Journal of Genetics (2021)100:29
https://doi.org/10.1007/s12041-021-01272-8

RESEARCH ARTICLE

SMRT sequencing of full-length transcriptome of birch-leaf pear (*Pyrus betulifolia* Bunge) under drought stress

QINGSONG YANG, JINXING WANG, ZHONGHUA WANG, XIAOGANG LI, JIALIANG KAN and JING LIN*

Institute of Pomology, Jiangsu Academy of Agricultural Sciences/Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Nanjing, Jiangsu 200014, People's Republic of China
*For correspondence. E-mail: lj84390224@126.com.

Received 29 January 2020; revised 23 September 2020; accepted 29 October 2020

Abstract. Drought limits the pear yield and quality. The birch-leaf pear (*Pyrus betulifolia* Bunge) is one of the most frequently used pear rootstocks. Identifying genes involved in drought resistance of *P. betulifolia* would suggest candidate genes for molecular breeding. We used single-molecule long-read sequencing technology to investigate the transcriptome of birch-leaf pear under drought stress. As a result, 362,139 consensus reads were identified using six databases, among which 342,162 genes were functionally annotated. Further, we identified 7094 long noncoding RNAs. The sequencing data contained 9891 alternative splicing and 100,836 alternative polyadenylation events. We report here the full-length sequence of birch-leaf pear, which can be used for breeding enhanced varieties.

Keywords. single-molecule real-time; full-length transcript; alternative splicing; drought stress; pear; *Pyrus betulifolia* Bunge.

Introduction

Short-read RNA sequencing is a powerful tool for analysing the gene expression (Nagalakshmi *et al*. 2008; Wang *et al*. 2009). However, it is unable to sequence the entire RNA molecule, and to identify full-length (FL) transcript isoforms (Sharon *et al*. 2013). Single-molecule real-time (SMRT) sequencing developed by Pacific Biosciences (PacBio, California, USA), is a third-generation sequencing technique that yield reads of at least several kilobases (Koren *et al*. 2013) and has several advantages compared to the first-generation and second-generation platforms (Schadt *et al*. 2010; Nakano *et al*. 2017). Although long-read sequencing technology produce reads with a high error rate, sequencing with high coverage enables self-correction of many errors (Chin *et al*. 2013; Allen *et al*. 2017). This technology has been used to obtain full-length (FL) cDNA sequences, it enhances genome annotation and enables the identification of RNA isoforms (Jiang *et al*. 2017). In *Zea mays*, PacBio sequencing yielded 111,151 unique FL transcript isoforms corresponding to 26,946 genes from six tissues (Wang *et al*. 2016). Also, 80,217 high-quality FL, and polished consensus transcripts were generated from *Camellia sinensis* (Xu *et al*. 2017). Moreover, a total of 160,468 nonredundant reads were produced by SMRT sequencing of root tissues from *Salvia miltiorrhiza* (Xu *et al*. 2015).

Long-chain noncoding RNAs (lncRNAs) are a class of RNAs that do not encode proteins (Kapranov *et al*. 2007). LncRNAs are >200 nucleotides in length with capped 5' ends and typically contain spliced introns (Guttman *et al*. 2009). lncRNAs regulate gene expression at the transcriptional and posttranscriptional levels by diverse mechanisms. A large number of lncRNAs have been identified using next-generation sequencing technologies. In wheat, 125 lncRNAs responded to powdery mildew infection and heat stress (Xin *et al*. 2011), and in *Populus*, 504 lncRNAs responded to drought stress (Shuai *et al*. 2014). In maize, 664 drought-responsive lncRNAs have been identified (Zhang *et al*. 2014). In birth-leaf pear, 251 lncRNAs were found to be drought-responsive (Wang *et al*. 2018).

Alternative splicing (AS), a major mechanism for generating transcript and proteome diversity occurs in the genomes of most eukaryotes (Vitulo *et al*. 2014; Liu *et al*. 2017). This phenomenon greatly increases the repertoire of proteins and regulates molecular, cellular, physiological and developmental processes/pathways (Kalsotra and Cooper

Qingsong Yang and Jinxing Wang contributed equally to this work.

Published online: 15 May 2021

2011; Wang *et al*. 2016). In *Arabidopsis*, 4707 (21.8%) genes with EST/cDNA evidence undergo 8264 AS events, ~56% of which are intron retention (IR). In rice, 6568 (21.2%) expressed genes exhibit 14,542 AS events, of which 53.5% are IR and 13.8% are exon skipping (ES) (Wang and Brendel 2006). Abiotic stresses significantly affect AS of pre-mRNAs in plants (Reddy 2007; Ali and Reddy 2008), and AS genes are over-represented among those related to stress responses (Filichkin *et al*. 2010; Abdel-Ghany *et al*. 2016).

