

Gene co-expression analysis and tissue-specific gene identification in *Liriodendron chinense* via hybrid sequencing

zhonghua tu

Nanjing Forestry University <https://orcid.org/0000-0002-2313-8550>

Yufang Shen

Nanjing Forestry University

Shaoying Wen

Nanjing Forestry University

Huanhuan Liu

Nanjing Forestry University

Lingmin Wei

Nanjing Forestry University

Huogen Li (✉ hgli@njfu.edu.cn)

<https://orcid.org/0000-0001-6455-9346>

Research article

Keywords: L. chinense, hybrid sequencing, transcriptome analysis, tissue-specific genes, gene co-expression modules, MIKC-type MADS-box genes

Posted Date: June 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-32884/v1>

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

[Read Full License](#)

Abstract

Background *Liriodendron chinense* (Hemsl.) Sarg. is an economically and ecologically important deciduous tree species that has been studied for many years. Although the complete *L. chinense* genome has been sequenced, the gene co-expression modules and tissue-specific genes of *L. chinense* remain unknown. **Results** Here, we used the bracts, petals, sepals, stamens, pistils, leaves, and the shoot apex of *L. chinense* as materials and analysed their gene co-expression modules and tissue-specific genes via hybrid sequencing. We identified 3,032 DEGs between the floral and vegetative tissues and 2,126 tissue-specific genes. By using WGCNA analysis, we identified 13 gene co-expression modules, and KEGG pathway enrichment analysis revealed that tissue-specific genes and genes from different modules were enriched in different pathways. Genes associated with plant defence were highly expressed in the bracts, genes participating in plant hormone signal transduction were highly expressed in the shoot apex, and genes participating in photosynthesis were highly expressed in the leaves, petals and sepals. Moreover, we identified 10 MIKC-type MADS-box genes that were classified as member of the AP3/PI, SVP, SEP, AG/SHP/STK, AGL12, SOC1 and TM8 subfamily. Phylogenetic analysis showed that the expression profiles of these ten genes were consistent with those reported in *Arabidopsis* and *Populus*, indicating that these genes are highly conserved evolutionarily and related to floral and vegetative tissue development. The small number of MIKC-type MADS-box genes in *L. chinense* was probably owing to its incomplete genome annotation. **Conclusions** In this work, we provided a reference transcriptome for *L. chinense* research by using hybrid sequencing. We identified 2,126 tissue-specific genes and 3,032 DEGs that contributed greatly to the functional differences between vegetative organs and floral organs. By using WGCNA analysis, 13 gene co-expression modules and 52 hub genes from six co-expression modules of interest were identified. Moreover, we identified 10 MIKC-type MADS-box genes that might be related to the development and growth regulation of floral and vegetative organs. These findings will improve our understanding of gene co-expression, tissue specific genes and flower development model of *L. chinense*.

Background

Liriodendron chinense (Hemsl.) Sarg (*L. chinense*) is a deciduous tree species with attractive flowers is a member of the magnolia family (Magnoliaceae). *L. chinense* is distributed throughout southern China and northern Vietnam, and it plays increasingly important economic and ecological roles because of its high value for greening, furniture manufacturing and pharmaceutical manufacturing [1-3]. In addition, owing to its crucial phylogenetic position that enables the study of the evolution of extant flowering plants, *L. chinense* has been the topic of an increasing amount of scientific research, such as research on gene flow among natural populations of *L. chinense* [4], transcriptomic analyses of *L. chinense* petals and leaves [5], the study of genes associated with *L. chinense* leaf shape development [6], and the whole-genome sequencing of *L. chinense* (the complete sequence of which was completed in 2018) [7]. However, a reference transcriptome of *L. chinense* has not been reported. A reference transcriptome can provide valuable information for researching gene expression and evolution; investigating alternative

splicing (AS) events and alternative polyadenylation (APA); and identifying long non-coding RNAs (lncRNAs), fusion transcripts and transcription factors (TFs) [8, 9]. The tissue-specific genes and gene co-expression modules in *L. chinense* remain unknown. The lack of such knowledge has hindered molecular breeding and studies of physiological and biochemical regulatory networks and metabolic pathways in *L. chinense*.

L. chinense is also referred to as tulip tree because of its brightly coloured and attractive flowers. Exploring flower development models can help improve the understanding of the phylogenetic position of *L. chinense* during the evolution of land flowering plants. To the best of our knowledge, the ABCDE model plays pivotal roles in the determination of floral organ identity, and with the exception of *APETALA2* (*AP2*), other ABCDE type genes belong to the MADS intervening keratin-like and C-terminal (MIKC)-type MADS-box gene family. In *Arabidopsis thaliana*, A-type genes include *AP1* and *FRUITFULL* (*FUL*), B-type genes include *AP3* and *PISTILLATA* (*PI*), C-type genes include *AGAMOUS* (*AG*), D-type genes include *SEEDSTICK* (*STK*), *SHATTERPROOF1* (*SHP1*) and *SHP2*, and E-type genes include *SEPALLATA1* (*SEP1*), *SEP2*, *SEP3* and *SEP4*; these genes belong to the MIKC-type MADS-box gene family [10]. Different types of genes together determine the identities of different floral organs; for example, sepals are determined by A-type and E-type genes, petals are determined by A-type, B-type and E-type genes, stamens are determined by B-type, C-type and E-type genes, carpels are determined by C-type and E-type genes, and ovules are determined by D-type and E-type genes [11-14]. Other MIKC-type MADS-box genes also play crucial roles in regulation of flowering time and floral initiation, such as *Flowering Locus C* (*FLC*), *Short Vegetative Phase* (*SVP*) and *Suppressor of Overexpression Of Constans1* (*SOC1*) [15-17]. However, MIKC-type MADS-box genes in *L. chinense* have not been studied, which hinders our understanding of floral organ identity in *L. chinense*.

Next-generation sequencing (NGS) technology has been widely applied to studies of metabolic pathways, differential gene expression, and regulatory mechanisms in model or non-model species, such as maize (*Zea mays*) [18], rice (*Oryza sativa*) [19], *Cucumis sativus* L. [20], *Cinnamomum camphora* [21], and *L. chinense* [6] because of their high throughput, high accuracy and low cost. However, NGS also has some weaknesses, such as the difficulty of sequence assembly, the possibility of producing low-quality transcripts, and the inability to capture full-length transcripts effectively [22-28].

Single-molecule long-read sequencing technology (SMRT), which was developed by Pacific Biosciences (PacBio), has recently been used effectively to discover lncRNAs, comprehensively analyse AS and APA, detect fusion genes, correct misannotated gene models and discover novel genes in animals and plants [23, 26-34]. However, the shortcomings of SMRT are its high error rate (~13%) and low throughput [26, 31]. Thus, the use of hybrid sequencing (combining NGS with SMRT) to study the transcriptome is an

effective strategy that has been widely applied in the study of various plants and animals, including *Salvia miltiorrhiza* [29], moso bamboo (*Phyllostachys edulis*) [35], sugarcane (*Saccharum officinarum*) [36], *Medicago sativa* L.[27], and *Oryctolagus cuniculus* [28].

To the best of our knowledge, biological traits and phenotypic characteristics are the result of the co-expression of multiple genes [37-40]. Co-expression network analysis, a powerful tool to discover connections between genes effectively and accurately, has been widely applied in plant research. For example, researchers found that genes encoding beta-glucosidases and sorbitol dehydrogenases in *Pyrus bretschneideri* were hub genes involved in a co-expression network [37]. By combining co-expression network analysis and functional enrichment tools, researchers identified 1,896 functional modules associated with moso bamboo development [38]. In maize, co-expression analysis was applied at the pathway level and enabled researchers to identify genes that potentially regulate the accumulation of gene transcripts associated with glossiness [39]. However, we have no knowledge of gene co-expression modules in *L. chinense*; thus, it is necessary to study gene co-expression in *L. chinense* to improve our understanding of gene expression patterns.

Gene co-expression and tissue-specific genes influence biological traits and phenotypic characteristics. Moreover, tissue-specific genes play fundamental roles in tissue differentiation and maintenance [41]. Researchers have found that the divergence of tissue-specific gene expression patterns modulates ion maintenance in tissues and improves salinity tolerance in *Populus euphratica* [42]. In pineapple (*Ananas comosus*), researchers detected 273 tissue-specific genes, which provided a basis for the study of the genetic mechanisms that regulate fruit phenotypes [43]. Although studies of *L. chinense* tissues (leaves and petals) have been reported, tissue-specific genes in *L. chinense* remain unknown, and this lack of knowledge impedes our understanding of tissue traits of this species [5, 6].

Researchers have studied the transcriptome of *L. chinense* but have focused on only one or two tissues [5, 6]; a high-quality reference transcriptome of *L. chinense* is still needed. More importantly, research on MIKC-type MADS-box genes in *L. chinense* has not been reported. To understand the transcriptome of *L. chinense* comprehensively and bridge the gap in knowledge concerning tissue-specific genes, gene co-expression modules and MIKC-type MADS-box genes in *L. chinense*, we used hybrid sequencing to study the transcriptomes of multiple tissues (bracts, sepal, petals, pistils, stamens, leaves, and the shoot apex) in *L. chinense*. By using hybrid sequencing, we identified TFs, lncRNAs, differentially expressed genes (DEGs), and fusion transcripts and investigated AS and APA in *L. chinense*. Importantly, gene co-expression modules and tissue-specific genes were discovered via hybrid sequencing. We analysed the phylogenetic relationship of MIKC-type MADS-box genes in *A. thaliana*, *P. trichocarpa* and *L. chinense*.

These findings will provide a basis for the study of the tissue characteristics and the establishment of a flower development model of *L. chinense* in the future.

Results

Global view of the PacBio sequencing data

Using PacBio SMRT, we obtained 611,876 polymerase reads from 21 pooled samples (Additional file 9: Table S1). We then extracted 10,437,029 subreads with an average length of 2,177 bp from the polymerase reads (Additional file 9: Table S1). After processing them further, we obtained 498,059 circular consensus sequences (CCSs), each of which contained a poly(A) tail and 5' and 3' adaptors, with an average length of 2,755 bp (Additional file 9: Table S1). Among the CCSs, we detected 430,554 full-length non-chimeric (FLNC) reads, with an average length of 2,568 bp (Additional file 9: Table S1). After polishing the reads with both iterative clustering for error correction (ICE) and Arrow, we obtained 227,276 polished consensus reads whose average length was 2,697 bp (Additional file 9: Table S1). By mapping the short reads generated from Illumina sequencing via LoRDEC, we corrected up to 100% of the sequencing errors [44]. We ultimately obtained 227,276 high-quality polished consensus reads with an average length of 2,696 bp for further analysis (Additional file 10: Table S2).

Isoform detection and characterization

Using TAPIS software, we classified and characterized the full-length isoforms [23]. In total, we obtained 89,640 isoforms, including 4,210 (4.70%) isoforms of known genes, 65,448 (73.01%) novel isoforms of known genes, and 19,982 (22.29%) isoforms of novel genes (Fig. 1a). We compared the lengths of the transcripts between the PacBio data and reference genome annotation, and found that the transcripts described in the reference genome annotation were shorter than those detected by PacBio Iso-Seq (Fig. 1b), which helps to reveal the complexity of the *L. chinense* transcriptome. In addition, we compared the numbers of exons in the transcripts between the PacBio data and reference genome annotation. We found that single-exon genes represented a large proportion (30.37%) of the PacBio data, while in the reference genome, single-exon genes accounted for only 7.29% (Fig. 1c). The proportion of multiple-exon genes (≥ 2 exons) in the reference genome annotation was greater than that in the PacBio data (Fig. 1c). Moreover, single-isoform genes constituted a large proportion of the PacBio data (49.22%), followed by two-isoform genes, which accounted for 15.98% (Fig. 1d).

