Identification of novel EST-SSR markers from the *de novo* transcriptome sequence and application in persimmon (*Diospyros* L.)

Zhaobin Jing^{1,2}, Xiaofeng Ruan¹ and Yong Yang¹

¹ College of Horticulture, Northwest A&F University, Yangling, Shaanxi, China

² Weinan Vocational and Technical College, Weinan Fruit Industry Institute of Shaanxi, Weinan, Shaanxi, China

Summary

Persimmon (*Diospyros* L.) is one of the most important commercially cultivated fruit crops in the world. However, the research on the molecular biology of persimmon is very less due to the lack of genomic and transcriptome data. In this study, the persimmon transcriptomes were sequenced from a new species of *Diospyros* L. (Deyangshi). Illumina sequencing generated 44,522,166 high-quality clean reads that were assembled into 48,012 unigenes with an average length of 835 bp. A total of 27,586 unigenes (57.7% of all unigenes) were annotated, and 14,275 unigenes were classified into 24 function classification using Clusters of Orthologous Groups of proteins (COG) database, 71,073 unigenes were assigned to 45 functional groups using Gene Ontology (GO) database, 7,186 annotated sequences were assigned to 124 specific metabolic pathways by Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Furthermore, 8,741 simple sequence repeats (SSRs) were identified from 7,074 unigenes. 100 pairs of primers were selected to evaluate the validity and universality in twenty genotypes of persimmon, and 89 pairs generated clear and abundant bands. This is the first report to develop SSR markers for persimmon by transcriptome information. These data could be used as an important resource for studying the genetics and molecular biology for persimmon in future work.

Keywords

persimmon, transcriptome sequence, functional classification, SSR markers

Introduction

The Ebenaceae is an important family of flowering plants belonging to the order Ericales. It consists of ebony and persimmon among about 768 species, distributed across the tropical and warmer temperate regions of the world (Bremer et al., 2009). Persimmon (*Diospyros* L.) is an important fruit crop and is cultivated extensively in China, Japan, Korea, Italy, Israel and Spain (Yang et al., 2015a). It belongs to the genus *Diospyros* in the family Ebenaceae. Persimmon included more than 400 species with ploidy levels ranging from diploid ($2n=2x=30$) to nonaploid ($2n=9x=135$) (Jing et al., 2013). Persimmon has existed for 10,000 years in China. China is one of the major gene centers of persimmon species origin. About 650 accessions of persimmon germplasm were

Significance of this study

What is already known on this subject?

- Persimmon (*Diospyros* L.) is an important fruit crop and is cultivated extensively in China, Japan, Korea, Italy, Israel and Spain. Persimmon plants have many excellent characteristics. In spite of their high nutrients importance, the research on the molecular biology and genetics is very lacking, which means that accurate molecular biology is necessary for future works.

What are the new findings?

- The persimmon transcriptomes were sequenced from a new species of *Diospyros* L. (Deyangshi). Illumina sequencing generated a total of 27,586 unigenes (57.7% of all unigenes) were annotated. Furthermore, 8,741 simple sequence repeats (SSRs) were identified from 7,074 unigenes. 100 pairs of primers were selected to evaluate the validity and universality in twenty genotypes of persimmon, and 89 pairs generated clear and abundant bands.

What is the expected impact on horticulture?

- We expected to provide a reference and resource for the study of the genetic diversity, and molecular biology for persimmon in future works, and enhance the molecular marker assistant breeding.

collected in the National Field Genebank for persimmon in Yangling, Shaanxi Province, China (Jing et al., 2013). Persimmon plants from China have many excellent characteristics. For example, the dietary fibers, total phenols, epicatechin, gallic and p-coumaric acids, and concentrations of Na, K, Mg, Ca, Fe, and Mn, are rich in whole persimmons, their pulps, and peels (Gorinstein et al., 2001). The seeds of date plum persimmon could be used as a medicine and its fruits can be eaten directly after destringing or for medical purposes (Yang et al., 2015b). In spite of its high nutrients importance, the research on the molecular biology and genetics is lacking. To date, the studies on persimmon only include a few fields, such as genetic diversity and relationships (Jing et al., 2013; Yang et al., 2015a, b), and marker development (Yang et al., 2015a). In addition, understanding the genetic diversity and relationships of the persimmon is important to put forward the conservation strategies and breeding for the persimmon. Simple sequence repeat (SSR) are widely used for genetic diversity and relationships of plants due to the

highly polymorphic, co-dominant and reproducible qualities. However, at present, only fifty-one persimmon simple sequence repeat markers have been developed (Guo and Luo, 2006, 2008; Yang et al., 2015a; Soriano et al., 2006).

