De novo sequencing of *Eucommia ulmoides* flower bud transcriptomes for identification of genes related to floral development

Huimin Liu [a,b,1], JianMin Fu [a,1], Hongyan Du [a,b], Jingjing Hu [c], Tana Wuyun [a,b,*]

[a] Non-timber Forestry Research and Development Center of Chinese Academy of Forestry, Zhengzhou 450003, Henan, China
[b] The Eucommia Engineering Research Center of State Forestry Administration, Zhengzhou 450003, Henan, China
[c] Total Genomics Solution Limited, Shenzhen 518081, China

ARTICLE INFO

ABSTRACT

Eucommia ulmoides Oliver is a woody perennial dioecious species native to China and has great economic value. However, little is known about flower bud development in this species. In this study, the transcriptomes of female and male flower buds were sequenced using the Illumina platform, a next-generation sequencing technology that provides cost-effective, highly efficient transcriptome profiling. In total, 11,558,188,080 clean reads were assembled into 75,065 unigenes with an average length of 1011 bp by de novo assembly using Trinity software. Through similarity comparisons with known protein databases, 47,071 unigenes were annotated, 146 of which were putatively related to the floral development of *E. ulmoides*. Fifteen of the 146 unigenes had significantly different expression levels between the two samples. Additionally, 24,346 simple sequence repeats were identified in 18,565 unigenes with 12,793 sequences suitable for the designed primers. In total, 67,447 and 58,236 single nucleotide polymorphisms were identified in male and female buds, respectively. This study provides a valuable resource for further conservation genetics and functional genomics research on *E. ulmoides*.

1. Introduction

Eucommia ulmoides Oliver, belonging to the monotypic family Eucommiaceae, is a woody perennial dioecious species that is native to China and is widely distributed across the temperate zone in central and eastern areas, such as Shanxi, Henan, Anhui, Zhejiang, Guangxi, Hunan, Guizhou, Sichuan, and Hubei provinces. *E. ulmoides* inhabits mixed mesophytic forest habitats of valleys, hills, and low mountains [1–4]. The fossil record of *Eucommia* indicates that several species are included in this genus and that *Eucommia* species were widely distributed in North America during the Cenozoic period [5]. A previous study also reported that *E. ulmoides* may now be extinct in the wild [6], and *E. ulmoides* was included in the Red List of Endangered Plant Species in China [7].

Practically, the bark of *E. ulmoides* has been used in Chinese medicinal preparations for at least 2000 years and is generally prepared as a general tonic to alleviate hypertension, strengthen muscles and bones, enhance liver and kidney function, and stimulate fetal movement [6,8, 9]. *E. ulmoides* is also called a "hard rubber tree" because of the abundant quantities of trans-polyisoprene rubber in their leaves, bark, and seed coats. In addition to its application as a Chinese herbal medicine and commercial rubber production, the plant is also used for ornamental purposes (as timber), nutrient tea (specifically the male flowers), and a source of chlorogenic acid (from leaves).

Owing to these important economic applications of *E. ulmoides*, researchers have become increasingly interested in this plant. Zhang et al. [10] recently investigated the genetic diversity of *E. ulmoides* using eight microsatellite markers. Additionally, with the increased popularity of expressed sequence tags (ESTs) in gene discovery in recent decades, Suzuki et al. [11] constructed and analyzed the EST libraries of *E. ulmoides* from inner and outer stem tissues. However, this approach is relatively low throughput, high cost, and lacks the capacity for gene quantification [12].

With the rapid development of next-generation DNA sequencing technologies, sequencing costs have decreased dramatically, and sequencing accuracy has improved significantly. Transcriptome sequencing (RNA-Seq) is based on next-generation sequencing technology, having the advantages of cost-effective and highly efficient transcriptome profiling [12]. Together with the development of improved software programs, RNA-Seq has become a popular and power tool for large scale sequencing in non-model plants without the requirement for a reference genome [13].

In this study, we applied the Illumina sequencing technology to study the flower buds transcriptome of *E. ulmoides* and to develop a set of simple sequence repeat (SSR) markers. The transcriptome data reported here will provide valuable resources for the development of molecular markers and the flower-related gene discovery in *E. ulmoides*.

2. Materials and methods

2.1. Plant material and RNA extraction

The flower buds from two individuals, "SNJ" (male) and "BJC" (female), were collected from a field nursery in Yuanyang County, China in September 2014. The samples were deposited in liquid nitrogen before use. Total RNA from these two samples was extracted using TRIzol (Invitrogen, USA) following the manufacturer's instructions and then treated with RNase-free Dnase I.