Read mapping and identification of novel genes and TFs

Using GMAP software, we aligned the high-quality polished consensus reads to the reference genome and found that 1.13% (2,572/227,276) of the reads were unmapped [45]. Among the remaining 224,704 alignments (98.87%), 9.96% (22,628/227,276) of the reads were mapped to multiple sequences, 88.91% (202,076/227,276) of the reads were uniquely mapped, 52.73% (119,840/227,276) of the reads mapped to the sense sequences in the genome ('+'), and 36.18% (82,236/227,276) of the reads mapped to the antisense sequences in the genome ('-') (Fig. 1e, Additional file 11: Table S3).

In the PacBio sequencing data set, 2,572 unique transcript clusters did not align to the reference genome. The function of the unmapped transcripts was annotated on the basis of the information within the non-redundant (NR) database, nucleotide (NT) database, Protein Family (Pfam) database, EuKaryotic Orthologous Groups (KOG) database, Swiss-Prot database, Kyoto Encyclopedia of Genes and Genomes (KEGG) database and Gene Ontology (GO) database [46-51]. In these databases, 2,022 (NR), 1,584 (NT), 1,066 (Pfam), 1,260 (KOG), 1,621 (Swiss-Prot), 1,995 (KEGG) and 1,066 (GO) unmapped transcripts were annotated (Fig. 1f). Among these unmapped transcripts, 619 were found in all 7 databases, and 2,163 unmapped transcripts were annotated in only one database (Fig. 1f).

We defined reads that mapped to unannotated regions of the reference genome as reads from novel genes, and a total of 13,139 novel genes were detected. To better understand the novel genes, they were functionally annotated. In total, 9,343 (NR), 5,833 (NT), 6,494 (Pfam), 5,478 (KOG), 5,883 (Swiss-Prot), 9,177 (KEGG) and 6,494 (GO) novel genes were annotated in the 7 databases (Fig. 1g). In addition, 1,945 novel genes were annotated across all seven databases, and 10,905 novel genes were annotated in only one database (Fig. 1g). Furthermore, 2,234 novel genes were unannotated in all 7 databases, which may indicate that these genes have little coding ability and thus may represent lncRNAs [52].

We then used iTAK software to predict TFs [53]. We obtained a total of 5,532 TFs from 95 families (Additional file 12: Table S4). The top 30 families detected are shown in Fig. 1h. The top ten TF families identified included FAR1 (303), C3H (286), bHLH (266), SNF2 (250), others (209), bZIP (209), MYB-related (179), PHD (176), NAC (174), and SET (167) (Additional file 12: Table S4).

Identification of lncRNAs and fusion transcripts.

lncRNAs are RNA molecules that are longer than 200 nt and that do not encode proteins. We used four tools, PLEK, CNCI, CPC, and the Pfam database, to identify lncRNAs from PacBio Iso-Seq data [47, 54-56]. The PLEK, CNCI, CPC and the Pfam database tools predicted 44,108 lncRNAs, 15,500 lncRNAs, 21,549 lncRNAs, and 43,102 lncRNAs, respectively (Fig. 2a). To improve the accuracy of the lncRNA prediction, each lncRNA was predicted by four tools, and a total of 7,527 lncRNAs were detected (Fig. 2a). According to the position of each lncRNA when mapped to the reference genome, we classified the lncRNAs into four types, lincRNAs (36.44%), antisense lncRNAs (16.99%), sense intronic lncRNAs (18.37%), and sense overlapping lncRNAs (28.19%) (Fig. 2b) [57]. Overall, compared with mRNAs, the lncRNAs had fewer exons and were shorter (Fig. 2c and d). The length of most lncRNAs was less than 1 kb, and single-exon lncRNAs accounted for the majority of the lncRNAs (72.34%) (Fig. 2c and d).

A fusion gene consists of multiple coding regions connected end to end and controlled by a set of regulatory sequences (including promoters, enhancers, and ribosome-binding sequences). A total of 887 fusion transcripts were identified in this study (Additional file 13: Table S5). By analysing the distribution of fusion transcripts on the scaffolds (because the reference genome has not yet been assembled at the chromosomal level), we found that these fusion events tended to occur between scaffolds (866) rather than within a scaffold (21) (Additional file 13: Table S5).

AS and APA analyses

AS is an important mechanism for regulating gene expression and proteome diversity, and it is also an important cause of the large differences between the number of genes and the number of proteins in eukaryotes [58, 59]. By using the SUPPA tool, we identified 8,503 alternatively spliced genes, including 3,748 genes with skipped exons (SEs), 4,269 genes with retained introns (RIs), 404 genes with mutually exclusive exons (MXs), 3,625 genes with alternative 5' splice sites (A5s), 1,521 genes with alternative first exons (AFs), 4,195 genes with alternative 3' splice sites (A3s), and 632 genes with alternative last exons (ALs) (Fig. 3a). Among the 7 basic alternatively spliced genes, RI genes were the most common, while MX genes were the least frequent (Fig. 3a).

Most genes in eukaryotes can generate a variety mRNA 3' ends through APA, which greatly increases the complexity of the transcriptome [60, 61]. By using the TAPIS pipeline [23], we found that 11,108 genes contained at least one poly(A) site, among which 4,387 genes contained a single poly(A) site and among which 6,271 genes contained more than one poly(A) sites (Fig. 3b). To identify the potential cis-element involved in polyadenylation, a motif enrichment analysis was performed to analyse the 50 nucleotides upstream from the poly(A) sites [23]. We discovered a conserved motif (AUAAA) upstream of the poly(A) cleavage site (Fig. 3c), and this motif also was present in maize, red clover (*Trifolium pratense* L.) and sorghum (*Sorghum bicolor*) [22-24]. Moreover, to investigate the preferential nucleotides at the poly(A) cleavage sites, we analysed the nucleotide composition of 50 downstream and 50 upstream nucleotides at all APA cleavage sites. We found that uracil (U) was enriched upstream of the APA cleavage sites while adenine (A) was enriched downstream of the cleavage sites (Fig. 3d). The same findings were reported in red clover and sorghum [23, 24].

Identification of DEGs and tissue-specific genes in *L. chinense*

To investigate gene expression patterns in the seven evaluated tissues of *L. chinense*, we used fragments per kilobase of transcript sequence per million mapped reads (FPKM) values to normalize the reads from Illumina sequencing. In total, we identified 3,032 DEGs between floral tissues (bracts, sepals, petals, stamens, and pistils) and vegetative tissues (leaves and the shoot apex), of which the expression of 878 DEGs was upregulated, whereas that of the other 2,154 was downregulated (Fig. 4a). We performed GO enrichment analysis to relate these 3,032 DEGs to their products, and the results showed that the top three terms containing categorized genes were 'catalytic activity', 'single-organism metabolic process', and 'transferase activity' (Additional file 1: Fig. S1a). Furthermore, KEGG pathway enrichment analysis revealed that the DEGs participated in 107 pathways, and the top 3 significantly enriched pathways were associated with biosynthesis of flavonoids, glycosphingolipids, stilbenoids, diarylheptanoids and gingerol (Additional file 1: Fig. S1b).

In the different tissues, the euclidean distance method was used to perform a clustering analysis of all the genes to identify their clustering patterns [62]. The results showed that these genes exhibited different expression patterns; for example, a number of genes tended to be highly expressed in vegetative tissues, and many genes tended to be highly expressed in floral tissues (Fig. 4c). The 21,944 genes (with different

FPKM values in different tissues) were then grouped into six subclusters via the H-cluster algorithm. The results showed that these genes presented different expression patterns (Additional file 2: Fig. S2). Interestingly, the expression pattern of the genes in subcluster 1 (10,124 genes), whose expression was upregulated in the stamens, pistils, shoot apex, and flowers (the bracts, sepals, petals, stamens, and pistils combined) but downregulated in the leaves, was opposite that in subcluster 2 (2,289) (Additional file 2: Fig. S2). Interestingly, genes in subcluster 6 (296 genes) showed a strong tissue preference, exhibiting patterns of which their expression highly upregulated in the stamens and flowers (Additional file 2: Fig. S2).

To better understand tissue-specific genes in *L. chinense*, we investigated the distribution of tissue-specific genes in *L. chinense* and the pathways with which these genes were associated. Genes with less than 1 FPKM in any tissue, were considered not expressed in that tissue [63]. In our study, we detected a total of 2,126 tissue-specific genes in 7 tissues. Stamen tissue contained the highest number of tissue-specific genes (819, 38.71%), followed by the leaf (372, 17.58%), shoot apex (343, 16.21%), bract (263, 12.43%), pistil (225, 10.63%) and petal (52, 2.46%) tissues, while the sepal tissue contained the minimum number of tissue-specific genes (42, 1.98%) (Fig. 4b). The expression patterns of the 2,126 genes were tissue specific (Fig. 4d), and KEGG pathway enrichment analysis also revealed that these tissue-specific genes may play roles in different pathways. For example, bract-specific genes were significantly enriched in plant-pathogen interactions (Additional file 3: Fig. S3a). Interestingly, the tissue-specific genes in the shoot apex were enriched not only in plant-pathogen interactions, but also in monoterpenoid biosynthesis (Additional file 3: Fig. S3b). The petal-specific genes were associated with thiamine metabolism (Additional file 3: Fig. S3c), while the stamen-specific genes were significantly enriched in glycosphingolipid biosynthesis and starch and sucrose metabolism (Additional file 3: Fig. S3d). The pistil-specific genes were enriched in flavonoid biosynthesis (Additional file 3: Fig. S3e). Interestingly, the sepal-specific genes were also enriched in monoterpenoid biosynthesis (Additional file 3: Fig. S3f).

Gene co-expression analysis

Weighted correlation network analysis (WGCNA) was used for gene co-expression analysis [64]. In total, 13 co-expression modules were obtained, with the number of genes involved ranging from 173 (tan module) to 4,503 (turquoise module) (Fig. 5a, Additional file 14: Table S6). In the shoot apex, genes from the black module were highly expressed, and KEGG enrichment analysis revealed that these genes were associated mainly with plant-pathogen interactions and plant hormone signal transduction (Fig. 5b, Additional file 4: Fig. S4). Genes from the green-yellow module tended to be expressed in the bracts and shoot apex, and these genes also participated in plant-pathogen interactions (Additional file 4: Fig. S4, Fig. 5c). Genes from the magenta module, which were associated with glycolysis and the tricarboxylic acid (TCA) cycle, tended to be expressed in the petals and sepals (Fig. 5d, Additional file 4: Fig. S4), while genes from the pink module were highly expressed in the leaves, petals, and sepals and were involved in porphyrin and chlorophyll metabolism and photosynthesis (Fig. 5e, Additional file 4: Fig. S4). Genes belonging to the purple module were preferentially expressed in the petals and stamens (Additional file 4: Fig. S4). By using KEGG pathway enrichment analysis, we found that these genes played a role in

oxidative phosphorylation and the TCA cycle (Fig. 5f). In the pistils, the expression of tan module genes was very high, and these genes were related to indole alkaloid biosynthesis (Fig. 5g, Additional file 4: Fig. S4).