Transcriptome is the complete collection of RNA that includes the full range of mRNA, tRNA, rRNA, and other non-coding RNA molecules, expressed by one or a group of cells, organs or tissues in a particular environment or a specific developmental stage. The Illumina paired-end sequencing technique has been widely used in transcriptome research of plants (Jiang et al., 2013; Zhou et al., 2015). Transcriptome sequencing could be used to the development of SSR molecular markers due to its reliable and efficient (Zhou et al., 2015).

In this study, the Illumina sequencing technology was used to analyze the transcriptome of Deyangshi (*Diospyros* L.) and to develop simple sequence repeat (SSR) markers. This is the first study attempting to characterize the transcriptome of Deyangshi. Our studies will provide the foundation for the molecular biology and molecular breeding in persimmon.

Materials and methods

Plant material and RNA extraction

Deyangshi (*Diospyros*) was grown in the National Field Genebank for persimmon at Yangling, Shaanxi Province of China, belonging to the College of Horticulture in Northwest A&F University. The tissues from leaves, flowers, stems and fruits were harvested from three male and three female of six Deyangshi individuals. The sampled tissues were frozen in liquid nitrogen immediately and brought back to laboratory and stored at -80°C for later RNA extraction. The total RNA of samples was isolated using the Trizol Kit (Promega, USA) according to the manufacturer's instructions. RNA quality and the concentration were tested by 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and RNase free agarose gel electrophoresis. The high quality RNA (260 nm/280 nm ratio is between 1.8 and 2.0) were used for the following analyses. Then the equal amounts of RNA from the different samples were mixed for further studies.

cDNA library construction and sequencing

cDNA library construction was carried out according to the previously established experiment methods by Zhou et al. (2015) and Wu et al. (2014). An RNA sample including fresh young leaves, stems, flowers and fruits was used for transcriptome sequencing analyses. The sequencing analyses were completed using the Illumina HiSeq TM2000 by Guangzhou Gene Denovo Biological Technology Co., Ltd. (Guangzhou, China). The sequencing data have been submitted to NCBI Sequence Read Archive (Accession number: PRJNA324726).

De novo assembly and gene annotation

First, adapter sequences, empty reads and low quality sequences were filtered using Perl program. After filtering raw reads, *de novo* assembly of the transcriptome were carried out with Trinity software (Grabherr et al., 2011). The detail process of *de novo* assembly were described by Wu et al. (2014). The generated sequences were used for BLASTX and annotation against several protein databases, including non-redundant (nr), Swiss-Port, KEGG, GO and COG protein database, with E-value < 0.00001. We used the Blast2GO program to analyze the GO annotation of the unigenes (Conesa et

al., 2005), and the functional classifications using the WEGO software (Ye et al., 2006). Pathway information of unigenes was obtained from KEGG databases (Kanehisa et al., 2008).

SSR detection, primer design and SSR validation

The MicroSATellite (MISA, <http://pgrc.ipk-gatersleben.de/misa/>) was used to identify microsatellites in the whole transcriptome. In this study, the 48,012 unigenes were searched for SSRs. The parameters of SSRs were considered according to Zhou et al. (2015). Primer pairs of each SSR loci were designed using the Primer premier 6.0 (Premier Biosoft International, Palo Alto, CA) in the flanking regions of SSRs. We selected 100 pairs of primers used to genetic diversity and relationships analysis for 20 persimmon varieties.

Genomic DNA extraction and SSR-PCR amplification

The total genomic DNA was extracted from fresh leaves using a CTAB method. The DNA quality and concentration were tested by 1% (w/v) agarose gel electrophoresis and UV-VIS spectrophotometer, respectively. PCR amplification reactions were carried out according to the previously established protocols by Yang et al. (2015a) in 25 µL volume, containing 50 ng DNA, 0.2 mM dNTPs, 1.5 pM aliquots of forward and reverse primers, 2 mM Mg²⁺, 1 U Taq DNA polymerase (TaKaRa Biotechnology Dalian Co., Ltd., China), and 1x Taq Buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl). PCR amplification was performed with following conditions: initial denaturation at 94°C for 5 min; 30 cycles at 94°C for 45 s, the primer-specific annealing temperature for 45 s, and 72°C for 90 s; and 72°C for 7 min. PCR products were separated by electrophoresis on denaturing 6% polyacrylamide gels and visualized by silver staining. The molecular size of the amplified fragments was estimated using a 10-bp DNA ladder (TaKaRa Biotechnology Dalian Co., Ltd., China).