The short reads of second-generation sequencing technology hamper unambiguous phasing of isoforms and mapping of highly repetitive sequences. Third-generation sequencing platforms, such as PacBio and Oxford Nanopore, utilize long reads to address this issue (Chen *et al*. 2018). Drought is the most common climatic disaster in China. It not only limits the crop yields but also has a major impact on crop quality. The birch-leaf pear (*Pyrus betulifolia* Bunge) originates from China and is one of the most widely used rootstocks for pear production, due to its drought and salt tolerance (Chang *et al*. 2012; Xu *et al*. 2015). Hence, identifying drought-resistance genes of *P. betulifolia* is critical. Improved assembly will link a larger number of contigs and scaffolds, thereby increasing the completeness and accuracy of gene annotation (Zimin *et al*. 2017). Here, we applied PacBio long-read technology to sequence pear transcriptome and obtained diverse FL cDNA sequences and isoforms. The results will enhance gene annotation and our understanding of the pear transcriptome.

Materials and methods

Plant materials

Birch-leaf pear (*P. betulifolia* Bunge) seedlings were grown in the seedling beds of the Institute of Pomology at JAAS, Nanjing, Jiangsu, China. They were placed in a growth chamber under a 24-h cycle: 14 h at 25°C in light and 10 h at 20°C in darkness (Xu *et al*. 2015). Seedlings at the six-leaf stage were transferred to half-strength Murashige and Skoog (MS) medium containing 15% polyethylene glycol (PEG) for 48 h. Seedlings in half-strength MS medium without PEG were used as controls. For SMRT sequencing, we collected samples from leaf, stem, and root of control and PEG-treated plants at 48 h. The samples were rinsed with cold distilled water and immediately frozen in liquid nitrogen at −80°C.

Library preparation and PacBio sequencing

Total RNA was extracted from pear using a Quick-RNA MiniPrep kit (TaKaRa, Dalian, China) according to the manufacturer's instructions. RNA degradation was analysed by agarose gel electrophoresis, and RNA purity using a NanoDrop instrument. First-strand cDNA was synthesized using a Clontech SMARTer PCR cDNA Synthesis kit with the isoform sequencing (iso-seq) protocol. Amplification was performed using Phusion DNA polymerase (NEB), and the resulting cDNAs were fractionated according to length (<4000 and >4000) using the BluePipin Size Selection System protocol. Next, amplification was performed using 12 PCR cycles. The cDNA products were used to generate SMRTbell template libraries. The SMRTbell templates were sequenced using the PacBio Sequel System by Beijing Novo (Beijing, China) using the Pacific platform.

Error correction of SMRT reads

To obtain clean reads, raw SMRT sequencing data were processed using SMRTlink 4.0 software by removing polymerase reads of <50 bp with a quality of <0.75. A circular consensus sequence (CCS) was generated from subread BAM files, CCS.BAM files were output and were classified into full-length (FL) and nonfull-length nonchimera (FLNC) reads. FLNC transcripts were identified by searching poly-A tail and 5′-tail cDNA primers. Iterative clustering of error correction (ICE) was used to produce consensus sequences by approaching clustering, and the FL consensus sequences from ICE were checked using Quiver. Consensus reads with a post-correction accuracy of >99% were selected for further analysis.

Prediction of lncRNAs

We used the coding–noncoding index (CNCI), coding potential calculator (CPC) (Li *et al*. 2014), and Pfam-scan tools to identify coding and noncoding transcripts. The CNCI profiles adjoining nucleotide triplets enabled discrimination of protein-coding and noncoding sequences independently of known annotations. CPC assesses the number and quality of ORFs in a transcript and searches protein sequence databases to identify coding and noncoding transcripts. We used Pfam-scan to identify protein family domains in the Pfam database. Any transcript with a Pfam hit was excluded from subsequent analysis. Transcripts predicted to encode proteins were filtered out, and noncoding transcripts were regarded as candidate lncRNAs.

Functional analysis

To functionally annotate the transcripts, we first aligned the consensus sequences to the pear reference genome using GMAP. Unmapped transcripts were functionally annotated using the following databases: NCBI nonredundant protein sequence (NR), NCBI nonredundant nucleotide sequence (NT); protein family (Pfam), clusters of orthologous groups of proteins (KOG/COG), Swiss-Prot (a manually annotated

and reviewed protein sequence database), KEGG ortholog (KO), and gene ontology (GO). We used BLAST and set the e-value to 1e−10 for the NT database analysis; Diamond BLASTX with an e-value of 1e−10 for the NR, KOG, Swiss-Prot, and KEGG database analyses; and Hmmscan for the Pfam database analysis.