2.2. cDNA library construction and Illumina sequencing

Library construction was performed according to the Illumina sample preparation for RNA-seq protocol (Illumina Inc., San Diego, USA; cat. no. RS-100-0801). After the total RNA extraction and DNase I treatment, mRNA was isolated using magnetic beads with Oligo (dT). mRNA was fragmented by mixing with fragmentation buffer, and cDNA was then synthesized using the mRNA fragments as templates. Short fragments were purified and resolved with ethidium bromide (EB) buffer for end reparation and single nucleotide A (adenine) addition. The short fragments were then connected with adapters. Suitable fragments were selected for PCR amplification as templates. During the quality control (QC) steps, an Agilent 2000 Bioanalyzer (G2939AA; Agilent) and Real-Time PCR System (StepOnePlus; ABI) were used for quantification and qualification of the sample library. Finally, the library was sequenced using HiSeq system (HiSeq 2000; Illumina). The sequencing data were deposited in the National Center for Biotechnology Information (NCBI, accession no.: SRA290287).

2.3. De novo assembly and annotation

Image data outputs from sequencing were transformed by base calling into raw data. The raw reads were then filtered to obtain clean reads by removing reads with adaptors, having unknown nucleotides accounting for >5%, and of low quality (>20% reads for which the quality value was ≤10). Short clean reads were finally assembled into unigenes using Trinity software [14]. The reads from two samples were assembled separately and then clustered together to acquire non-redundant unigenes that were as long as possible. The unigenes could then be divided into classes by gene family clustering. In this study, we divided the genes into clusters, in which several unigenes had high similarity (>70%), and singletons.

All the unigenes were first aligned to databases, such as NR, Swiss-Prot, KEGG, and COG (e-value < 0.00001) using BLASTx [15] and nucleotide database NT (e-value < 0.00001) using BLASTn [16]. The sequence direction and amino sequences of the unigenes were determined according to the best alignment results. Unigenes that could not be aligned to any database were scanned by ESTScan [17], producing nucleotide sequence direction and amino sequence data for the predicted coding regions. With NR annotations, we used the Blast2GO program [18] to obtain gene ontology (GO) annotations of unigenes, followed by WEGO software [19] to determine GO functional classifications for all unigenes and to elucidate the distribution of gene functions in species at the macro level.

2.4. SSR detection and primer design

SSRs were detected using MicroSAtellite (MISA) software [20] with all unigenes as references. We used sequences with SSRs for which the lengths of both ends on the unigene were >150 bp to design primers and then filtered the primers as follows. First, we ensured that there were no SSRs in the primer. Next, we aligned the primers to unigene sequences with three mismatches allowed in the 5′ site and one mismatch allowed in the 3′ site. Finally, we removed the primers that aligned to more than one unigene.

2.5. Detection of single nucleotide polymorphisms (SNPs)

The consensus sequence for each sample was assembled separately based on the alignment of the raw sequences on the unigenes. Then the SNPs were identified on the consensus sequence through the comparison with the unigenes.

2.6. EST data

The assembled EST data of *E. ulmoides* used in this study were downloaded from GenBank (accession no.: FY896671-FY925126).

3. Results and discussion

3.1. Sequencing and de novo assembly

We sequenced the genomes for the flower buds using the Illumina 2000 platform. In total, 11,558,188,080 clean reads were obtained with a mean length of 100 bp. The percentages of Q20, 'N', and GC were 97.84%, 0.01%, and 46.73%, respectively. The clean reads were assembled into 75,065 unigenes using Trinity, with a total length of 75,898,028 bp, a mean length of 1011 bp, and an N50 of 1653 bp. Most (48,129) of the unigenes ranged from 201 to 1000 bp, accounting for 64% of the total unigenes. Twenty-three percent (16,984) of the unigenes ranged from 1001 to 2000 bp. Nine percent (6445) of the unigenes ranged from 2001 to 3000 bp, and 3507 (5%) of the unigenes were longer than 3000 bp (Fig. 1).

Previous EST libraries constructed from the outer and inner stem tissues of *E. ulmoides* represented 10,520 unigenes, with an average length of 559 bp and an N50 of 634 bp [11]. Compared with EST libraries, our transcriptome sequenced using the Illumina platform indicated a better assembly result with more unigenes and a longer average length and N50, thereby offering a reliable data resource for further analyses. The average length (1011 bp) of the unigenes was longer than that reported for the *Jasminum sambac* (846 bp) transcriptome, using the Illumina sequencing system, and the Chinese jujube (473 bp), using the 454 GS FLX Titanium genomic sequencer platform [21,22]. The N50 length (1653 bp) was also longer than that reported for the Chinese jujube [22].

