# Single molecule, full-length transcript sequencing provides insight into the *TPS* gene family in *Paeonia ostii*

Jing Sun[*], Tian Chen[*] and Jun Tao

College of Horticulture and Plant Protection, Yangzhou University, Yangzhou, China
Joint International Research Laboratory of Agriculture and Agri-Product Safety, the Ministry of Education of China, Yangzhou University, Yangzhou, China
[*] These authors contributed equally to this work.

## ABSTRACT

**Background**. The tree peony (*Paeonia* section *Moutan* DC), one of the traditional famous flowers with both ornamental and medicinal value, was widely used in China. Surprisingly little is known about the full-length transcriptome sequencing in tree peony, limiting the research on its gene function and molecular mechanism. The trehalose phosphate phosphatase (*TPS*) family genes has been found to affect plant growth and development and the function of *TPS* genes in *Paeonia ostii* is unknown.
**Methods**. In our study, we performed single molecule, full-length transcript sequencing in *P. ostii*. 10 *TPS* family members were identified from PacBio sequencing for bioinformatics analysis and transcriptional expression analysis.
**Results**. A total of 230,736 reads of insert (ROI) sequences and 114,215 full-Length non-chimeric reads (FLNC) were obtained for further ORFs and transcription factors prediction, SSR analysis and lncRNA identification. NR, Swissprot, GO, COG, KOG, Pfam and KEGG databases were used to obtain annotation information of transcripts. 10 *TPS* family members were identified with molecular weights between 48.0 to 108.5 kD and isoelectric point between 5.61 to 6.37. Furthermore, we found that *TPS* family members contain conserved TPP or TPS domain. Based on phylogenetic tree analysis, PoTPS1 protein was highly similar to AtTPS1 protein in *Arabidopsis*. Finally, we analyzed the expression levels of all *TPS* genes in *P. ostii* and found *PoTPS5* expressed at the highest level. In conclusion, this study combined the results of the transcriptome to systematically analyze the 10 *TPS* family members, and sets a framework for further research of this important gene family in development of tree peony.

**Subjects** Agricultural Science, Genomics, Molecular Biology, Plant Science
**Keywords** *Paeonia ostii*, Single molecule, Full-length transcript sequencing, TPS family members, Functional genes

## INTRODUCTION

As a traditional ornamental and medicinal flower with a long history of cultivation, tree peony (*Paeonia* section *Moutan* DC) is known as the 'king of flowers' in China and belongs to the family Paeoniaceae, symbolizing wealth, prosperity and happiness. Tree peony consists of nine species (Table S1), with *Paeonia ostii* receiving the most extensive promotion at the national level (*Ren et al., 2018*). Currently, *P. ostii* not only has

the characteristics of large flowering, less tillering, high seed yield and strong ecological adaptability, but also is an important new oil crop integrating ornamental, medicinal and oil use (*Wang et al., 2019*) compared to other nine species. *P. ostii* can be generally found in the Gansu, Anhui, Shanxi and Henan provinces of China. It is a diploid cultivated tree peony species, with 2n =2x =10 chromosomes, clearly characterized by white or pale rose colored petals without a base blotch (*Hong, Pan & Pan, 2004*). At present, research on tree peony has mainly focused on its morphological characteristics, cytological characteristics and germplasm resources. Surprisingly, full-length transcriptome sequencing in tree peony has not been performed.

The trehalose phosphate phosphatase (*TPS*) gene family in higher plants is a small gene family that is divided into two distinct classes, class I and class II (*Lunn, 2007*). Obviously, they are different in copy number, gene expression patterns, enzyme activity and physiological functions (*Ping et al., 2019*). Currently, in *Arabidopsis* only class I *TPS* genes have been reproducibly shown to encode catalytically active TPS enzymes (*Blazquez et al., 1998*; *Van Dijck et al., 2002*; *Vandesteene et al., 2010*), and the functions of noncatalytic class II TPS-like proteins are poorly understood (*Harthill et al., 2006*; *Lunn et al., 2014*). Although significant progress has been made in the study of *TPS* genes, a number of *TPS* gene families have been identified in only some model plants. For example, there are 11 *TPS* members in *Arabidopsis* (*AtTPS1-11*) (*Shima et al., 2007*) and in rice (*OsTPS1-11*) (*Vandesteene et al., 2010*), while 12 members were found in poplar (*PtTPS1-12*) (*Tuskan et al., 2006*; *Vandesteene et al., 2010*).

The *TPS* gene family also shows widespread functional diversification. Catalyzed by TPS, the biosynthetic precursor of trehalose, trehalose-6-phosphate (T6P) plays a crucial role, participating in the physiological processes of plant embryo development, flower induction, senescence regulation and seed filling, and in the respond to biotic and abiotic stresses (*Gomez et al., 2010*; *Kumar et al., 2019*; *Wahl et al., 2013*; *Wingler et al., 2012*; *Zhao et al., 2019*). For example, overexpressing *OsTPS1* in rice enhanced its tolerance to salt, drought and low temperatures (*Jang et al., 2003*). Moreover, *SlTPS1* in the *TPS* gene family of *Selaginella lepidophylla* participates in the response to heat and salinity by enhancing T6P biosynthesis (*Zentella et al., 1999*). When *AtTPS1* was deleted, it caused embryo arrest, hindered vegetative growth and delayed flowering of *Arabidopsis thaliana* (*Eastmond et al., 2002*; *Gomez et al., 2006*; *Gomez et al., 2010*; *Van Dijken, Schluepmann & Smeekens, 2004*; *Wahl et al., 2013*). The product of the TPS reaction, T6P, functions as a signal for sucrose availability (*Yadav et al., 2014*) and it is involved in triggering axillary bud outgrowth in garden pea (*Pisum sativum* L) (*Fichtner et al., 2017*). In potato tubers, T6P is a pivotal mechanism linking growth, development and metabolism by SnRK1 signaling (*Debast et al., 2011*). Similar studies also suggested that when T6P is deleted, the size of the pea seeds will decrease, and the starch yield will decrease because auxin acts downstream of T6P to facilitate seed filling (*Meitzel et al., 2019*). However, whether the members of *TPS* gene family have an effect on the development of *P. ostii* and their underlying molecular mechanisms remains unknown.

Transcriptome research is one of the essential tools to understand the life process. However, most of the previous technologies are not capable of generating reads representing

entire transcripts due to sequencing read shortness. The second-generation high-throughput sequencing platforms, such as Illumina sequencing platforms often cannot obtain or assemble the complete transcript entirely and recognize isoforms, homologous genes, and the transcripts of superfamily genes and alleles, making it challenging to accurately detect transcript isoforms generated by alternative splicing (AS) events (*Bessa et al., 2020*) and structural variations (SVs) (*Sharon et al., 2013*; *Wang et al., 2016b*). Moreover, NGS has difficulty in distinguishing among different transcripts possessing identical exons (*Tilgner et al., 2013*) and has the disadvantage of high GC preference in analyzing the original sequencing reads.

Currently, the outcomes of third-generation sequencing, single-molecule, real-time (SMRT) sequencing developed by Pacific Biosciences has enabled the generation of kilobase-sized sequencing reads (*Gordon et al., 2014*), showing more RNA molecules (*Sharon et al., 2013*). The RACE technique requires no interruption of the RNA fragments or reverse transcription of the full-length cDNA (*Thomas et al., 2014*). By employing appropriate methods for cDNA preparation, bona fide full-length transcript sequences can be generated. SMRT sequencing reads exhibit high sequencing error rates, most notably base insertions or deletions. However, due to the random nature of the encountered errors, the construction of highly accurate consensus by reiterated sequencing of the same fragment is straightforward. Currently, SMRT sequencing has been applied to investigate the full-length transcriptome in diverse species, such as wheat (*Dong et al., 2015*), *Salvia miltiorrhiza* (*Xu et al., 2015*), sorghum (*Abdel-Ghany et al., 2016*), maize (*Wang et al., 2016a*), sugarcane (*Hoang et al., 2017b*), perennial ryegrass (*Xie et al., 2020*), *Trifolium pratense* L. (*Chao et al., 2018*), Chinese cabbage (*Tan et al., 2019*) and *Crocus sativus* L (*Qian et al., 2019*).

The objective of this study was to apply PacBio full-length sequencing to provide a new reference gene database for the further study of *P. ostii*. In this paper, young leaves, roots, stems, seeds, flowers and fruit pods were collected for Single molecule, full-length transcript sequencing to analyze the *P. ostii* transcriptome. Then, we identified the members of the *TPS* gene family, providing valuable gene resources for further research into the evolution and biological functions in *P. ostii*.