Results

Transcriptome generation and *de novo* assembly

A total of 5,565,270,750 raw reads were generated from Hi-Seq™ 2000 (Illumina) paired-end sequencing of the persimmon. The total length of the reads was approximately 5.6 Gigabase pairs (Gb); a total 44,522,166 high-quality clean reads were obtained after removing the low-quality reads, the coverage percentage of reads blasted unigenes are listed in Supplementary Figure S1. The percentage of Q20 (the bases number > 20), N and GC% were 95.62%, 0.00% and 47.80%, respectively. Then all high-quality reads were assembled by the Trinity program, and a total of 48,012 unigenes with an average length of 835 bp and N50 length of 1,449 bp were obtained. The unigenes ranged from 201 bp to 13,906 bp (Table 1). 34,687 unigenes (72.25%) ranged from 201 to 1,000 bp, and 8,701 unigenes (18.12%) were longer than 1,000 bp, 4,624 unigenes (9.63%) more than 2,000 bp. The coverage percentage of reads blasted unigenes were 53.43% (more than 10 reads), 28.03% (more than 100 reads), and 18.54% (more than 1,000 reads).

Sequence functional annotation

Unigenes function annotation could provide information on protein sequence similarities, COG clusters, gene ontology (GO) analysis and KEGG pathway. Function annotation of persimmon unigenes were performed in Nr, Swiss-Prot, KEGG, and COG databases by BLASTX (E value < 10⁻⁵). The E-value distribution of the top hits in Nr, Swiss-Prot, KEGG, and COG databases were listed in Supplementary Figure S2.

A total of 27,586 (57.7% of all unigenes) unigenes were annotated. 27,546 (57.4% of all unigenes) unigenes of the largest match were annotated at the Nr database, followed by Swiss-Prot (19,315, 40.2%), COG (9,250, 19.3%) and KEGG (7,186, 15.0%) databases. And others (20,426, 42.5%) not annotated to four databases, indicating these unigenes may

be the novel genes (Table 1). For the annotated unigenes in the Nr databases, the homologous sequences belong to species being analyzed and listed in Supplementary Table S1. The largest match was '*Vitis vinifera*' (7,288 unigenes), followed by '*Theobroma cacao*' (5,439 unigenes), '*Solanum lycopersicum*' (2,311 unigenes), '*Cucumis sativus*' (1,498

TABLE 1. Summary of transcriptome data for persimmon.

Item	Number
Raw sequences and Assembly statistics	
Total reads	44522166
Total nucleotides (nt)	5565270750
GC content percentage	47.80%
Total unigenes (average length; N50; min-max length)	48012 (835; 1449; 201-13906)
Mean RPKM value of unigenes (min-max RPKM value)	15 (0-5044)
Bioinformatics annotations of mango fruit unigenes	
Protein coding sequence (CDS)	27291
Gene annotation against Nr (%)	27546 (57.7)
Gene annotation against Swiss-Prot (%)	19315 (40.2)
Gene annotation against KEGG (%)	7186 (15.0)
Gene annotation against COG (%)	9250 (19.3)
All unigenes annotated (%)	27586 (57.5)
Not annotation unigenes	20426 (42.5)
GO Ontology (%)	
Biological process category	31633
Cellular component category	25123
Molecular function category	14317

FIGURE 1. COG function classification of the persimmon transcriptome. There are twenty-four categories. The X-axis indicates function class, and the Y-axis indicates the number of unigenes.

FIGURE 2. CO function classification of the persimmon transcriptome. There are three main GO categories: biological process, cellular component, and molecular function. The X-axis indicates the categories, and the Y-axis indicates the number of the unigenes.

unigenes), '*Oryza sativa Japonica Group*' (1,474 unigenes), '*Fragaria vesca* subsp. *vesca*' (1,155 unigenes), '*Arabidopsis thaliana*' (1,026 unigenes), '*Medicago truncatula*' (963 unigenes), '*Cicer arietinum*' (861 unigenes), '*Glycine max*' (860 unigenes) and other species (4,671 unigenes).

COG and GO functional classifications

In order to predict and classify possible functions of the persimmon unigenes, COG classification was searched based on BLAST against COG database. A total of 14,275 unigenes were assigned to 24 function classifications (Figure 1, Supplementary Table S2). For the 24 COG categories, 'General function prediction only' was the largest group (2,570, 18.00%). 'Replication, recombination and repair' (1,702, 11.92%), 'transcription' (1,295, 9.07%), 'posttranslational modification, protein turnover, chaperones' (1,200, 8.41%), 'signal transduction mechanisms' (1,152, 8.07%) also had a high percentage of genes. 'Translation, ribosomal structure and biogenesis' (748, 5.24%), 'carbohydrate transport and metabolism' (726, 5.09%), 'amino acid transport and metabolism' (647, 4.53%), 'energy production and conversion' (515, 3.61%) shared a few genes, and 3,720 unigenes were assigned to other functional categories.