3.2. Sequence annotation and classification

After aligning the unigene sequences to protein databases, nearly 47,071 of the unigenes were annotated, leaving 27,994 unigenes that were not aligned to any database. The numbers and percentage of unigenes annotated within the Non Redundant (NR), Nucleotide (NT), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Clusters of Orthologous Group of proteins (COG), and gene ontology (GO) databases were 44,205 (59%), 37,061 (49%), 28,419 (38%), 26,312 (35%), 17,441 (23%), and 32,228 (43%), respectively (Table 1). A total of 74.8% of the unigenes in previous EST libraries of *E. ulmoides* were annotated within the NR database [11], which is higher than that of our annotation. This result may be explained by the shorter unigenes assembled using ESTs; shorter assembled unigenes were easier to match with items in the database.

In the unigenes annotated in the NR database, the majority (39.4%) matched proteins from *Vitis vinifera*, followed by those from *Lycopersicon esculentum* (15.9%), *Amygdalus persica* (8.5%), *Ricinus communis* (6.7%), *Populus balsamifera* subsp. *Trichocarpa* (6.0%), *Fragaria*
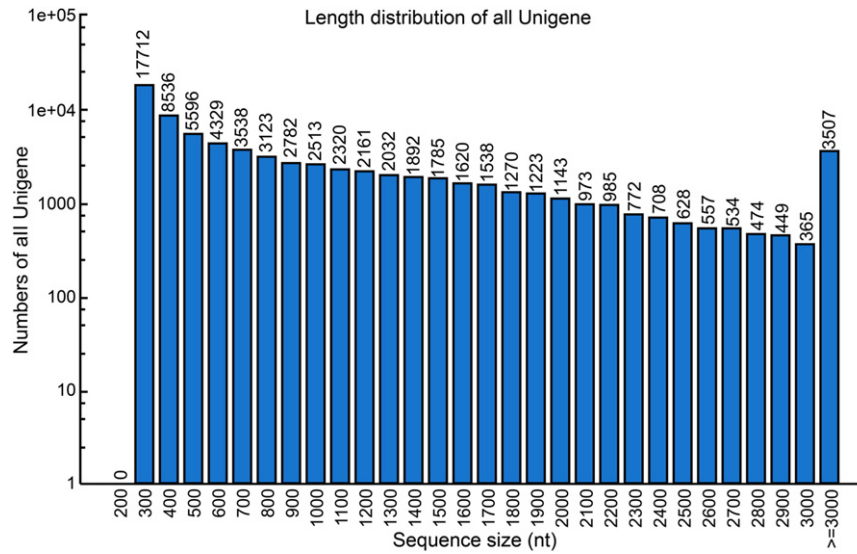
Fig. 1. Length distributions of all the assembled unigenes. X-axis: length distribution of all assembled unigenes, e.g., 300 indicates the length range ≥ 200 bp and <300 bp. Y-axis: numbers of all unigenes in different length ranges.

vesca subsp. vesca (3.2%), Glycine max (2.7%), and other species (17.7%), as shown in Fig. 2.

With annotations in the NR database, 32,228 unigenes were assigned to GO categories with 5316 unique functional terms (Fig. 3). From this analysis, 24,744, 25,402, and 24,231 unique unigenes were assigned to the GO categories of biological processes, cellular components, and molecular functions, respectively. Of all the unigenes assigned GO categories, 16,830 unigenes were shared by the three categories, whereas 1236, 3490, and 2183 unigenes were uniquely assigned to the GO categories of biological processes, cellular components, and molecular functions, respectively. In the biological processes category, the categories of cellular processes (20,098) and metabolic processes (19,393) were the largest groups. The numbers of genes related to reproduction, reproductive processes, rhythmic processes, and signaling were 3862, 3485, 229, and 2720, respectively. In the cellular component category, most of the unigenes were related to the category of cells and cell parts. In the molecular functions category, more unigenes were assigned to the binding and catalytic activity category than to any other category. In total, 54 unigenes were annotated with GO terms related to carpel development (GO:0048440), identity (GO:0010094), and morphogenesis (GO:0048445); 146 unigenes were annotated with GO terms related to stamen development (GO: 0048443), filament development (GO: 0080086), and morphogenesis (GO: 0048448); and 891 unigenes were annotated with GO terms related to flower development, morphogenesis, regulation, and flowering photoperiodism (GO:0009909,00048439,0048573,0009910,0009911,00048578,00048574,00048586, and 0009908).