## MATERIALS & METHODS

### Plant materials and sample preparation

Young leaves, stems, roots, flowers, seeds and fruit pods of three different 3-year-old *P. ostii* from the germplasm repository of the Horticulture and Plant Protection College, Yangzhou University, Jiangsu Province, P.R. China (32°23′31′N,119°24′50′E) were collected and stored in liquid nitrogen for RNA extraction. They had been grown under the same conditions for three years. Seeds, leaves and stems used for qRT-PCR were collected from the same strain of three different *P. ostii* at the developmental stages of 30, 50 and 70 days after flowering (DAF) in 2019, because the endosperm of seeds begins to develop after 30 DAF and becomes cellular after 70 DAF (*Yao, 2020*). Leaves, stems, roots, and flowers in the bloom period of *P. ostii* were collected in April 2019. Fruit pods and seeds were
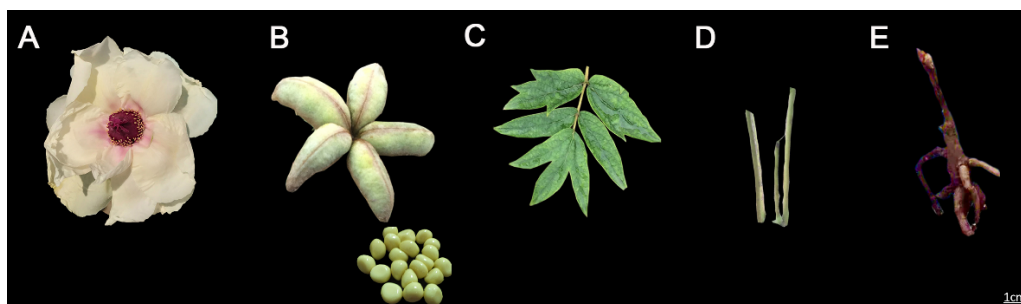
**Figure 1** **The tissues of *Paeonia ostii* used in this study.** (A) The flower in full bloom. (B) Pods and seeds. (C) Leaves. (D) Stems. (E) Roots.

collected at 70 DAF in June 2019. RNA was extracted separately from leaf, stem, root, flower, seed and pod tissues (Figs. 1A–1E) using the CTAB method. To ensure the accuracy of the sequencing data, all RNA samples were checked using a Nanodrop 2000C (Thermo Scientific). The RNA integrity was checked using an Agilent 2100 bioanalyzer (Agilent Technologies), which included RIN, 28S/18S and 5S peaks. Electrophoresis was used to detect whether the RNA samples contained gDNA contamination and assess RNA quality by identifying the ribosomal bands. gDNA were digested by using the Exonuclease I (NEB, M0293S).

## PacBio library construction and sequencing

After RNA quality inspection, we first mixed equal quantities of high-quality RNA from different organs of one *P. ostii*. Then, the mixed samples from three different *P. ostii* were mixed as one pool (Fig. S1). Ten micrograms of RNA from the sample pool was manipulated as follows: (1) reverse transcribed into cDNA using the SMARTer PCR cDNA Synthesis Kit to prepare full-length, high-quality cDNAs with VN Primer (2 μM Nanopore) and Strand-Switching Primer (10 μM Nanopore) (Table S2); (2) a BluePippin$^{TM}$ Size Selection System (Sage Science, Beverly, MA) was used to screen full-length cDNA fragments and construct cDNA libraries of different sizes (1–2 kb, 2–3 kb, and 3–6 kb); (3) the amplified full-length cDNA was reamplified by PCR and end-repair of the full-length cDNA; (4) then, each SMRT bell library (500 ng size-selected cDNA) was constructed using the Pacific Biosciences DNA Template Prep Kit 2.0; (5) after digestion by Exonuclease I (NEB, M0293), secondary screening was carried out with BluePippin to obtain a sequencing library. After the completion of the library construction, the quality of the library was tested: Qubit 2.0 was use for accurate quantification. The library size was tested using the Agilent 2100, and the library size was as expected before on-board sequencing (the concentration of Qubit >2 ng/μL and the total amount >50 ng/μL). After the library was purified by AMpure PB Beads, full-length transcriptome sequencing was performed using PacBio RSII according to the target data volume.

## Full-length corrected transcript collection in SMART sequencing

Full-length transcriptome sequencing based on PacBio SMRT (Sequencing By Synthesis, SMRT) single-molecule real-time sequencing technology does not require the disruption

of RNA fragments, and the full-length cDNA had been obtained by reverse transcription using RACE technology. The long read (median 10 kb) of the platform contains a single complete transcript sequence information without post assembly analysis (*Gordon et al., 2015*; *Thomas et al., 2014*). The analysis process for acquiring the full-length transcriptome mainly includes three stages (*Sharon et al., 2014*): (1) Full-length sequence identification. Convert all original sequences to ROI (Reads Of Insert) sequences based on the adaptor. According to whether there are 3′ primers, 5′ primers, or Poly A (optional), they are divided into two categories: full-length and non-full-length sequences. (2) Isoform level clustering to obtain consensus sequences. ROI sequences from the same transcript were clustered using the ICE (Iteratively Clustering and Error Correction) algorithm, and each cluster obtained a consensus sequence. (3) Consistent sequence polishing. The non-full-length sequence was used to perform the polishing on the obtained consensus sequence to obtain a high-quality sequence for subsequent analysis. The resulting transcript sequences can be directly used for subsequent isoform, homologous gene, gene family, SSR, alternative splicing, lncRNA and other analyses.

The PacBio RS II used for SMRT sequencing contained 150,000 Zero-Mode Waveguides (ZMWs) per cell. The reads were sequenced into ZMW wells, and one ZMW contained reads as valid data. The ROI sequences satisfying the conditions of full passes $\geq 0$ and accuracy $\geq 0.75$ were extracted. To obtain the iteratively clustered sequences, the iterative clustering for error correction algorithm was performed by the SMRT Analysis (v2.3.0). Combined with a non-full-length sequence, high-quality transcripts (HQs) with accuracies greater than 99% were obtained. In the process of full-length transcript sequencing, the 3′ end has a poly-A structure, and it can be determined that the 3′ end is relatively intact, and the 5′ end sequence may be degraded, resulting in different copies of the same transcript being assigned to different clusters. CD-HIT (*Li & Godzik, 2006*) software was used to combine sequences with high similarity and to remove redundant sequences from among the high-quality transcripts.

## Functional annotation, ORFs, SSR analysis and lncRNA identification

We used BLAST to combine the obtained nonredundant transcript sequences with the NR (http://www.ncbi.nlm.nih.gov) (April, 2021), SwissProt (http://www.ebi.ac.uk/swissprot) (April, 2021), GO (http://www.geneontology.org) (April, 2021), COG (http://www.ncbi.nlm.nih.gov/COG) (March, 2021), KOG (https://mycocosm.jgi.doe.gov/help/kogbrowser.jsf) (April, 2021), Pfam (http://pfam.xfam.org/) (March, 2021) and KEGG (http://www.genome.jp/kegg) (April, 2021) databases to obtain annotation information of the transcripts (*Altschul, 1999*).

TransDecoder software 2.0.1 (https://transdecoder.github.io/) is based on the length of the open reading frame (ORF), log-likelihood score and the comparison of amino acid sequences and protein domain sequences in the Pfam database, which can identify reliable coding sequences (CDSs) from the transcript sequence.

Simple repetitive sequences were identified through MISA software. Seven types of SSRs were identified by analyzing the sequences of the transcripts: mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide and compound SSRs.

We filtered out the transcripts with coding potential to obtain the predicted lncRNAs. In this study, the most widely used coding potential analysis method was used to predict the lncRNAs among the transcripts, including four methods: CPC analysis (*Kong et al., 2007*), CNCI analysis, Pfam protein domain analysis, and CPAT analysis.

## Quantitative Real-Time PCR anlaysis

Each organ was represented by three biological replicates with three technical replicates. Total RNA from the seeds, leaves and stems from three *P. ostii* was extracted by a MiniBEST Plant RNA Extraction Kit (TaKaRa, Japan). Then, the cDNA was synthesized by the PrimeScript RT reagent Kit with gDNA Eraser (TaKaRa, Japan) (*Zhao et al., 2020*). Before qRT-PCR, we used water instead of cDNA and ran the PCR. The agarose gel electrophoresis showed clear objective band, so there was no pollution in the PCR system. A BIO-RAD CFX Connect Optics Module (Bio-Rad, Des Plaines, IL, U.S.A.) was used to analyze the *PoTPS* expression levels. Expression values were calculated by the $2^{-\Delta\Delta Ct}$ comparative threshold cycle ($C_t$) method (*Livak & Schmittgen, 2001*). SYBR Premix Ex Taq (Perfect Real Time) (TaKaRa, Japan) with 12.5 µL 2× SYBR Premix Ex Taq, 2 µL cDNA solution, 8.5 µL ddH$_2$O, and 2 µL solution of mix primers were the system to perform qRT-PCR. Amplification conditions: 95 °C for 30 s, 40 cycles at 95 °C for 5 s, 52 °C for 30 s, and 72 °C for 30 s. The *ubiquitin* gene (JN699053) was used as an internal reference for this experiment and the expression level of this reference gene was stable in all organs of *Paeonia suffruticosa* (*Wang et al., 2012*). All primers were listed in Table S2.

