

# Genome-wide profiling of circular RNAs, alternative splicing, and R-loops in stem-differentiating xylem of *Populus trichocarpa*

Xuqing Liu<sup>1†</sup>, Yubang Gao<sup>2†</sup>, Jiakai Liao<sup>2</sup>, Miao Miao<sup>1</sup>, Kai Chen<sup>2</sup>, Feihu Xi<sup>2</sup>, Wentao Wei<sup>1</sup>, Huihui Wang<sup>1</sup>, Yongsheng Wang<sup>2</sup>, Xi Xu<sup>2</sup>, Anireddy S.N. Reddy<sup>3</sup> and Lianfeng Gu<sup>1\*</sup>

1. Basic Forestry and Proteomics Research Center, College of Forestry, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou 350002, China

2. Basic Forestry and Proteomics Research Center, College of Life Science, Fujian Agriculture and Forestry University, Fuzhou 350002, China

3. Department of Biology and Program in Cell and Molecular Biology, Colorado State University, Fort Collins Colorado, 80523, USA

<sup>†</sup>These authors contributed equally to this work.

\*Correspondence: Lianfeng Gu (lfgu@fafu.edu.cn)



Xuqing Liu



Lianfeng Gu

## ABSTRACT

Circular RNAs (circRNAs) are a recently discovered type of non-coding RNA derived from pre-mRNAs. R-loops consist of a DNA:RNA hybrid and the associated single-stranded DNA. In *Arabidopsis thaliana*, circRNA:DNA R-loops regulate alternative splicing (AS) of *SEPALLATA3* (*SEP3*). However, the occurrence and functions of circRNAs and R-loops in *Populus trichocarpa* are largely unexplored. Here, we performed circRNA-enriched sequencing in the stem-differentiating xylem (SDX) of *P. trichocarpa* and identified 2,742 distinct circRNAs, including circ-*CESA4*, circ-*IRX7*,

and circ-*GUX1*, which are generated from genes involved in cellulose, and hemicellulose biosynthesis, respectively. To investigate the roles of circRNAs in modulating alternative splicing (AS), we detected 7,836 AS events using PacBio Iso-Seq and identified 634 circRNAs that overlapped with 699 AS events. Furthermore, using DNA:RNA hybrid immunoprecipitation followed by sequencing (DRIP-seq), we identified 8,932 R-loop peaks that overlapped with 181 circRNAs and 672 AS events. Notably, several SDX-related circRNAs overlapped with R-loop peaks, pointing to their possible roles in modulating AS in SDX. Indeed, overexpressing circ-*IRX7* increased the levels of R-loop structures and decreased the frequency of intron retention in linear *IRX7* transcripts. This study provides a valuable R-loop atlas resource and uncovers the interplay between circRNAs and AS in SDX of *P. trichocarpa*.

Keywords: alternative splicing, circular RNA, R-loop, *Populus trichocarpa*, stem-differentiating xylem

Liu, X., Gao, Y., Liao, J., Miao, M., Chen, K., Xi, F., Wei, W., Wang, H., Wang, Y., Xu, X., Reddy A.S.N., and Gu, L. (2021). Genome-wide profiling of circular RNAs, alternative splicing, and R-loops in stem-differentiating xylem of *Populus trichocarpa*. *J. Integr. Plant Biol.* **00**: 1–15.

## INTRODUCTION

Circular RNAs (circRNAs) are covalently closed loop RNAs generated by back splicing that connects downstream 5' donor sites to upstream 3' acceptor sites (Capel et al., 1993). CircRNAs perform important functions, such as regulating exon skipping (Conn et al., 2017), acting as vehicles for

RNA-binding proteins (Qu et al., 2015) and as miRNA sponges (Hansen et al., 2013), and regulating their corresponding linear transcripts via localization in the nucleus (Li et al., 2015). In plants, circRNAs inhibit the binding of DICER-Like1 (DCL1)/HYPONASTIC LEAVES1 (HYL1) to pri-miRNAs (Li et al., 2016) and are responsive to drought stress (Zhang et al., 2019a). With the increasing development of

detection methodologies, the roles and mysteries of circRNAs are gradually being unraveled. CircRNAs can be identified using rRNA-depleted libraries (Wang et al., 2019) or RNase R-treated libraries (Li et al., 2016; Zhang et al., 2019b) to enrich circular transcripts and then sequencing on the Illumina (Li et al., 2016; Cheng et al., 2018) or nanopore-based sequencing platforms (Wang et al., 2020). Analysis on these data using bioinformatics algorithms can reliably detect circRNAs based on the presence of back-splicing junctions (Zhang et al., 2016).

Alternative splicing (AS) modulates transcriptome diversity and proteome complexity (Chaudhary et al., 2019). Until recently, most AS events were identified using RNA-Seq (Deng et al., 2010; Wu et al., 2016; Huang et al., 2017). However, during the past few years Pacific Biosciences (PacBio) single-molecule real-time (SMRT) long-read isoform sequencing (Iso-Seq) has been applied in *Zea mays* (Wang et al., 2016), *Sorghum bicolor* (Abdel-Ghany et al., 2016), *Phyllostachys edulis* (Wang et al., 2017), *Populus trichocarpa* (Filichkin et al., 2018), and *Populus bolleana* (Hu et al., 2020) to investigate the developmental dynamics of AS or changes in AS in response to abiotic stresses in different tissues.

DNA:RNA hybrids and the associated single-stranded DNAs form endogenous three-stranded nucleic acid structures called R-loops, which are generated due to the thermodynamic stability of binding between guanine-rich RNA and cytosine-rich DNA templates (Roberts and Crothers, 1992; Ginno et al., 2012). The widespread formation of genomic R-loops at CpG islands can provide protection against *de novo* DNA methylation to regulate transcription (Ginno et al., 2012). Genome-wide R-loop structures can be captured by DNA:RNA immunoprecipitation sequencing and high-throughput sequencing (DRIP-seq) using the structure-specific S9.6 monoclonal antibody (Ginno et al., 2013; Xu et al., 2017; Sanz and Chedin, 2019). R-loop signals are further identified using peak-calling methods designed for ChIP-seq (Zhang, 2008). A recent DRIP-seq study revealed that these chromosomal structures are prevalent in plants (Xu et al., 2017). R-loop patterns change dynamically in response to environmental stimuli (Xu et al., 2020). *SEPALLATA3* (*SEP3*) exon 6 circRNA regulates the exon skipping of *SEP3* via the formation of circRNA:DNA R-loops in *Arabidopsis thaliana* (Conn et al., 2017). However, whether this is a common regulatory mechanism in plants remains unknown.

At least 36% of genes in *P. trichocarpa* undergo AS (Bao et al., 2013); these genes encode important dominant-negative regulators (Li et al., 2012; Lin et al., 2017) of processes such as fiber cell wall thickening (Zhao et al., 2014). However, the regulatory role of circRNAs in splicing has not been characterized in *P. trichocarpa*. In the current study, we identified circRNAs in stem-differentiating xylem (SDX) of *P. trichocarpa* enriched using RNase R-treated samples. Moreover, we used a hybrid sequencing strategy involving RNA-Seq and Iso-Seq to generate a high-resolution transcriptome map of AS isoforms to address the interplay between circRNAs and AS. Finally, we performed DRIP-seq to explore

the potential role of circRNAs in regulating AS via the formation of R-loop structures.

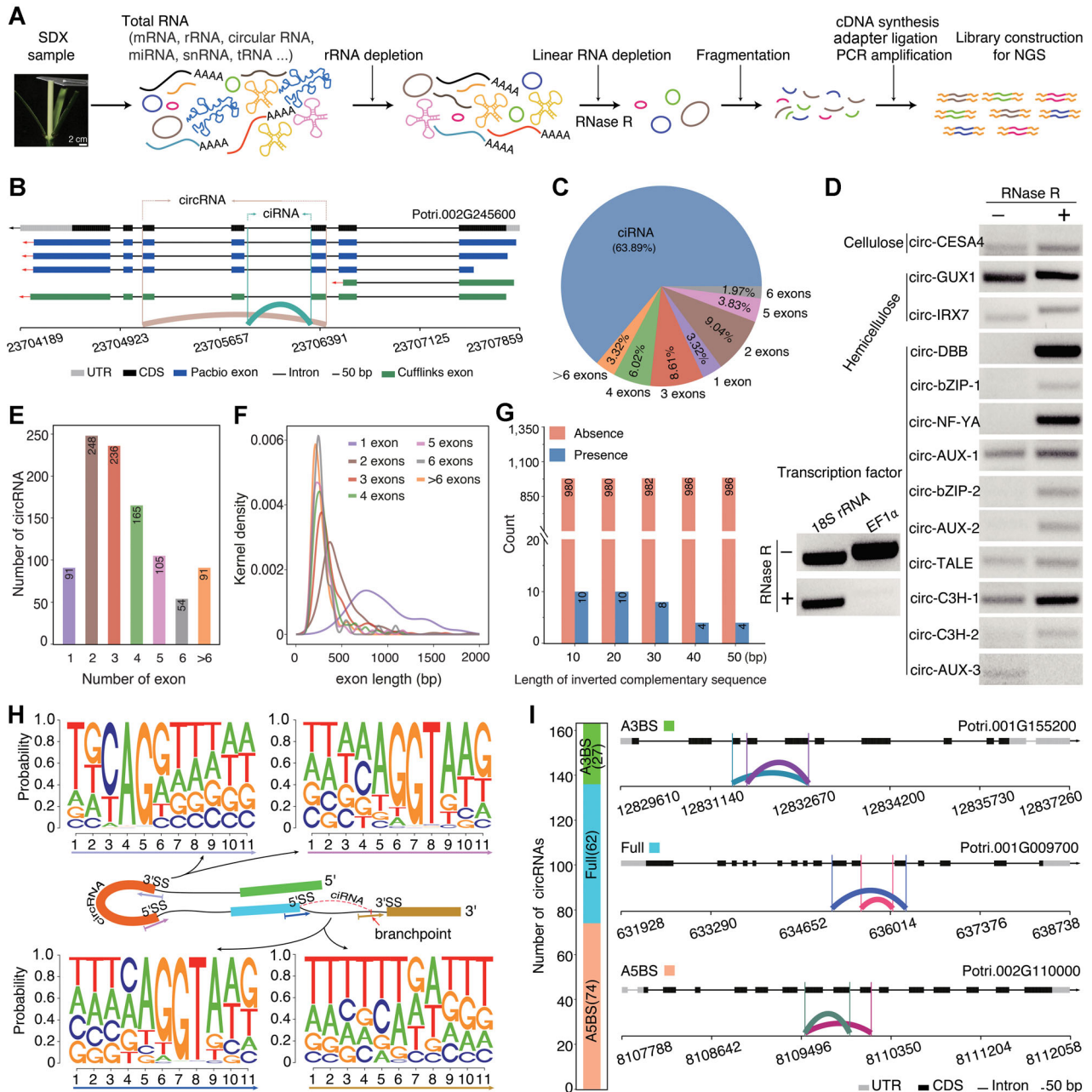
## RESULTS

### Characterization of circRNAs in SDX

CircRNAs arise through a form of AS known as back-splicing. CircRNAs are abundant, stable, conserved, and expressed in a stage-specific manner in plants (Ye et al., 2015; Tan et al., 2017; Zhao et al., 2017; Zeng et al., 2018; Wang et al., 2019). To identify and characterize circRNAs during wood formation on a genome-wide scale, we performed rRNA removal and RNase R treatment to enrich circRNAs from total RNAs, followed by Illumina paired-end sequencing (Figure 1A). We obtained 267,553,488 150-bp paired-end (PE) reads (Table S1). In total, we identified 2,742 circRNAs, including 1,752 circRNAs derived from lariat introns (ciRNAs) and 990 generated from exon circularization (Figure 1B, full list in Table S2). The percentage of ciRNAs (63.89%) was higher than the percentage of circRNAs (Figure 1C), suggesting that ciRNAs are more widespread in *P. trichocarpa*, which is consistent with findings in soybean (Zhao et al., 2017), human, and mouse (Panda et al., 2017).