COG is a database that classifies orthologous gene products. We mapped all the unigenes to the COG database to predict the possible functions and statistics and to elucidate gene function distribution characteristics of species at the macro level. In total, 17,441 unigenes were assigned to the COG database and classified into 25 COG categories (Fig. 4). Of the 25 categories, six categories, i.e., general function prediction only (5551, 31.8%); transcription (2977, 17.1%); replication, recombination and repair (2974, 17.1%); posttranslational modification, protein turnover, chaperones (2263, 13.0%); signal transduction mechanisms (2239, 12.8%); and translation, ribosomal structure, and biogenesis (2130, 12.2%), contained >2000 unigenes each. Moreover, a total of 1929 unigenes were mapped to the carbohydrate transport and metabolism category, of which 21, 20, and 35 unigenes were annotated as glycogen synthases (COG0297), 6-phosphofructokinases (COG0205), and glyceraldehyde-3-phosphate dehydrogenases/erythrose-4-phosphate dehydrogenases (COG0057), respectively. Of the 1072 unigenes assigned to the energy production and conversion category, only two unigenes (Unigene21954 and Unigene45295) were annotated as fumarases (COG0114), whereas seven and 11 unigenes were annotated as glycerol-3-phosphate dehydrogenases (COG0240) and phosphoenolpyruvate carboxykinases (ATP) (COG1866), respectively. The proteins encoded by these genes are involved in basal metabolism.

KEGG is a database that is able to analyze gene products during metabolic processes and related gene functions in cellular processes. Using the KEGG database, we studied the complex biological behaviors of genes in more detail and obtained pathway annotations for unigenes. In total, 26,312 unigenes were annotated in the KEGG database and were assigned to 128 pathways. The top five pathways were metabolic pathways (ko01100), biosynthesis of secondary metabolites (ko01110), plant-pathogen interaction (ko04626), plant hormone signal transduction (ko04075), and RNA transport (ko03013), consisting of 6114 (23.24%), 3073 (11.68%), 1677 (6.37%), 1187 (4.51%), and 975 (3.66%) unigenes, respectively. Unigenes mapped to the plant hormone signal transduction pathway were involved in the biosynthesis of auxins, cytokinins, gibberellins, abscisic acid, ethylene, brassinosteroid, jasmonic
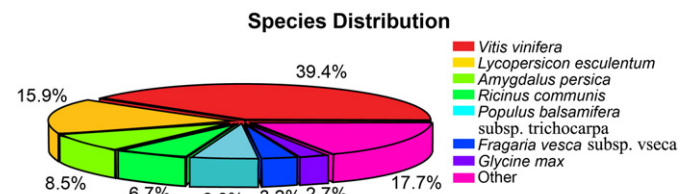
Table 1
The number of unigenes annotated to different databases.

NR	NT	Swiss-Prot	KEGG	COG	GO	All
44,205	37,061	28,419	26,312	17,441	32,228	47,071

Fig. 2. Percent distribution of unigenes assigned to different species in the NR database. 39.4% indicates that 39.4% of the unigenes matched the proteins from Vistis vinifera.
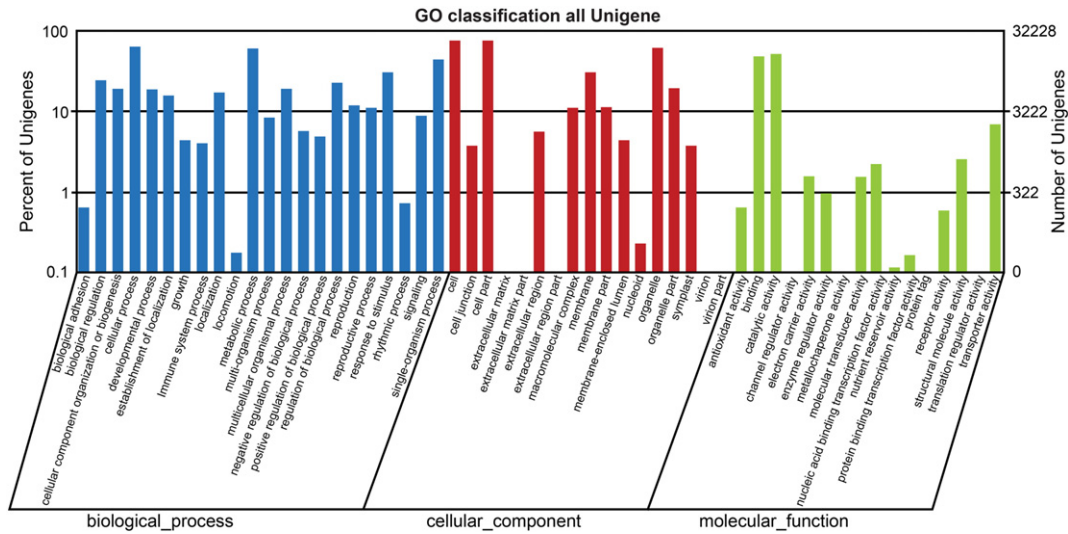
Fig. 3. Percent and number distributions of unigenes assigned to the GO database. X-axis: GO categories. Y-axis: percentage (left) and number (right) of unigenes assigned to different GO categories.