## Analysis of the *TPS* genes family in *Paeonia ostii*

To classify the *PoTPS* genes in *P. ostii*, Cluster X 2.0.12 software (http://www.cluster-x.org/) was applied for multiple sequence alignment by using protein sequences of *Arabidopsis*. To construct the phylogenetic tree, neighbor-joining (NJ) method was used by MEGA 7.0 software, and bootstrap values were set as 1000 bootstrap replicates (*Saitou & Nei, 1987*). The conserved motifs of the *PoTPS* sequences were identified by the MEME program (https://meme-suite.org/meme/), and the parameters were set as a maximum of 10 motifs and an optimum motif width of 6-200 amino acid residues (*Zhao et al., 2018*). The conserved domains were visualized using the TBtools software.

## RESULTS

### *Paeonia ostii* transcriptome sequencing with SMRT

The transcriptome of the pooled samples was sequenced and analyzed with the PacBio Sequel platform to accurately capture the full-length sequences. A total of three libraries divided by cDNA lengths of 1–2 kb, 2–3 kb and 3–6 kb were constructed. Then, 3,568,378 subreads (8.49 Gb) and 343,264 polymerase reads were obtained after filtering. These data have been deposited in the National Center for Biotechnology Information (NCBI) (BioProject accession: PRJNA688625). The ROI sequence was extracted from the original sequence according to the condition that full passes ≥0 and the sequence accuracy ≥0.75. To evaluate the offline date, the number of ROIs, the number of bases of the ROIs and the mean length of the inserted sequence in each library were counted. In total, we obtained

230,736 ROI sequences and the mean read quality of the insert of each size-selected library was 0.92 (Table S3). The ROI read length distribution of each size-selected library had three consistent peaks (Figs. 2A–2C). By filtering out short fragments <300 bp, the sequences containing both the 3′primers and 5′ primers at the same time and having a poly A tail before the 3′ primers were defined as full-length sequences. After further filtering and analysis, 114,215 full-length nonchimeric (FLNC) reads were obtained. The FLNC read length distribution of each size-selected library is shown in Figs. 2D–2F. The trends of the three peaks were consistent, which was in line with the expectations. The percentage of FLNC reads in all ROI sequences was 49.5% (Fig. 3A), accounting for half of the number of all ROI sequences, combined with the ROI and FLNC sequences, which showed that the quality of the transcriptome was better.

By using IsoSeq of SMRT Analysis software to perform cluster analysis on the full-length sequence and combining the non-full-length sequences, 45,006 identical transcripts were obtained (Fig. 4). The consensus isoform read length distribution of each size bin (<1, 1–2, 2–3, 3–6 and >6 kb) is shown in Figs. 3B–3F, and the average consensus isoform read lengths were 912, 1399, 2346, 3665 and 9297 (Table S4). After correcting the consistent sequences in each cluster by quiver program, only 8,239 low-quality isoforms (LQs) and 36,767 high-quality isoforms (HQs) were obtained. Among them, 0–1 kb high-quality transcripts accounted for the largest proportion (91.58%) of consensus isoform reads (Table S4), indicating that the shorter the cDNA length, the more high-quality reads were obtained. Finally, by using CD-HIT to remove the redundant sequences of the high-quality sequences, 30,100 transcript sequences were obtained for subsequent analysis (Fig. 4).

## ORF and transcription factor prediction, SSR analysis and lncRNA identification

TransDecoder software was used to predict the coding protein sequences and to identify full-length ORFs. In our study, a total of 28,415 ORFs were obtained, of which 17,904 were complete ORFs. Sequences with start and stop codons were defined as complete coding sequences. The predicted length distribution of the complete ORFs is as follows (Fig. 5A). Transcripts >500 bp in length were selected. By using MISA software to analyze 30,086 transcripts, 9,789 SSR- containing sequences from among the 13,632 evaluated SSRs were identified. Large numbers of SSR containing sequences were mono-, di-, or trinucleotide repeats (Table S5). Currently, long noncoding RNAs (lncRNAs) play an important role in many life activities, such as dose compensation effects, epigenetic regulation, cell cycle regulation and cell differentiation regulation, and they have become a hotspot of genetic research. Therefore, it is very meaningful to analyze lncRNAs through transcriptome analysis. Four methods were used to make lncRNA predictions for the transcripts CPC, CNCI, Pfam and CPAT. To visually display the analysis results, the noncoding transcripts identified by the above four analysis methods were taken into intersection for subsequent lncRNA analysis (Fig. 5B). A total of 827 lncRNAs were predicted by filtering out the transcripts <300 bp in length.
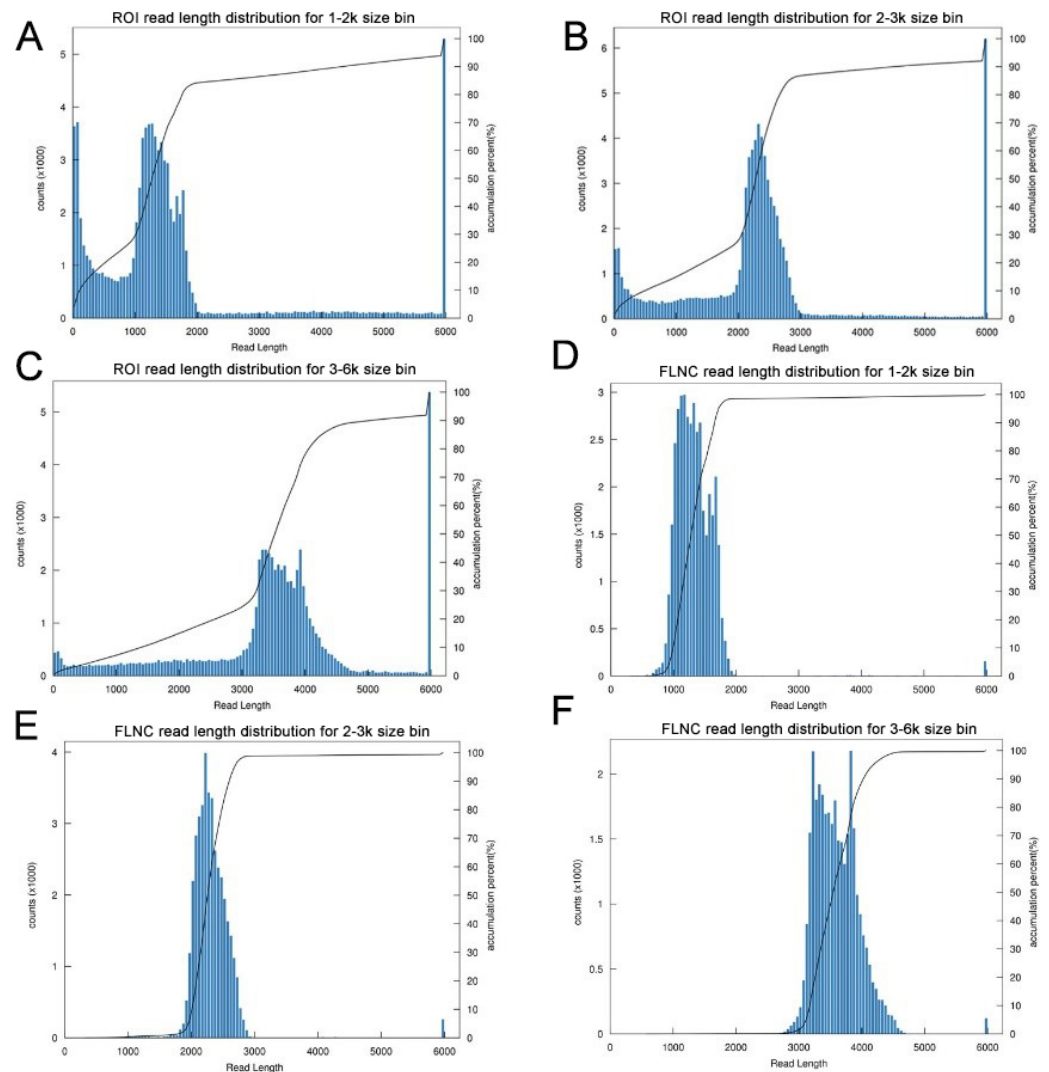
**Figure 2** **Summary of PacBio SMRT ROI and FLNC sequences.** (A–C) Number and length distributions of 230,736 ROI sequences from 1–2 k, 2–3 k and 3–6 k sized libraries. (D–F) Number and length distributions of 114,908 FLNC reads from 1–2 k, 2–3 k and 3–6 k sized libraries.