GO is an international standardized gene functional classification system that has three ontologies: molecular function, cellular component and biological process. GO functional analysis could provide classification annotation for differentially expressed genes. The GO analyses were conducted using the Blast2GO program. We used GO assignments to classify the functions of the persimmon unigenes, 71,073 unigenes (biological process: 31,633 unigenes; molecular function: 14,317; cellular component: 25,123) were assigned to 45 functional groups (Figure 2, Supplementary Table S3). For the biological process, the most groups included 'metabolic process' (7,610, 24.06%), 'cellular process' (7,410, 23.42%), 'response to stimulus' (3,109, 9.83%). 'binding' (6,517, 45.52%) and 'catalytic activity' (6,465, 45.16%) were the most highly represented groups in the molecular function category. For the cellular component category, 'cell' (7,809, 31.08%), 'cell part' (7,809, 31.08%), and 'organelle' (5,669, 22.56%) represented the largest proportion.

Metabolic pathway annotations

The pathway annotations could analyze the biological functions of genes. In this study, 7,186 annotated sequences were assigned to 124 KEGG pathways (Supplementary Table S4). The largest number of pathways was 'metabolic pathways' (1,802, 25.08%), followed by 'biosynthesis of secondary metabolites' (881, 12.26%), and 'ribosome' (287, 3.99%). These pathways may provide a valuable resource for studying functions and pathways in persimmon research. 3,989 unigenes were mapped to other pathways. Additionally, 227 (3.16%) unigenes involved in plant hormone signal transduction were found, which included 8 pathways (Supplementary Table S5), controlling the signal transduction of auxin, cytokine, gibberellin, abscisic acid, ethylene, brassinosteroid, jasmonic acid, and salicylic acid. These hormones regulate plant growth, shoot initiation, seed dormancy, fruit ripening, and stress response and disease resistance. We also found an important circadian rhythm pathway (49 unigenes), controlling the photoperiod physiological activities. The circadian clock is an important part of the photoperiod pathway, and also plays an important role for controlling plant physiology. In this study, we found a circadian rhythm pathway including 21 (0.29%) unigenes. This pathway may help us to understand the photoperiod pathway on persimmon flowering-related processes.

Searching the genes putatively related to floral development

Due to Deyangshi might be independent new species of *Diospyros* L. (Ebenaceae). The main feature of Deyangshi is its short juvenility and that it has a different attractive flower color that is red from other species of *Diospyros* L. (Figure 3). Therefore, in this study we mainly focused on the genes related to flower color development. According to the annotation of unigenes, we obtained one key unigene of anthocyanin biosynthesis pathway (Unigene0033663), this gene involved with three flavonoid biosynthesis, including pelargonidin, cyanidin, and delphinidin, at the same time. Another important unigene of flavonoid biosynthesis pathway was found, including thirty unigenes. In addition, the genes of flavone and flavonol biosynthesis pathway were also identified in our Deyangshi transcriptome database. For example, the

FIGURE 3. Flower colour of Deyangshi. A. Male flower; B. Female flower; C. The branch of male in flower period; D. Ovary of female.

unigenes of anthocyanin biosynthesis, flavone and flavonol biosynthesis pathway were listed in Supplementary Table S6. This is important information for the study of flower color development in the future.

Development of SSR markers in persimmon

In this study, 48,012 unigenes were used to develop new SSR markers with MISA software. A total of 8,741 SSRs were identified from 7,074 unigenes. 1,325 unigenes included more than one SSR. For the 8,741 SSRs, di-nucleotide motifs were the most abundant form (5,166, 59.10%), followed by tri-nucleotide (2,563, 29.32%), tetra-nucleotide (660, 7.55%), penta-nucleotide (190, 2.17%), and hexa-nucleotide (162, 1.85%). In addition, the number of repeat units of SSRs was analyzed. The number of repeat units of the di-nucleotide motifs ranged from 6 to 12, and the tri-nucleotide, tetra-nucleotide, penta-nucleotide, and hexa-nucleotide motifs included 5 to 10, 4 to 10, 4 to 5, and 4 to 8, respectively. SSRs with six tandem repeats were the most (2,284, 26.13%), followed by five tandem repeats (1,595, 18.25%), seven tandem repeats (1,461, 16.71%), eight tandem repeats (994, 11.37%), and so on (Table 2). Furthermore, the motif types of SSRs were analyzed. The most type of SSRs was AG/CT (44.2%), followed by AAG/CTT (10.5%), AT/AT (9.6%), AC/CT (5.1%), AGG/CCT (3.6%), AGC/CTG (3.4%), ATC/ATG (2.8%) and CCG/CGG (2.8%).