To determine the hub genes in the black, green-yellow, magenta, pink, purple, and tan modules, six algorithms, degree, edge percolated component (EPC), maximal clique centrality (MCC), closeness, radiality, and stress, in the CytoHubba package of Cytoscape software were used [65]. To improve reliability, only the genes ranked in the top 20 by all seven algorithms were considered hub genes. We ultimately identified 9, 8, 9, 5, 9, and 12 hub genes in the black, green-yellow, magenta, pink, purple, and tan modules, respectively (Additional file 5: Fig. S5, Additional file 15: Table S7). These hub genes play important roles in different pathways. For example, *Lchi02285* (which encodes the MADS-box protein JOINTLESS, black module) played a role in regulating transcription (Additional file 15: Table S7) [66]. *Novelgene6677* (which encodes the disease resistance protein RPM1, green-yellow module) had a substantial role in plant resistance (Additional file 15: Table S7) [67]. *Lchi18683* (which encodes plastocyanin, pink module) was indispensable for photosynthesis (Additional file 15: Table S7).

MIKC-type MADS-box genes in *L. chinense*

Using HMMER software and the BLASTP tool, we identified 10 MIKC-type MADS-box genes (7 known genes and 3 novel genes) in the *L. chinense* genome [68]. By comparing MIKC-type MADS-box genes in *A. thaliana* and *P. trichocarpa* with those in *L. chinense*, we determined the phylogenetic relationship among these 105 genes and conducted a neighbour-joining phylogenetic tree (Fig. 6a). Among these genes, *Lchi23168*, *Lchi01744* and *Novelgene12125* were classified as member of the AP3/PI subfamily, *Lchi02285* was classified as member of the SVP subfamily, *Lchi20361* were classified as member of the SEP subfamily, *Novelgene10568* were classified as member of the AGL12 subfamily, *Lchi01587* and *Lchi04024* were classified as member of the AG/SHP/STK subfamily, *Lchi17876* was classified as member of the SOC1 subfamily, and *Novelgene0718* was classified as member of the TM8 subfamily (Fig. 6a). We found that the number of MIKC-type MADS-box genes in *L. chinense* (red circle) was less than that in *A. thaliana* (yellow circle) and *P. trichocarpa* (blue circle), and no *L. chinense* MIKC-type MADS-box gene was classified as member of the AGL6, AGL17, AGL15, FLC or AP1/FUL subfamily (Fig. 6a). We believed that the poor quality of the genome annotation may limit identification of additional MIKC-type MADS-box genes in *L. chinense*, as three of the ten MIKC-type MADS-box genes were identified via hybrid sequencing, which indicated that the information of *L. chinense* genome annotation was incomplete.

By using hybrid sequencing, we determined the expression profiles of ten MIKC-type MADS-box *L. chinense* genes. *Lchi23168* and *Lchi01744* were B-type genes that were related to the determination of stamen and petal identity and were highly expressed in the stamens and petals (Fig. 6b). As an E-type gene involved in the determination of the identify of all floral organs, *Lchi20361* was highly expressed in the pistils, petals and stamens (Fig. 6b). Furthermore, *Lchi04024* was highly expressed in the pistils and stamens, *Lchi01587* was highly expressed in the pistils, and these two genes were C/D-type genes related

to the determination of carpel and stamen identity (Fig. 6b). Interestingly, we found that the expression levels of four MIKC-type MADS-box genes, *Lchi17876* (SOC1 subfamily), *Lchi02285* (SVP subfamily), *Novelgene0718* (TM8 subfamily) and *Novelgene10568* (AGL12 subfamily), were greater in vegetative tissues than in floral tissues (Fig. 6b). Notably, not all of the MIKC-type MADS-box genes were highly expressed only in floral organs, as some of them were also highly expressed in vegetative organs. For example, *AGL12* is highly expressed in the *A. thaliana* root meristem, and the expression levels of *SVP-like* genes in *Prunus mume* and *SOC1-like* genes in *P. bretschneideri* are greater in vegetative organs than in floral organs [69-71].

RT-qPCR validation

To validate the accuracy of the gene expression levels detected by RNA sequencing, we performed RT-qPCR of twenty randomly selected DEGs. As shown in Additional file 6: Fig. S6 and Additional file 7: Fig. S7, the expression levels of the ten DEGs measured by Illumina sequencing and RT-qPCR were correlated ($R^2 = 0.947$, p -value < 0.01) despite some differences in transcript abundance. These results indicated that the gene expression levels detected by Illumina sequencing were reliable.

Discussion

L. chinense, an economically and ecologically important tree species is continually being studied. From the study of the relationship between the geographical distribution of *L. chinense* and climate to the study of the characteristics and development of its different tissues, research on *L. chinense* is intensifying [1, 5, 6]. Especially since the genome of *L. chinense* was published, the understanding of the position of *L. chinense* in evolutionary history has improved⁷. However, the quality of the genome sequencing data is unsatisfactory, the *L. chinense* samples studied have been limited to only a few tissues, and no full-length transcriptome of *L. chinense* has been reported. Recently, SMRT sequencing combined with NGS has been used to improve the quality of genomes, discover fusion genes and lncRNAs, and determine the full-length transcriptome of many species, such as moso bamboo, *S. miltiorrhiza* and *Populus* spp. [25, 29, 72]. For these purposes, we used hybrid sequencing to obtain a reference transcriptome of *L. chinense* from 21 samples. We detected 13,139 novel genes, 5,532 TFs from 95 families, 7,527 lncRNAs and 887 fusion genes (Fig. 1g, Additional file 12: Table S4, Additional file 13: Table S5). Moreover, we also found that the lncRNAs were shorter and smaller than the mRNAs in terms of length and exon number (Fig. 2c and d). This phenomenon has also been observed in humans, *Sus scrofa*, *Populus* and *Gossypium hirsutum* [30, 34, 72-74]. In addition, in the study of maize and red clover, fusion events tended to arise inter-chromosomally rather than intra-chromosomally [22, 24]. In *L. chinense*, it is difficult to determine the regularity of the occurrence of fusion events at the chromosomal level because the genome has not yet been assembled at this level. However, these results still improve our understanding of the transcriptome of *L. chinense*.

Notably, our transcriptome analysis not only included the discovery of lncRNAs, TFs, and fusion genes but also included the discovery of AS and APA. AS and APA are important post-transcriptional regulatory mechanisms that can increase the complexity and flexibility of the proteome and transcriptome [75]. In maize and sorghum, AS events occur in 45% and 38.5% of the genes, respectively [32]. In our study, 17.6% (8,503/48,408) of alternatively spliced genes were detected, and RI genes were the most frequent alternatively spliced genes (Fig. 3a). This result was consistent with the findings in sorghum, red clover, and maize [22-24]. We suggest that these alternatively spliced genes may play important roles in maintaining homeostasis between transcripts and proteins in *L. chinense*. Because AS can cause alternatively spliced genes to produce different AS variants, some AS variants can encode novel proteins, some of which are degraded in different ways before they encode proteins [59, 76]. Moreover, APA also plays an important regulatory role in maintaining RNA stability, ensuring accurate RNA localization and translation, and in plant development especially flowering [60, 77, 78]. In *A. thaliana*, approximately 60% of the genes have multiple poly(A) sites [60]. In *L. chinense*, 13.0% (6,271/48,408) of the genes had more than one poly(A) site (Fig. 3b). As far as we know, the number and source of sequencing samples, sequencing depth, and analysis methods have an important impact on the analysis results of AS and APA, so it was not uncommon for the proportion of alternatively spliced genes and APA genes in *L. chinense* to be low [60, 61, 79]. Moreover, we found that the distribution of nucleotides upstream and downstream of the APA cleavage sites was consistent with the reports of previous poly(A) analyses in sorghum and red clover (Fig. 3d) [23, 24]. We also found that a motif (AUAAA) upstream of the poly(A) cleavage sites was present in maize, red clover, sorghum and other plant species (Fig. 3c) [22-24]. This indicated that the motif was conserved. These findings indicated that AS and APA made great contributions to complexity and flexibility of the *L. chinense* transcriptome.

The discovery of DEGs was important for the study of plant development, metabolic pathways and the stress response. Researchers identified 3,118 DEGs that were related to *L. chinense* leaf shape development [6]. In addition, 962 DEGs were discovered to be associated with the alkaline stress response in rice [19]. By analysing the DEGs in the root, researchers revealed the pathway of tanshinone biosynthesis in *S. miltiorrhiza* [29]. Tissue-specific genes, play important roles in plant defence, stress response, plant development and material metabolism [41, 42, 80-83]. In *Ferula asafetida*, genes involved in terpenoid and phenylpropanoid metabolism tended to be expressed in the flowers [83]. In the study of tomato (*Solanum pennellii*), researchers concluded that tissue-specific genes play important regulatory roles during fruit development [41]. In addition, tissue-specific genes can enhance *Petunia hybrida* tolerance under salt conditions [81]. In our study, we identified 3,032 DEGs between floral tissues and vegetative tissues and detected 2,116 tissue-specific genes in *L. chinense* (Fig. 4a and b). Interestingly, we found that bract-specific genes were enriched in plant-pathogen interactions (Additional file 3: Fig. S3a). According to a previous study, bracts have multiple functions, such as rain protection, pollinator attraction, and photosynthetic surfaces [84]. However, there are few reports on plant-pathogen interactions in bracts. We also found that the stamen-specific genes were enriched in starch and sucrose metabolism (Additional file 3: Fig. S3d); it is easy to understand that stamen development is

accompanied by starch accumulation and degradation [85]. Shoot-apex-specific genes are involved in plant hormone signal transduction (Additional file 3: Fig. S3b). The shoot apex contains different primordia, and phytohormones participate in the regulation of primordial development. These results showed that tissue-specific genes and DEGs may play roles in the functional diversity of tissues in *L. chinense*. Overall, these findings greatly increase our understanding of tissue-specific genes and DEGs in *L. chinense*.

We were interested not only in DEGs and tissue-specific genes but also in co-expressed genes. Gene co-expression has an important impact on biological traits and phenotypic characteristics and has basic functions in plant growth, development, and environmental adaptability [86-88]. Co-expressed genes play important roles in the growth and development of moso bamboo [38]. In addition, co-expression of *salt overly sensitive 1 (SOS1)* and *PM-localized H⁺-ATPase (AHA1)* from *Sesuvium portulacastrum* in transgenic *A. thaliana* can enhance salinity resistance in transgenic plants [88]. By analyzing the co-expression of genes, we obtained additional data and provided new insight into the functions of organs in *L. chinense*. In our study, we identified 13 co-expression modules, and the genes in these modules participated in different pathways (Fig. 5, Additional file 4: Fig. S4). KEGG pathway enrichment analysis revealed that the genes from the green-yellow module were highly expressed in the shoot apex and bracts and were involved in plant-pathogen interactions (Fig. 5c, Additional file 4: Fig. S4). In pineapple, genes co-expressed in bracts participate in plant-pathogen interactions and these genes may play roles in plant defence [43]. Therefore, we can infer that the bract and shoot apex may function in plant defence. We also found that leaves, petals, and sepals were the main organs for photosynthesis in *L. chinense* because the genes from the pink module were highly expressed in these organs, and those genes participated mainly in photosynthesis (Fig. 5e, Additional file 4: Fig. S4). These findings indicated that the function of different *L. chinense* tissues was the result of gene co-expression.