Full-size ◳ DOI: 10.7717/peerj.11808/fig-2

## Functional annotation of the transcripts

We have submitted the final polished consensus mRNA sequences to NCBI (accession number GJFW00000000). Blast software was used to compare the nonredundant transcripts with the NR, SwissProt, GO, COG, KOG, Pfam, and KEGG databases. A total of 28,850 transcript annotation information points were obtained (Table S6). We searched for homologous species through sequence alignment. The 28,790 transcripts in Nr were aligned as shown below, showing that the largest number of transcripts were distributed in *Vitis vinifera* (37.64%) (Fig. 6A). The GO annotation system includes three main branches, namely biological process, molecular function and cellular component. Among them, biological processes involved metabolism, cellular and single-organism processes,
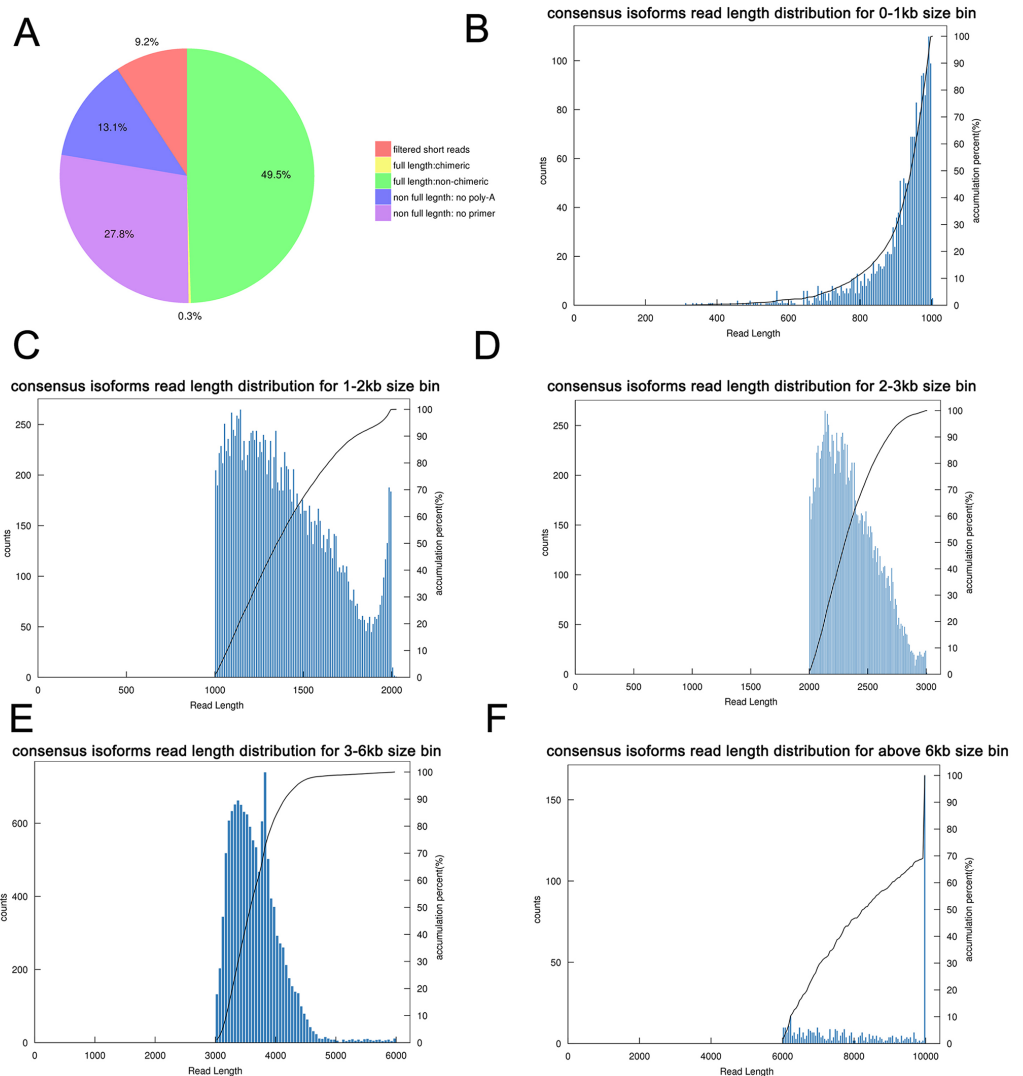
**Figure 3  Summary of PacBio SMRT sequencing.** (A) Proportion of different types of PacBio reads. (B–F) Number and length distributions of 45,006 consensus isoforms reads from < 1 kb, 1–2 kb, 2–3 kb, 3–6 kb and > 6 kb.

responses to stimulus and biological regulation. Cell components mainly involved cells, cell parts, organelles, membranes and other parts. Molecular functions involve catalysis, binding, transport, transcription activities and other activities. The statistical results of 20,052 transcripts of GO are shown in Fig. 6B. The COG (Cluster of Orthologous Groups of proteins) database was constructed based on the phylogenetic relationships of the bacteria, algae and eukaryotes. The COG database can be used to classify gene products as orthologues. The COG classification statistics of 12,892 transcripts are shown in Fig. 6C.
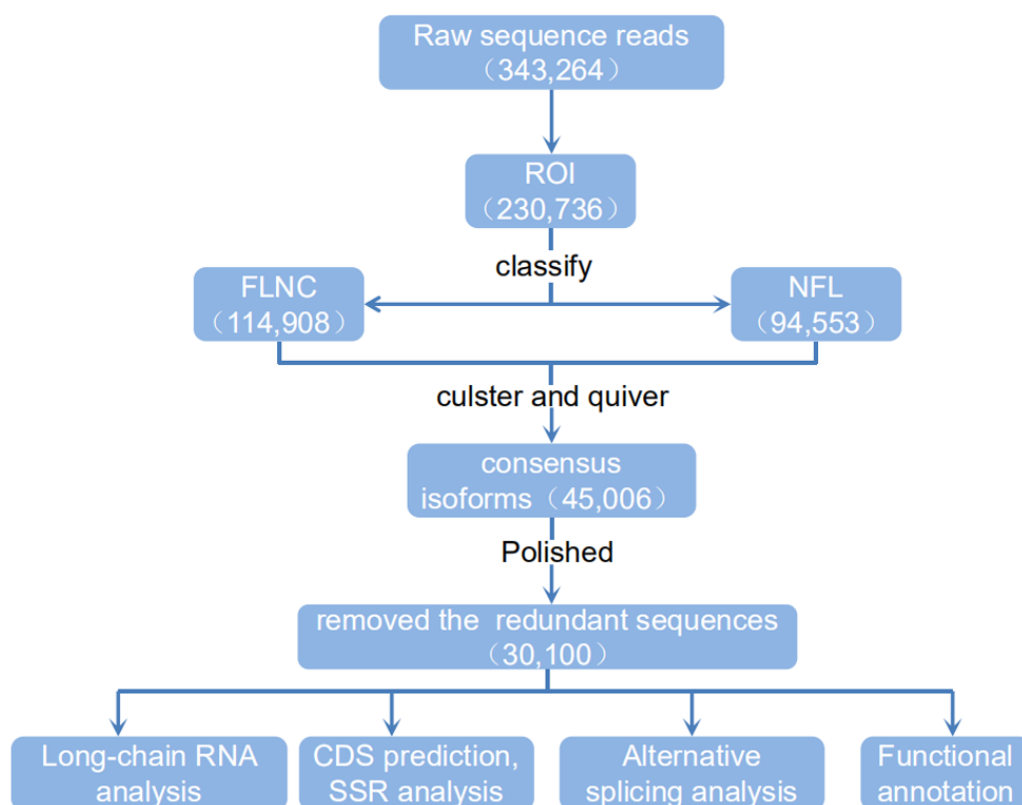
**Figure 4** **Flow chart of bioinformation analysis.** The raw sequence reads from a PacBio Rs sequencer were sorted into full and non-full length reads using a classification algorithm that identified full length reads with forward and reverse primers, as well as a poly A tail. Iterative clustering for isoforms (ICE) was performed on full length reads, and non-full-length reads were recruited to perform ARROW polished on the consensus isoforms. Polished sorted reads into high and low quality bins, and either high quality data, all sequence data or both sets of data, were carried on to further applications.

Full-size 🖼 DOI: 10.7717/peerj.11808/fig-4

## Identification of the *TPS* gene family

Ten *TPS* family members were selected and identified from the transcriptome sequencing with the SMRT database for *P. ostii*, and the accession numbers for the 10 *TPS* genes are MW700299–MW700308. According to the percentage of similarity between these *TPS* family members and *AtTPS1-11* in *Arabidopsis*, we named these genes *PoTPS1*, *PoTPS3-PoTPS11* (Table S6). To clarify the related protein information of the TPS family in *P. ostii*, ProtParam was used to analyze the physical and chemical properties and secondary structure elements of the TPS family members in this study. The number of amino acids was approximately 429-959, the isoelectric point ranged between 5.61−6.37 and the protein molecular weight was between 48.0–108.5 kD. The secondary structure is mainly composed of a helix, extended strand and random coil. Specifically, the random coil structure of 10 TPS members was above 29%, which is beneficial to maintaining the stability of TPS proteins (Table 1). Using Pfam and SMART to analyze protein domains, it was found that these TPS proteins all have conserved domains, which contain two typical characteristics of TPS (Pfam: Glyco-transf-20) and TPP (Pfam: Trehalose-PPase) structural domains. In
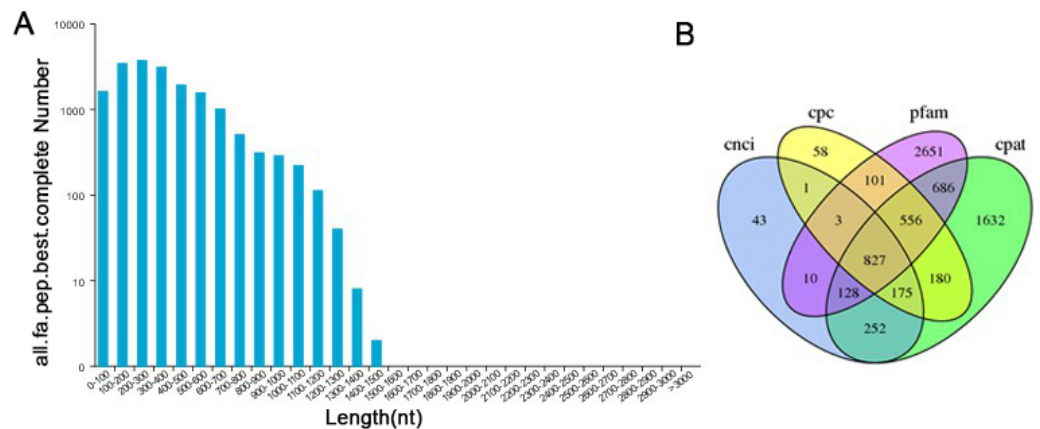
Sun et al. (2021), *PeerJ*, DOI 10.7717/peerj.11808