To validate the identified circRNAs, we selected 13 circRNAs generated from exon circularization and performed an RNase R resistance experiment. We designed a set of divergent primers (Table S3) for RT-PCR of each circRNA and convergent primers to amplify linear transcripts of *EF1 $\alpha$*  and 18S rRNA as controls. Most circRNAs were resistant to RNase R digestion compared to linear transcripts of 18S rRNA and *EF1 $\alpha$*  (Figure 1D). However, circ-*AUX-3* generated from the transcription factor gene *AUX* (Potri.003G001000) was susceptible to RNase R and did not show enrichment after RNase R treatment, and 18S rRNA transcript was not completely digested, likely because it is the most stable housekeeping gene and is not sensitive to RNase R (Bas et al., 2004). Sanger sequencing of gel-purified bands confirmed the presence of authentic back-splicing junction regions, which is consistent with the finding that some circRNAs are susceptible to RNase R (Jeck et al., 2013; Li et al., 2019).

Among the 990 circRNAs generated from exon circularization, 91 (9.19%) originated from only one annotated exon, and the 899 remaining circRNAs included multiple exons. CircRNA with two circularized exons was the most common type (Figure 1E), which is similar to findings for *Phyllostachys edulis* (Wang et al., 2019), *Poncirus trifoliata* (Zeng et al., 2018), *Glycine max* (Zhao et al., 2017), and *A. thaliana* (Ye et al., 2015). Interestingly, exons from circRNAs with only one circularized exon were much longer than exons from circRNAs containing multiple circularized exons (Figure 1F). A gradual reduction in the length of circularized exons was observed with increasing number of circularized exons, suggesting that exon length is one condition affecting circRNA biogenesis. Repetitive elements are responsible for



**Figure 1. General characteristics of circular RNAs (circRNAs) in stem-differentiating xylem (SDX) of *Populus trichocarpa***

**(A)** Construction of circRNA libraries using rRNA-depleted RNA treated with RNase R before sequencing. **(B)** Schematic representation of circRNAs derived from lariat introns (ciRNA) (blue green) generated from the lariat intron and circRNA (brown) derived from back-splicing through the joining of 5' donor and 3' acceptor sites. **(C)** Pie chart showing the percentage of ciRNAs and circRNAs identified in this study. **(D)** Validation of circRNAs in RNase R<sup>-</sup> and ddH<sub>2</sub>O (RNase R<sup>+</sup>) treated samples from SDX of *Populus trichocarpa* using divergent primers. The linear transcripts of 18S rRNA and EF1 $\alpha$  were used as controls. **(E)** Number of exons in circRNAs. **(F)** Length distribution of exons in circRNAs. **(G)** Number of circRNAs with (blue) or without (orange) inverted complementary flanking intron sequences. **(H)** Sequence logos showing splicing signals around circRNAs. **(I)** The number and structures of alternative 5' back-spliced circRNA isoforms (A5BS), alternative 3' back-spliced circRNA isoforms (A3BS), and fully inclusive circRNAs (Full).

most circRNA formation in humans (Dong et al., 2017). However, a few inverted complementary sequences were detected around the circularized exons in *P. trichocarpa* (Figure 1G). This result indicates that the flanking reverse complementary sequence is not the major factor involved in the biogenesis of circRNAs in *P. trichocarpa*, suggesting that other mechanisms might be involved, which is consistent

with previous findings in *Phyllostachys edulis* (Wang et al., 2019), *Oryza sativa* (Lu et al., 2015; Ye et al., 2015), and *A. thaliana* (Ye et al., 2015; Sun et al., 2016).

To investigate the splicing signal of circRNAs, we extracted and analyzed 11-nt DNA sequences around 5' splicing sites (5' SS), 3' splicing sites (3' SS), and in regions close to branchpoints for ciRNAs. Motif analysis revealed



canonical GT-AG splicing signals involved in the formation of circRNAs (Figure 1H), indicating that circRNA formation depends on the spliceosome machinery. For ciRNAs, the sequence logo for the region close to the branchpoint was enriched for U nucleosides. The mean distance from the branchpoint to the 3' SS for ciRNAs was approximately 25 bp, which is similar to that of *Phyllostachys edulis* (Wang et al., 2019). Back-splicing/splicing sites play decisive roles in the origination of circRNA, and can result in the generation of multiple circRNAs from a single gene locus (Jeck et al., 2013; Zhang et al., 2016). Among the 990 circRNAs in SDX, we identified 27 alternative 3' back-splicing (A3BS) circRNAs and 74 alternative 5' back-splicing (A5BS) circRNAs (Table S4). In addition, 62 circRNAs were identified with fully overlapping region across the back-splicing junction (Figure 1I). GO enrichment analysis revealed that genes generating ciRNAs are involved in the Golgi apparatus, intracellular protein transport, and ATP-dependent RNA helicase activity. Genes generating circRNAs were enriched for microtubule binding, intracellular protein transport, and microtubule-based movement.

### Global survey of post-transcriptional regulation in SDX

PacBio sequencing produces longer reads than short-reads methods, facilitating the generation of a well-annotated transcriptome, including splicing isoforms (Rhoads and Au, 2015; Abdel-Ghany et al., 2016; Wang et al., 2016). We constructed full-length first strand cDNA libraries with insert sizes of 1–2, 2–3, 3–6, and 5–10 kb and sequenced them on the PacBio RSII Single Molecule, Real-Time (SMRT) sequencer, yielding 298,716 reads of inserts (ROIs) (Figure 2A; Table S5). To assess the data quality and obtain FLNC reads, we filtered the polymerase reads, yielding 161,486 FLNC reads (54.1%). We also obtained 120,818,725 paired-end reads by performing dUTP mRNA-seq on the Illumina HiSeq 2500 platform to correct the single-molecule long reads. Finally, 161,334 error-corrected FLNC reads were generated, and 160,622 mapped FLNC reads were obtained (Figure 2A). A total of 7,959 genes containing two or more polyadenylation sites were identified (Figure 2B; Table S6), suggesting that alternative polyadenylation (APA) events are a common phenomenon in *P. trichocarpa*. Proximal and distal poly(A) cleavage sites showed very similar nucleotide distributions (Figure 2C), indicating that proximal sites are likely common genuine poly(A) cleavage sites in plants. We analyzed the nucleotide composition in the upstream (–50 nt) and downstream (+50 nt) regions of all poly(A) cleavage sites and found that the AAGAAA motif and the UGUA motif were the two most common polyadenylation signals, which were enriched in the regions 5 nt downstream and 40 nt upstream of poly(A) sites, respectively (Figure 2D).

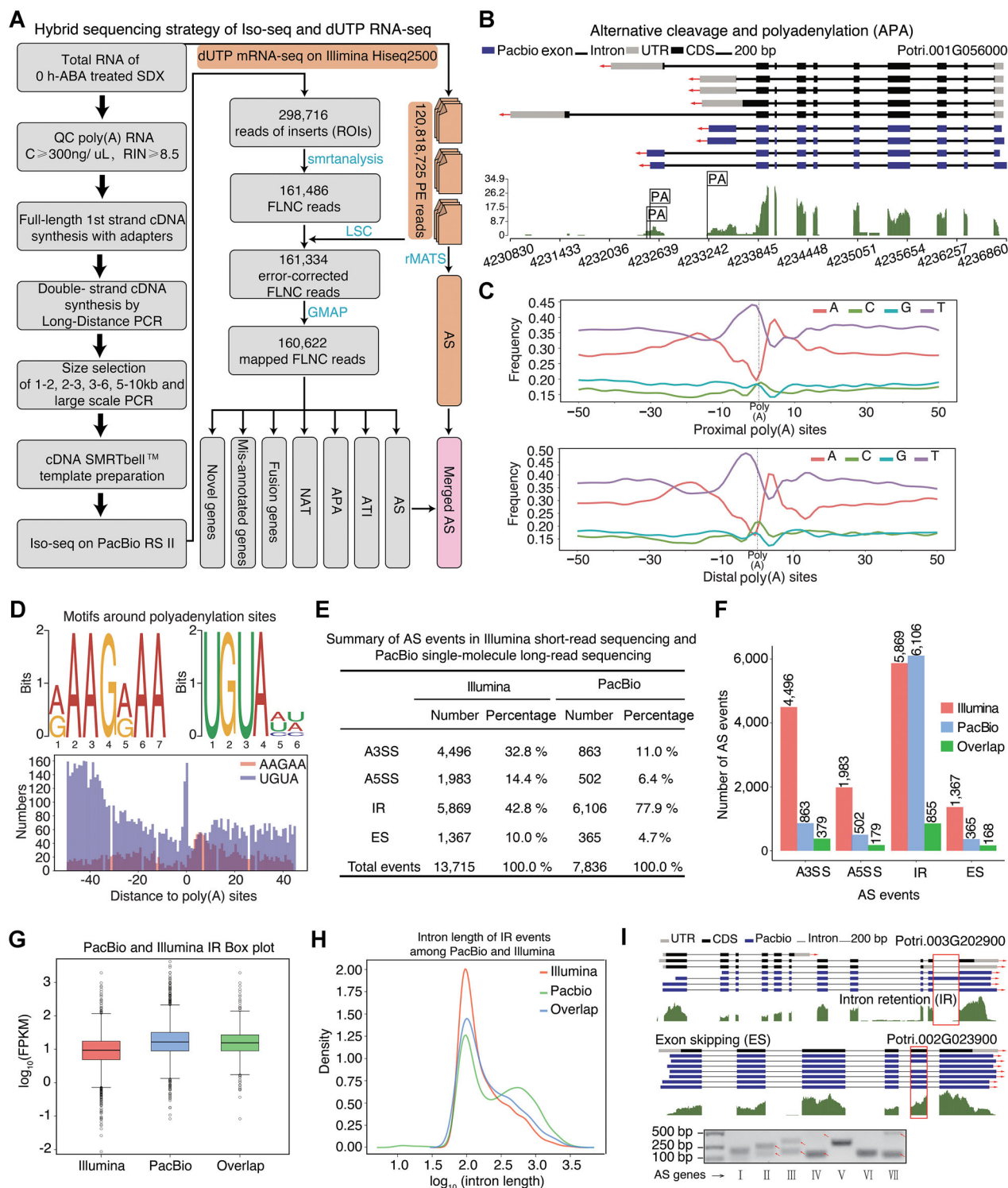
To confirm the expression of the four modes of AS in *P. trichocarpa*, including alternative 3' splice site (A3SS), alternative 5' splice site (A5SS), intron retention (IR), and exon skipping (ES), we analyzed AS events using PacBio long-read