To the best of our knowledge, MIKC-type MADS-box genes include most ABCDE-type genes that play pivotal roles in flower development regulation and the determination of floral organ identity [89, 90]. Many MIKC-type MADS-box genes have been identified in different species [10, 91-93]. In this study, we identified 10 MIKC-type MADS-box genes from 7 subfamilies (Fig. 6a). Phylogenetic analysis revealed that these ten genes were highly conserved evolutionarily. However, no AP1/FUL (A-type genes) subfamily genes were identified in this study, and the number of MIKC-type MADS-box genes was lower in *L. chinense* less than in *A. thaliana*, *P. trichocarpa* and *Amborella trichopoda* [94]. In actuality, *L. chinense* originated earlier than *A. thaliana* and *P. trichocarpa* but later than *A. trichopoda*, and in angiosperms, the number of ABCE-type genes has increased [7, 94]. Theoretically, the number of A-type genes in *L. chinense* should be greater than zero. We believed that the poor quality of the *L. chinense* genome annotation prevented us from finding more MIKC-type MADS-box genes. Moreover, MIKC-type MADS-box genes function not only in reproductive organ development and flowering regulation but also in

vegetative organ growth and the stress response. For example, *OsMADS26* (AGL12 subfamily) is related to stress response, *AGL12* regulates the proliferation of root meristem cells in *A. thaliana*, and *SVP-like* genes control dormancy and budbreak in apple [69, 95, 96]. Similarly, among the 10 MIKC-type MADS-box genes, 6 genes were highly expressed in reproductive tissues, and 4 genes were highly expressed in vegetative tissues (Fig. 6b). These findings indicated that *L. chinense* MIKC-type MADS-box genes might be related to the development and growth regulation of floral and vegetative organs.

Conclusions

In this work, 13,139 novel genes, 5,532 TFs, 7,527 lncRNAs, 887 fusion transcripts, 8,503 alternatively spliced genes, and 11,108 genes with APA sites were identified via hybrid sequencing. On the basis of hybrid sequencing data, we used WGCNA to identify 2,126 tissue-specific genes and 13 gene co-expression modules. KEGG pathway enrichment analysis further revealed that tissue-specific genes and co-expressed genes functioned in different pathways. We also identified 52 hub genes from six co-expression modules of interest, and these genes played important roles in specific pathways. Moreover, we identified 10 MIKC-type MADS-box genes that might be related to the development and growth regulation of floral and vegetative organs. These findings will support future research on the differentiation of tissue functions in *L. chinense*.

Methods

Plant materials and total RNA isolation

Twenty-six-year-old *L. chinense* trees in a provenance trial plantation in Xiashu, Jurong County, Jiangsu Province (119°13'E, 32°7'N), were used as materials. The provenance of sample trees in Lushan, Jiangxi Province (116°0'E, 29°32'N) (Specimen No. 20010020016, deposited in the specimen room of Nanjing Forestry University). Plant samples were collected from the shoot apices, expanded flower buds and mature leaves (Additional file 8: Fig. S8a, g and h). The flower bud samples were divided into bracts, sepals, petals, stamens and pistils (Additional file 8: Fig. S8 b – f). Each sample included three replicates. Twenty-one samples were frozen in liquid nitrogen and then stored at -80 °C.

An RNAPrep Pure Plant Kit (Tiangen, China) was used to extract total RNA according to the detailed steps shown in the product description. The concentration and integrity of the total RNA were detected by a NanoDrop 2000 c spectrophotometer (Thermo Scientific, USA) and an Agilent 2100 TapeStation instrument (Agilent Technologies, USA).

PacBio library construction and sequencing

Twenty-one samples were pooled together with equimolar ratios. One microgram of total RNA was used to synthesize cDNA via SMARTer™ PCR cDNA Synthesis Kit (TaKaRa, Japan). After PCR amplification, we used a portion of cDNA for size fractionation via a BluePippin™ Size Selection System (Sage Science, MA) to enrich fragments that were longer than 4 kb. Two complete SMRT bell libraries (≤ 4 kb and > 4 kb)

were then constructed after full-length cDNA amplification, damage repair, end repair, ligation of SMRT linkers, exonuclease digestion, application of binding primers, and application of binding DNA polymerase. Sequencing was then performed on the PacBio Sequel platform.

Illumina sequencing and quality control

Twenty-one samples from 7 tissues (bracts, sepals, petals, stamens, pistils, leaves and the shoot apex) were sequenced on the Illumina HiSeq 2500 platform. For each sample, three micrograms of total RNA were used for library construction with a NEBNext® Ultra™ RNA Library Prep Kit. In-house Perl scripts were used to process the raw reads. We then obtained clean data after removing the reads with adaptors, reads containing poly(-N) sequences, and low-quality reads from the raw data.

PacBio data processing

SMRT Link 6.0 software was used to produce effective subreads (length ≥ 200 , read score ≥ 0.75). The CCSs were then selected from the subreads (passes ≥ 1 , predicted accuracy ≥ 0.8). The pbclassify.py script was subsequently used to classify the CCSs into full-length and non-full-length reads (ignoring false poly(A) results, sequencing length < 200). Each CCS had a poly(A) tail and 5' and 3' adaptors and was considered an FLNC read. The full-length and non-full-length reads were then used to perform ICE, followed by a final polishing by Arrow. Polished consensus reads were ultimately produced. The Illumina sequencing data were used to correct additional nucleotide errors in the consensus reads with LoRDEC software [44].

PacBio Iso-Seq data analysis

GMAP software was applied to align the polished consensus reads (long reads) to the reference genome [45]. The polished reads were classified into five categories: (a) unmapped, (b) multiply mapped, (c) uniquely mapped, (d) mapped to '+' sequences (sense sequences of the genome), and (e) mapped to '-' sequences (antisense sequences of the genome). According to the mapping results, reads that were aligned to the unannotated regions of the reference genome were defined as novel genes. Seven databases, the NT, GO, Pfam, KEGG, NR, KOG, and Swiss-Prot databases, were used to annotate the functions of unmapped transcripts and novel genes [46-51]. Moreover, we investigated the number of fusion transcripts in the PacBio data. The fusion transcripts had to meet the following requirements: (a) a full-length transcript mapped to two or more gene loci in the reference genome; (b) each locus must cover 10% of the transcript; (c) the total coverage of the transcript with respect to the reference genome must be more than 99%; (d) each locus must be separated by more than 100 kb in the reference genome; and (e) the gene loci must be supported by at least two NGS reads.

We used iTAK software to predict TFs and used four additional tools (PLEK, CNCI, CPC, and the Pfam database) to predict lncRNAs, with the default parameters [47, 53-56]. Transcripts with no potential coding sequence were retained as our set of candidate lncRNAs.

SUPPA software (<https://bitbucket.org/regulatorygenomicsupf/suppa>) was used for AS analysis with the default parameters. SUPPA classifies AS events into 7 categories: SEs, MXs, A5s, A3s, RIs, AFs, ALs. We then used the TAPIS pipeline to predict APA sites [23], and we used MEME to analyse the nucleotide composition of the sequences upstream (- 50 nt) and downstream (+ 50 nt) of all the APA sites for nucleotide bias [97].

Differential expression and GO/KEGG enrichment analyses

The number of reads that mapped to each gene was calculated by Cuffdiff software [98]. The FPKM value of each gene was calculated on the basis of the read count mapped to the gene and the length of the gene.

DEseq software (<http://www.bioconductor.org/packages/release/bioc/html/DESeq.html>) was used for differential expression analysis. Genes with adjusted p-values that were less than 0.05 according to DEGseq were defined as differentially expressed. GO enrichment analysis was performed by Goseq software ([http://www.bioconductor.org/packages/release/bioc/html/goseq.h-](http://www.bioconductor.org/packages/release/bioc/html/goseq.html)

[tml](http://www.bioconductor.org/packages/release/bioc/html/goseq.html)), and we used KOBAS software (<http://kobas.cbi.pku.edu.cn/download.php>) to perform KEGG pathway enrichment analysis. Different modules were obtained by pruning the gene expression network.

Gene co-expression network analysis and identification of hub genes

Gene co-expression analysis was performed via WGCNA [64]. First, on the basis of the expression level of each gene in the different samples, the correlation between every pair genes was calculated. The calculated results were then used to construct a gene expression network, and the gene expression network map was subsequently pruned to obtain different modules. The CytoHubba package in Cytoscape software (<http://www.cytoscape.org/download.html>) was used to identify hub genes in the different modules [65]. We used the degree, EPC, MCC, closeness, radiality, and stress algorithms to select genes ranked in the top 20 by all algorithms as hub genes.

Identification and phylogenetic analysis of MIKC-type MADS-box genes

We used the amino acid sequences of *A. thaliana* MIKC-type MADS-box genes as queries to against the *L. chinense* genome database via the BLASTP algorithm [7, 10]. We then used HMMER software to search genes with SRF-TF (PF00319) and K-box (PF01486) domains on the basis of hidden Markov model (HMM) [68]. The genes identified in the results of the above two analyses were further identified via the Pfam database online tool (<http://pfam.xfam.org/>).

Proteins of MIKC-type MADS-box genes from *A. thaliana*, *P. trichocarpa* and *L. chinense* were used to construct phylogenetic tree [10, 93]. We used MEGA 6 software and EVOLVIEW online software (<https://evolgenius.info/evolview-v2/>) to construct phylogenetic tree based on neighbour-joining algorithm, and the number of bootstrap replications was set to 1,000 [99].

RT-qPCR validation

We used RT-qPCR to validate the gene expression levels detected by Illumina sequencing. In each sample, one microgram of total RNA was used to synthesize cDNA by using PrimeScript™ RT Master Mix (TaKaRa, Japan). The Oligo 7 algorithm (<https://en.freedownloadmanager.org/Wi-ndows-PC/OLIGO.html>) was then used to design RT-qPCR primers for the twenty selected DEGs (Additional file 16: Table S8). RT-qPCR was performed on a StepOnePlus™ system (Applied Biosystems) with SYBR® *Premix Ex Taq*™ (TaKaRa, Japan) as directed by the instructions. The internal control gene used was *eukaryotic translation initiation factor 3 (eIF3)* [100].

Abbreviations

A3s: alternative 3' splice sites; A5s: alternative 5' splice sites; AFs: alternative first exons; ALs: alternative last exons; AS: alternative splicing; APA: alternative polyadenylation; AP2: APETALA2; BLAST: Basic local alignment search tool; CCSs: circular consensus sequences; DEGs: differentially expressed genes; EPC: edge percolated component; FLC: Flowering Locus C; FLNC: full-length non-chimeric; FUL: FRUITFULL; GO: Gene Ontology; ICE: iterative clustering for error; KEGG: Kyoto Encyclopedia of Genes and Genomes; KOG: EuKaryotic Orthologous Groups; lncRNAs: long non-coding RNAs; MCC: maximal clique centrality; MXs: mutually exclusive exons; NGS: next-generation sequencing; NR: non-redundant; NT: nucleotide; Pfam: Protein Family; PI: PISTILLATA; RIs: retained introns; SEP: SEPALLATA; SEs: skipped exons; SHP1: SHATTERPROOF1; SMRT: single-molecule long-read sequencing technology; SOC1: Suppressor of Overexpression Of Constans1; STK: SEEDSTICK; SVP: Short Vegetative Phase; TCA: tricarboxylic acid; TFs: transcription factors; WGCNA: weighted correlation network analysis;

Declarations

Ethics approval and consent to participate: We confirm that the material collection presented here were conducted in accordance with the wild plant care regulations and natural reserves regulations set forth by the Decree of the state council of the People's Republic of China.