Analysis of AS

After removing redundancy, transcripts were subjected to AS analysis. The AS events comprise of intron retention (IR), exon skipping (ES), alternative 5′ splice site (Alt.5′), and alternative 3′ splice site (Alt.3′). IR is defined as retention of one intron within a longer exon flanked by two shorter exons. ES is the absence of an exon in some transcripts but its presence in others. When an intron is excised at more than one site and linked to its 5′ or 3′ exons with different boundaries, the exons are considered Alt.5′ and Alt.3′, respectively. AStalavista tool was used to analyse the four major types of AS events.

RT-PCR validation

To examine the expression profiles of drought-responsive lncRNAs, 10 lncRNAs were randomly selected for expression analysis through RT-qPCR. Total RNA, used as template, was extracted from leaves using the total RNA kit (Tiangen, Beijing, China) according to the manufacturer's instructions. The first cDNA strand was synthesized from 1000 ng total RNA in a volume of 20 μL using the PrimeScript RT reagent kit with gDNA eraser (Perfect Real Time) (Clontech, Shiga, Japan) according to the manufacturer's protocol. Primers were designed using Primer Premier 5.0 software and the primer sequences are shown in table 1. RT-qPCR was performed on a 7500 real-time PCR system (Applied Biosystems, USA). The total reaction volume was 20 μL containing 10 μL 2x SYBR Premix Ex Taq (TaKaRa Bio, Japan), 1 μL complementary DNA (cDNA) reaction mixture, 0.5 μL of each primer with concentration is 10 pmol/μL, 0.5 μL ROX Reference DyeII (50×), and 7.5 μL double distilled H_2O. The primers sequence used in our PCR experiments are described in table 1. PCR was performed as follows: predenaturation at 95°C for 30 s, denaturation at 95°C for 3 s, annealing at 60°C for 30 s, and 55–95°C for melting curve analysis. All reactions were performed using biological triplicates. The $2^{-\Delta\Delta CT}$ method was used to calculate relative changes in gene expression between control and treatment plants (Livak and Schmittgen 2001). Ubiquitin (*UBQ*) gene was used as housekeeping gene for normalization.

Results

Sequencing of the FL pear transcriptome

To obtain the sequences of as many expressed transcripts as possible from pear tissue, we constructed a cDNA library using pooled RNA from PEG-treated and control pears. We prepared two size-fractionated cDNA libraries (<4000 and >4000) and subjected them to PacBio single-molecule, long-read sequencing. A total of 17.86 G subreads were yielded by four SMRT cells, and the sequences were processed using ToTu after filtering low-quality and chimeric regions; 986,205 high-quality CCS of average length 1502 nt were obtained (table 2). Also, 118,566 FL consensus sequences were classified as FL based on the presence of barcoded primers and poly-A tails. The high-quality FL consensus sequences contained 750,693 full-length non-chimeric cDNA sequences (FLNCs); after removing redundant sequences, 362,139 consensus reads were subjected to gene annotation and function analyses. The length distribution of consensus reads of birth-leaf pear by this sequencing is shown in figure 1.

Functional annotations of protein-coding genes

To functionally annotate the transcripts, we first mapped the SMRT consensus sequences onto the pear reference genome

Table 1. Sequence of the primers used in qRT-PCR analysis.

Gene	Forward primer	Reverse primer	Amplicon length (pb)
461	TCGTGCTTTCACCCTTAG	GATTGTTCGATCCTTTGT	96
1867	ACCCAATAAAACAAACTG	TTAATAATCCTCACGCAT	94
4629	ACCTTATCTTGCTCTCTC	GTAAACGCACTACTCATT	118
4907	TACCCAGTCCCTCTTCT	GGCTTATTGTGTCCATT	91
5338	TAAATAAGTGACAGCCC	TGAAGATAAAGGAGCAG	105
6041	ATTTTCCACGTCGCTTT	TGCTTCTTTGGGTCTGC	95
5961	AGAAAAATTGTTGTCGC	GGCTGATTATGGTGGTA	101
4536	GCCAATGAAATCTGTTC	CTCTCTCCTGTGTACGA	103
2066	GCAGTTTGGTCAAACCT	ACTTACAATCTCCCCGT	111
1805	GAGGGATGGAAAGGGAA	AACACGCCAAGGGACAC	107
UBQ	TGGTGTGAACGAGAAGGA	CCCTCAACAATCCCAAAC	108

Table 2. Summary statistics for SMRT-seq data in the PEG-treated and control birth-leaf pear.