FIGURE 4. Examples of polymorphic products amplified by different SSR primer pairs. 1-25 represent different primer pairs. M represents DL2000 DNA markers.

TABLE 2. Summary of the different repeat units of identified EST-SSRs.

Number of repeat unit	Di-	Tri-	Tetra-	Penta-	Hexa-	Total	Percentage (%)
4	0	0	472	176	145	793	9.07
5	0	1,414	157	14	10	1,595	18.25
6	1,450	802	27	0	5	2,284	26.13
7	1,138	320	2	0	1	1,461	16.71
8	966	26	1	0	1	994	11.37
9	961	0	0	0	0	961	10.99
10	554	1	1	0	0	556	6.36
11	94	0	0	0	0	94	1.08
12	3	0	0	0	0	3	0.03
Total	5,166	2,563	660	190	162	8,741	100

Confirm validity of SSR markers

In order to confirm the validity of SSRs, we randomly synthesized 100 pairs of primers and assessed in *Diospyros*. Twenty genotypes of persimmon were selected to evaluate the validity and universality of primers (Supplementary Table S7). These genotypes of persimmon showed abundant diversity in fruit size, shape and texture, which belong to the *Diospyros* spp., *D. kaki*, *D. sylvestris*, *D. lotus*, *D. virginiana*, *D. oleifera*, *D. cathayensis*, *D. spp.*, *D. rhombifolia*. The results showed that 89 pairs generated clear and abundant bands and 11 pairs of primers were able to produce bands without polymorphism or blurred bands. PCR amplification results of some different primer pairs, shown in Figure 4.

Discussion

Persimmon is one of the most important commercially cultivated fruit crops in the world (Akaqi et al., 2014), due to its high nutrients or for medical purposes (Yang et al., 2015a). It existed for 10,000 years in China (Jing et al., 2013). As we know, previous studies on persimmon focused on genetic diversity and relationships of different species. However, there is no report on construction of genetic linkage map and QTL mapping, for which the reasons may be the lack of genomic data, especially SSRs markers. Until now, the genome sequencing of some important fruit were completed, such as kiwifruit (Huang et al., 2013), jujube (Liu et al., 2014) and so on. Due to persimmon containing more than 400 species with ploidy levels ranging from diploid ($2n=2x=30$) to nonaploid ($2n=9x=135$) and the high costs of sequencing, it is difficult to complete the genome sequencing of persimmon. The transcriptome analysis based on Illumina HiSeq™ 2000 sequencing platform have provided an efficient tool to obtain lots of genomic data for some plants without a reference genome, such as wax gourd (Jiang et al., 2013), and pumpkin (Wu et al., 2014). Therefore, we sequenced the persimmon transcriptome to obtain some genomic data and then developed many SSR markers. These transcriptome data will provide some information for future studies on genetic breeding, and molecular biology. In the previous similar studies, the average length of the unigenes were 709 bp (wax gourd, Jiang et al., 2013), 765 bp (pumpkin, Wu et al., 2014), 652 bp (Siberian apricot, Dong et al., 2014), 473.4 bp (Chinese jujube, Li et al., 2014), 446 bp (safflower, Huang et al., 2012), respectively (Li et al., 2014; Zhang et al., 2013, 2016). In this study, the average length of unigenes showed a high level (835 bp); these results indicated that the quality of persimmon transcriptome sequencing and *de novo* assembly was high.

For gene annotation, the sequences of unigenes were blasted against several protein databases, including non-redundant (nr), Swiss-Port, KEGG, GO and COG protein database. A total of 27,586 unigenes (57.7% of all unigenes) were annotated in four protein databases, indicating that the transcriptomic data of persimmon may have a large diversity of transcripts. Additional, about 20,426 (42.5%) of unigenes were not annotated to four databases, suggesting that some unigenes may be new genes and unique to persimmon. For metabolic pathway annotations, we found that 227 (3.16%) unigenes were involved in plant hormone signal transduction. These unigenes were important for Deyangshi, because one feature of Deyangshi is its short juvenility and this feature of the flora may be regulated by these different hormones.