acid, and salicylic acid, which are related to cell enlargement and plant growth, cell division and shoot initiation, stem growth and induced germination, stomatal closure and seed dormancy, fruit ripening and senescence, cell elongation and cell division, senescence and stress response, and disease resistance, respectively. Sixty-one unigenes were involved in the brassinosteroid biosynthesis pathway (ko00905). These unigenes may be involved in the process of flowering.

3.3. Genes putatively related to flower development

All *E. ulmoides* individuals are dioecious. Male flowers are fascicled without perianth and degenerated pistils, whereas female flowers are solitary. We identified several unigenes that may be involved in floral development (Table S1), including seven *PHYA* genes encoding phytochrome A, 13 *PHYB* genes encoding phytochrome B, one *CRY1* gene encoding cryptochrome 1, two *CRY2* genes encoding cryptochrome 2, one *FKF1* gene encoding flavin-binding kelch repeat F-box protein 1,

one *ZTL2* gene encoding ZEITLUPE 2, two *SPA* genes encoding suppressor of PHYA-105, six *TOC1* genes encoding timing of cab expression 1/pseudo-response regulator 1, three *ELF3* genes encoding early flowering 3, five *ELF4* genes encoding early flowering 4, 14 *LHY* genes encoding late elongated hypocotyl, two *GI* genes encoding GIGANTEA, two *CO* genes encoding CONSTANS, two *FT* genes encoding flowering locus T, 21 *FLC* genes encoding flowering locus C, six *VIN3* genes encoding vernalization insensitive 3, 15 *FCA* genes encoding flowering time control, four *LFY* genes encoding leafy, 10 *FRI* genes encoding FRIGIDA, one *HOS1* gene encoding high expression of osmotically responsive protein 1, eight *LUG* genes encoding LEUNIG, and 20 *SOC1* genes encoding suppressor of overexpression of constans 1.

By comparing the unigenes that first assembled separately in male and female flower buds, we found that only 15 of all the abovementioned unigenes, i.e., CL745.Contig1, 745.Contig2, and unigene9068 (*PHYB*); unigene11188 (*ELF4*); unigene4137 (*FT*); CL2716.Contig3 (*VIN3*); CL1895.Contig1 (*FCA*); unigene10446 (*LFY*); CL7732.Contig1
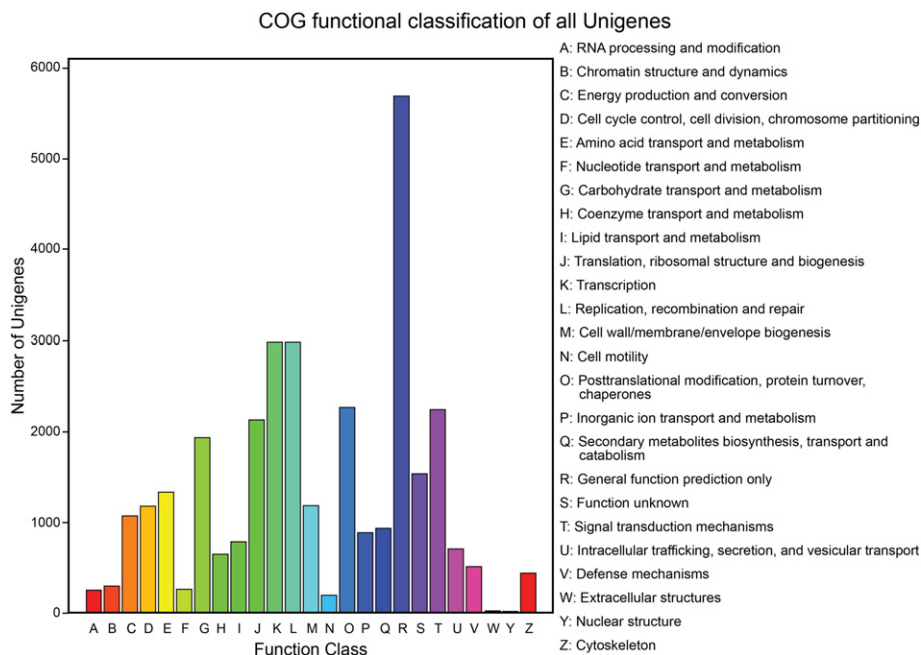