CCS	5′ Primer	3′ Primer	Poly-A	FL	Flnc	Average Flnc	Consensus reads
986,205	873,256	904,681	893,216	778,566	750,693	1918	362,139

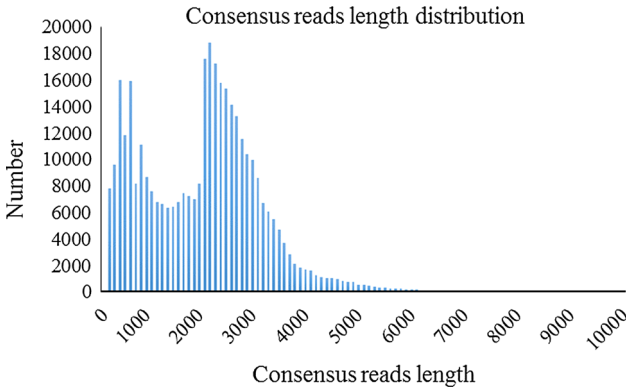

Figure 1. Length distribution of all subreads extracted from the complete dataset by SMRT. Subreads are in bp. The most abundant length was 2000–3000 bp.

using TopHat v. 2.0.12 and mapped 128,880 consensus sequences using BLAST. The remaining 233,259 consensus sequences were subjected to functional annotation using the Nr, Nt, SwissProt, KOG/COG, KEGG, and GO databases (table 3). A total of 145,942 consensus sequences were annotated using the Nr database. A total of 118,436 sequences were classified into 25 KOG categories, and 68,503 (11%) had GO annotations (figure 2). A total of 141,823 sequences were involved in 111 KEGG pathways. Also, 170,543, 140,996 and 68,503 genes were annotated using the SwissProt, Nt and Pfam databases, respectively. In summary, 213,282 genes had matches in at least one of the protein databases, accounting for ∼95% of the genes. We examined the taxonomic distribution of the best BLASTx hits, 60,377 of which matched the sequences of *P. bretschneideri* as well as *Malus domestica* (32,739) and *Anthurium amnicola* (6005) (figure 3).

Identification of lncRNAs

LncRNAs are defined as ncRNAs of >200 nt. To obtain a high-confidence set of lncRNA genes and to avoid false positives, we used three independent methods of identification and retained only the consensus sequences.

A total of 7198 reads were identified as putative lncRNAs using the CPC for nonredundant long reads. Analysis using CNCI software identified 20,994 putative lncRNAs according to homology searches in the reference protein database. Using Pfam-scan, after excluding those with hits in Pfam, 25,111 transcripts were identified as lncRNAs. Finally, 7094 transcripts were identified as potential lncRNAs based on the results of CPC, CNCI, and Pfam-scan (figure 4).

The lncRNA transcripts were classified into four groups based on their positions relative to the RefGen_v3 annotations: 6430 (92.09%) were from intergenic regions, and 181 (1.16%) from intronic regions, while 471 were antisense (figure 5). In addition, 5491 (78.6%) of the lncRNAs had a single exon.

Analysis of AS

AS occurs in most eukaryotic cells and can get significantly impacted by stress. The long reads generated by SMRT sequencing can be used to identify AS events by hybrid-seq isoform detection and a prediction pipeline. Of the 22,687 mapped genes, 9891 exhibited AS isoforms. Using ASTALAVISTA, we found that the most common AS event was Alt.3′ (43.3%) (figure 6). Alt.5′ and IR accounted for 38.8% and 14.7%, respectively, and ES for 3.1%, of AS events.

RT-qPCR validation

To examine the expression profiles of lncRNA, 10 lncRNAs were randomly selected for expression analysis through RT-qPCR. It can be seen from the figure 7, the relative expression of lncRNA461, lncRNA1867, lncRNA6041 and lncRNA2066 increase obviously after drought stress, but the others have no evident change. These results indicate the drought stress affected the expression of some lncRNAs.

Table 3. Annotations of consensus reads of *P. betulifolia*.

Gene	Nr	SwissProt	KEGG	KOG	GO	Nt	Pfam
Annotation	145,942	170,543	141,823	118,436	68,503	140,996	68,503

Using seven database, we annotated 213,282 transcripts that did not map onto the pear reference genome.