sequencing and Illumina short-read sequencing. Using hybrid sequencing, we identified 7,836 (Table S7) and 13,715 (Table S8) AS events on the PacBio Iso-seq and Illumina platforms, respectively (Figure 2E, F). In total, the two sequencing platforms generated 1,581 overlapping AS events, primarily in highly expressed genes (Figure 2F, G). The Illumina platform generated 42.8% (5,869) IR events and 10% (1,367) ES events, whereas the PacBio platform generated 77.9% (6,106) IR events and only 4.7% (365) ES events. Further analysis of the intron lengths of IR events revealed that there were longer introns in the PacBio versus Illumina data (Figure 2H). Thus, the PacBio platform is better able to detect IR in long introns, which is consistent with disadvantage for the Illumina-based short read assembly. However, Illumina is better able to detect low-abundance genes due to the better depth of sequencing.

To validate the results of long-read sequencing, we performed RT-PCR to detect seven predicted AS events. Four of these AS events were confirmed, as two bands of the expected size were produced (Figure 2I). However, the primers for the three remaining AS events failed to amplify the expected bands for long isoforms, perhaps due to the low expression levels of these long isoforms in SDX tissue. The successfully validated AS genes are all homologous to *Arabidopsis* genes related to cellulose development, such as homologs of *FE11*, which functions in cell wall biosynthesis (Xu et al., 2008), and *GH9B8*, encoding a cell-wall-loosening protein (Xu et al., 2014).

### R-loop atlas in SDX

R-loops comprising one single-stranded DNA and a DNA:RNA hybrid are ubiquitous functional structures in *Arabidopsis* (Xu et al., 2017). However, to date, an atlas of R-loop peaks in *P. trichocarpa* has not been generated. Therefore, we characterized R-loops in *P. trichocarpa* by performing DRIPc-seq and DRIP-seq (negative control with RNase H treatment) on SDX samples (Figure 3A). This analysis identified 4,834 R-loop plus (+) peaks and 4,098 R-loop minus (–) peaks (Table S9). Among the 8,932 R-loop peaks, only 785 peaks (8.79%) were also detected in samples subjected to RNase H treatment. The 8,147 remaining R-loop peaks (91.21%) were undetectable in samples treated with RNase H, suggesting that R-loops were enriched in these samples compared to the negative control. Analysis of the distribution of R-loop peaks across the transcripts revealed that R-loop structures were enriched near the transcript start and gene body regions (Figure 3B), which is consistent with previous findings (Xu et al., 2017). Most peaks were 250–500 bp long (Figure 3B, C). In total, 56.55% and 35.81% of the DRIPc signals mapped onto CDSs and UTRs, respectively (Figure 3D). Furthermore, 27.5% of the R-loop peaks contained the sequence motif GARGAAG (R = A or G) (Figure 3E), which is consistent with previous findings in *Arabidopsis* (Xu et al., 2017). These results confirm the reliability of DRIP-seq in *P. trichocarpa* and suggest that the GARGAAG motif is conserved in plants.



**Figure 2. A global survey of post-transcriptional regulation in stem-differentiating xylem (SDX) using single-molecule long-read sequencing** (A) Flowchart of the hybrid sequencing strategy involving PacBio sequencing and Illumina RNA sequencing of SDX. (B) Gene structure showing alternative polyadenylation (APA). The blue boxes, solid lines, and gray boxes indicate exons, introns, and UTRs, respectively. Wiggle track represents the expression levels using Illumina based RNA-sequencing. (C) Nucleotide composition profile around proximal poly(A) sites and distal poly(A) sites. (D) Sequence logos of motifs around polyadenylation sites. (E, F) The number of alternative splicing (AS) events from Illumina mRNA-seq and PacBio Iso-seq. Alternative 3' back-spliced circRNA isoforms (A3SS), alternative 5' back-spliced circRNA isoforms (A5SS), intron retention (IR), and exon skipping (ES) represent alternative 3' splice site, alternative 5' splice site, intron retention, and exon skipping, respectively. (G) Box plot of FPKM values for genes with IR events based on Illumina mRNA-seq and PacBio Iso-seq. (H) Density distribution of intron length from Illumina mRNA-seq and PacBio Iso-seq data. (I) Upper panel shows the gene structures of IR (Potri.003G020900) and ES (Potri.002G023900) events, respectively. The bottom panel shows the validation of seven AS events. Red arrows represent the expected bands from AS events (II, III, IV, VII) from the short and long isoform, respectively.

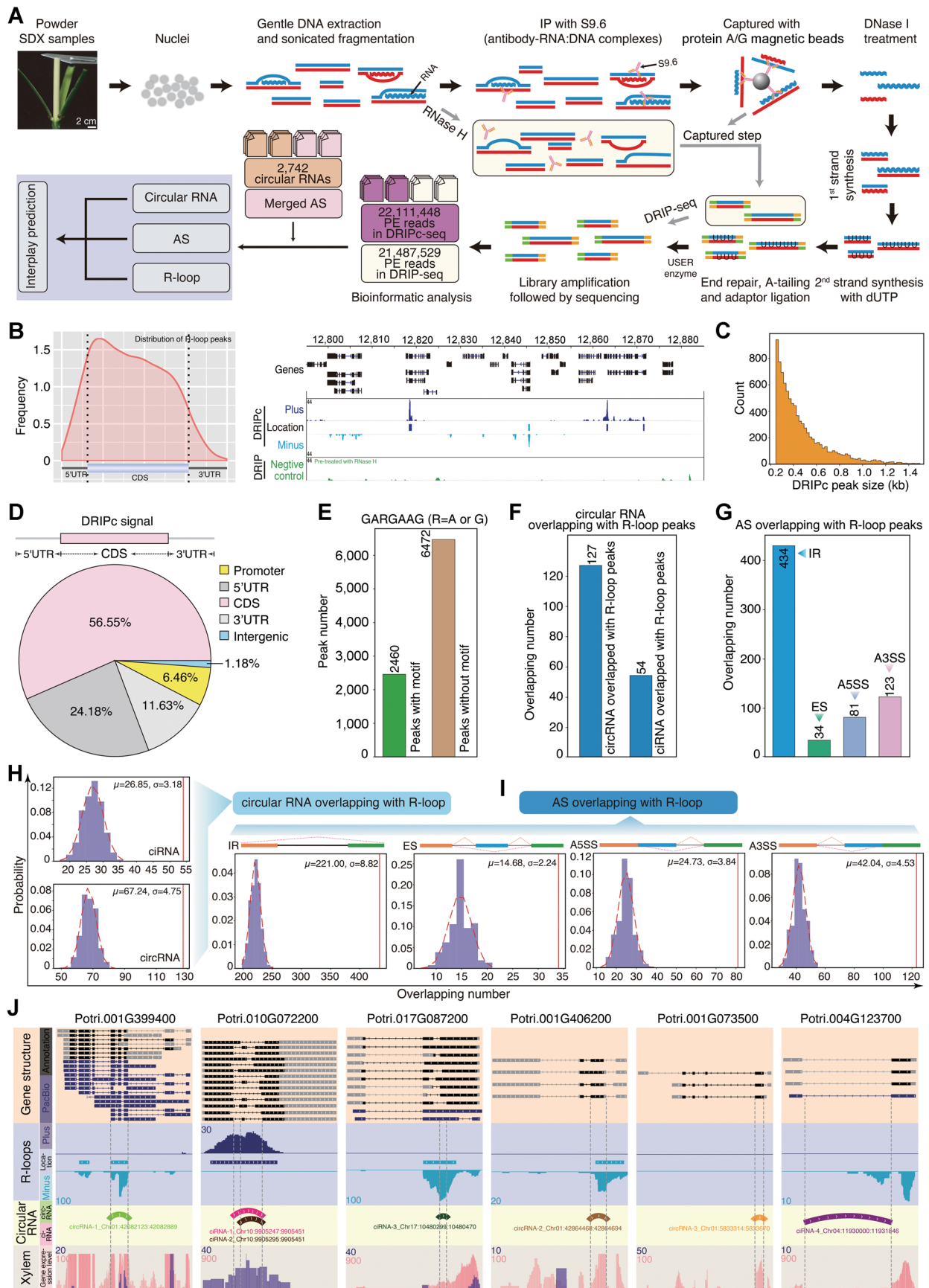


FIGURE 3. Continued



### R-loops associated with circRNAs and AS

To identify R-loops associated with circRNA and AS in SDX, we calculated the number of circRNAs and AS that overlapped with R-loop peaks. In total, 181 circRNAs (Figure 3F) and 672 AS events (Figure 3G) overlapped with R-loop peaks. The AS events that overlapped with R-loops included 3.8% IR, 2.1% ES, 3.5% A5SS, and 2.5% A3SS events. In total, 7.2% circRNAs and 5.5% ciRNA overlapped with AS events. We randomly selected gene regions that included R-loop peaks of the same length and detected circRNAs/AS events that overlapped with these regions. After repeating the process 1,000 times, we calculated the probability that circRNAs/AS events overlapped with R-loops events, including specific mean value ( $\mu$ ) and standard deviation ( $\sigma$ ). The probability that R-loop peaks overlapped with circRNAs (Figure 3H) and AS events (Figure 3I) was higher than that of random simulation. Here, we selected seven circRNAs as examples, in which two circRNAs (circRNA-1–2) and four ciRNAs (ciRNA-1–4) overlapped with R-loop peaks (Figure 3J). Among these, four overlapping circRNAs (circRNA-1, ciRNA-1, ciRNA-2, and ciRNA-3) were located exactly where AS had occurred in the parent genes.

