

Article

De Novo Transcriptome Analysis of *Dalbergia odorifera* T. Chen (Fabaceae) and Transferability of SSR Markers Developed from the Transcriptome

Fu-Mei Liu ^{1,2} , Zhou Hong ^{3,*}, Zeng-Jiang Yang ³, Ning-Nan Zhang ³, Xiao-Jin Liu ³  and Da-Ping Xu ³

¹ State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin 150040, China; liufumei1115@163.com

² The Experimental Centre of Tropical Forestry, Chinese Academy of Forestry, Pingxiang 532600, China

³ Research Institute of Tropical Forestry, Chinese Academy of Forestry, Longdong, Guangzhou 510520, China; yzengjiang@126.com (Z.-J.Y.); ningnanzhang@126.com (N.-N.Z.); xjliucafe@163.com (X.-J.L.); gzfsrd@163.com (D.-P.X.)

* Correspondence: hzhou1981@caf.ac.cn; Tel.: +86-20-8703-1037

Received: 15 December 2018; Accepted: 24 January 2019; Published: 26 January 2019



Abstract: *Dalbergia odorifera* T. Chen (Fabaceae), indigenous to Hainan Island, is a precious rosewood (Hainan hualimu) in China. However, only limited genomic information is available which has resulted in a lack of molecular markers, limiting the development and utilization of the germplasm resources. In this study, we aim to enrich genomic information of *D. odorifera*, and develop a series of transferable simple sequence repeat (SSR) markers for *Dalbergia* species. Therefore, we performed transcriptome sequencing for *D. odorifera* by pooling leaf tissues from three trees. A dataset of 138,516,418 reads was identified and assembled into 115,292 unigenes. Moreover, 35,774 simple sequence repeats (SSRs) were identified as potential SSR markers. A set of 19 SSR markers was successfully transferred across species of *Dalbergia odorifera* T. Chen, *Dalbergia tonkinensis* Prain, and *Dalbergia cochinchinensis* Pierre ex Laness. In total, 112 alleles (3–13 alleles/locus) were presented among 60 *Dalbergia* trees, and polymorphic information content ranged from 0.38 to 0.75. The mean observed and mean expected heterozygosity was 0.34 and 0.40 in *D. odorifera*, 0.27 and 0.32 in *D. tonkinensis*, and 0.29 and 0.33 in *D. cochinchinensis*, respectively. The cluster analysis classified these 60 trees into three major groups according to the three *Dalbergia* species based on the genetic similarity coefficients, indicating these newly developed transferable markers can be used to explore the relationships among *Dalbergia* species and assist genetic research. All these unigenes and SSR markers will be useful for breeding programs in the future.

Keywords: *Dalbergia odorifera* T. Chen; *Dalbergia tonkinensis* Prain; *Dalbergia cochinchinensis* Pierre ex Laness; molecular marker; cross-amplification; rosewood

1. Introduction

The genus *Dalbergia* belongs to the Fabaceae which is a member of Angiosperms, and comprises about 100 species of trees, shrubs, and woody lianas [1]. Many tree species in this genus are of great values due to their precious heartwood, which are used in the high-grade furniture and crafts markets. Their population sizes have been dramatically reduced due to overexploitation and illegal logging. Often, the most highly valued timber trees are the most threatened species in their native habitats [2]. Examples include *Dalbergia tonkinensis* Prain, *Dalbergia cochinchinensis* Pierre ex Laness, and *Dalbergia*

odorifera T. Chen. All the three *Dalbergia* tree species have been listed on the International Union for Conservation of Nature's (IUCN) red list of threatened species since 1998 [3–5]. *Dalbergia tonkinensis* is a medium-sized floral species with a height of 5–13 m, located in Vietnam and Hainan Island of China [5]. *Dalbergia cochinchinensis* is a hardwood tree species, distributed in the remaining forested areas of Thailand, Laos, Cambodia, and Vietnam [6]. *Dalbergia odorifera* is a semi-deciduous perennial tree species, confined to a relatively narrow tropical geographic area of Hainan Island in China [7]. The three endangered species play important roles in the biodiversity of their habitats, let alone their ecological and economic importance. Therefore, it is imperative to establish efficient strategies for conservation and sustainable use of the three *Dalbergia* tree species. The first essential part is a comprehensive knowledge of their genetic diversity, especially for *D. odorifera* where only limited numbers of individuals are found in parts of their original habitat [7].

