited the main lines of the text. DH, ZCC, and JL performed some experiments for this paper. XL and SHM helped to prepare the plant material. JG designed this research. All of the authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Ali SM, Chee SK, Yuen GY, Olivo M (2001) Hypericin and hypocrellin induced apoptosis in human mucosal carcinoma cells. J Photochem Photobiol B 65:59–73

Arash K, Cheng SO, Gunnar R, Armin H (2014) Genome-wide analysis of alternative splicing in *Volvox carteri*. BMC Genom 15:1117

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MS, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk A, Ritland C, Boyle B, Bousquet J, Ritland K, Mackay J, Bohlmann J, Jones SJM (2013) Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics 29:1492–1497

Birzele F, Csaba G, Zimmer R (2008) Alternative splicing and protein structure evolution. Nucleic Acids Res 36:550–558

Blencowe BJ (2006) Alternative splicing: new insights from global analyses. Cell 26:37–47

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

Cheng ZC, Ge W, Li L, Hou D, Ma YJ, Liu J, Bai QS, Li XP, Mu SH, Gao J (2017) Analysis of MADS-box gene family reveals conservation in floral organ ABCDE model of moso bamboo (*Phyllostachys edulis*). Front Plant Sci 8:656

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinforma Oxf Engl 21:3674–3676

Ding F, Cui P, Wang ZY, Zhang SD, Ali S, Xiong LM (2014) Genome-wide analysis of alternative splicing of pre-mRNA under salt stress in *Arabidopsis*. BMC Genom 15:431

Emilio P, Milvia LR, Lieven S, Emanuela C, Elisa A, Silvio CET, Riccardo V, Yves VP, Alessandro C (2015) Exploration of alternative splicing events in ten different grapevine cultivars. BMC Genom 16:706

Filichkin SA, Mockler TC (2012) Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. Biol Direct 7:20

Filichkin SA, Priest HD, Givan SA, Shen R, Bryant DW, Fox SE, Wong WK, Mockler TC (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res 20:45–58

Gao J, Zhang Y, Zhang CL, Qi FY, Li XP, Mu SH, Hou CL (2014) Characterization of the floral transcriptome of moso bamboo (*Phyllostachys edulis*) at different flowering developmental stages by RNA-seq and digital gene expression analysis. PLoS One 9:e98910

Gilbert W (1978) Why genes in pieces? Nature 271:501

Graveley BR (2009) Alternative splicing: regulation without regulators. Nat Struct Mol Biol 16:13–15

Hartmuth K, Barta A (1986) In vitro processing of a plant pre-mRNA in a HeLa cell nuclear extract. Nucleic Acids Res 14:7513–7528

Howard BE, Hu Q, Babaoglu AC, Chandra M, Borghi M, Tan X, He L, Winter-Sederoff H, Gassmann W, Veronese P, Heber S (2013) High-throughput RNA sequencing of pseudomonas-infected *Arabidopsis* reveals hidden transcriptome complexity and novel splice variants. PLoS One 8:e74183

Huang J, Gao YJ, Jia HT, Liu L, Zhang D, Zhang ZX (2015) Comparative transcriptomics uncovers alternative splicing changes and signatures of selection from maize improvement. BMC Genom 16:363

Iida K, SekiM Sakurai T, SatouM Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2004) Genome-wide analysis of alternative pre-mRNA splicing in *Arabidopsis thaliana* based on full-length cDNA sequences. Nucleic Acids Res 32:5096–5103

Iwata H, Gotoh O (2011) Comparative analysis of information contents relevant to recognition of introns in many species. BMC Genom 12:45

Kalyna M, Barta A (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? Biochem Soc Trans 32:561

Kranthi KM, Karen-Beth GS (2015) Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in *Brachypodium distachyon*. Plant Cell 27:71–85

Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. Science 324:255–258

Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res 18:1851–1858

Li L, Cheng ZC, Ma YJ, Bai QS, Li XY, Cao ZH, Wu ZN, Gao J (2018) The association of hormone signaling, transcription and anatomy during shoot growth in moso bamboo. Plant Biotechnol J 16:72–85

Marchler-Bauer ACDD (2015) NCBI's conserved domain database. Nucleic Acids Res 43:222–226

Marquez Y, Brown JWS, Simpson C, Barta A, Kalyna M (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. Genome Res 22:1184–1195