### CircRNAs associated with R-loops regulate splicing in SDX

We randomly selected gene regions showing AS events of the same length and detected ciRNAs/circRNAs that overlapped with these regions. After repeating the process 1,000 times, we found that the probability that ciRNAs/circRNAs overlapped with AS events was higher than that of random simulation (Figure 4A). Conn et al. (2017) provided the first example of the co-regulation of circRNA-R-loop formation and AS in plants, but whether this is a widespread phenomenon in plants had been unclear. We also performed overlap analysis among the three events (R-loops, circRNAs, and AS). Random simulation revealed that R-loops are generally correlated with circRNAs and AS simultaneously (Figure 4A). In total, we identified 418 R-loop peaks containing regions complementary to circRNAs by cis- or trans-regulation in 234 AS regions (Figure 4B; Table S10). These regions were enriched in GO terms such as ‘microtubule-based process’, ‘hydrolase activity’, and so on (Figure 4C).

Finally, to investigate the effects of circRNAs associated with R-loops on the regulation of AS of their parent genes in SDX, we developed high efficiency transfection of protoplasts for circRNAs. Protoplasts transfected with an empty vector (*pUC19-35S-sGFP*) were used as a control (Figure 4D). We selected circ-*IRX7* from a hemicellulose synthase gene (Potri.009G006500), which presented IR events (Figure 4E). Circ-*IRX7* derived from single exon2 and showed enrichment of R-loop. We overexpressed circ-*IRX7* in protoplasts derived from SDX using the *pUC19-35S-circ-IRX7-35S-sGFP* vector (Figure 4F; Table S3). RT-PCR analysis of total RNA from transfected protoplasts using divergent primers of circ-*IRX7* confirmed the overexpression of circ-*IRX7* (OE) compared to the empty vector (EV) control in protoplasts after 12 h of incubation. First, we measured the levels of R-loops by DRIP-quantitative PCR (DRIP-qPCR), revealing that overexpressing circ-*IRX* increased the levels of R-loop structures compared to EV (Figure 4F, left panel). Second, we examined the effect of increasing both circRNA and R-loop formation on splicing. The overexpression of circ-*IRX7* in *P. trichocarpa* SDX protoplasts resulted in a reduction in the levels of the long transcript (*PtIRX7-1l*) and an increase in the levels of the short transcript (*PtIRX7-s*) (Figure 4F, right panel). We also selected circ-*GUX1* without R-loop structure to do the same experiment. Circ-*GUX1* derived from exon2 and exon3 and overlapped with IR events in linear transcripts (Figure 4G; Table S3). We did not observe obvious alteration of R-loop level in overexpression of circ-*GUX1* (OE) sample based on DRIP-qPCR (Figure 4H). Semi-quantitative PCR using convergent primers only detected the up-regulation of the most abundant full spliced transcript, which suggested that circ-*GUX1* (OE) might affect the expression of corresponding parental transcripts. The long transcript including retention of first intron was not detected in both EV and circ-*GUX1* (OE), which suggested circ-*GUX1* without R-loop structure did not affect AS (Figure 4H). CRISPR-Cas9-generated null alleles of circRNA loci were recently obtained, providing the opportunity to investigate the functions of circRNAs with minimal interference from cognate linear transcripts (Zhou et al., 2021). Future technological advances should make it possible to investigate the roles of circRNAs in regulating splicing via R-loop formation in more detail.

### Figure 3. A global exploration of the interplay between R-loop peaks, circular RNA (circRNA), and alternative splicing (AS) in stem-differentiating xylem (SDX)

(A) Workflow of DRIPc-seq to investigate the interplay among R-loops, circRNA, and AS. Genomic DNA was extracted from nuclei and S9.6 antibodies were used to pull down the R-loop structures. The RNA in R-loops was subjected to Illumina RNA sequencing (DRIPc-seq); samples treated with RNase H were used as a negative control for Illumina RNA sequencing (DRIP-seq). The interplay of R-loops was analyzed by identifying overlap among R-loops, circRNAs, and AS event. (B) Metaplots of R-loop peaks across the transcripts, and snapshot of a representative genomic region. DRIPc data are shown in purple (+strand) and blue (–strand). DRIP-seq data (pre-treated with RNase H) are shown in green. (C) Distribution of DRIPc peak numbers (y-axis) for different peak sizes (x-axis). (D) DRIPc peak distribution within different gene contexts. (E) Number of GARGAAG motifs from R-loop peaks. (F, G) Number of overlapping circRNAs (F) and AS events (G) with R-loops. (H) Random simulation of overlap between circRNA with R-loop peaks.  $\mu$  and  $\sigma$  represent the mean value and standard deviation, respectively. The x-axis shows the number of circRNAs derived from lariat introns (ciRNAs) or circRNAs randomly overlapping with R-loops. The y-axis shows the probability of ciRNAs or circRNAs overlapping with R-loops. The dotted red lines are the best fit normal distribution curves of random overlap. The vertical red lines are the observed results. (I) Random simulation of overlap between AS with R-loop peaks. (J) Snapshots of the overlap between R-loops and circRNAs. The top panels represent gene structures. R-loops peaks are shown dark blue (+) and light blue (–) tracks. Semicircles in the circRNA tracks represent the genome coordinates of circRNAs. The bottom wiggle track shows the expression based on RNA-seq data.

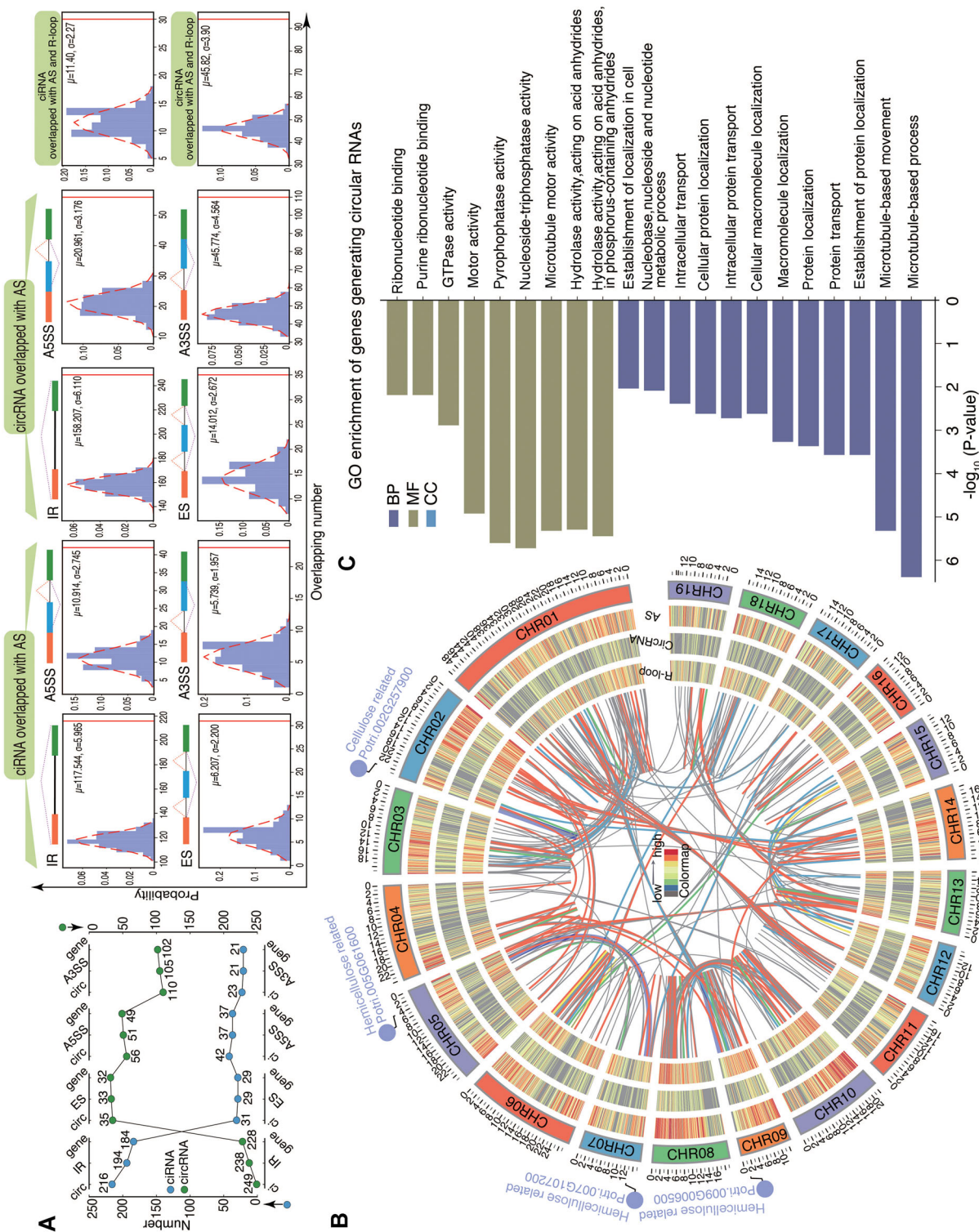
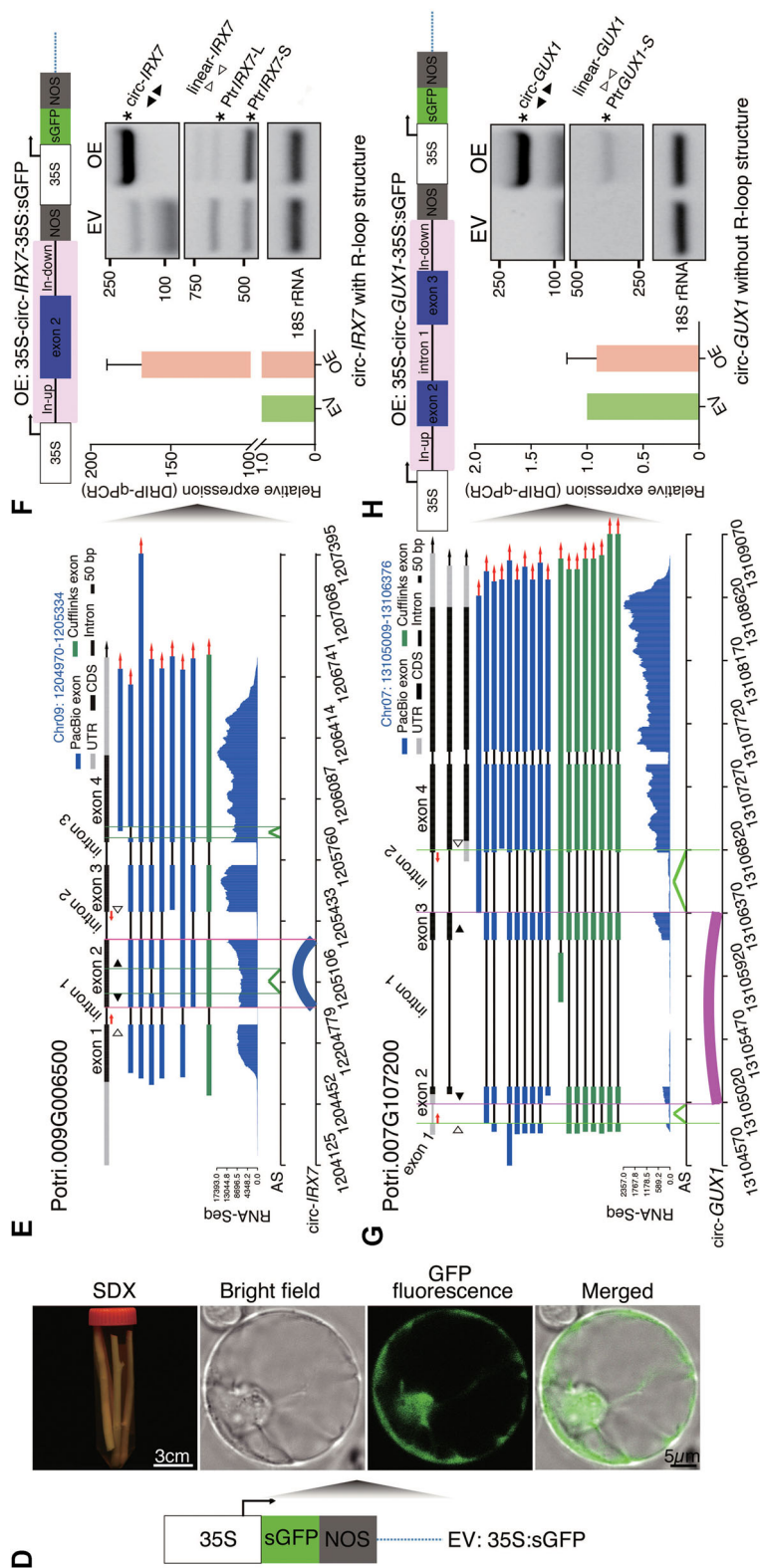