10/22

**Figure 5** **Information on the distribution of ORFs and lncRNAs.** (A) The length distribution of 17,904 completed ORFs. (B) Venn diagram of lncRNAs predicted by CPC, CNCI, CPAT and Pfam methods.
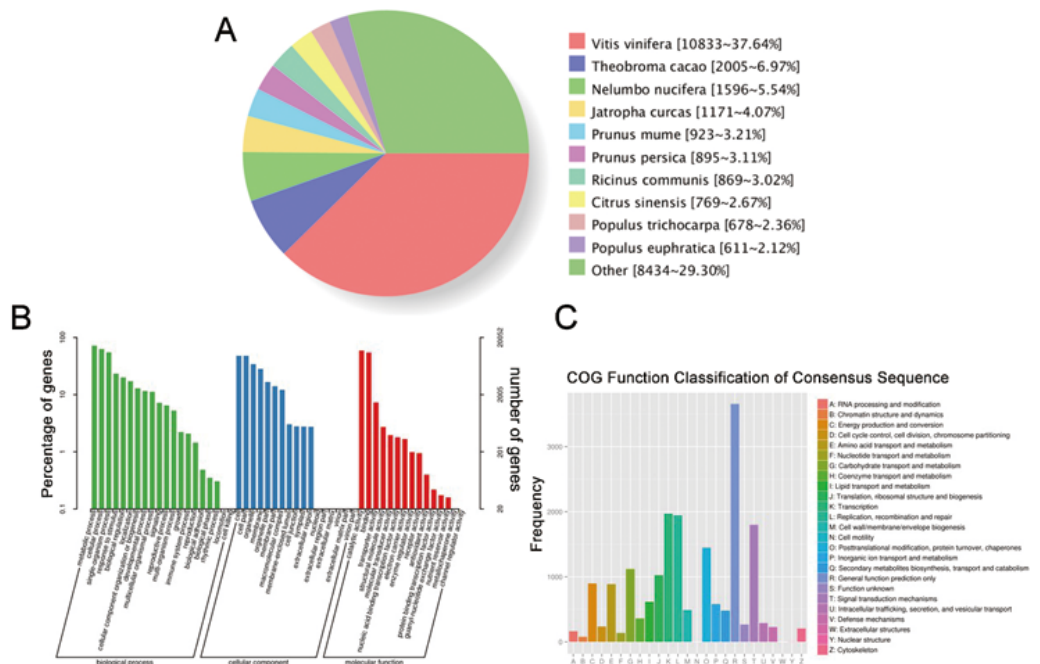
Full-size ☒ DOI: 10.7717/peerj.11808/fig-5



**Figure 6** **Function annotation of transcripts.** (A) Nr Homologous species distribution. (B) Distribution of GO terms for all annotated transcripts in biological process, cellular component and molecular function. (C) COG function classification of consensus sequence.

Full-size ☒ DOI: 10.7717/peerj.11808/fig-6

particular, *PoTPS1* contains only one TPS domain, and the remaining nine TPS family members all contain one TPS domain and one TPP domain. In addition, it was also found that the distribution of the domains of the TPS family members is roughly the same as that of *Arabidopsis*, rice and soybean, showing that the TPS and TPP domains of this family are positionally conserved among different species. However, whether these family members

**Table 1  The physical and chemical properties and secondary structure elements of TPS family members.**

| Gene name | Number of amino acids | Molecular weight (kD) | Isoelectric point (pI) | Aliphatic index | Protein hydrophobicity | Helix (%) | Extended strand (%) | Random coil (%) |
|---|---|---|---|---|---|---|---|---|
| PoTPS1 | 565 | 64.17 | 6.33 | 89.20 | −0.291 | 51.50 | 13.81 | 29.73 |
| PoTPS3 | 959 | 108.46 | 5.77 | 85.29 | −0.337 | 41.40 | 14.08 | 39.10 |
| PoTPS4 | 561 | 62.63 | 5.81 | 89.66 | −0.249 | 42.42 | 14.44 | 36.90 |
| PoTPS5 | 856 | 97.06 | 5.61 | 90.95 | −0.214 | 43.81 | 16.94 | 34.11 |
| PoTPS6 | 674 | 76.82 | 6.37 | 94.12 | −0.138 | 45.10 | 16.62 | 31.75 |
| PoTPS7 | 858 | 96.65 | 6.32 | 85.77 | −0.227 | 42.66 | 16.90 | 35.78 |
| PoTPS8 | 807 | 91.49 | 6.19 | 89.63 | −0.223 | 44.24 | 16.23 | 34.70 |
| PoTPS9 | 429 | 48.01 | 6.15 | 85.38 | −0.206 | 44.76 | 17.02 | 33.33 |
| PoTPS10 | 754 | 86.10 | 6.16 | 90.49 | −0.224 | 44.69 | 16.84 | 33.16 |
| PoTPS11 | 474 | 53.09 | 5.85 | 84.56 | −0.299 | 45.36 | 16.24 | 33.76 |

are functionally similar is not yet known. To analyze the genetic relationship and evolution characteristics of the *TPS* members in *P. ostii*, a phylogenetic tree was constructed by using MEGA 7.0 software, finding that the PoTPS1 protein has a high similarity to AtTPS1 and that the PoTPS7 protein has a high similarity to AtTPS7 (Fig. 7). In addition, the motif is a small conserved sequence fragment. In biology, a mathematical statistical model based on data can help predict the reliability of phylogenetic analysis, so MEME was used to analyze the motifs of all TPS proteins and it was found that only PoTPS5, PoTPS7 and PoTPS8 had ten conserved motifs, while the other proteins had 5-9 motifs each (Fig. 7).

## Gene expression analysis

We analyzed the expression levels of 10 *TPS* family members on the stems, leaves and seeds at 30, 50, and 70 DAF to explore whether the *TPS* gene expression in different tissues and different periods follows certain expression patterns and whether *TPS* genes were specifically expressed in different tissues. TBtools software was used to analyze the obtained PCR results and the red color from light to dark represents the expression level from low to high. The results showed that in seeds the relative expression levels of *PoTPS5* and *PoTPS6* were relatively high at 30 DAF, and the relative expression levels of *PoTPS5* and *PoTPS11* were high at 50 DAF. The relative expression levels of *PoTPS1* and *PoTPS5* were relatively high at 70 DAF. Obviously, the relative expression level of *PoTPS1* in seeds tended to increase with the development of the seeds. In leaves, the expression of *PoTPS11* was higher at 30 DAF, the expression of *PoTPS5* and *PoTPS8* was higher at 50 DAF, and the expression of *PoTPS5* was highest at 70 DAF. In stems, the relative expression level of *PoTPS5* was high regardless of the period (Fig. 8). In summary, the relative expression of *PoTPS5* was always been at a high level in different tissues at different periods.