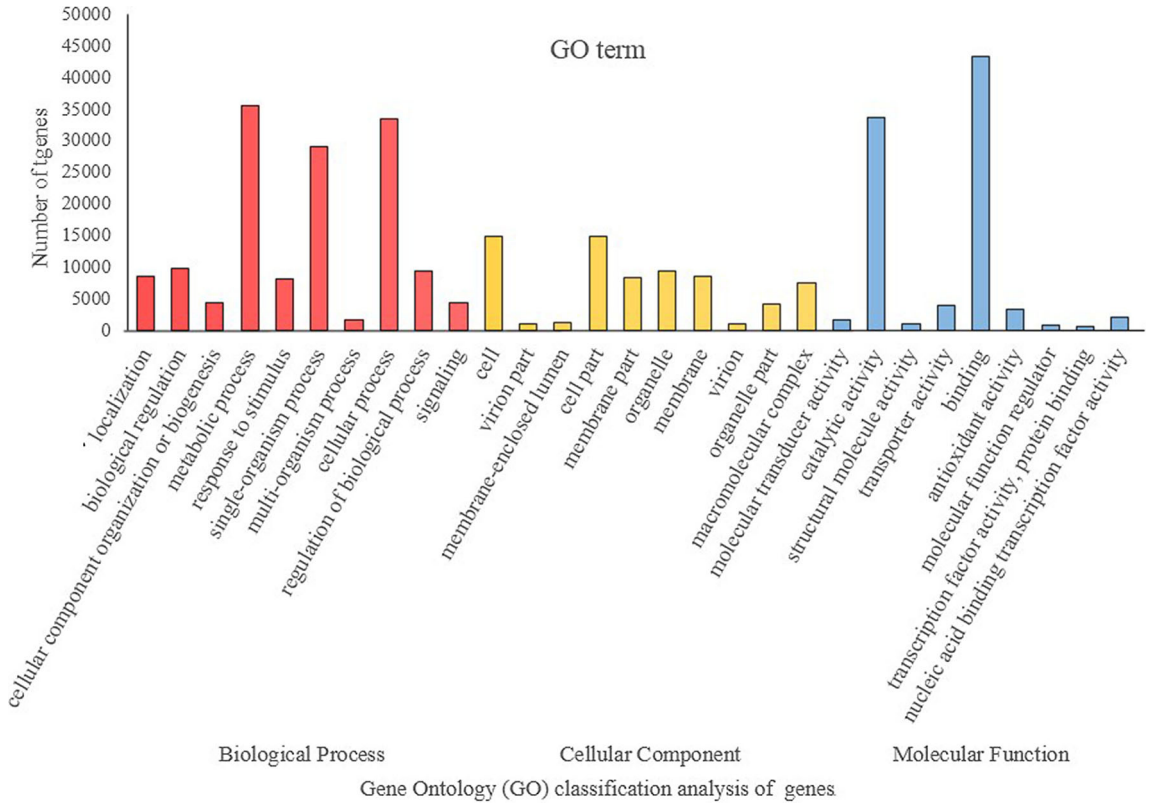

Figure 2. GO functional classification. The *x*-axis shows the GO function classed and the right *y*-axis shows the number of genes having the GO function.

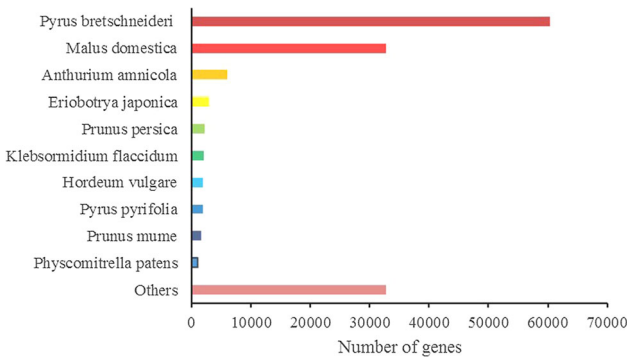

Figure 3. Distribution of the new transcript annotations based on the top BLASTX hits in the species statistics.

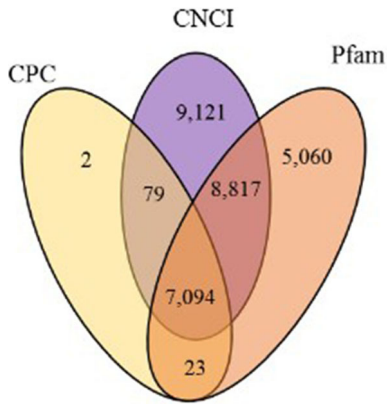

Figure 4. Overlap between the numbers of lncRNAs according to three reference protein databases, namely, coding–noncoding index (CNCI), coding potential calculator (CPC), Pfam-scan (Pfam).