FIGURE 4. Continued





**Figure 4. Circular RNAs (circRNAs) associated with alternative splicing (AS) by R-loop peaks in stem-differentiating xylem (SDX)**

(A) The overlap between circRNAs derived from lariet introns (circRNAs)/circRNAs and AS events. Blue and green dots represent the number of circRNAs and circRNAs that overlap with AS events (IR, ES, A3SS, A5SS), respectively. The middle panel shows a random simulation of overlap between AS and circRNAs. The right panel shows a random simulation of overlap between R-loop peaks and both circRNAs and AS. A3SS, alternative 3' back-spliced circRNA isoforms; A5SS, alternative 5' back-spliced circRNA isoforms; ES, exon skipping; IR, intron retention. (B) Circos diagram showing (from outer to inner rings) the chromosomes and the heat map density of AS, circRNAs, and R-loop peaks. The colored lines in the center represent the complementary regions between R-loop peaks and circRNAs. Red and blue represent the links with and without AS events, respectively. Several hemicellulose and cellulose-related genes are indicated in the Circos diagram. (C) GO enrichment of genes generating circular RNAs including complementary regions with R-loop peaks. (D) The schematic diagram at the left panel represents constructs overexpressing GFP (35S::sGFP) as the empty vector (EV) line. The right panel shows (from top to bottom) debarked stems in cell wall digestion enzyme solution, protoplast morphology in a bright-field image, protoplast morphology in a GFP-fluorescence image, and a merged image of bright field and GFP fluorescence. (E) Visualization of gene structures, which represented linear-IRX7 and circ-IRX7, respectively. Convergent primers (red arrow pairs) were used to amplify circ-IRX7 and linear-IRX7 including the flanking intron sequences. Divergent primers (black back-to-back triangle pairs) and convergent primers (white opposing triangle pairs) were used to amplify circ-IRX7 and linear-IRX7, respectively. (F) Construction of the overexpression vector for GFP-circ-IRX7 (35S-circ-IRX7-35S::sGFP, OE), which was generated by back-splicing of the second exon of Potri.009G006500. The entire exon 2 with the flanking intron sequences was inserted into a transient expression vector that included two independent CaMV 35S promoters, one to overexpress circ-IRX7 and the other to express GFP. DRIP-qPCR was used to detect the expression of R-loops in the IRX7 OE and EV lines. Histogram in the left panel shows R-loop level based DRIP-qPCR. The right panel shows the transfection of SDX protoplasts and semi-quantitative PCR of circRNA and splicing isoforms of the host gene. (G) Visualization of gene structures, which represented linear-GUX1 and circ-GUX1 without R-loop structure. (H) Construction of the overexpression vector for GFP-circ-GUX1 (35S-circ-GUX1-35S::sGFP, OE). Histogram in the left panel shows the expression level of circ-GUX1 and linear transcript of GUX1.

## DISCUSSION

The circular form of viral RNA was discovered approximately 40 years ago (Hsu and Coca-Prados, 1979). CircRNAs were initially regarded as splicing errors produced by a non-canonical mode of RNA splicing and were dismissed due to their low levels (Salzman et al., 2012; Haddad and Lorenzen, 2019). Intron circularization and exon circularization are two major types of circRNA biogenesis. Here, we determined that the percentage of canonical GT-AG splice sites for ciRNAs and circRNAs in the SDX of *P. trichocarpa* was 99.7% and 98.7%, respectively. Thus, the biogenesis of both ciRNAs and circRNAs requires the spliceosomal machinery, which is consistent with previous findings (Starke et al., 2015). We also determined that a greater number of ciRNAs (63.89%) was generated from lariat circularization compared to circRNAs (36.11%) generated from exonic circularization, suggesting that intron circularization is more efficient than exon circularization in SDX. It is also possible that lariat RNAs generated from excised introns in SDX degrade slowly, giving them more time for further trimming of the lariat tail downstream from the branchpoint to generate ciRNAs. Debranching enzyme1 (DBR1) linearizes intron lariats, which accumulate in the *db1-2* mutant (Li et al., 2016; Zhang et al., 2019b). The expression of *DBR1* also could be detected in SDX of *P. trichocarpa*, suggesting that the high percentage of ciRNAs might not have been due to slower than usual RNA lariat debranching. It would be interesting to investigate why the percentage of ciRNAs in *P. trichocarpa* is different from that in most other plants (Yin et al., 2018; Wang et al., 2019; Zhu et al., 2019; Zhang et al., 2019a; Fan et al., 2020), which have higher percentages of circRNAs vs. ciRNAs. Additional sequencing of circRNAs in other tissues in *P. trichocarpa* could help determine whether ciRNAs are more widespread in SDX than in other tissues.

PacBio sequencing generates longer reads than Illumina sequencing, which could facilitate the identification of splicing isoforms (Rhoads and Au, 2015; Abdel-Ghany et al., 2016; Wang et al., 2016). Illumina sequencing data were poorly correlated with PacBio SMRT data in a previous study since the sequencing depth of the Illumina platform is higher than that of the SMRT platform (Li et al., 2017). In the current study, we performed a quantitative comparison of sequencing data from SDX samples using the PacBio vs. Illumina platform. We identified 13,715 and 7,836 AS events using the Illumina and PacBio platforms, respectively. However, the percentage of overlapping data between the two platforms was low since PacBio more readily identifies long introns and the Illumina platform is better at detecting isoforms with low abundance. The performance was better for identifying complex isoforms from long reads versus short reads from the Illumina platform. However, the use of short reads from the Illumina platform not only corrects the errors from long reads, but it also provides very high throughput to

improve estimations of the abundance of isoforms (2017). Thus, hybrid-seq (long reads + short reads) strategies could be used to discover AS events that would be missed using either platform alone, such as isoforms with complex splicing or isoforms with low abundance.

A previous study provided a novel mechanistic insight into the role of circRNA in regulating splicing via R-loops to modulate floral homeotic phenotypes (Conn et al., 2017). However, whether this is a common mechanism remains unknown. Therefore, we performed DRIPc-Seq and identified 4,834 R-loop plus (+) peaks and 4,098 R-loop minus (–) peaks in *P. trichocarpa* SDX. In total, there are 24,451 genes with R-loops in *Arabidopsis* (Xu et al., 2017). We identified only 6,416 genes with R-loop peaks, fewer than was found in a previous study (Xu et al., 2017). This deviation might be caused by the different methods used for library construction, which focused on DNA strands in the *Arabidopsis* study and RNA strands in the current study. However, most R-loop genes in *P. trichocarpa* share high levels of homology. Among these genes, 72.55% (4,655/6,416) included R-loop peaks, suggesting that the formation of this structure might be a conserved regulatory mechanism in both *Arabidopsis* and *P. trichocarpa*. In addition, the GARGAAG motif is conserved in *P. trichocarpa*, pointing to a common mechanism of R-loop formation. A genome-wide analysis of R-loop peaks revealed enrichment in active histone modifications and negative correlation with CG hypermethylation in the model plant *Arabidopsis* (Xu et al., 2017). It would be interesting to investigate the role of the interplay between R-loop structure and chromatin structure via epigenetic modifications, including DNA methylation and histone modifications, in the transcriptional regulation of gene expression in *P. trichocarpa*.

In total, 672 AS events and 181 circRNAs directly overlapped with R-loop peaks. It is also possible that circRNAs from one locus regulates the splicing of another locus by trans-regulation. However, it is difficult to distinguish between cis- and trans-regulated R-loop formation using DRIPc-Seq. Thus, we could only detect possible trans-regulation based on the identification of R-loops and the complementary sequences between circRNAs and splicing regions. Further experiments are required to validate the trans-regulation of circRNAs.