## DISCUSSION

Tree peony, as a traditional Chinese flower, is worthy of research by gardeners and landscape researchers to improve its breeding. The current understanding of the tree peony transcriptome is mainly devoted only to the transcriptome of specific tissues such
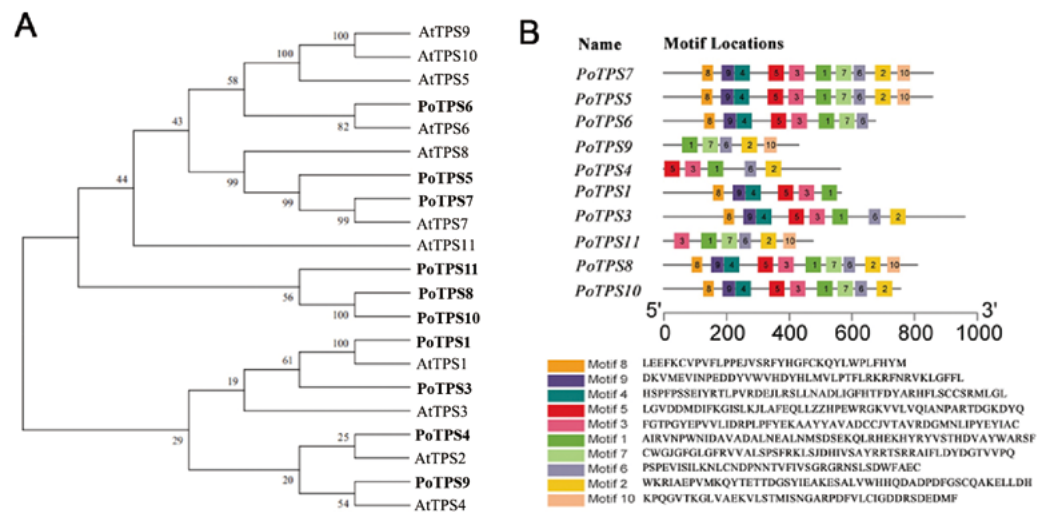
**Figure 7** **Bioinformatics analysis of *PoTPSs* members.** (A) Phylogenetic relationships of the TPS proteins between *Paeonia ostii* and *Arabidopsis*. Bootstrap values only above 60 are provided above branches. (B) Schematic diagram of amino acid motifs of PoTPS protein. The TPS domain includes 1, 3, 4, 5, 8, 9 modifs and TPP domain includes 2, 6, 7, 10 modifs.
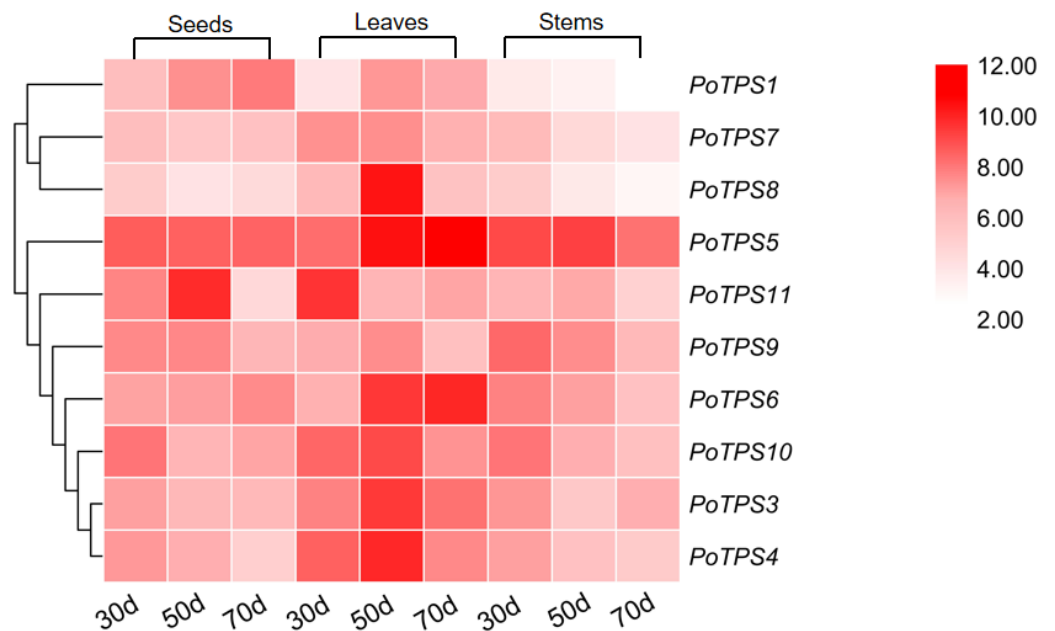
**Figure 8** **Expression profiles of the 10 *PoTPS* genes in different tissues and different periods.** Seeds, leaves and stems of 30d, 50d, 70d after flowering were used to analyze the expression levels of 10 *PoTPS* genes in *Paeonia ostii*. The red from light to dark represents the expression level from low to high.

as petals (*Li et al., 2017*). Surprisingly, little full-length transcriptome sequencing has been conducted for tree peony. In our study, SMRT-seq allowed us to obtain information about the full-length transcriptome sequence of tree peony. We obtained 230,736 ROI sequences

and 114,215 FLNC transcripts for further ORF and transcription factor prediction, SSR analysis and lncRNA identification. We circumvented the shortcomings of short reads in RNA-Seq2.0 and obtained long reads that can represent the entire transcript, confirming that SMRT-seq was more accurate in recovering full-length transcripts (*Hoang et al., 2017a*). Since the complete peony genome has not yet been disclosed, the massive amount of full-length transcripts obtained in this study provide a reliable biological information background resource for the subsequent in-depth study of the molecular mechanism of peony growth and development. It has laid an important foundation for molecular breeding.

*TPS* family members in *Arabidopsis* (*Vandesteene et al., 2010*), poplar (*Yang et al., 2012*), rice (*Zang et al., 2011*), cotton (*Xie et al., 2015*), lotus (*Jin et al., 2016*), potato (*Xu et al., 2017*) and apple (*Du et al., 2017*) have been genome-wide identified. Based on the transcriptome result, a total of 10 *TPS* family members in *P. ostii* were identified. Except for PoTPS1 which only contains the TPS domain, other PoTPS proteins contain both conserved TPS and TPP domains, similar to those in *Arabidopsis*, poplar, rice and drumstick trees (*Lin et al., 2018*; *Yang et al., 2012*). The TPS and TPP domain of this family are positionally conserved across different species. Studies have shown that genes lacking a TPP conserved domain in the C-terminus in *Arabidopsi* s can cause catastrophic consequences for the plant (*Fichtner et al., 2020*). Through anlaysis of the phylogenetic tree, it was found that the PoTPS1 protein is very similar to AtTPS1 (Fig. 7). AtTPS1 is a key enzyme for Tre6P synthesis in *Arabidopsis* and is essential for embryogenesis and normal postembryonic growth and development (*Fichtner et al., 2020*). Therefore, whether the deletion of the TPP domain of *PoTPS1* has a good or bad effect on the growth and development of *P. ostii* is not yet known.

The expression levels of *PoTPS* genes are totally different in seeds, stems and leaves at different periods. Surprisingly, *PoTPS5* showed the highest transcript level in different organs at different periods (Fig. 8). In *Arabidopsis*, *TPS5* is involved in affecting the trehalose content and functions as a negative regulator of ABA signaling (*Tian et al., 2019*). Whether the high expression of *PoTPS5* represents a positive effect on the growth of *P.ostii* needs further study. Moreover, it is worth noting that expression levels of *PoTPS1* increase during seed development (Fig. 8). Based on this qRT-PCR result, we can predict the pathway of *PoTPS1* to participate in the biosynthesis and accumulation of trehalose in seeds. When T6P synthesized by TPS is lacking, the size of pea seeds will decrease, and the starch yield will decrease because auxin acts downstream of T6P to facilitate seed filling (*Meitzel et al., 2019*). It can also be hypothesized that *PoTPS1* may be involved in the development and filling process of seeds in *P. ostii*. The cell-specific pattern of TPS expression remains to be examined. Fortunately, previous studies have found that *AtTPS1* is normally expressed in reproductive organs, roots, leaves of all major tissues (*Schmid et al., 2005*). For example, it was shown that the CsTPS1 protein is contained in phloem sap exudates in *Cucumis sativus* (*Hu et al., 2016*). In contrast, in *Arabidopsis*, the number of *AtTPS1* transcripts is quite large as shown by ribosome pull-down experiments in bundle sheath cells (*Aubry et al., 2014*). Therefore, to better explore the specific functions of the *PoTPSs*, it is worthwhile

to study the relationship between *PoTPSs* and biomass accumulation in different tissues in future studies.

## CONCLUSIONS

In this study, we analyzed the full-length transcriptome of *P. ostii* with SMRT-seq technology. Our findings provide valid information for improving tree peony draft genome annotation and full characterization of its transcriptome. Moreover, we identified 10 TPS family members in *P. ostii* and provided a comprehensive analysis of their physical and chemical properties, conserved protein motifs and phylogenetic tree analysis. Finally, we analyzed the expression patterns of *PoTPSs*, helping us to better understand the functions of these *TPS* family members.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Jing Sun conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Tian Chen performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Jun Tao conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:
The PCR data and annotations regarding GO, KEGG are available in the Supplemental Files.

The data are available at the Sequence Read Archive (SRA): PRJNA688625 and the 10 PoTPS genes are available at NCBI: MW700299–MW700308.

The final polished consensus mRNA sequences are available at NCBI: GJFW00000000.

## Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11808#supplemental-information.

## REFERENCES

**Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS. 2016.** A survey of the sorghum transcriptome using single-molecule long reads. *Nature Communications* **7**:11706 DOI 10.1038/ncomms11706.

**Altschul S. 1999.** Hot papers - Bioinformatics - Gapped BLAST and PSI-BLAST: a new generation of protein database search programs by S.F Altschul, T.L. Madden, A.A. Schaffer, J.H. Zhang, Z. Zhang, W. Miller, D.J. Lipman - Comments. *Scientist* **13**:15–15.

**Aubry S, Smith-Unna RD, Boursnell CM, Kopriva S, Hibberd JM. 2014.** Transcript residency on ribosomes reveals a key role for the Arabidopsis thaliana bundle sheath in sulfur and glucosinolate metabolism. *The Plant Journal* **78**:659–673 DOI 10.1111/tpj.12502.