Fig. 4. Number distribution of unigenes assigned to the COG database. X-axis: COG functional classification. Y-axis: numbers of unigenes assigned to different COG functional classifications.

and CL7732.Contig3 (*FRI*); unigene32289 (*LUG*); and CL5214.Contig1, CL5214.Contig2, CL5214.Contig3, and CL8558.Contig3 (*SOC1*), had significantly different expression levels in the two samples. *PHYB* acts in a partially redundant manner with *PHYD* and *PHYE*, mediating the inhibition of flowering by R light [23–25]. A previous study reported that *ELF4* from *Arabidopsis thaliana* is involved in photoperiod perception and circadian regulation, promotes clock accuracy, and is required for sustained rhythms in the absence of daily light/dark cycles [26]. *FT* promotes the transition to reproductive development and flowering [27, 28]. Proteins encoded by *VIN3* genes in *A. thaliana* function to collectively repress different members of the *FLC* gene family during the course of vernalization [29]. The autonomous pathway component *FCA* is the founding member of the thermosensory pathway [30]. Multiple alleles of *FCA* exhibit insensitive flowering phenotypes to different ambient temperatures in an *FT*-dependent manner [31]. *LFY* encodes a plant-specific transcription factor that plays dual roles in determining floral meristem identity and floral organ patterning via *AP1* and other floral homeotic genes [32]. In *Arabidopsis*, dominant alleles of *FRI* confer the late flowering phenotype, which is reversed to the early flowering phenotype by vernalization [33]. *LUG* is a putative transcriptional corepressor that regulates *AGAMOUS* expression during flower development [34]. *SOC1* is a flowering integrator that acts partially downstream of *FT* [35].

3.4. SSR identification

SSR detection was carried out using MIcroSAtellite (MISA) software with unigenes as the reference. In total, 24,346 SSRs were identified in 18,565 unigenes with 12,793 sequences suited for primer design and 4349 sequences containing more than one SSR. Only 1629 SSRs were present in compound formation. The number of repeats ranged from four to 23, and the number of different repeat unit sizes were 7432 for mono-, 12,390 for di-, 3677 for tri, 162 for tetra-, 262 for penta-, and 423 for hexanucleotide repeats (Fig. 5). Di- and trinucleotide repeats were the most types among all the repeats, consistent with results in other angiosperms [36–38]. Most of the mononucleotide repeats had 12 or 13 repeats, whereas the majority of nucleotide repeats had six

repeats, most of the tri- and tetranucleotide repeats had five repeats, the pentanucleotide repeats had four or five repeats, and the hexanucleotide repeats only had four repeats. Most of the mononucleotide repeats were A/T motifs. The AG/CT motifs included 74.6% of the dinucleotide repeats. The most frequent motifs of trinucleotide repeats were AAG/CTT (31.5%).

Compared with the flower bud transcriptomes of *E. ulmoides*, only 8794 SSRs were identified in 7859 unigenes in the EST libraries of *E. ulmoides* [11], indicating that the transcriptome sequenced by the Hiseq system may contain more information than EST libraries.

3.5. SNP analysis

We also found the SNPs for each sample using all unigenes as references. In total, 67,447 and 58,236 SNPs were identified in "SNJ" and "BJC," respectively. In "SNJ," 44,466 SNPs were transitions, and 22,981 SNPs were transversions. In "BJC," the numbers of transition and transversion SNPs were 38,472 and 19,784, respectively. In both of these samples, the majority of transition SNPs were A to G transitions, and most of the transversion SNPs were A to C transversions. Comparing "SNJ" with "BJC," we found that 17,171 SNPs were the same (with consistent loci and SNP types), whereas 11,554 SNPs had only consistent loci.

4. Conclusion

In summary, we analyzed the transcriptome of *E. ulmoides* using Illumina sequencing-by-synthesis technology. After de novo assembly and sequence annotation, we obtained 75,065 unigenes and identified 146 unigenes putatively related to the floral development of *E. ulmoides*. In addition, we also identified 24,346 SSRs and detected 67,447 and 58,236 SNPs in "SNJ" and "BJC," respectively. Notably, we only sequenced one period of floral development; more samples from different periods of development are needed to analyze the expression profiles of genes related to floral development in order to identify key genes and sex-related genes. Further analysis of these SSRs and SNPs will provide useful resources for conservation genetics and functional genomics research on *E. ulmoides* in the future.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2016.07.001.

Author contributions

Tana Wuyun and Hongyan Du conceived and designed the experiments. Huimin Liu and Jianmin Fu performed the experiments and wrote the paper. Jingjing Hu analyzed the data.

Conflicts of interest

The authors declare no conflicts of interest.