R-loop peaks associated with circRNAs were validated using DRIP-qPCR of SDX protoplasts overexpressing circRNAs. Indeed, *Arabidopsis* plants overexpressing circRNAs showed more obvious R-loop formation than plants overexpressing their linear counterparts (Conn et al., 2017). However, we still cannot rule out the possibility that some linear RNAs also contribute to the biogenesis of R-loop peaks. More experiments are needed to reveal the exact roles of circRNAs and linear RNAs in regulating R-loop structure to affect AS in the SDX of *P. trichocarpa*.

Overexpressing circGORK enhanced the splicing of two major isoforms (ID6 and ID7) from cognate linear transcripts of GORK in *Arabidopsis* (Zhang et al., 2019a). Phenotypic analysis showed that these transgenic plants were more

tolerant to drought stress than control plants (Zhang et al., 2019a). Overexpressing circ-0003418 altered the expression of the *PtoXBAT32.5* transcript variant and inhibited lateral root development (Song et al., 2020), highlighting the role of circRNA in regulating splicing. These stable transformation experiments revealed the roles of circRNAs in regulating the splicing of their cognate linear RNAs. However, the biological functions of most circRNAs in *P. trichocarpa* remain unknown. In the current study, we transfected protoplasts from *P. trichocarpa* SDX with circRNAs and identified differential splicing events. However, the true biological functions of circRNAs in splicing regulation can only be revealed by phenotypic analysis of stable transformants harboring distinct splicing isoforms and circRNAs such as circ-*IRX7*. Nonetheless, the current data represent a useful resource for the stable transformation of circRNAs in plants to further investigate the interplay between circRNAs and AS in the future.

## MATERIALS AND METHODS

### Plant Materials

*Populus trichocarpa* plants were grown in a greenhouse with an average temperature of 22°C under a 16-h light/8-h dark cycle. All plant samples were wrapped in aluminum foil, immediately frozen in liquid nitrogen, and stored at -80°C (Shen et al., 2019). Total RNA was isolated from SDX using an RNAprep Pure Plant Kit (polysaccharides and polyphenolics-rich) (no. DP441; Tiangen, China) and treated with DNase I to remove DNA. Genomic DNA was extracted from SDX using a Plant Genomic DNA Kit (no. DP305; Tiangen). The quality and concentration of RNA and DNA were determined using a NanoDrop 2000 spectrophotometer (Thermo scientific).

### Library construction and bioinformatics analysis of circRNA

In total, 2-μg total RNA samples from SDX were used for circRNA library construction. A Ribo-Zero Magnetic Kit (Epicentre) was used to remove ribosomal RNAs (rRNAs), and 1 μl (20 U) RNase R was added to the rRNA-depleted RNA samples to digest linear RNAs at 37°C for one hour. The enriched circRNAs were randomly fragmented and DNA fragments 400-500 bp in size were selected as templates for PCR amplification. A strand-specific RNA-Seq library was then constructed as previously described (Wu et al., 2014; Wang et al., 2017), and the dUTP strand-specific library was sequenced to generate 150-nt paired-end reads.

Raw reads from RNase R-treated RNA sequencing were filtered using ht2-filter (1.92.1) from the HTQC package (Yang et al., 2013) using the default option. These reads were mapped to *P. trichocarpa* genome v3.0 using TopHat (Trapnell et al., 2009) with the option -r 50 -a 8. Unmapped reads including circRNA junction sequences were remapped to the reference genome using fusion-search with the following option: tophat2 -p 50 --fusion-search --keep-fasta-order --no-coverage-search --bowtie1. Candidate circRNAs were identified using CIRCexplorer (1.1.10) (Zhang et al., 2016). The flanking

complementary sequences of the circRNAs were extracted using the CSI method (Dong et al., 2017).

### Validation of circRNA by RT-PCR

Divergent primers for the circRNAs were designed to amplify the back-spliced junction using PRAPI (Gao et al., 2018). A 20-μg sample of RNA was dissolved in 54-μL RNase-free water and equally divided into control and RNase R-treated samples. For RT-PCR, 10 μL total RNA (RNase R treated or untreated sample) was reverse transcribed into cDNA in a 20-μL reaction mixture using a PrimeScript™ II first Strand cDNA Synthesis Kit (Cat# 6210A; TaKaKa) with random primers. PCR products of the predicted sizes were visualized and excised from a 2% agarose gel. Divergent primers for circRNAs and convergent primers for *EF1α* and 18S rRNA for RT-PCR are listed in Table S3. In addition to RNase R enrichment experiments, Sanger sequencing was performed to validate the selected circRNAs.

### Overexpressing circRNA in SDX protoplasts

Genomic DNA of circ-*IRX7* or circ-*GUX1* generated from the hemicellulose synthase gene was amplified using convergent primers covering the circRNA sequence and two flanking intron sequences with two linkers (*attB* sides) on the 5' and 3' ends (Table S3). In total, approximately 200 ng of genomic DNA in a reaction volume of 50 μL was subjected to PCR (30 cycles: 98°C, 10 s; 50°C, 20 s; 68°C, 1 min) to amplify the target *attB*-PCR products, which were mixed with the pDONR207 plasmid and *attP* sides to generate *attL*-containing entry clones. Finally, the LR recombination reaction was performed with the *pUC19-35S-sGFP* vector to generate the expression vector *pUC19-35S-circ-IRX7-35S-sGFP* and *pUC19-35S-circ-GUX1-35S-sGFP*, which was prepared using an EndoFree Maxi Plasmid Kit (no. DP117; Tiangen) for SDX protoplast transfection.

Protoplast isolation and transfection were performed as previously described (Lin et al., 2013; Lin et al., 2014). In brief, 10-cm de-barked stem segments were submerged into 40 mL of cell wall enzyme digestion solution (1.5% (w/v) Cellulase R-10, 0.4% (w/v) Macerozyme R-10, 20 mM MES (pH 5.7), 0.5 M mannitol, 20 mM KCl, 10 mM CaCl<sub>2</sub> and 0.1% (w/v) BSA) in a 50-mL tube for 3 h in the dark at room temperature. Protoplasts from SDX were released by gently shaking the 50-mL tube for 1-2 min in MMG solution (0.25 M mannitol, 4 mM MES (pH 5.7), and 15 mM MgCl<sub>2</sub>). The concentrations of protoplasts were measured with a hemocytometer (no. 3110; Sigma-Aldrich, China) and adjusted to  $2 \times 10^5$ - $8 \times 10^5$  cells/mL with MMG solution. For protoplast transfection, protoplasts, plasmid DNA (*pUC19-35S-circ-IRX7-35S-sGFP* and *pUC19-35S-circ-GUX1-35S-sGFP*), and PEG-4000 solution (40% (w/v) PEG 4000, 0.2 M mannitol and 100 mM CaCl<sub>2</sub>) were combined at a volume ratio of 10:1:11. The transformation system was incubated for 10 min at room temperature. Protoplasts from the same batch transfected with empty vector (*pUC19-35S-sGFP*) were used as a control. Transfected protoplasts were incubated in WI solution (0.25 M mannitol, 4 mM MES (pH 5.7), 20 mM KCl) in a Petri dish coated with 1% (w/v) BSA



solution for 12 h at room temperature in the dark. To generate cDNA, each 0.5- $\mu$ g RNA sample from transfected SDX protoplasts was reverse transcribed using random primers with a PrimeScript™ II first Strand cDNA Synthesis Kit. The cDNA samples were diluted 5 $\times$ . Divergent primers (circ-*IRX7* and circ-*GUX1*) and convergent primers (linear-*IRX7* and linear-*GUX1*, Table S3) were used to detect the overexpression of circRNA and the AS of host genes, respectively. The PCR products were visualized on a 1.5% agarose gel stained with GelStain (no. GS101-03; TransGen, China).

### PacBio Iso-Seq, bioinformatics analysis, and validation

Total RNA concentrations were measured using an Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA). RNA samples with RNA integrity number (RIN) > 8.5 were reserved for PacBio Iso-Seq library construction. Double-stranded cDNA synthesis was performed using a SMARTer PCR cDNA Synthesis Kit (no. 634925; Clontech). During the cDNA size selection step, we used Blue Pippin (Sage Science) to select cDNA fragments 1-2, 2-3, 3-6, and 5-10 kb in size for downstream library construction. The selected cDNA fragments were further amplified by PCR to obtain sufficient cDNA for subsequent library construction using a Template Prep Kit (no. 100-259-100; PacBio). Adaptor dimer and other contaminating sequences were removed. The small insert libraries (1-2 and 2-3 kb) were purified twice with PB beads. Large insert libraries (3-6 and 5-10 kb) were obtained using Blue Pippin selection.

To pre-process the PacBio Iso-Seq data, we used ConsensusTools.sh from smrtanalysis\_2.3.0 to obtain circular consensus sequences (CCS) with the following parameter: --minFullPasses 0 --minPredictedAccuracy 80. Full-length non-chimeric reads (FLNC) were generated by pbtranscript.py classify with the following command: --min\_seq\_len 300. The FLNC reads were corrected by LSC (Au et al., 2012) with the default option. The FLNC reads were aligned to *P. trichocarpa* genome v3.0 using GMAP (Wu and Watanabe, 2005) with the following option: --no-chimeras --cross-species --expand-offsets 0 -K 8000 -f 2 -n 1 -t 40. The AS from FLNC reads were identified using rMATS (Shen et al., 2014). Motif analysis was performed by MEME-CHIP (Machanick and Bailey, 2011) with the following option: meme-chip -norc -oc output-meme-minw 6 -meme-maxw 20 -meme-p 40.

Each 1- $\mu$ g RNA sample from SDX was reverse transcribed into cDNA using Oligo dT as the primer performed using a PrimeScript™ II first Strand cDNA Synthesis Kit. 5 $\times$  cDNA was used as a template for the identification of AS (Table S11). RT-PCR was carried out using 15- $\mu$ L Premix Taq (TaKaRa) in a 30- $\mu$ L reaction system (1  $\mu$ L 5 $\times$  cDNA, 0.5  $\mu$ L each F/R primer, 13  $\mu$ L ddH<sub>2</sub>O) for 33 cycles (94°C, 45 s; 55°C, 35 s; 72°C, 30 s). The PCR products were visualized on a 1.5% agarose gel stained with GelStain (TransGen).