Matthew AC, Brian JH, John PH, Stephen MM (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. BMC Genom 7:327

Modrek B, Lee C (2002) A genomic view of alternative splicing. Nat Genet 30:13–19

Nicola V, Claudio F, Elisa CC, Andrea T, Davide C, Michela D, Rosanna Z, Massimiliano C, Alessandro V, Claudio B, Margherita L, Giorgio V (2014) A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. BMC Plant Biol 14:99

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson Å, Rilakovic N, Ritland C, Rosselló JA, Sena J (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584

Palusa SG, Ali GS, Reddy AS (2007) Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant J 49:1091–1107

Panahi B, Mohammadi MS, Khaksefidi RE, Mehrabadi JF, Ebrahimie E (2015) Genome-wide analysis of alternative splicing events in *Hordeum vulgare*: highlighting retention of intron-based splicing and its possible function through network analysis. FEBS Lett 589:3564–3575

Peng ZH, Lu TH, Li LB, Liu XH, Gao ZM, Tao H, Yang XW, Feng Q, Guan JP, Weng QJ, Fan DL, Zhu CR, Lu Y, Han B, Jiang ZH (2010) Genome-wide characterization of the biggest grass, bamboo, based on 10, 608 putative full length cDNA sequences. BMC Plant Biol 10:116

Peng ZH, Zhang CL, Zhang Y, Hu T, Mu SH, Li XP, Gao J (2013a) Transcriptome sequencing and analysis of the fast growing shoots of moso bamboo (*Phyllostachys edulis*). PLoS One 8:e78944

Peng ZH, Lu Y, Li LB, Zhao Q, Feng Q, Gao ZM, Lu HY, Hu T, Yao N, Liu KY, Li Y, Fan DL, Guo JL, Li WJ, Lu YQ, Weng QJ, Zhou CC, Zhang L, Huang T, Zhao Y, Zhu CR, Liu XG, Yang XW, Wang T, Miao K, Zhuang CY, Cao XL, Tang WL, Liu GS, Liu YL, Chen J, Liu ZJ, Yuan LC, Liu ZH, Huang XH, Lu TT, Fei BH, Ning ZM, Han B, Jiang ZH (2013b) The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). Nat Genet 45:456–461

Rauch HB, Patrick TL, Klusman KM (2014) Discovery and expression analysis of alternative splicing events conserved among plant SR proteins. Mol Biol Evol 31(3):605–613

Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. Plant Biol 58:267–294

Reddy ASN, Marquez Y, Kalyna M, Barta A (2013) Complexity of the alternative splicing landscape in plants. Plant Cell 25:3657–3683

Sammeth M, Foissac S, Guigó R (2008) A general definition and nomenclature for alternative splicing events. PLoS Comput Biol 4:e1000147

Shen YT, Zhou ZK, Wang Z, Li WY, Fang C, Wu M, Ma YM, Liu TF, Kong LA, Peng DL, Tian ZX (2014) Global dissection of alternative splicing in paleopolyploid soybean. Plant Cell 26:996–1008

Shawn T, Zhou WG, April L, Wang BB, Mary B, Gina ZH, Zhao XY, Andy B, Li BL (2014) Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation. Plant Cell 26:3472

Staiger D, Brown JW (2013) Alternative splicing at the intersection of biological timing, development, and stress responses. Plant Cell 25:3640–3656

Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H (2005) Function of alternative splicing. Gene 344:1–20

Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960

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

Tomasz EK, Konstantin VK (2011) Evolution of exon-intron structure and alternative splicing. PLoS One 6(3):e18055

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

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476

Xu P, Kong YM, Song DL, Huang C, Li X, Li LG (2014) Conservation and functional influence of alternative splicing in wood formation of *Populus* and *Eucalyptus*. BMC Genom 15:780

Yoshimura K, Yabuta Y, Ishikawa T, Shigeoka S (2002) Identification of a cis element for tissue-specific alternative splicing of chloroplast ascorbate peroxidase pre-mRNA in higher plants. J Biol Chem 277:40623–40632

Zhang G (2010) Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. Genome Res 20:646–654

Zhu W, Schlueter SD, Brendel V (2003) Refined annotation of the Arabidopsis genome by complete expressed sequence tag mapping. Plant Physiol 132:469–484