**Bessa C, Matos P, Jordan P, Gonçalves V. 2020.** Alternative splicing: expanding the landscape of cancer biomarkers and therapeutics. *International Journal of Molecular Sciences* **21**:9032 DOI 10.3390/ijms21239032.

**Blazquez MA, Santos E, Flores CL, Martinez-Zapater JM, Salinas J, Gancedo C. 1998.** Isolation and molecular characterization of the Arabidopsis TPS1 gene, encoding trehalose-6-phosphate synthase. *The Plant Journal* **13**:685–689 DOI 10.1046/j.1365-313x.1998.00063.x.

**Chao YH, Yuan JB, Li SF, Jia SQ, Han LB, Xu LX. 2018.** Analysis of transcripts and splice isoforms in red clover (Trifolium pratense L.) by single-molecule long-read sequencing. *BMC Plant Biology* **18**:300 DOI 10.1186/s12870-018-1534-8.

**Debast S, Nunes-Nesi A, Hajirezaei MR, Hofmann J, Sonnewald U, Fernie AR, Bornke F. 2011.** Altering trehalose-6-phosphate content in transgenic potato tubers affects tuber growth and alters responsiveness to hormones during sprouting. *Plant Physiology* **156**:1754–1771 DOI 10.1104/pp.111.179903.

**Dong LL, Liu HF, Zhang JC, Yang SJ, Kong GY, Chu JSC, Chen NS, Wang DW. 2015.** Single-molecule real-time transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genomics* **16**:1039 DOI 10.1186/s12864-015-2257-y.

**Du LS, Qi SY, Ma JJ, Xing LB, Fan S, Zhang SW, Li YM, Shen YW, Zhang D, Han MY. 2017.** Identification of TPS family members in apple (Malus x domestica Borkh.) and the effect of sucrose sprays on TPS expression and floral induction. *Plant Physiology and Biochemistry* **120**:10–23 DOI 10.1016/j.plaphy.2017.09.015.

**Eastmond PJ, Van Dijken AJH, Spielman M, Kerr A, Tissier AF, Dickinson HG, Jones JDG, Smeekens SC, Graham IA. 2002.** Trehalose-6-phosphate synthase 1, which catalyses the first step in trehalose synthesis, is essential for Arabidopsis embryo maturation. *The Plant Journal* **29**:225–235 DOI 10.1046/j.1365-313x.2002.01220.x.

**Fichtner F, Barbier FF, Feil R, Watanabe M, Annunziata MG, Chabikwa TG, Hofgen R, Stitt M, Beveridge CA, Lunn JE. 2017.** Trehalose 6-phosphate is involved in triggering axillary bud outgrowth in garden pea (Pisum sativum L.). *The Plant Journal* **92**:611–623 DOI 10.1111/tpj.13705.

**Fichtner F, Olas JJ, Feil R, Watanabe M, Krause U, Hoefgen R, Stitt M, Lunn JE. 2020.** Functional features of TREHALOSE-6-PHOSPHATE SYNTHASE1, an essential enzyme in Arabidopsis. *The Plant Cell* **32**:1949–1972 DOI 10.1105/tpc.19.00837.

**Gomez LD, Baud S, Gilday A, Li Y, Graham IA. 2006.** Delayed embryo development in the ARABIDOPSIS TREHALOSE-6-PHOSPHATE SYNTHASE 1 mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. *The Plant Journal* **46**:69–84 DOI 10.1111/j.1365-313X.2006.02662.x.

**Gomez LD, Gilday A, Feil R, Lunn JE, Graham IA. 2010.** AtTPS1-mediated trehalose 6-phosphate synthesis is essential for embryogenic and vegetative growth and responsiveness to ABA in germinating seeds and stomatal guard cells. *The Plant Journal* **64**:1–13 DOI 10.1111/j.1365-313X.2010.04312.x.

**Gordon S, Tseng E, Salamov A, Zhang J, Meng X, Zhao Z, Kang DD, Underwood J, Grigoriev IV, Figueroa M, Schilling JS, Chen F, Wang Z. 2014.** Widespread polycistronic transcripts in mushroom-forming fungi revealed by single-molecule long-read mRNA sequencing. *PLOS ONE* **10**:e0132628 DOI 10.1101/012542.

**Gordon SP, Tseng E, Salamov A, Zhang JW, Meng XD, Zhao ZY, Kang DW, Underwood J, Grigoriev IV, Figueroa M, Schilling JS, Chen F, Wang Z. 2015.** Widespread polycistronic transcripts in fungi revealed by single-molecule mRNA sequencing. *PLOS ONE* **10**:e0132628 DOI 10.1371/journal.pone.0132628.

**Harthill JE, Meek SEM, Morrice N, Peggie MW, Borch J, Wong BHC, MacKintosh C. 2006.** Phosphorylation and 14-3-3 binding of Arabidopsis trehalose-phosphate synthase 5 in response to 2-deoxyglucose. *The Plant Journal* **47**:211–223 DOI 10.1111/j.1365-313X.2006.02780.x.

**Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ. 2017a.** A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* **18**:395 DOI 10.1186/s12864-017-3757-8.

**Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ. 2017b.** A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* **18**:395 DOI 10.1186/s12864-017-3757-8.

**Hong DY, Pan KY, Pan ZQ. 2004.** Circumscription of Paeonia suffruticosa Andrews and identification of cultivated tree peonies. *Journal of Systematics and Evolution, Beijing. Sin* **42**:275–283 (In Chinese).

**Hu CY, Ham BK, El-Shabrawi HM, Alexander D, Zhang DB, Ryals J, Lucas WJ. 2016.** Proteomics and metabolomics analyses reveal the cucurbit sieve tube system as a complex metabolic space. *The Plant Journal* **87**:442–454 DOI 10.1111/tpj.13209.

**Jang IC, Oh SJ, Seo JS, Choi WB, Song SI, Kim CH, Kim YS, Seo HS, Do Choi Y, Nahm BH, Kim JK. 2003.** Expression of a bifunctional fusion of the Escherichia coli genes for trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase in transgenic rice plants increases trehalose accumulation and abiotic stress tolerance without stunting growth. *Plant Physiology* **131**:516–524 DOI 10.1104/pp.007237.

**Jin QJ, Hu X, Li X, Wang B, Wang YJ, Jiang HW, Mattson N, Xu YC. 2016.** Genome-wide identification and evolution analysis of Trehalose-6-Phosphate synthase gene family in nelumbo nucifera. *Frontiers in Plant Science* **7**:1445 DOI 10.3389/fpls.2016.01445.

**Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007.** CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* **35**:W345–W349 DOI 10.1093/nar/gkm391.

**Kumar R, Bishop E, Bridges WC, Tharayil N, Sekhon RS. 2019.** Sugar partitioning and source–sink interaction are key determinants of leaf senescence in maize. *Plant Cell and Environment* **42**:2597–2611 DOI 10.1111/pce.13599.

**Li WZ, Godzik A. 2006.** Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**:1658–1659 DOI 10.1093/bioinformatics/btl158.

**Li Y, Lu JX, Chang YH, Tang WL, Yang QS. 2017.** Comparative analysis of tree peony petal development by transcriptome sequencing. *Acta Physiologiae Plantarum* **39**:216 DOI 10.1007/s11738-017-2520-8.

**Lin MF, Jia RH, Li JC, Zhang MJ, Chen HB, Zhang D, Zhang JJ, Chen XY. 2018.** Evolution and expression patterns of the trehalose-6-phosphate synthase gene family in drumstick tree (Moringa oleifera Lam.). *Planta* **248**:999–1015 DOI 10.1007/s00425-018-2945-3.

**Livak KJ, Schmittgen TD. 2001.** Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**:402–408 DOI 10.1006/METH.2001.1262.

**Lunn JE. 2007.** Gene families and evolution of trehalose metabolism in plants. *Functional Plant Biology* **34**:550–563 DOI 10.1071/Fp06315.

**Lunn JE, Delorge I, Figueroa CM, Van Dijck P, Stitt M. 2014.** Trehalose metabolism in plants. *The Plant Journal* **79**:544–567 DOI 10.1111/tpj.12509.

**Meitzel T, Radchuk R, McAdam EL, Thormählen I, Feil R, Munz E, Hilo A, Geigenberger P, Ross JJ, Lunn JE, Borisjuk L. 2019.** Trehalose 6-phosphate controls seed filling by inducing auxin biosynthesis. *BioRxiv* 752915 DOI 10.1101/752915.

**Ping X, Wang T, Lin N, Di F, Li Y, Jian H, Wang H, Lu K, Li J, Xu X, Liu L. 2019.** Genome-wide identification of the LAC gene family and its expression analysis under stress in brassica napus. *Molecules* **24**:1985 DOI 10.3390/molecules24101985.

**Qian XD, Sun YP, Zhou GF, Yuan YM, Li J, Huang HL, Xu LM, Li LQ. 2019.** Single-molecule real-time transcript sequencing identified flowering regulatory genes in Crocus sativus. *BMC Genomics* **20**:857 DOI 10.1186/s12864-019-6200-5.

**Ren XX, Xue JQ, Wang SL, Xue YQ, Zhang P, Jiang HD, Zhang XX. 2018.** Proteomic analysis of tree peony (Paeonia ostii 'Feng Dan') seed germination affected by low temperature. *Journal of Plant Physiology* **224–225**:56–67 DOI 10.1016/j.jplph.2017.12.016.