Discussion

Because of the complexity of transcriptomes of higher eukaryotes, short-read approaches suffer when it comes to precise reconstruction of transcript structures (Tilgner *et al*. 2014). Long-read RNA sequencing facilitates precise characterization of FL transcripts, thus is an indispensable tool in transcriptomics (Balazs *et al*. 2017). Although the genome and transcriptome sequences of Chinese pear are available (Wu *et al*. 2013), genomic information on *P. betulifolia* is lacking. We obtained 362,139 transcriptome-wide FL consensus sequences from pear. Similar analyses have been carried out for human (Sharon *et al*. 2013), rat (Hartley *et al*. 2016), sorghum, maize (Wang *et al*. 2016), and other eukaryotes. Compared to second-generation techniques, our

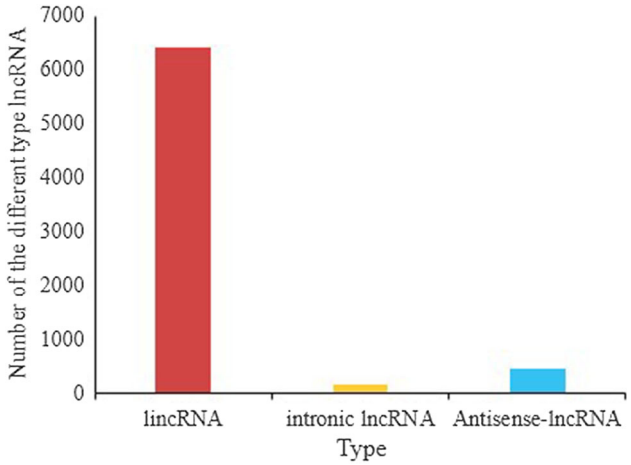

Figure 5. Numbers of different types of lncRNAs, including intergenic lncRNA (lincRNA), intronic lncRNA and antisense-lncRNA.

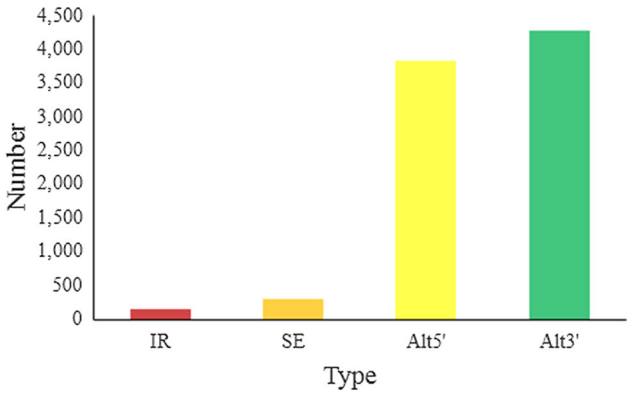

Figure 6. Numbers of different types of alternative splicing events in birth-leaf pear.

method is more effective in terms of generating FL transcripts. Our data provide insight into the transcriptome of *P. betulifolia* Bunge.

Protein-coding DNA sequences comprise 2–25% of the genome, depending on the organism in question (Liu *et al.* 2015). Eukaryotic genomes encode a large number of non-coding RNAs, including lncRNAs, which regulate gene expression by recruiting factors that activate transcription or modify chromatin, serving as precursors of small RNAs and affecting the nuclear architecture (Mach 2017). LncRNAs also play vital roles in the responses of plants to abiotic and biotic stresses. We previously reported 251 drought-responsive lncRNAs in *P. betulifolia* (Wang *et al.* 2018). In this study, 7094 transcripts were ncRNAs, which facilitate further studies of the lncRNAs of pear.

AS greatly diversifies the proteome by promoting the formation of new genes (Jiang *et al.* 2017). Many genes undergo regulated AS (Vitulo *et al.* 2014). In *Arabidopsis*, 4707 (21.8%) genes with EST/cDNA evidence underwent 8264 AS events. In rice, 6568 (21.2%) expressed genes exhibited 14,542 AS events (Wang and Brendel 2006). AS in plants can also be regulated by environmental factors (Filichkin *et al.* 2010). Of the multiexon genes in strawberry, 66.43% underwent AS, and the percentage of intron retention is markedly reduced, whereas that of alternative acceptor sites is significantly increased, postcompared with prefertilization (Li *et al.* 2017).