### DRIPc-seq and DRIP-seq to identify R-loop peaks

DRIP was performed as described previously (Xu et al., 2017; Sanz and Chedin, 2019) with minor modifications. In brief, nuclei were isolated from 2.5–3 g of SDX powder using

Honda buffer (0.44 M sucrose, 1.25% Ficoll 400, 2.5% Dextran T40, 20 mM HEPES-KOH (pH 7.4), 10 mM MgCl<sub>2</sub>, 0.5% Triton X-100, 5 mM  $\beta$ -mercaptoethanol, 1 mM PMSF, and 1% protease inhibitors) (Sun et al., 2013). The isolated nuclei were resuspended in nuclei lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) and incubated at 37°C overnight (12–14 h) after adding proteinase K. Genomic DNA was extracted from the samples using the phenol–chloroform method and precipitated with ethanol. The DNA was fragmented to 200–500-bp pieces with an automatic high-throughput DNA sonicator (Bioruptor Pico, Diagenode, BEL). The negative control was treated with RNase H (New England Biolabs) at 37°C overnight, and both the experimental group and negative control were subjected to DRIP using the S9.6 anti-RNA-DNA antibody at 4°C overnight. After antibody-RNA:DNA complexes were bound to Protein A/G Magnetic Beads and thoroughly washed, the DNA pellets were eluted in 30–50  $\mu$ L RNase-free ddH<sub>2</sub>O. After DRIP, the eluted DNA of the experimental group was treated with DNase I (New England Biolabs) at 37°C for 45 min to degrade all DNA. The RNA strands in R-loops were ethanol-precipitated and reverse-transcribed into cDNA using random hexamers. Second-strand synthesis was performed using deoxyuridine triphosphate (dUTP). A USER enzyme treatment was added before the PCR amplification step to ensure strand specificity in the experimental group. Sequencing libraries of the experimental group and control were checked on a StepOne Plus Real-time PCR system (ABI) prior to sequencing on the NovaSeq 6000 platform (Illumina, San Diego, CA, USA). R-loop levels were detected using DRIP-qPCR as described previously (Xu et al., 2017; Sanz and Chedin, 2019).

R-loop data were pre-processed as described previously (Xu et al., 2017) with minor modifications. Reads were aligned to *P. trichocarpa* genome v3.0 with bowtie2 (v2.2.1) using --local --phred33. Duplicated reads were removed with picard 2.21.6 (<http://picard.sourceforge.net/>). MACS2 (2.2.6) was used to identify peaks (Zhang, 2008). The aligned reads files were converted to normalized coverage files (bigWig) via Reads Per Genomic Content of 5 bp bins (Ramirez et al., 2014). R-loop peaks were visualized using IGV (Thorvaldsdóttir et al., 2013).

### Data availability

All clean reads were deposited in the SRA under the following accession numbers: CircRNA (SRR8517644, SRR8517643, and SRR8517642), PacBio Iso-Seq (SRR8447264), strand-specific RNA-Seq libraries (SRR13481183, SRR13481184, SRR13481185), and R-loop (SRR11845458). All R-loop peaks, circRNA coordinates, and AS events can be found at [forestry.fafu.edu.cn/db/Ptr\\_Circular\\_RNA](http://forestry.fafu.edu.cn/db/Ptr_Circular_RNA).

## ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2016YFD0600106).

## AUTHOR CONTRIBUTIONS

L.G. and A.R. conceived and designed the study. X.L., J.L., M.M., K.C., F.X., W.W., and H.W. performed the experiments. Y.G., Y.W., and X.X. analyzed the high-throughput sequencing data. X.L., Y.G., A.R., and L.G. wrote the manuscript. All authors have read and approved of the final version of the manuscript.

**Edited by:** Binglian Zheng, Fudan University, China

**Received** Nov. 21, 2020; **Accepted** Feb. 8, 2021; **Published** Feb. 11, 2021

## REFERENCES

- Abdel-Ghany, S.E., Hamilton, M., Jacobi, J.L., Ngam, P., Devitt, N., Schilkey, F., Ben-Hur, A., and Reddy, A.S. (2016). A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **7**: 1–11.
- Au, K.F., Underwood, J.G., Lee, L., and Wong, W.H. (2012). Improving PacBio long read accuracy by short read alignment. *PLoS One* **7**: e46679.
- Bao, H., Li, E., Mansfield, S.D., Cronk, Q.C., Elkassaby, Y.A., and Douglas, C.J. (2013). The developing xylem transcriptome and genome-wide analysis of alternative splicing in *Populus trichocarpa* (black cottonwood) populations. *BMC Genomics* **14**: 359.
- Bas, A., Forsberg, G., Hammarström, S., and Hammarström, M.L. (2004). Utility of the housekeeping genes 18S rRNA, beta-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. *Scand. J. Immunol.* **59**: 566–573.
- Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P., and Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell* **73**: 1019–1030.
- Chaudhary, S., Jabre, I., Reddy, A.S., Staiger, D., and Syed, N.H. (2019). Perspective on alternative splicing and proteome complexity in plants. *Trends Plant Sci.* **24**: 496–506.
- Cheng, J., Zhang, Y., Li, Z., Wang, T., Zhang, X., and Zheng, B. (2018). A lariat-derived circular RNA is required for plant development in *Arabidopsis*. *Sci. China: Life Sci.* **61**: 204–213.
- Conn, V.M., Hugouvieux, V., Nayak, A., Conos, S.A., Capovilla, G., Cildir, G., Jourdain, A., Tergaonkar, V., Schmid, M., and Zubieta, C. (2017). A circRNA from SEPALLATA3 regulates splicing of its cognate mRNA through R-loop formation. *Nat. Plants* **3**: 17053.
- Deng, X., Gu, L., Liu, C., Lu, T., Lu, F., Lu, Z., Cui, P., Pei, Y., Wang, B., and Hu, S. (2010). Arginine methylation mediated by the *Arabidopsis* homolog of PRMT5 is essential for proper pre-mRNA splicing. *Proc. Natl. Acad. Sci. USA* **107**: 19114–19119.
- Dong, R., Ma, X.K., Chen, L.L., and Yang, L. (2017). Increased complexity of circRNA expression during species evolution. *RNA Biol.* **14**: 1064–1074.
- Fan, J., Quan, W., Li, G.B., Hu, X.H., Wang, Q., Wang, H., Li, X.P., Luo, X., Feng, Q., Hu, Z.J., Feng, H., Pu, M., Zhao, J.Q., Huang, Y. Y., Li, Y., Zhang, Y., and Wang, W.M. (2020). circRNAs are involved in the rice-*Magnaporthe oryzae* interaction. *Plant Physiol.* **182**: 272–286.
- Filichkin, S.A., Hamilton, M., Dharmawardhana, P.D., Singh, S.K., Sullivan, C., Ben-Hur, A., Reddy, A.S. and Jaiswal, P. (2018). Abiotic stresses modulate landscape of poplar transcriptome via alternative splicing, differential intron retention, and isoform ratio switching. *Frontiers in Plant Science* **9**: 5.
- Gao, Y., Wang, H., Zhang, H., Wang, Y., Chen, J., and Gu, L. (2018). PRAP1: Post-transcriptional regulation analysis pipeline for Iso-Seq. *Bioinformatics* **34**: 1580–1582.
- Ginno, P.A., Lim, Y.W., Lott, P.L., Korf, I., and Chédin, F. (2013). GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Res.* **23**: 1590–1600.
- Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**: 814–825.
- Haddad, G., and Lorenzen, J.M. (2019). Biogenesis and function of circular RNAs in health and in disease. *Front. Pharmacol.* **10**: 428.
- Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* **495**: 384–388.
- Hsu, M.T., and Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* **280**: 339–340.
- Hu, H., Yang, W., Zheng, Z., Niu, Z., Yang, Y., Wan, D., Liu, J., and Ma, T. (2020). Analysis of alternative splicing and alternative polyadenylation in *Populus alba* var. *pyramidalis* by single-molecule long-read sequencing. *Front. Genet.* **11**: 48.
- Huang, J., Gu, L., Zhang, Y., Yan, T., Kong, G., Kong, L., Guo, B., Qiu, M., Wang, Y., and Jing, M. (2017). An oomycete plant pathogen reprograms host pre-mRNA splicing to subvert immunity. *Nat. Commun.* **8**: 2051.
- Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**: 141–157.
- Li, Q., Lin, Y.C., Sun, Y.H., Song, J., Chen, H., Zhang, X.H., Sederoff, R.R., and Chiang, V.L. (2012). Splice variant of the SND1 transcription factor is a dominant negative of SND1 members and their regulation in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. USA* **109**: 14699–14704.
- Li, S., Lin, Y.J., Wang, P., Zhang, B., Li, M., Chen, S., Shi, R., Tunlaya-Anukit, S., Liu, X., Wang, Z., Dai, X., Yu, J., Zhou, C., Liu, B., Wang, J.P., Chiang, V.L., and Li, W. (2019). The AREB1 transcription factor influences histone acetylation to regulate drought responses and tolerance in *Populus trichocarpa*. *Plant Cell* **31**: 663–686.
- Li, Y., Dai, C., Hu, C., Liu, Z., and Kang, C. (2017). Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry. *Plant J.* **90**: 164–176.
- Li, Z., Huang, C., Bao, C., Chen, L., Lin, M., Wang, X., Zhong, G., Yu, B., Hu, W., and Dai, L. (2015). Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* **22**: 256–264.
- Li, Z., Wang, S., Cheng, J., Su, C., Zhong, S., Liu, Q., Fang, Y., Yu, Y., Lv, H., and Zheng, Y. (2016). Intron lariat RNA inhibits microRNA biogenesis by sequestering the dicing complex in *Arabidopsis*. *PLoS Genet.* **12**: e1006422.
- Lin, Y.C., Li, W., Sun, Y.H., Kumari, S., Wei, H., Li, Q., Tunlaya-Anukit, S., Sederoff, R.R., and Chiang, V.L. (2013). SND1 transcription factor-directed quantitative functional hierarchical genetic regulatory network in wood formation in *Populus trichocarpa*. *Plant Cell* **25**: 4324–4341.
- Lin, Y.C., Li, W., Chen, H., Li, Q., Sun, Y.H., Shi, R., Lin, C.Y., Wang, J. P., Chen, H.C., Chuang, L., Qu, G.Z., Sederoff, R.R., and Chiang, V. L. (2014). A simple improved-throughput xylem protoplast system for studying wood formation. *Nat. Protoc.* **9**: 2194–2205.
- Lin, Y.J., Chen, H., Li, Q., Li, W., Wang, J.P., Shi, R., Tunlaya-Anukit, S., Shuai, P., Wang, Z., Ma, H., Li, H., Sun, Y.H., Sederoff, R.R., and Chiang, V.L. (2017). Reciprocal cross-regulation of VND and SND multigene TF families for wood formation in *Populus trichocarpa*. *Proc. Natl. Acad. Sci. USA* **114**: E9722–E9729.
- Lu, T., Cui, L., Zhou, Y., Zhu, C., Fan, D., Gong, H., Zhao, Q., Zhou, C., Zhao, Y., Lu, D., Luo, J., Wang, Y., Tian, Q., Feng, Q., Huang, T., and Han, B. (2015). Transcriptome-wide investigation of circular RNAs in rice. *RNA* **21**: 2076–2087.