Fig. 5. SSRs identified in the transcriptome of *E. ulmoides*. X-axis: motif types. Y-axis: numbers of SSRs matched to different motif types.

References

[1] H.T. Chang, Eucommiaceae. Flora Repub. Popular. Sin. 35 (1979) 116–118.
[2] C.W. Wang, The Forests of China With a Survey of Grassland and Desert Vegetation, in: Maria Moors Cabot Foundation Publication Number 5. Harvard University, Cambridge, MA, 1961.
[3] H.S. Wang, A study of the distribution and origin of endemic genera of spermatophytes in China. in: P. Whyte, J.S. Aigner, N.G. Jablonski, G. Taylor, D. Walker, W. Pinxian, S. Chak Lam (Eds.), The Paleoenvironment of East Asia From the Mid-tertiary, Geology, Sea Level Changes, Paleoclimatology and Paleobotany, 1, Centre of Asian Studies, University of Hong Kong, Hong Kong 1988, pp. 605–620.
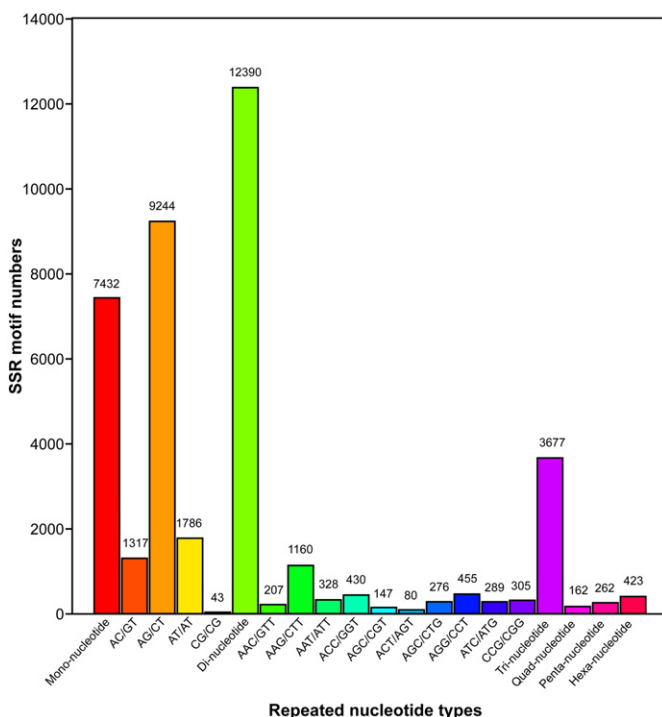
[4] T.S. Ying, Y.L. Zhang, D.E. Boufford, The Endemic Genera of Seed Plants of China. Science Press, Beijing, 1993 337–340.

[5] V.B. Call, D.L. Dilcher, The fossil record of *Eucommia* (Eucommiaceae) in North America. Am. J. Bot. 84 (1997) 798–814.

[6] D.J. Mabberley, The Plant Book. Cambridge University Press, Cambridge, MA, 1989 270.

[7] L.G. Fu, J.M. Jin, Red List of Endangered Plants in China. Science Press, Beijing, 1992.

[8] Chinese Pharmacopoeia Committee of Ministry of Health of the People's Republic of China, Chinese Pharmacopoeia. People's Medical Publishing House and Chemical Industry Press, Beijing, 1990 131.

[9] H. Huang, Plant diversity and conservation in China: planning a strategic bioresource for a sustainable future. Bot. J. Linn. Soc. 166 (2011) 282–300.

[10] J. Zhang, C. Xing, H. Tian, X. Yao, Microsatellite genetic variation in the Chinese endemic *Eucommia ulmoides* (Eucommiaceae): implications for conservation. Bot. J. Linn. Soc. 173 (2013) 775–785.

[11] N. Suzuki, H. Uefuji, T. Nishikawa, Y. Mukai, A. Yamashita, M. Hattori, et al., Construction and analysis of EST libraries of the *trans*-polyisoprene producing plant, *Eucommia ulmoides*, Oliver. Planta 236 (2012) 1405–1417.

[12] Z. Wang, M. Gerstein, M. Snyder, RNA-seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10 (2009) 57–63.

[13] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 29 (2011) 644–652.

[14] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, et al., *De novo* transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat. Protoc. 8 (2013) 1494–1512.

[15] B.J. Robert, D.E. Ranta, C.E. Joachim, BlastX Code, Version 4.2, User's Manual, ERDC/GSL TR-01-2. US Army Engineer Research and Development Center, Vicksburg, MS, 2001.

[16] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, et al., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (1997) 3389–3402.