Deyangshi might be an independent new species of *Diospyros*. Deyangshi is a type of dioecious deciduous peren-

nial tree. The color of female plant and male plant flowers is red, however, the color of females is lighter than that of males; this feature is very different from other species. According to field survey, seeds of Deyangshi were sown in the following spring after low-temperature stratification, and seedlings began to develop their blossoms in the next year. This unusual morphology reveals its specificity in *Diospyros* genus. Further, on the shoots of top-grafted Deyangshi, female flowering is earlier than that of males, which may indicate that reproductive isolation has occurred and that the declining selfing depression phenomenon appeared (Zhang et al., 2016). Therefore, in this study we mainly focused on the floral development, such as flowering time and flower colors. The photoperiod and vernalization pathways are two important genetic pathways of flowering control (Zhang et al., 2013). In our study, circadian rhythm as one of the photoperiod pathways could be identified. The circadian rhythm could adjust plant development and regulates flowering time by providing a seasonal signal. SCF-dependent proteasomal ubiquitin-dependent protein catabolic is a key process in the circadian rhythm pathway. In this study, 16 persimmon SCF unigenes were found. Of them, some genes involved in regulation of SCF process (such as COL genes) could be found. Previous studies showed that COL genes could control flowering time (Griffiths et al., 2003).

Flavonols play an important role in determining flower color (Tan et al., 2013). Flavonoids are the major flower pigments (Martin and Gerats, 1993). The metabolism of the flavonoid pathway has been studied in some plants, such as *Arabidopsis*, *Petunia* and *Antirrhinum* (Tan et al., 2013). In addition, anthocyanin is the major component with effect on the flower development and color changes (Weiss, 2000). Interestingly, the metabolism of flavone and flavonol biosynthesis pathway, flavonoid biosynthesis and anthocyanin biosynthesis pathways were found in our Deyangshi transcriptome database that included a total of thirty-seven unigenes (Table 3). Flavonoids are catalyzed by phenylalanine ammonia-lyase (PAL). Then it was mediated by chalcone synthase (CHS), flavanone 3-hydroxylase (F3H), and fluxed into anthocyanin biosynthesis by dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS) and UDP-glucose: flavonoid 3-O-glucosyltransferase (UGT); fluxed into flavonol biosynthesis by flavonol synthase (FLS) following F3H; or fluxed into proanthocyanidin biosynthesis through anthocyanidin reductase (ANR) following ANS (Tan et al., 2013; Buer et al., 2010). In this study, sequence homologues of Deyangshi involved in regulation of flavonoids described above (F3H and UGT) could be identified in our database (Table 3). Therefore, these pathways may be tightly linked with flower development and color changes in persimmon, and these unigenes would facilitate more studies on the mechanism of floral development for Deyangshi.

Previous studies showed that SSR markers are highly polymorphic, codominant and easily reproducible (Yang et al., 2015a). Due to traditional methods for SSR marker development, they are time-consuming and expensive, nowadays only a set of thirty-one persimmon SSRs obtained through simple methods (Yang et al., 2015a). The Illumina paired-end sequencing technique is an efficient tool for transcriptome sequencing, and further these transcriptome data were used to develop SSR markers. The SSR markers for some plants have been reported based on transcriptome sequencing (Dong et al., 2014; Huang et al., 2012; Li et al., 2014; Wu et al., 2014). In our study, we identified a total of 8,741 SSRs from transcriptome data of persimmon. In order to assess

TABLE 3. Deyangshi unigenes that include homology with reported flower color genes.

Category	Gene ID	Homologous gene	Nr-accession
Anthocyanin biosynthesis pathway	Unigene0033663	f3galtase-1	AB435084.1
Flavone and flavonol biosynthesis pathway	Unigene0008687	O-methyltransferase	KJ830748
	Unigene0013849	Not found	Not found
	Unigene0019161	Not found	Not found
	Unigene0025311	F3'5'H	AB472676
	Unigene0029783	DkF3'H	AB472369
	Unigene0033663	f3galtase-1	AB435084

the quality of these SSR markers, we randomly selected 100 pairs of primers and assessed in twenty genotypes of persimmon, and 89% showed polymorphisms; these results were higher than in other plants, such as gourd, sweet-potato, and lentil. We believed that there new SSR markers will be used in genetic diversity, genetic mapping, and particularly marker-assisted breeding for persimmon.

Conclusions

This paper reports on the transcriptome characterizations of persimmon trees and provides a large number of SSR markers to elucidate the molecular biology of almond trees. To our knowledge, these developed SSR markers will significantly contribute to genetic diversity studies, QTL mapping, and marker-assisted selection breeding for persimmon. Notably, due to high transferability, these SSR markers may provide an efficient tool to accelerate molecular breeding in other *Diospyros* species.

Acknowledgments

This work was supported by Special Fund for Agro-scientific Research in Public Interest (201203047), "Project for Crop Germplasm Conservation" from China Agriculture Ministry (2015NWB021), Science and Technology Research and Development Program of Shaanxi Province, China (2016KTTSNY01-02, 2016KTTSNY01-06, 2016KJXX-58) and China Postdoctoral Science Foundation (2015M582712), Postdoctoral Science Foundation of Shaanxi Province (2016BSHYDZZ07).