Consent for publication: Not applicable.

Availability of data and materials: All the raw data in this study have been uploaded to the public database of National Center for Biotechnology Information under PRJNA559687 (Release date:2019-08-12). The experimental materials were collected from a trial plantation (belonging to Nanjing Forestry University) in Xiashu, Jurong County, Jiangsu Province (119°13'E,32°7'N).

Competing interests: The authors declare that they have no competing interests.

Funding: The design of the study, sample collection, NGS and PacBio sequencing, data analysis and interpretation were supported by the National Natural Science Foundation of China (31770718,

31470660) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Authors' contributions: Experimental design: ZT and HL; plant material collection and performing of the experiments: ZT, YS, HL and SW; data analysis: ZT, YS and LW; manuscript writing: ZT and HL. All authors have read and approved the final manuscript.

Acknowledgements: We wish to thank Meng Xu (Nanjing Forestry University, Nanjing, China) for suggestions concerning the data analysis and writing. The authors would also like to thank Novogene Company (Tianjin, China) for RNA sequencing. In addition, the authors would also like to thank other colleagues in the laboratory for their help in data analysis and writing, they are Ziyuan Hao, Weiping Zhong, Lichun Yang, Xinyu Zhai, Xujia Wu, Shan Hu, Kang Xu, Shenghua Zhu and Longjie Ni.

Supplementary Information

Additional file 1: Fig. S1. GO and KEGG enrichment analyses of DEGs between floral tissues and vegetative tissues.

Additional file 2: Fig. S2. Gene clustering results.

Additional file 3: Fig. S3. KEGG enrichment analysis of tissue-specific genes in bracts, shoot apex, petals, stamens, pistils, and sepals.

Additional file 4: Fig. S4. Heat map of the expression of genes from 13 modules in 21 samples.

Additional file 5: Fig. S5. Co-expression network analysis of hub genes from six co-expression modules.

Additional file 6: Fig. S6. Correlation between Illumina sequencing and RT-qPCR results ($R^2 = 0.9647$, $p < 0.01$).

Additional file 7: Fig. S7. Comparison of the expression results of 20 DEGs between the RT-qPCR and RNA-seq data. BR: bracts. SA: shoot apexes. PE: petals. PI: pistils. SE: sepals. ST: stamens.

Additional file 8: Fig. S8. *L. chinense* materials used for sequencing.

Additional file 9: Table S1 Information about the PacBio sequencing data.

Additional file 10: Table S2 Transcript lengths before and after correction by Illumina reads.

Additional file 11: Table S3 Statistics of the GMAP mapping results.

Additional file 12: Table S4 Transcription factor families detected in the PacBio Iso-Seq data.

Additional file 13: Table S5 Statistics of the fusion transcripts.

Additional file 14: Table S6 Numbers of genes in 13 modules.

Additional file 15: Table S7 Hub genes of the black, green-yellow, magenta, pink, purple, tan, blue, green, red, and turquoise modules.

Additional file 16: Table S8 Primers used for RT-qPCR validation.

References

1. Xu X, Zhang H, Xie T, Xu Y, Zhao L, Tian W: **Effects of Climate Change on the Potentially Suitable Climatic Geographical Range of *Liriodendron chinense***. *Forests* 2017, **8**(10):399-412.
2. Zhang WW, Niu JF, Wang XK, Tian Y, Yao FF, Feng ZZ: **Effects of ozone exposure on growth and photosynthesis of the seedlings of *Liriodendron chinense* (Hemsl.) Sarg, a native tree species of subtropical China**. *Photosynthetica* 2011, **49**(1):29-36.
3. Hao R, He S, Tang S, Wu S: **Geographical distribution of *Liriodendron chinense* in China and its significance**. *J Plant Resour and Environ* 1995, **4**(4):1-6.
4. Li K, Chen L, Feng Y, Yao J, Li B, Xu M, Li H: **High genetic diversity but limited gene flow among remnant and fragmented natural populations of *Liriodendron chinense* Sarg**. *Biochem Syst Ecol* 2014, **54**:230-236.
5. Yang Y, Xu M, Luo Q, Wang J, Li H: **De novo transcriptome analysis of *Liriodendron chinense* petals and leaves by Illumina sequencing**. *Gene* 2014, **534**(2):155-162.
6. Ma J, Wei L, Li J, Li H: **The Analysis of Genes and Phytohormone Metabolic Pathways Associated with Leaf Shape Development in *Liriodendron chinense* via De Novo Transcriptome Sequencing**. *Genes (Basel)* 2018, **9**(12):577-592.
7. Chen J, Hao Z, Guang X, Zhao C, Wang P, Xue L, Zhu Q, Yang L, Sheng Y, Zhou Y *et al*: ***Liriodendron* genome sheds light on angiosperm phylogeny and species-pair differentiation**. *Nat Plants* 2018, **5**:18-25.
8. Feng X, Jia Y, Zhu R, Chen K, Chen Y: **Characterization and analysis of the transcriptome in *Gymnocypris selincuoensis* on the Qinghai-Tibetan Plateau using single-molecule long-read sequencing and RNA-seq**. *DNA Res* 2019, **0**(0):1-11.
9. Jeong HB, Kang MY, Jung A, Han K, Lee JH, Jo J, Lee HY, An JW, Kim S, Kang BC: **Single-molecule real-time sequencing reveals diverse allelic variations in carotenoid biosynthetic genes in pepper (*Capsicum spp.*)**. *Plant Biotechnol J* 2019, **17**(6):1081-1093.
10. Parenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B *et al*: **Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world**. *Plant Cell* 2003, **15**(7):1538-1551.
11. Silva CS, Puranik S, Round A, Brennich M, Jourdain A, Parcy F, Hugouvieux V, Zubieta C: **Evolution of the Plant Reproduction Master Regulators LFY and the MADS Transcription Factors: The Role of Protein Structure in the Evolutionary Development of the Flower**. *Front Plant Sci* 2015, **6**:1193.

12. Bowman JL, Smyth DR, Meyerowitz EM: **Genetic interactions among floral homeotic genes of *Arabidopsis***. *Development* 1991, **112**:1-20.
13. Coen ES, Meyerowitz EM: **The war of the whorls: genetic interactions controlling flower development**. *Nature* 1991, **353**:31-37.
14. Theißen G, Saedler H: **Floral quartets**. *Nature* 2001, **409**:469-471.
15. Searle I, He Y, Turck F, Vincent C, Fornara F, Krober S, Amasino RA, Coupland G: **The transcription factor FLC confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis***. *Genes Dev* 2006, **20**(7):898-912.
16. Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH: **Role of SVP in the control of flowering time by ambient temperature in *Arabidopsis***. *Genes Dev* 2007, **21**(4):397-402.
17. Shelley R, Hepworth, Federico Valverde, Dean Ravenscroft, Aidyn Mouradov, Coupland G: **Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs**. *Embo J* 2002, **16**(12):4327-4337.
18. Yu Y, Shi J, Li X, Liu J, Geng Q, Shi H, Ke Y, Sun Q: **Transcriptome analysis reveals the molecular mechanisms of the defense response to gray leaf spot disease in maize**. *BMC Genomics* 2018, **19**(1):742-758.
19. Li N, Liu H, Sun J, Zheng H, Wang J, Yang L, Zhao H, Zou D: **Transcriptome analysis of two contrasting rice cultivars during alkaline stress**. *Sci Rep* 2018, **8**(1):9586-9601.
20. Han Y, Wang X, Zhao F, Gao S, Wei A, Chen Z, Liu N, Zhang Z, Du S: **Transcriptomic analysis of differentially expressed genes in flower-buds of genetic male sterile and wild type cucumber by RNA sequencing**. *Physiol Mol Biol Plants* 2018, **24**(3):359-367.
21. Chen C, Zheng Y, Zhong Y, Wu Y, Li Z, Xu LA, Xu M: **Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora***. *BMC Genomics* 2018, **19**(1):550-564.
22. Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Z, Olson A, Stein JC, Ware D: **Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing**. *Nat Commun* 2016, **7**:11708-11720.
23. Abdel-Ghany SE, Hamilton M, Jacobi JL, Ngam P, Devitt N, Schilkey F, Ben-Hur A, Reddy AS: **A survey of the sorghum transcriptome using single-molecule long reads**. *Nat Commun* 2016, **7**:11706-11716.
24. Chao Y, Yuan J, Li S, Jia S, Han L, Xu L: **Analysis of transcripts and splice isoforms in red clover (*Trifolium pratense* L.) by single-molecule long-read sequencing**. *BMC Plant Biol* 2018, **18**(1):300.
25. Wang T, Wang H, Cai D, Gao Y, Zhang H, Wang Y, Lin C, Ma L, Gu L: **Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*)**. *Plant J* 2017, **91**(4):684-699.
26. Zhang J, Liu C, He M, Xiang Z, Yin Y, Liu S, Zhuang Z: **A full-length transcriptome of *Sepia esculenta* using a combination of single-molecule long-read (SMRT) and Illumina sequencing**. *Marine Genomics* 2019, **43**:54-57.

27. Chao Y, Yuan J, Guo T, Xu L, Mu Z, Han L: **Analysis of transcripts and splice isoforms in *Medicago sativa* L. by single-molecule long-read sequencing.** *Plant Mol Biol* 2019, **99**(3):219-235.
28. Chen SY, Deng F, Jia X, Li C, Lai SJ: **A transcriptome atlas of rabbit revealed by PacBio single-molecule long-read sequencing.** *Sci Rep* 2017, **7**(1):7648-7657.
29. Xu Z, Peters RJ, Weirather J, Luo H, Liao B, Zhang X, Zhu Y, Ji A, Zhang B, Hu S *et al.*: **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.** *Plant J* 2015, **82**(6):951-961.
30. Li Y, Fang C, Fu Y, Hu A, Li C, Zou C, Li X, Zhao S, Zhang C, Li C: **A survey of transcriptome complexity in *Sus scrofa* using single-molecule long-read sequencing.** *DNA Res* 2018, **25**(4):421-437.
31. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C: **The Third Revolution in Sequencing Technology.** *Trends Genet* 2018, **34**(9):666-681.
32. Wang B, Regulski M, Tseng E, Olson A, Goodwin S, McCombie WR, Ware D: **A comparative transcriptional landscape of maize and sorghum obtained by single-molecule sequencing.** *Genome Res* 2018, **28**(6):921-932.
33. Wang M, Wang P, Liang F, Ye Z, Li J, Shen C, Pei L, Wang F, Hu J, Tu L *et al.*: **A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation.** *New Phytol* 2018, **217**(1):163-178.
34. Liu S, Sun Z, Xu M: **Identification and characterization of long non-coding RNAs involved in the formation and development of poplar adventitious roots.** *Ind Crops Prod* 2018, **118**:334-346.
35. Zhao H, Gao Z, Wang L, Wang J, Wang S, Fei B, Chen C, Shi C, Liu X, Zhang H *et al.*: **Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*).** *Gigascience* 2018, **7**(10).
36. Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ: **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* 2017, **18**(1):395-416.
37. Li M, Dunwell JM, Zhang H, Wei S, Li Y, Wu J, Zhang S: **Network analysis reveals the co-expression of sugar and aroma genes in the Chinese white pear (*Pyrus bretschneider*).** *Gene* 2018, **677**:370-377.
38. Ma X, Zhao H, Xu W, You Q, Yan H, Gao Z, Su Z: **Co-expression Gene Network Analysis and Functional Module Identification in Bamboo Growth and Development.** *Front Genet* 2018, **9**:574-588.
39. Zheng J, He C, Qin Y, Lin G, Park WD, Sun M, Li J, Lu X, Zhang C, Yeh CT *et al.*: **Co-expression analysis aids in the identification of genes in the cuticular wax pathway in maize.** *Plant J* 2019, **97**(3):530-542.
40. Zou X, Liu A, Zhang Z, Ge Q, Fan S, Gong W, Li J, Gong J, Shi Y, Tian B *et al.*: **Co-Expression Network Analysis and Hub Gene Selection for High-Quality Fiber in Upland Cotton (*Gossypium hirsutum*) Using RNA Sequencing Analysis.** *Genes (Basel)* 2019, **10**(2):119-141.