[17] C. Iseli, V.C. Jongeneel, P. Bucher, ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. ISMB 99 (1999) 138–148.

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

[19] J. Ye, L. Fang, H. Zheng, Y. Zhang, J. Chen, Z. Zhang, et al., WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34 (2006) W293–W297.

[20] T. Thiel, MISA-microsatellite Identification Tool, 2003, URL: http://pgrc.ipk-gatersleben.de/misa/misa.html.

[21] Y.H. Li, W. Zhang, Y. Li, Transcriptomic analysis of flower blooming in *Jasminum sambac* through *de novo* RNA sequencing. Molecules 20 (2015) 10734–70747.

[22] Y. Li, C. Xu, X. Lin, B. Cui, R. Wu, X. Pang, De novo assembly and characterization of the fruit transcriptome of Chinese jujube (*Ziziphus jujuba* Mill.) using 454 pyrosequencing and the development of novel tri-nucleotide SSR markers. PLoS One 9 (2014), e106438.

[23] P.F. Devlin, S.R. Patel, G.C. Whitelam, Phytochrome E influences internode elongation and flowering time in *Arabidopsis*. Plant Cell 10 (1998) 1479–1488.

[24] P.F. Devlin, P.R. Robson, S.R. Patel, L. Goosey, R.A. Sharrock, G.C. Whitelam, Phytochrome D acts in the shade avoidance syndrome in *Arabidopsis* by controlling elongation growth and flowering time. Plant Physiol. 119 (1999) 909–915.

[25] T.C. Mockler, H. Guo, H. Yang, H. Duong, C. Lin, Antagonistic actions of *Arabidopsis* cryptochromes and phytochrome B in the regulation of floral induction. Development 126 (1999) 2073–2082.

[26] M.R. Doyle, S.J. Davis, R.M. Bastow, H.G. McWatters, L. Kozma Bognar, F. Nagy, et al., The *ELF4* gene controls circadian rhythms and flowering time in *Arabidopsis thaliana*. Nature 419 (2002) 74–77.

[27] Y. Kobayashi, H. Kaya, K. Goto, M. Iwabuchi, T. Araki, A pair of related genes with antagonistic roles in mediating flowering signals. Science 286 (1999) 1960–1962.

[28] J.H. Ahn, D. Miller, V.J. Winter, M.J. Banfield, J.H. Lee, S.Y. Yoo, et al., A divergent external loop confers antagonistic activity on floral regulators *FT* and *TFL1*. EMBO J. 25 (2006) 605–614.

[29] D.H. Kim, S. Sung, Coordination of the vernalization response through a *VIN3* and *FLC* gene family regulatory network in *Arabidopsis*. Plant Cell 25 (2013) 454–469.

[30] K. Jang, H.G. Lee, S.J. Jung, N.C. Paek, P.J. Seo, The E3 ubiquitin ligase *COP1* regulates thermosensory flowering by triggering *GI* degradation in *Arabidopsis*. Sci. Report. 5 (2015) 12071.

[31] M.A. Blazquez, J.H. Ahn, D. Weigel, A thermosensory pathway controlling flowering time in *Arabidopsis thaliana*. Nat. Genet. 33 (2003) 168–171.

[32] E. Moyroud, E. Kusters, M. Monniaux, R. Koes, F. Parcy, LEAFY blossoms. Trends Plant Sci. 15 (2010) 346–352.

[33] U. Johanson, J. West, C. Lister, S. Michaels, R. Amasino, C. Dean, Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science 290 (2000) 344–347.

[34] J. Conner, Z. Liu, LEUNIG, a putative transcriptional corepressor that regulates *AGAMOUS* expression during flower development. Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 12902–12907.

[35] S.K. Yoo, K.S. Chung, J. Kim, J.H. Lee, S.M. Hong, S.J. Yoo, et al., Constans activates suppressor of overexpression of constans 1 through flowering locus T to promote flowering in *Arabidopsis*. Plant Physiol. 139 (2005) 770–778.

[36] T. Wu, S. Luo, R. Wang, Y. Zhong, X. Xu, Y.E. Lin, et al., The first Illumina-based *de novo* transcriptome sequencing and analysis of pumpkin (*Cucurbita moschata* Duch.) and SSR marker development. Mol. Breed. 34 (2014) 1437–1447.

[37] D. Li, Z. Deng, B. Qin, X. Liu, Z. Men, *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). BMC Genomics 13 (2012) 192.

[38] K. Triwitayakorn, P. Chatkulkawin, S. Kanjanawattanawong, S. Sraphet, T. Yoocha, D. Sangsrakru, et al., Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. DNA Res. 18 (2011) 471–482.