References

Akaqi, T., Henry, I.M., and Tao, R. (2014). Comai L, an Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* 346(6209), 646–650. <https://doi.org/10.1126/science.1257225>.

Bremer, B., Bremer, K., Chase, M., Fay, M., Reveal, J., Soltis, D., Soltis, P., and Stevens, P. (2009). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. <https://doi.org/10.1111/j.1095-8339.2009.00996.x>.

Buer, C.S., Imin, N., and Djordjevic, M.A. (2010). Flavonoids: new roles for old molecules. *J. Integr. Plant. Biol.* 52, 98–111. <https://doi.org/10.1111/j.1744-7909.2010.00905.x>.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18), 3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>.

Dong, S., Liu, Y., Niu, J., Ning, Y., Lin, S., and Zhang, Z. (2014). *De novo* transcriptome analysis of the Siberian apricot (*Prunus sibirica* L.) and search for potential SSR markers by 454 pyrosequencing. *Gene* 544, 220–227. <https://doi.org/10.1016/j.gene.2014.04.031>.

Gorinstein, S., Zachwieja, Z., Folta, M., Barton, H., Piotrowicz, J., Zemser, M., Weisz, M., Trakhtenberg, S., and Martín-Belloso, O. (2001). Comparative contents of dietary fiber, total phenolics, and minerals in persimmons and apples. *J. Agric. Food. Chem.* 49(2), 952–957. <https://doi.org/10.1021/jf000947k>.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, Z., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>.

Griffiths, S., Dunford, R.P., Coupland, G., and Laurie, D.A. (2003). The evolution of *CONSTANS*-like gene families in barley, rice, and Arabidopsis. *Plant Physiol.* 131, 1855–1867. <https://doi.org/10.1104/pp.102.016188>.

Guo, D.L., and Luo, Z.R. (2006). Development of SSR primers using ISSR-PCR in *Diospyros kaki* Thunb. *Mol. Ecol. Notes.* 6, 886–887. <https://doi.org/10.1111/j.1471-8286.2006.01387.x>.

Guo, D.L., and Luo, Z.R. (2008). Microsatellite isolation and characterization in Japanese persimmon (*Diospyros kaki*). *Biochem. Genet.* 46, 323–328. <https://doi.org/10.1007/s10528-008-9143-0>.

Huang, L.L., Yang, X., Sun, P., Tong, W., and Hu, S.Q. (2012). The first Illumina based *de novo* transcriptome sequencing and analysis of safflower flowers. *PLoS ONE* 7, e38653. <https://doi.org/10.1371/journal.pone.0038653>.

Huang, S., Ding, J., Deng, D., Tang, W., Sun, H., Liu, D., Zhang, L., Niu, X., Zhang, X., Meng, M., Yu, J., Liu, J., Han, Y., Shi, W., Zhang, D., Cao, S., Wei, Z., Cui, Y., Xia, Y., Zeng, H., Bao, K., Lin, L., Min, Y., Zhang, H., Miao, M., Tang, X., Zhu, Y., Sui, Y., Li, G., Sun, H., Yue, J., Sun, J., Liu, F., Zhou, L., Lei, L., Zheng, X., Liu, M., Huang, L., Song, J., Xu, C., Li, J., Ye, K., Zhong, S., Lu, B.R., He, G., Xiao, F., Wang, H.L., Zheng, H., Fei, Z., and Liu, Y. (2013). Draft genome of the kiwifruit *Actinidia chinensis*. *Nat. Commun.* 4, 2640. <https://doi.org/10.1038/ncomms3640>.

Jiang, B., Xie, D., Liu, W., Peng, Q., and He, X. (2013). *De novo* assembly and characterization of the transcriptome, and development of SSR Markers in Wax gourd (*Benicasa hispida*). *PLoS ONE* 8(8), e71054. <https://doi.org/10.1371/journal.pone.0071054>.

Jing, Z.B., Ruan, X.F., Wang, R.Z., and Yang, Y. (2013). Genetic diversity and relationships between and within persimmon (*Diospyros* L.) wild species and cultivated varieties by SRAP markers. *Plant. Syst. Evol.* 299, 1485–1492. <https://doi.org/10.1007/s00606-013-0810-1>.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–484. <https://doi.org/10.1093/nar/gkm882>.

Li, Y., Xu, C., Lin, X., Cui, B., Wu, R., and Pang, X. (2014). *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, e106438. <https://doi.org/10.1371/journal.pone.0106438>.