41. Pattison RJ, Csukasi F, Zheng Y, Fei Z, van der Knaap E, Catala C: **Comprehensive Tissue-Specific Transcriptome Analysis Reveals Distinct Regulatory Programs during Early Tomato Fruit Development.** *Plant Physiol* 2015, **168**(4):1684-1701.
42. Yu L, Ma J, Niu Z, Bai X, Lei W, Shao X, Chen N, Zhou F, Wan D: **Tissue-Specific Transcriptome Analysis Reveals Multiple Responses to Salt Stress in *Populus euphratica* Seedlings.** *Genes (Basel)* 2017, **8**(12):372-385.
43. Mao Q, Chen C, Xie T, Luan A, Liu C, He Y: **Comprehensive tissue-specific transcriptome profiling of pineapple (*Ananas comosus*) and building an eFP-browser for further study.** *PeerJ* 2018, **6**:e6028-6043.
44. Salmela L, Rivals E: **LoRDEC: accurate and efficient long read error correction.** *Bioinformatics* 2014, **30**(24):3506-3514.
45. Wu TD, Watanabe CK: **GMAP: a genomic mapping and alignment program for mRNA and EST sequences.** *Bioinformatics* 2005, **21**(9):1859-1875.
46. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35**(Database issue):D61-65.
47. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: **The Pfam protein families database: towards a more sustainable future.** *Nucleic Acids Res* 2016, **44**(D1):D279-285.
48. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN *et al*: **The COG database: an updated version includes eukaryotes.** *BMC Bioinf* 2003, **4**:41-54.
49. The UniProt C: **UniProt: the universal protein knowledgebase.** *Nucleic Acids Res* 2017, **45**(D1):D158-169.
50. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 1999, **28**(1):27-30.
51. Young MD, Wakefield MJ, Smyth GK, Oshlack A: **Gene ontology analysis for RNA-seq: accounting for selection bias.** *Genome Biol* 2010, **11**:R14-25.
52. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG *et al*: **The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression.** *Genome Res* 2012, **22**(9):1775-1789.
53. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ *et al*: **iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases.** *Mol Plant* 2016, **9**(12):1667-1670.
54. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y: **Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts.** *Nucleic Acids Res* 2013, **41**(17):e166-173.

55. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G: **CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine.** *Nucleic Acids Res* 2007, **35**(Web Server issue):W345-349.
56. Li. A, Zhang. J, Zhou. Z: **PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme.** *BMC Bioinf* 2014, **15**:311-321.
57. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S *et al.* **GENCODE: the reference human genome annotation for The ENCODE Project.** *Genome Res* 2012, **22**(9):1760-1774.
58. Zhang M, Liu YH, Chang CS, Zhi H, Wang S, Xu W, Smith CW, Zhang HB: **Quantification of gene expression while taking into account RNA alternative splicing.** *Genomics* 2018, **111**(6):1517-1528.
59. Szakonyi D, Duque P: **Alternative Splicing as a Regulator of Early Plant Development.** *Front Plant Sci* 2018, **9**:1174-1182.
60. Shen Y, Venu RC, Nobuta K, Wu X, Notibala V, Demirci C, Meyers BC, Wang GL, Ji G, Li QQ: **Transcriptome dynamics through alternative polyadenylation in developmental and environmental responses in plants revealed by deep sequencing.** *Genome Res* 2011, **21**(9):1478-1486.
61. Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG: **Genome-wide landscape of polyadenylation in *Arabidopsis* provides evidence for extensive alternative polyadenylation.** *Proc Natl Acad Sci USA* 2011, **108**(30):12533-12538.
62. Yang L, Jin Y, Huang W, Sun Q, Liu F, Huang X: **Full-length transcriptome sequences of ephemeral plant *Arabidopsis pumila* provides insight into gene expression dynamics during continuous salt stress.** *BMC Genomics* 2018, **19**(1):717-730.
63. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**(7):621-628.
64. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinf* 2008, **9**:559-571.
65. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T: **Cytoscape 2.8: new features for data integration and network visualization.** *Bioinformatics* 2011, **27**(3):431-432.
66. Nakano T, Kato H, Shima Y, Ito Y: **Apple SVP Family MADS-Box Proteins and the Tomato Pedicel Abscission Zone Regulator JOINTLESS have Similar Molecular Activities.** *Plant Cell Physiol* 2015, **56**(6):1097-1106.
67. Tornero P, Chao RA, Luthin WN, Goff SA, Dangl JL: **Large-scale structure-function analysis of the *Arabidopsis* RPM1 disease resistance protein.** *Plant Cell* 2002, **14**(2):435-450.
68. Finn RD, Clements J, Eddy SR: **HMMER web server: interactive sequence similarity searching.** *Nucleic Acids Res* 2011, **39**(Web Server issue):W29-37.
69. Tapia-Lopez R, Garcia-Ponce B, Dubrovsky JG, Garay-Arroyo A, Perez-Ruiz RV, Kim SH, Acevedo F, Pelaz S, Alvarez-Buylla ER: **An AGAMOUS-related MADS-box gene, XAL1 (AGL12), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*.** *Plant Physiol* 2008, **146**(3):1182-1192.

70. Li Y, Zhou Y, Yang W, Cheng T, Wang J, Zhang Q: **Isolation and functional characterization of SVP-like genes in *Prunus mume***. *Scientia Horticulturae* 2017, **215**:91-101.
71. Liu Z, Wu X, Cheng M, Xie Z, Xiong C, Zhang S, Wu J, Wang P: **Identification and functional characterization of SOC1-like genes in *Pyrus bretschneideri***. *Genomics* 2020, **112**(2):1622-1632.
72. Chao Q, Gao ZF, Zhang D, Zhao BG, Dong FQ, Fu CX, Liu LJ, Wang BC: **The developmental dynamics of the *Populus* stem transcriptome**. *Plant Biotechnol J* 2019, **17**(1):206-219.
73. Sharon D, Tilgner H, Grubert F, Snyder M: **A single-molecule long-read survey of the human transcriptome**. *Nat Biotechnol* 2013, **31**(11):1009-1014.
74. Li Y, Qin T, Dong N, Wei C, Zhang Y, Sun R, Dong T, Chen Q, Zhou R, Wang Q: **Integrative Analysis of the lncRNA and mRNA Transcriptome Revealed Genes and Pathways Potentially Involved in the Anther Abortion of Cotton (*Gossypium hirsutum* L.)**. *Genes (Basel)* 2019, **10**(12).
75. Li Y, Dai C, Hu C, Liu Z, Kang C: **Global identification of alternative splicing via comparative analysis of SMRT- and Illumina-based RNA-seq in strawberry**. *Plant J* 2017, **90**(1):164-176.
76. Turner RE, Pattison AD, Beilharz TH: **Alternative polyadenylation in the regulation and dysregulation of gene expression**. *Semin Cell Dev Biol* 2018, **75**:61-69.
77. Zhang Y, Carrion SA, Zhang Y, Zhang X, Zinski AL, Michal JJ, Jiang Z: **Alternative polyadenylation analysis in animals and plants: newly developed strategies for profiling, processing and validation**. *Int J Biol Sci* 2018, **14**(12):1709-1714.
78. Liu. F, Marquardt. S, Lister. C, Swiezewski. S, Dean. C: **Targeted 3' Processing of Antisense Transcripts Triggers Arabidopsis FLC Chromatin Silencing**. *Science* 2010, **327**:94-97.
79. Ruan J, Guo F, Wang Y, Li X, Wan S, Shan L, Peng Z: **Transcriptome analysis of alternative splicing in peanut (*Arachis hypogaea* L.)**. *BMC Plant Biol* 2018, **18**(1):139.
80. Celedon JM, Yuen MMS, Chiang A, Henderson H, Reid KE, Bohlmann J: **Cell-type- and tissue-specific transcriptomes of the white spruce (*Picea glauca*) bark unmask fine-scale spatial patterns of constitutive and induced conifer defense**. *Plant J* 2017, **92**(4):710-726.
81. Villarino GH, Hu Q, Scanlon MJ, Mueller L, Bombarely A, Mattson NS: **Dissecting Tissue-Specific Transcriptomic Responses from Leaf and Roots under Salt Stress in *Petunia hybrida* Mitchell**. *Genes (Basel)* 2017, **8**(8):195-214.
82. Alonso-Serra J, Safronov O, Lim KJ, Fraser-Miller SJ, Blokhina OB, Campilho A, Chong SL, Fagerstedt K, Haavikko R, Helariutta Y *et al*: **Tissue-specific study across the stem reveals the chemistry and transcriptome dynamics of birch bark**. *New Phytol* 2019, **222**(4):1816-1831.
83. Amini H, Naghavi MR, Shen T, Wang Y, Nasiri J, Khan IA, Fiehn O, Zerbe P, Maloof JN: **Tissue-Specific Transcriptome Analysis Reveals Candidate Genes for Terpenoid and Phenylpropanoid Metabolism in the Medicinal Plant *Ferula assafoetida***. *G3 (Bethesda)* 2019, **9**(3):807-816.
84. Zhang L, Li HT, Gao LM, Yang JB, Li DZ, Cannon CH, Chen J, Li QJ: **Phylogeny and evolution of bracts and bracteoles in *Tacca* (Dioscoreaceae)**. *J Integr Plant Biol* 2011, **53**(11):901-911.