We identified 90,952 AS events, the most common of which was Alt.3′ (43.3%). This is in agreement with a prior report that Alt.3′ is the predominant form of AS in plants (Walters *et al.* 2013). We applied PacBio single-molecule real time sequencing technology to pear tissue and generated a large number of gene models and alternative isoforms. The transcriptome obtained will facilitate further studies of the genetics of birch-leaf pear and related species.

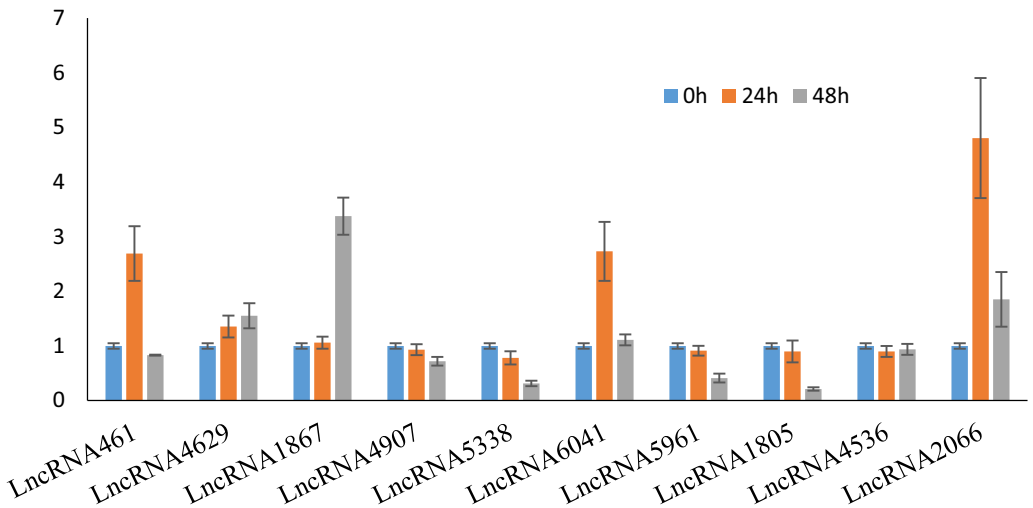

Figure 7. Validation of lncRNAs expression patterns using quantitative RT-PCR.

Acknowledgements

This work was supported by Pear technical system of modern agricultural industry in Jiangsu—genetic and breeding (JATS[2018]269). Jiangsu Agricultural Science and Technology Innovation fund (Cx(19)2028), supported by Fearmarked Fund for Jiang Agricultural Industry Technology System (JATS [2020]390) and Jiangsu Agricultural Science and Technology Innovation Fund (SCx(19)3030).

Reference

Abdel-Ghany S. E., Hamilton M., Jacobi J. L., Ngam P., Reddy A. S. N. *et al*. 2016 A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **7**, 11706.

Ali G. S. and Reddy A. S. N. 2008 Regulation of alternative splicing of pre-mRNAs by stresses. *Curr. Topics Microbiol. Immunol.* **326**, 257.

Allen S. L., Delaney E. K., Kopp A. and Chenoweth S. F. 2017 Single-molecule sequencing of the *Drosophila serrata* genome. *G3 Genes Genom. Genet.* **7**, 116.037598.

Balazs Z., Dora Tombacz, Szucs A., Snyder M. and Boldogki Z. 2017 Long-read sequencing of the human cytomegalovirus transcriptome with the pacific biosciences RSII platform. *Sci. Data* **4**, 170194.

Chang Y. H., Li H., Cong Y., Lin J., Sheng B. L. *et al*. 2012 Characterization and expression of a phytochelatin synthase gene in birch-leaf pear (*pyrus betulaefoliabunge*). *Plant Mol. Biol. Rep.* **30**, 1329–1337.

Chen X., Houten S. M., Allette K., Sebra R. P., Stolovitzky G., Losic B. *et al*. 2018 Characterization of drug-induced splicing complexity in prostate cancer cell line using long read technology. *Pac. Symp. Biocomput.* **23**, 8–19.

Chin C. S., Alexander D. H., Marks P., Klammer A. A., Drake J., Heiner C. *et al*. 2013 Nonhybrid, finished microbial genome assemblies from long-read smrt sequencing data. *Nat. Methods* **10**, 563–569.

Filichkin S. A., Priest H. D., Givan S. A., Shen R., Bryant D. W., Fox S. E. *et al*. 2010 Genome-wide mapping of alternative splicing in arabidopsis thaliana. *Genome Res.* **20**, 45–58.

Guttman M., Amit I., Garber M., French C., Lin M. F., Feldser D. *et al*. 2009 Chromatin signature reveals over a thousand highly conserved large non-coding rnas in mammals. *Nature* **458**, 223–227.