Liu, M.J., Zhao, J., Cai, Q.L., Liu, G.C., Wang, J.R., Zhao, Z.H., Liu, P., Dai, L., Yan, G.J., Wang, W.J., Li, X.S., Chen, Y., Sun, Y.D., Liu, Z.G., Lin, M.J., Xiao, J., Chen, Y.Y., Li, X.F., Wu, B., Ma, Y., Jian, J.B., Yang, W., Yuan, Z., Sun, X.C., Wei, Y.L., Yu, L.L., Zhang, C., Liao, S.G., He, R.J., Guang, X.M., Wang, Z., Zhang, Y.Y., and Luo, L.H. (2014). The complex jujube genome provides insights into fruit tree biology. *Nat. Comm.* 5, 5315. <https://doi.org/10.1038/ncomms6315>.

Martin, C., and Gerats, T. (1993). Control of pigment biosynthesis genes during petal development. *Plant Cell* 5, 1253–1364. <https://doi.org/10.2307/3869778>.

Soriano, J.M., Pecchioli, S., Romero, C., Vilanova, S., Llacer, G., Giordani, E., and Badenes, M.L. (2006). Development of microsatellite markers in polyploid persimmon (*Diospyros kaki* L.) from an enriched genomic library. *Mol. Ecol. Notes* 6(2), 368–370. <https://doi.org/10.1111/j.1471-8286.2006.01236.x>.

Tan, J., Wang, M., Tu, L., Nie, Y., Lin, Y., and Zhang, X.L. (2013). The flavonoid pathway regulates the petal colors of cotton flower. *PLoS ONE* 8(8), e72364. <https://doi.org/10.1371/journal.pone.0072364>.

Weiss, D. (2000). Regulation of flower pigmentation and growth: Multiple signaling pathways control anthocyanin synthesis in expanding petals. *Physiol. Plant.* 110, 152–157. <https://doi.org/10.1034/j.1399-3054.2000.110202.x>.

Wu, T.Q., Luo, S.B., Wang, R., Zhong, Y.J., Xu, X.M., Lin, T.E., He, X.M., Sun, B.J., and Huang, H.X. (2014). The first Illumina-based *de novo* transcriptome sequencing and analysis of pumpkin (*Cucurbita moschata* Duch.) and SSR marker development. *Mol. Breeding* 34, 1437–1447. <https://doi.org/10.1007/s11032-014-0128-x>.

Yang, Y., Jing, Z.B., Ruan, X.F., and Cheng, J.M. (2015a). Development of simple sequence repeat markers in persimmon (*Diospyros* L.) and their potential use in related species. *Genet. Mol. Res.* 14(1), 609–618. <https://doi.org/10.4238/2015.January.30.2>.

Yang, Y., Yang, T.T., and Jing, Z.B. (2015b). Genetic diversity and taxonomic studies of date plum (*Diospyros lotus* L.) using morphological traits and SCoT markers. *Biochem. Syst. Ecol.* 61, 253–259. <https://doi.org/10.1016/j.bse.2015.06.008>.

Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., and Wang, J. (2006). WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 34, W293–297. <https://doi.org/10.1093/nar/gkl031>.

Zhang, J.X., Wu, K.L., Zeng, S.J., de Silva, J.A.T., Zhao, X.L., Tian, C.E., Xia, H.Q., and Duan, J. (2013). Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. *BMC Genomics* 14, 279. <https://doi.org/10.1186/1471-2164-14-279>.

Zhang, Y.F., Yang, Y., Guo, J., Hu, C.Q., and Zhu, R.S. (2016). Taxonomic status of Deyangshi based on chromosome number and SRAP markers. *Sci. Hortic.* 207, 57–64. <https://doi.org/10.1016/j.scienta.2016.05.006>.

Zhou, X.J., Wang, Y.Y., Xu, Y.N., Yan, R.S., Zhao, P., and Liu, W.Z. (2015). *De novo* characterization of flower bud transcriptomes and the development of EST-SSR markers for the endangered tree *Tapiscia sinensis*. *Int. J. Mol. Sci.* 16, 12855–12870. <https://doi.org/10.3390/ijms160612855>.

Addresses of authors:

Zhaobin Jing^{1,2}, Xiaofeng Ruan¹ and Yong Yang^{1,*}

¹ College of Horticulture, Northwest A&F University, Yangling 712100, Shaanxi, China

² Weinan Vocational and Technical College, Weinan Fruit Industry Institute of Shaanxi, Weinan 714026, Shaanxi, China

* Corresponding author; E-mail: Yang_Yong@nwsuaf.edu.cn
Tel.: 86-15353590637; Fax: 86-29-87081351

Supplementary Tables S1–S7 and Supplementary Figures S1–S2 are published on www.ishs.org.

Received: Mar. 9, 2018

Accepted: Jul. 26, 2019