85. Julian C, Rodrigo J, Herrero M: **Stamen development and winter dormancy in apricot (*Prunus armeniaca*)**. *Ann Bot* 2011, **108**(4):617-625.
86. Liu W, Lin L, Zhang Z, Liu S, Gao K, Lv Y, Tao H, He H: **Gene co-expression network analysis identifies trait-related modules in *Arabidopsis thaliana***. *Planta* 2019, **249**(5):1487-1501.
87. Wang R, Xu S, Jiang C, Sun H, Feng S, Zhou S, Zhuang G, Bai Z, Zhuang X: **Transcriptomic Sequencing and Co-Expression Network Analysis on Key Genes and Pathways Regulating Nitrogen Use Efficiency in *Myriophyllum aquaticum***. *Int J Mol Sci* 2019, **20**(7):1587-1606.
88. Fan Y, Yin X, Xie Q, Xia Y, Wang Z, Song J, Zhou Y, Jiang X: **Co-expression of SpSOS1 and SpAHA1 in transgenic *Arabidopsis* plants improves salinity tolerance**. *BMC Plant Biol* 2019, **19**(1):74-86.
89. Kwantes M, Liebsch D, Verelst W: **How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes**. *Mol Biol Evol* 2012, **29**(1):293-302.
90. Liu J, Fu X, Dong Y, Lu J, Ren M, Zhou N, Wang C: **MIKC(C)-type MADS-box genes in *Rosa chinensis*: the remarkable expansion of ABCDE model genes and their roles in floral organogenesis**. *Hortic Res* 2018, **5**:25.
91. Diaz-Riquelme J, Lijavetzky D, Martinez-Zapater JM, Carmona MJ: **Genome-wide analysis of MIKCC-type MADS box genes in grapevine**. *Plant Physiol* 2009, **149**(1):354-369.
92. Ren Z, Yu D, Yang Z, Li C, Qanmber G, Li Y, Li J, Liu Z, Lu L, Wang L *et al*: **Genome-Wide Identification of the MIKC-Type MADS-Box Gene Family in *Gossypium hirsutum* L. Unravels Their Roles in Flowering**. *Front Plant Sci* 2017, **8**:384.
93. Leseberg CH, Li A, Kang H, Duvall M, Mao L: **Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa***. *Gene* 2006, **378**:84-94.
94. Chen F, Zhang X, Liu X, Zhang L: **Evolutionary Analysis of MIKC(c)-Type MADS-Box Genes in Gymnosperms and Angiosperms**. *Front Plant Sci* 2017, **8**:895.
95. Lee S, Woo YM, Ryu SI, Shin YD, Kim WT, Park KY, Lee IJ, An G: **Further characterization of a rice AGL12 group MADS-box gene, OsMADS26**. *Plant Physiol* 2008, **147**(1):156-168.
96. Wu R, Tomes S, Karunairetnam S, Tustin SD, Hellens RP, Allan AC, Macknight RC, Varkonyi-Gasic E: **SVP-like MADS Box Genes Control Dormancy and Budbreak in Apple**. *Front Plant Sci* 2017, **8**:477.
97. Alamancos GP, Pages A, Trincado JL, Bellora N, Eyras E: **Leveraging transcript quantification for fast computation of alternative splicing profiles**. *RNA* 2015, **21**(9):1521-1531.
98. Sreya G, Chon-Kit Kenneth C: **Analysis of RNA-Seq Data Using TopHat and Cufflinks**. *Methods Mol Biol* 2016, **1374**:339-361.
99. Hall BG: **Building phylogenetic trees from molecular data with MEGA**. *Mol Biol Evol* 2013, **30**(5):1229-1235.
100. Tu Z, Hao Z, Zhong W, Li H: **Identification of Suitable Reference Genes for RT-qPCR Assays in *Liriodendron chinense* (Hemsl.) Sarg.** *Forests* 2019, **10**(5).

Figures

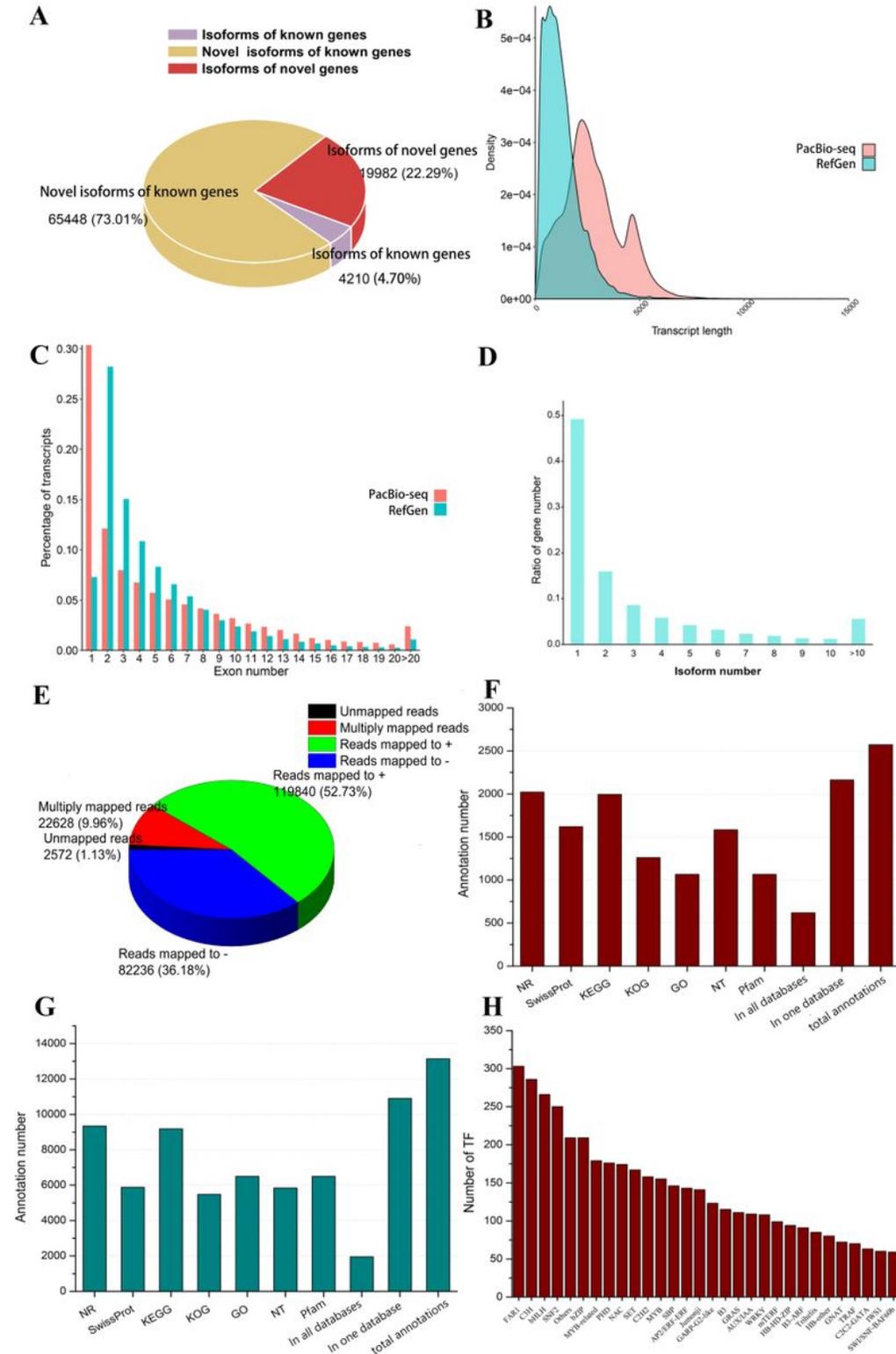


Figure 1

Statistics of Iso-Seq isoforms. a Classification of transcripts on the basis of the reference genome. b Comparison of isoform lengths between the PacBio data and reference genome annotation. c Comparison of the number of exons in the transcripts between the PacBio data and reference genome

annotation. d Number of transcripts produced by each gene according to PacBio Iso-Seq. e Statistics of the GMAP mapping results ('+' represents the sense sequence of the genome, '-' represents the antisense sequence of the genome). f Results of the functional annotation of the unmapped transcripts in seven databases. g Results of the functional annotation of the novel genes in seven databases. h Prediction of TFs by iTAK software (only the top 30 TF families are shown).

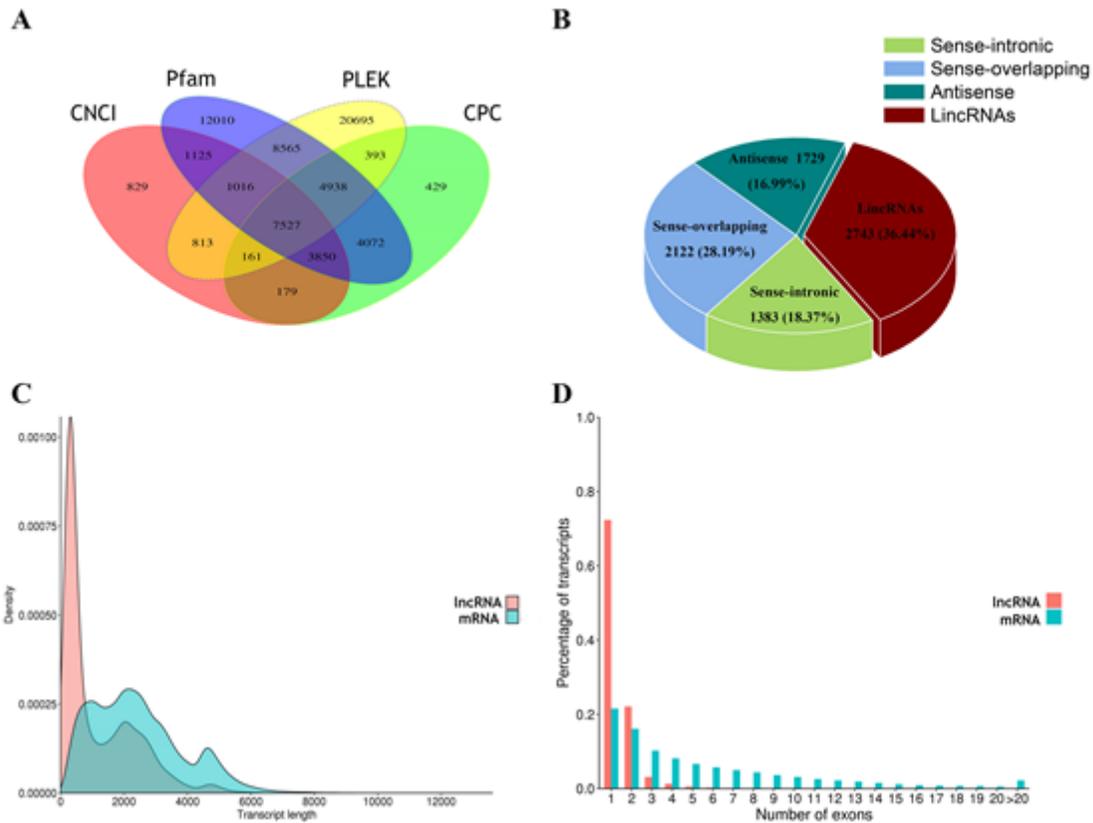


Figure 2

Analysis of lncRNAs and fusion transcripts. a Venn diagram of lncRNAs predicted by the CNCI, Pfam, PLEK, and CPC tools. b Classification of lncRNAs. c Comparison of the lengths of mRNAs and lncRNAs. d Comparison of the numbers of exons of mRNAs and lncRNAs.