Hartley S. W., Mullikin J. C., Klein D. C., Park M. and Coon S. L. 2016 Alternative isoform analysis of Ttc8 expression in the rat pineal gland using a multi-platform sequencing approach reveals neural regulation. *PLoS One* **11**, e0163590.

Jiang X., Hall A. B., Biedler J. K. and Tu Z. 2017 Single molecule RNA sequencing uncovers trans-splicing and improves annotations in Anopheles stephensi. *Insect Mol. Biol.* **26**, 298–307.

Liu J., Wang H. and Chua N. H. 2015 Long noncoding RNA transcriptome of plants. *Plant Biotechnol. J.* **13**, 319–328.

Kalsotra A. and Cooper T. A. 2011 Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* **12**, 715–729.

Kapranov P., Cheng J., Dike S., Nix D. A., Duttagupta R., Willingham A. T. *et al*. 2007 RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488.

Koren S., Harhay G. P., Smith T. P., Bono J. L., Harhay D. M. and Mcvey S. D. 2013 Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol.* **14**, R101.

Li L., Eichten S. R., Shimizu R., Petsch K., Yeh C. T. and Evans M. M. 2014 Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* **15**, R40.

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.

Livak K. and Schmittgen T. 2001 Analysis of relative gene expression data using real time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Method* **25**, 402–408.

Mach J. 2017 The long-noncoding RNA ELENA1 functions in plant immunity. *Plant Cell* **29**, 916.

Nakano K., Shiroma A., Shimoji M., Tamotsu H., Ashimine N. and Ohki S. 2017 Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Hum. Cell* **30**, 149–161.

Nagalakshmi U., Wang Z., Waern K., Shou C., Raha D. and Gerstein M. 2008 The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349.

Reddy A. S. 2007 Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.* **58**, 267–294.

Schadt E. E., Turner S. and Kasarskis A. 2010 A window into third-generation sequencing. *Hum. Mol. Genet.* **19**, R227–R240.

Sharon D., Tilgner H., Grubert F. and Snyder M. 2013 A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* **31**, 1009–1014.

Shuai P., Liang D., Tang S., Zhang Z., Ye C. Y., Su Y. Y. *et al*. 2014 Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J. Exp. Bot.* **65**, 4975–4983.

Tilgner H., Grubert F., Sharon D. and Snyder M. P. 2014 Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc. Natl. Acad. Sci. USA* **111**, 9869–9874.

Vitulo N., Forcato C., Carpinelli E. C., Telatin A., Campagna D. and Dangelo M. 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.

Walters B., Lum G., Sablok G. A. and Min X. J. 2013 Genome-wide landscape of alternative splicing events in *Brachypodium distachyon*. *DNA Res.* **20**, 163–171.

Wang B. B. and Brendel V. 2006 Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. USA* **103**, 7175–7180.

Wang B., Tseng E., Regulski M., Clark T. A., Hon T. and Jiao Y. 2016 Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **7**, 11708.

Wang J., Lin J., Kan J., Wang H., Li X., Yang Q. *et al*. 2018 Genome-wide identification and functional prediction of novel drought-responsive lncRNAs in pyrus betulifolia. *Genes* **9**, 311.

Wang Z., Gerstein M. and Snyder M. 2009 RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63.

Wu J., Wang Z., Shi Z., Zhang S., Ming R., Zhu S. *et al*. 2013 The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* **23**, 396–408.

Xin M., Wang Y., Yao Y., Song N., Hu Z., Qin D. *et al*. 2011 Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biol.* **11**, 61.

Xu Q., Zhu J., Zhao S., Hou Y., Li F., Tai Y. *et al*. 2017 Transcriptome profiling using single-molecule direct RNA sequencing approach for in-depth understanding of genes in

secondary metabolism pathways of *Camellia sinensis*. *Front. Plant. Sci.* **8**, 1205.

Xu Z., Peters R. J., Weirather J., Luo H., Liao B., Zhang X. *et al*. 2015 Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of S alvia miltiorrhiza and tanshinone biosynthesis. *Plant J.* **82**, 951–961.

Zhang W., Han Z., Guo Q., Liu Y., Zheng Y., Wu F. *et al*. 2014 Identification of maize long non-coding RNAs responsive to drought stress. *PLoS One* **9**, e98958.

Zimin A. V., Stevens K. A., Crepeau M. W., Puiu D., Wegrzyn J. L., Yorke J. A. *et al*. 2017 An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *GigaScience* **6**, 1.

Corresponding editor: H. A. RANGANATH