- Machanick, P., and Bailey, T.L.** (2011). MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics* **27**: 1696–1697.
- Panda, A.C., De, S., Grammatikakis, I., Munk, R., Yang, X., Piao, Y., Dudekula, D.B., Abdelmohsen, K., and Gorospe, M.** (2017). High-purity circular RNA isolation method (RPAD) reveals vast collection of intronic circRNAs. *Nucleic Acids Res.* **45**: e116.
- Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., Sun, W., Dou, K., and Li, H.** (2015). Circular RNA: A new star of noncoding RNAs. *Cancer Lett.* **365**: 141–148.
- Ramirez, F., Dündar, F., Peter, S., Grüning, B., and Manke, T.** (2014). DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**: W187–W191.
- Rhoads, A., and Au, K.F.** (2015). PacBio sequencing and its applications. *Genom. Proteom. Bioinf.* **13**: 278–289.
- Roberts, R.W., and Crothers, D.M.** (1992). Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science* **258**: 1463–1466.
- Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O.** (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**: e30733.
- Sanz, L.A., and Chedin, F.** (2019). High-resolution, strand-specific R-loop mapping via S9.6-based DNA-RNA immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **14**: 1734–1755.
- Shen, L., Liang, Z., Wong, C.E., and Yu, H.** (2019). Messenger RNA modifications in plants. *Trends Plant Sci.* **24**: 328–341.
- Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., and Xing, Y.** (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. USA* **111**: E5593–E5601.
- Song, Y., Bu, C., Chen, P., Liu, P., and Zhang, D.** (2020). Miniature inverted repeat transposable elements (MITEs) cis-regulate circular RNA expression and promote ethylene biosynthesis, reducing poplar heat tolerance. *J. Exp. Bot. eraa570*.
- Starke, S., Jost, I., Rossbach, O., Schneider, T., Schreiner, S., Hung, L. H., and Bindereif, A.** (2015). Exon circularization requires canonical splice signals. *Cell Rep.* **10**: 103–111.
- Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N.J., and Dean, C.** (2013). R-loop stabilization represses antisense transcription at the *Arabidopsis* FLC locus. *Science* **340**: 619–621.
- Sun, X., Wang, L., Ding, J., Wang, Y., Wang, J., Zhang, X., Che, Y., Liu, Z., Zhang, X., Ye, J., Wang, J., Sablok, G., Deng, Z., and Zhao, H.** (2016). Integrative analysis of *Arabidopsis thaliana* transcriptomics reveals intuitive splicing mechanism for circular RNA. *FEBS Lett.* **590**: 3510–3516.
- Tan, J., Zhou, Z., Niu, Y., Sun, X., and Deng, Z.** (2017). Identification and functional characterization of tomato circRNAs derived from genes involved in fruit pigment accumulation. *Sci. Rep.* **7**: 8594.
- Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P.** (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**: 178–192.
- Trapnell, C., Pachter, L., and Salzberg, S.L.** (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105–1111.
- Wang, B., Tseng, E., Regulski, M., Clark, T.A., Hon, T., Jiao, Y., Lu, Z., Olson, A., Stein, J.C., and Ware, D.** (2016). Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **7**: 1–13.
- Wang, T., Wang, H., Cai, D., Gao, Y., Zhang, H., Wang, Y., Lin, C., Ma, L., and Gu, L.** (2017). Comprehensive profiling of rhizome associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). *Plant J.* **91**: 684–699.
- Wang, Y., Gao, Y., Zhang, H., Wang, H., Liu, X., Xu, X., Zhang, Z., Markus, V.K., Kaiqiang, H., Wang, H., Xi, F., Zhao, L., Lin, C., and Gu, L.** (2019). Genome-wide profiling of circular RNAs in the rapidly growing shoots of moso bamboo (*Phyllostachys edulis*). *Plant Cell Physiol.* **60**: 1354–1373.
- Wang, Y., Wang, H., Xi, F., Wang, H., Han, X., Wei, W., Zhang, H., Zhang, Q., Zheng, Y., Zhu, Q., Kohnen, M.V., Reddy, A.S.N., and Gu, L.** (2020). Profiling of circular RNA N(6)-methyladenosine in moso bamboo (*Phyllostachys edulis*) using nanopore-based direct RNA sequencing. *J. Integr. Plant Biol.* **62**: 1823–1838.
- Weirather, J.L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.J., Buck, D., and Au, K.F.** (2017). Comprehensive comparison of pacific biosciences and oxford nanopore technologies and their applications to transcriptome analysis. *F1000 Res.* **6**: 100.
- Wu, B., Suo, F., Lei, W., and Gu, L.** (2014). Comprehensive analysis of alternative splicing in *Digitalis purpurea* by strand-specific RNA-Seq. *PLoS One* **9**: e106001.
- Wu, T.D., and Watanabe, C.K.** (2005). GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875.
- Wu, Z., Zhu, D., Lin, X., Miao, J., Gu, L., Deng, X., Yang, Q., Sun, K., Zhu, D., and Cao, X.** (2016). RNA binding proteins RZ-1B and RZ-1C play critical roles in regulating pre-mRNA splicing and gene expression during development in *Arabidopsis*. *Plant Cell* **28**: 55–73.
- Xu, P., Cai, X.T., Wang, Y., Xing, L., Chen, Q., and Xiang, C.B.** (2014). HDG11 upregulates cell-wall-loosening protein genes to promote root elongation in *Arabidopsis*. *J. Exp. Bot.* **65**: 4285–4295.
- Xu, S.L., Rahman, A., Baskin, T.I., and Kieber, J.J.** (2008). Two leucine-rich repeat receptor kinases mediate signaling, linking cell wall biosynthesis and ACC synthase in *Arabidopsis*. *Plant Cell* **20**: 3065–3079.
- Xu, W., Li, K., Li, S., Hou, Q., Zhang, Y., Liu, K., and Sun, Q.** (2020). The R-loop atlas of *Arabidopsis* development and responses to environmental stimuli. *Plant Cell* **32**: 888–903.
- Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., and Sun, Q.** (2017). The R-loop is a common chromatin feature of the *Arabidopsis* genome. *Nat. Plants* **3**: 704–714.
- Yang, X., Liu, D., Liu, F., Wu, J., Zou, J., Xiao, X., Zhao, F., and Zhu, B.** (2013). HTQC: A fast quality control toolkit for Illumina sequencing data. *BMC Bioinform.* **14**: 33.
- Ye, C.Y., Chen, L., Liu, C., Zhu, Q.H., and Fan, L.** (2015). Widespread noncoding circular RNAs in plants. *New Phytol.* **208**: 88–95.
- Yin, J., Liu, M., Ma, D., Wu, J., Li, S., Zhu, Y., and Han, B.** (2018). Identification of circular RNAs and their targets during tomato fruit ripening. *Postharvest Biol. Technol.* **136**: 90–98.
- Zeng, R.F., Zhou, J.J., Hu, C.G., and Zhang, J.Z.** (2018). Transcriptome-wide identification and functional prediction of novel and flowering-related circular RNAs from trifoliate orange (*Poncirus trifoliata* L. Raf.). *Planta* **247**: 1191–1202.
- Zhang, P., Fan, Y., Sun, X., Chen, L., Terzaghi, W., Bucher, E., Li, L., and Dai, M.** (2019a). A large-scale circular RNA profiling reveals universal molecular mechanisms responsive to drought stress in maize and *Arabidopsis*. *Plant J.* **98**: 696–713.
- Zhang, X., Zhang, Y., Wang, T., Li, Z., Cheng, J., Ge, H., Tang, Q., Chen, K., Liu, L., Lu, C., Guo, J., Zheng, B., and Zheng, Y.** (2019b). A comprehensive map of intron branchpoints and lariat RNAs in plants. *Plant Cell* **31**: 956–973.
- Zhang, X.O., Dong, R., Zhang, Y., Zhang, J.L., Luo, Z., Zhang, J., Chen, L.L., and Yang, L.** (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res.* **26**: 1277–1287.
- Zhang, Y.** (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**: 1–9.



**Zhao, W., Cheng, Y., Zhang, C., You, Q., Shen, X., Guo, W., and Jiao, Y.** (2017). Genome-wide identification and characterization of circular RNAs by high throughput sequencing in soybean. *Sci. Rep.* **7**: 1–11.

**Zhao, Y., Sun, J., Xu, P., Zhang, R., and Li, L.** (2014). Intron-mediated alternative splicing of WOOD-ASSOCIATED NAC TRANSCRIPTION FACTOR1B regulates cell wall thickening during fiber development in *Populus* species. *Plant Physiol.* **164**: 765–776.

**Zhou, J., Yuan, M., Zhao, Y., Quan, Q., Yu, D., Yang, H., Tang, X., Xin, X., Cai, G., Qian, Q., Qi, Y., and Zhang, Y.** (2021). Efficient deletion of multiple circle RNA loci by CRISPR-Cas9 reveals Os06circ02797 as a putative sponge for OsMIR408 in rice. *Plant Biotechnol. J.* <https://doi.org/10.1111/pbi.13544>

**Zhu, Y.X., Jia, J.H., Yang, L., Xia, Y.C., Zhang, H.L., Jia, J.B., Zhou, R., Nie, P.Y., Yin, J.L., Ma, D.F., and Liu, L.C.** (2019). Identification of cucumber circular RNAs responsive to salt stress. *BMC Plant Biol.* **19**: 1–18.

## SUPPORTING INFORMATION

**Table S1.** Statistics for reads from circular RNAs sequencing

**Table S2.** List of circRNAs and ciRNAs

**Table S3.** Divergent and convergent primers for circRNAs and linear transcripts

**Table S4.** Alternative back-splicing events of circRNAs

**Table S5.** Summary for reads from PacBio Iso-seq

**Table S6.** List of alternative polyadenylation events based on PacBio Iso-seq

**Table S7.** List of alternative splicing events based on long reads from PacBio Iso-seq

**Table S8.** List of alternative splicing events based on short reads from Illumina RNA-seq

**Table S9.** List of R-loop peaks based on DRIPc-seq

**Table S10.** List of R-loop peaks containing regions complementary to circRNAs

**Table S11.** Convergent primers for validation of alternative splicing events



Scan using WeChat with your smartphone to view JIPB online



Scan with iPhone or iPad to view JIPB online