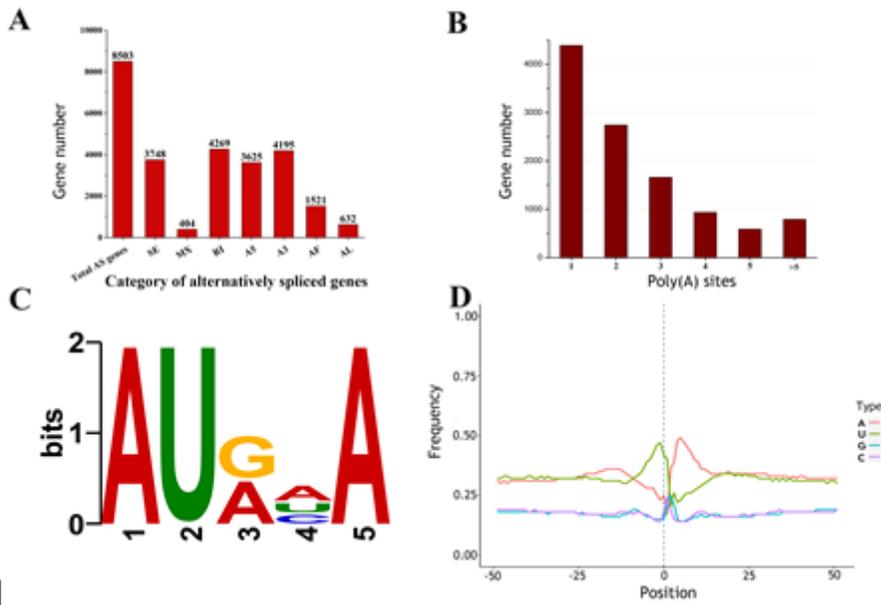


Figure 3

AS and APA analyses. a Number of seven different types of alternatively spliced genes. b Statistics of the number of poly(A) sites with genes. c Motif analysis around poly(A) cleavage sites. d Nucleotide distribution around poly(A) cleavage sites.

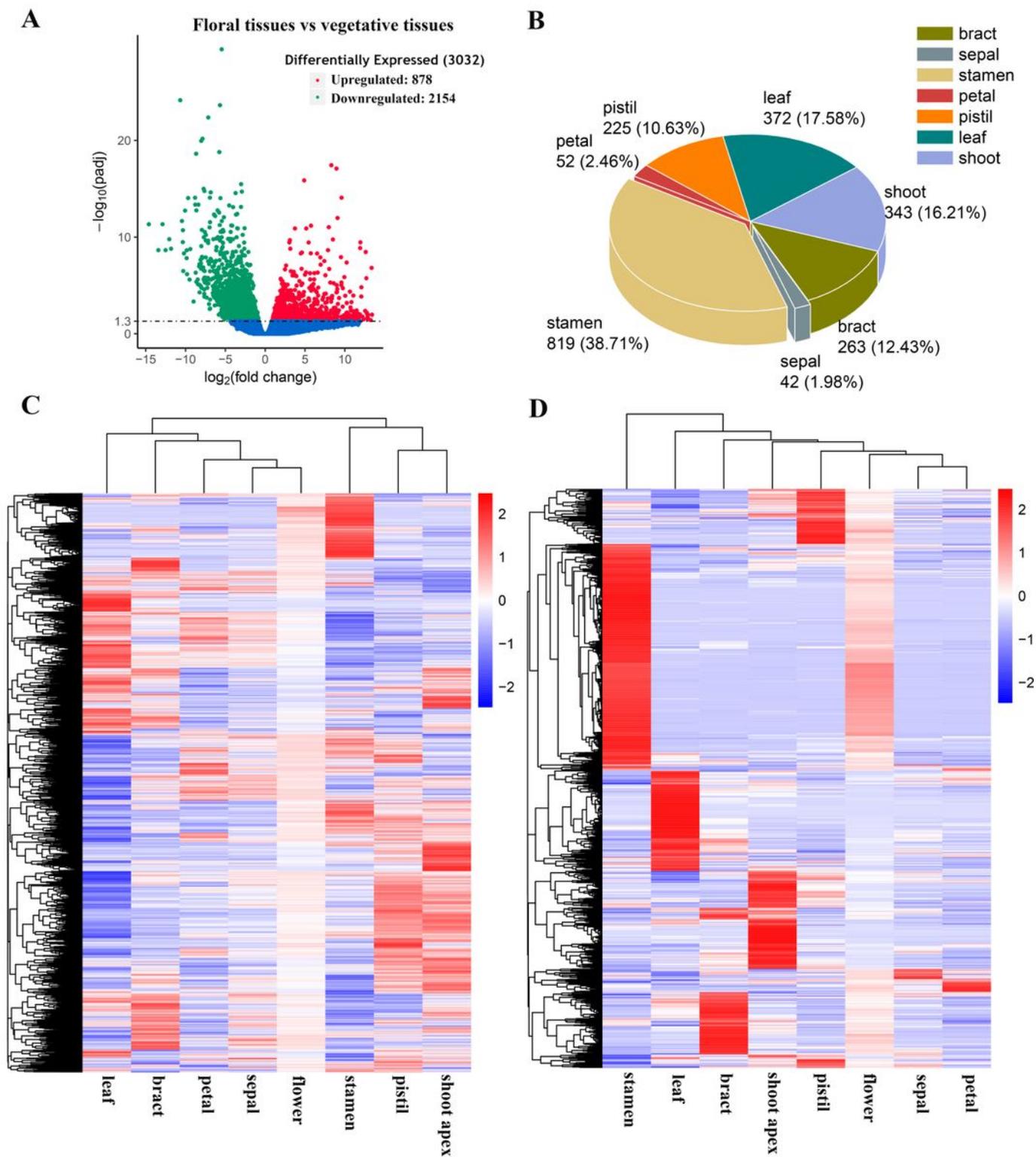


Figure 4

Analysis of DEGs and tissue-specific genes. a Volcano plot of DEGs between floral tissues and vegetative tissues (the upregulated genes are represented by red points; the downregulated genes are represented by green points). b Distribution of tissue-specific genes in seven tissues. c Hierarchical clustering analysis of genes based on the normalized $\log_{10}(\text{FPKM} + 1)$ values of all the genes. The floral tissues include bracts,

sepals, petals, stamens, and pistils. d Hierarchical clustering analysis of tissue-specific genes based on the normalized $\log_{10}(\text{FPKM} + 1)$ values of all tissue-specific genes.

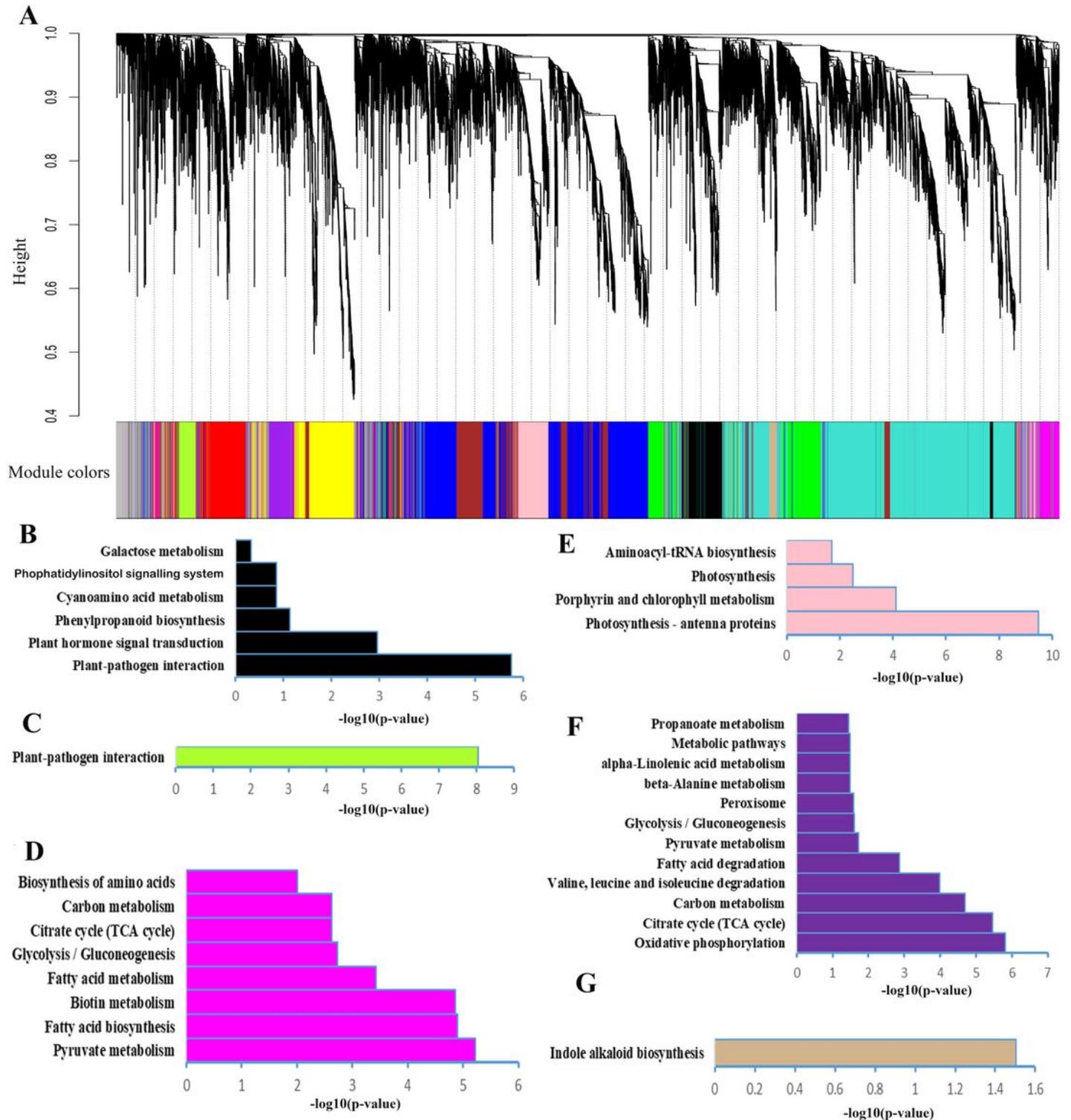


Figure 5

Dendrogram of co-expression gene modules in *L. chinense* and KEGG pathway enrichment analysis of six modules. a Dendrogram of co-expression gene modules in *L. chinense*. The genes are represented by vertical lines in the upper panel, and in the bottom panel, rectangles of thirteen colours represent the 13

blue circle represents *P. trichocarpa* protein, the yellow circle represents *A. thaliana* protein, and red circle represents *L. chinense* protein. b Expression profile of MIKC-type MADS-box genes in *L. chinense*.

Supplementary Files

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

- [Additionalfile9TableS1.docx](#)
- [Additionalfile1FigureS1.doc](#)
- [Additionalfile2FigureS2.doc](#)
- [Additionalfile10TableS2.docx](#)
- [Additionalfile3FigureS3.doc](#)
- [Additionalfile11TableS3.docx](#)
- [Additionalfile4FigureS4.doc](#)
- [Additionalfile12TableS4.docx](#)
- [Additionalfile5FigureS5.doc](#)
- [Additionalfile6FigureS6.doc](#)
- [Additionalfile13TableS5.docx](#)
- [Additionalfile14TableS6.docx](#)
- [Additionalfile7FigureS7.doc](#)
- [Additionalfile15TableS7.docx](#)
- [Additionalfile16TableS8.docx](#)
- [Additionalfile8FigureS8.doc](#)