**Saitou N, Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406–425 DOI 10.1093/oxfordjournals.molbev.a040454.

**Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005.** A gene expression map of Arabidopsis thaliana development. *Nature Genetics* **37**:501–506 DOI 10.1038/ng1543.

**Sharon D, Tilgner H, Grubert F, Snyder M. 2013.** A single-molecule long-read survey of the human transcriptome. *Nature Biotechnology* **31**:1009–1014 DOI 10.1038/nbt.2705.

**Sharon D, Tilgner H, Grubert F, Snyder M. 2014.** A single-molecule long-read survey of the human transcriptome (vol 31, pg 1009, 2013). *Nature Biotechnology* **32**:291–291 DOI 10.1038/nbt0314-291b.

**Shima S, Matsui H, Tahara S, Imai R. 2007.** Biochemical characterization of rice trehalose-6-phosphate phosphatases supports distinctive functions of these plant enzymes. *The FEBS Journal* **274**:1192–1201 DOI 10.1111/j.1742-4658.2007.05658.x.

**Tan C, Liu HX, Ren J, Ye XL, Feng H, Liu ZY. 2019.** Single-molecule real-time sequencing facilitates the analysis of transcripts and splice isoforms of anthers in Chinese cabbage (Brassica rapa L. ssp. pekinensis). *BMC Plant Biology* **19**:517 DOI 10.1186/s12870-019-2133-z.

**Thomas S, Underwood JG, Tseng E, Holloway AK, on behalf of the Bench To Basinet CvDC Informatics Subcommittee. 2014.** Long-read sequencing of chicken transcripts and identification of new transcript isoforms. *PLOS ONE* **9**:e94650 DOI 10.1371/journal.pone.0094650.

**Tian LF, Xie Zj, Lu CQ, Hao XH, Wu S, Huang Y, Li DP, Chen LB. 2019.** The trehalose-6-phosphate synthase *TPS5* negatively regulates ABA signaling in *Arabidopsis thaliana*. *Plant Cell Reports* **38**:869–882 DOI 10.1007/s00299-019-02408-y.

**Tilgner H, Raha D, Habegger L, Mohiuddin M, Gerstein M, Snyder M. 2013.** Accurate identification and analysis of human mRNA isoforms using deep long read sequencing. *G3-Genes Genomes Genetics* **3**:387–397 DOI 10.1534/g3.112.004812.

**Tuskan GA, Di Fazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier**

J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, VandePeer Y, Rokhsar D. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science* **313**:1596–1604 DOI 10.1126/science.1128691.

Van Dijck P, Mascorro-Gallardo JO, De Bus M, Royackers K, Iturriaga G, Thevelein JM. 2002. Truncation of Arabidopsis thaliana and Selaginella lepidophylla trehalose-6-phosphate synthase unlocks high catalytic activity and supports high trehalose levels on expression in yeast. *Biochemical Journal* **366**:63–71 DOI 10.1042/BJ20020517.

Van Dijken AJH, Schluepmann H, Smeekens SCM. 2004. Arabidopsis trehalose-6-phosphate synthase 1 is essential for normal vegetative growth and transition to flowering. *Plant Physiology* **135**:969–977 DOI 10.1104/pp.104.039743.

Vandesteene L, Ramon M, Le Roy K, Van Dijck P, Rolland F. 2010. A single active Trehalose-6-P Synthase (TPS) and a family of putative regulatory TPS-like proteins in Arabidopsis. *Molecular Plant* **3**:406–419 DOI 10.1093/mp/ssp114.

Wahl V, Ponnu J, Schlereth A, Arrivault S, Langenecker T, Franke A, Feil R, Lunn JE, Stitt M, Schmid M. 2013. Regulation of flowering by Trehalose-6-phosphate signaling in Arabidopsis thaliana. *Science* **339**:704–707 DOI 10.1126/science.1230406.

Wang YJ, Dong L, Zhang C, Wang XQ. 2012. Reference Gene Selection for Real-time Quantitative PCR Normalization in Tree Peony (Paeonia suffruticosa Andr.). *Journal of Agricultural Biotechnology* **8**:521–528 (In Chinese).

Wang XJ, Liang HY, Guo DL, Guo LL, Duan XG, Jia QS, Hou XG. 2019. Integrated analysis of transcriptomic and proteomic data from tree peony (*P. ostii*) seeds reveals key developmental stages and candidate genes related to oil biosynthesis and fatty acid metabolism. *Horticulture Research* **6**:111 DOI 10.1038/s41438-019-0194-7.

Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D. 2016a. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Communications* **7**:11708 DOI 10.1038/ncomms11708.

Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao YP, Lu ZY, Olson A, Stein JC, Ware D. 2016b. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature Communications* **7**:11708 DOI 10.1038/ncomms11708.

**Wingler A, Delatte TL, O'Hara LE, Primavesi LF, Jhurreea D, Paul MJ, Schluepmann H. 2012.** Trehalose 6-Phosphate is required for the onset of leaf senescence associated with high carbon availability. *Plant Physiology* **158**:1241–1251 DOI 10.1104/pp.111.191908.

**Xie LJ, Teng K, Tan PH, Chao YH, Li YRZ, Guo WE, Han LB. 2020.** PacBio single-molecule long-read sequencing shed new light on the transcripts and splice isoforms of the perennial ryegrass. *Molecular Genetics and Genomics* **295**:475–489 DOI 10.1007/s00438-019-01635-y.

**Xie DW, Wang XN, Fu LS, Sun J, Zheng W, Li ZF. 2015.** Identification of the trehalose-6-phosphate synthase gene family in winter wheat and expression analysis under conditions of freezing stress. *Journal of Genetics* **94**:55–65 DOI 10.1007/s12041-015-0495-z.

**Xu Z, Peters RJ, Weirather J, Luo H, Liao B, Zhang X, Zhu Y, Ji A, Zhang B, Hu S, Au KF, Song J, Chen S. 2015.** Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. *The Plant Journal* **82**:951–961 DOI 10.1111/tpj.12865.

**Xu YC, Wang YJ, Mattson N, Yang L, Jin QJ. 2017.** Genome-wide analysis of the Solanum tuberosum (potato) trehalose-6-phosphate synthase (TPS) gene family: evolution and differential expression during development and stress. *BMC Genomics* **18**:926 DOI 10.1186/s12864-017-4298-x.

**Yadav UP, Ivakov A, Feil R, Duan GY, Walther D, Giavalisco P, Piques M, Carillo P, Hubberten HM, Stitt M, Lunn JE. 2014.** The sucrose-trehalose 6-phosphate (Tre6P) nexus: specificity and mechanisms of sucrose signalling by Tre6P. *The Journal of Experimental Botany* **65**:1051–1068 DOI 10.1093/jxb/ert457.

**Yang HL, Liu YJ, Wang CL, Zeng QY. 2012.** Molecular evolution of trehalose-6-phosphate synthase (TPS) Gene family in populus, Arabidopsis and rice. *PLOS ONE* **7**:e42438 DOI 10.1371/journal.pone.0042438.

**Yao LJ. 2020.** Seed development, dormancy and germination characteristic of *Paeonia ostii*. Doctoral dissertation, Yangzhou University (in Chinese with English abstract).

**Zang BS, Li HW, Li WJ, Deng XW, Wang XP. 2011.** Analysis of trehalose-6-phosphate synthase (TPS) gene family suggests the formation of TPS complexes in rice. *Plant Molecular Biology* **76**:507–522 DOI 10.1007/s11103-011-9781-1.

**Zentella R, Mascorro-Gallardo JO, Van Dijck P, Folch-Mallol J, Bonini B, Van Vaeck C, Gaxiola R, Covarrubias AA, Nieto-Sotelo J, Thevelein JM, Iturriaga G. 1999.** A Selaginella lepidophylla trehalose-6-phosphate synthase complements growth and stress-tolerance defects in a yeast tps1 mutant. *Plant Physiology* **119**:1473–1482 DOI 10.1104/pp.119.4.1473.

**Zhao ML, Ni J, Chen MS, Xu ZF. 2019.** Ectopic expression of jatropha curcas TREHALOSE-6-PHOSPHATE PHOSPHATASE J causes late-flowering and heterostylous phenotypes in Arabidopsis but not in Jatropha. *International Journal of Molecular Sciences* **20**:2165 DOI 10.3390/ijms20092165.

**Zhao P, Wang D, Wang R, Kong N, Zhang C, Yang C, Wu W, Ma H, Chen Q. 2018.**
Genome-wide analysis of the potato Hsp20 gene family: identification, genomic organization and expression profiles in response to heat stress. *BMC Genomics* **19**:61 DOI 10.1186/s12864-018-4443-1.

**Zhao X, Yang G, Liu X, Yu Z, Peng S. 2020.** Integrated analysis of seed microRNA and mRNA transcriptome reveals important functional genes and microRNA-targets in the process of walnut (Juglans regia) seed oil accumulation. *International Journal of Molecular Sciences* **21** DOI 10.3390/ijms21239093.