Molecular markers have proven to be an important and effective tool for genetic diversity analysis and marker-assisted selection [8–12]. However, there is a significant shortage of molecular markers for *D. odorifera*, resulting from the scarce genomic information with only 112 expressed sequence tag (EST) sequences recorded in the public GenBank sequence database (<http://www.ncbi.nlm.nih.gov/genbank/>). Currently, reported molecular markers are restricted to six random amplified polymorphic DNA (RAPD) markers [7] and 25 sequence-related amplified polymorphism (SRAP) markers [13]. Yet there are no reports on microsatellite markers (SSR markers) or simple sequence repeats (SSRs) which are widely used in diverse areas, particularly in genetic diversity and structure studies [14–17].

Microsatellite markers are the ideal choice compared to other molecular markers for their advantages: abundance, high polymorphism, co-dominance, reproducibility, and transferability to related species [18]. According to the origins of SSR markers, they are generally categorized into genomic SSRs of genomic sequences and EST-SSRs of transcribed RNA sequences. However, the discovery of genomic SSRs and development using traditional methods are laborious, time-consuming, and costly [18]. The identification and development of EST-SSRs have been revolutionized along with the development of next-generation sequencing-based transcriptome sequencing (RNA-seq), especially for species with no reference genome, and extensive similar studies have been performed for many tree species [19–21]. Additionally, transcriptomics is also a powerful tool for genome-wide analysis of RNA transcripts, allowing not only genomic data mining, but also facilitating genetic and molecular breeding approaches for endangered species [22,23]. Therefore, we presented the first transcriptome of *D. odorifera* prepared using the Illumina Hi Seq sequencing platform. The main objectives of this study were to (1) provide high-quality transcriptome data and enrich the current knowledge of the genome background for *D. odorifera*, (2) identify a large number of SSR markers based on the *D. odorifera* transcriptome information obtained, (3) develop, validate, and transfer the identified SSR markers across *Dalbergia* species, and (4) evaluate the genetic relationships among 60 individuals from the three *Dalbergia* species (*D. odorifera*, *D. tonkinensis*, and *D. cochinchinensis*) using these validated SSR markers. And all these findings will provide useful information for breeding, hybridization, and conservation of *Dalbergia* germplasm.

2. Materials and Methods

2.1. Plant Materials and Preparation

Leaf samples were collected from three *Dalbergia* tree species including *D. odorifera*, *D. tonkinensis*, and *D. cochinchinensis*. The sampled *D. odorifera* individuals were located in the Hainan Island of China, the *D. tonkinensis* individuals were in Vietnam, and the *D. cochinchinensis* individuals were in Thailand. The latitudes and longitudes of the collection sites were recorded with a portable GPS (PokeNavi map21EX; Empex Instruments, Tokyo, Japan) (Table 1). In total, 60 *Dalbergia* individuals (20 samples from each *Dalbergia* species) were used to analyze the genetic diversity. Ten leaves were harvested from each individual and sealed in plastic bags with desiccant. Then total genomic DNA was extracted using the Hi-DNA-secure Plant Kit (Tiangen, Beijing, China) according to the

manufacturer's instructions. After determining the quality and concentration using a NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA), the DNA was finally stored at -20°C for use.

Table 1. Geographical location of the three *Dalbergia* species.

Species	Code	location	Size	Latitude (N)	Longitude (E)	Altitude (m)
<i>D. odorifera</i>	H1-H20	Hainan Island, China	20	18°40'–20°32'	108°37'–110°45'	5–250
<i>D. tonkinensis</i>	1-20	Vietnam	20	13°49'–21°52'	104°55'–108°07'	8–324
<i>D. cochinchinensis</i>	J1-J20	Thailand	20	13°41'–16°49'	100°16'–102°49'	7–280

2.2. Transcriptome Sequencing, Assembly and Annotation

Three young leaf-samples from three *D. odorifera* trees were used for transcriptome sequencing on April 30, 2017. The three samples were DO27, DO98 and DO100, and collected from three different wild populations in Haikou city, Dongfang city, and Changjiang autonomous county of Hainan island (China), respectively. Total RNA was isolated from each sample using TRIzol reagent (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's instructions. The RNA quality and integrity were accurately confirmed using the NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA) and the Bioanalyzer 2100 system (Agilent Technologies, Columbia, MD, USA). Three cDNA libraries were built from each RNA sample following the Illumina protocol. The transcriptome sequencing was conducted at Beijing Novogene Biological Information Technology Co., Ltd., Beijing, China (<http://www.novogene.com/>). The generated 100 bp paired-end reads were filtered by strict principles [24,25], and the assembly of these high-quality reads were carried out by Trinity [26,27] with min_kmer_cov set to 2 and other parameters set to default. The longest transcripts of each gene were pooled as “unigenes”, and unigene annotations were performed and analyzed by searching and comparing the public databases [25]. Additionally, the transcript sequence data used here were submitted to the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>). The accession number is SRP175426 (SRR8398210 for DO27, SRR8398212 for DO98, and SRR8398211 for DO100).

2.3. SSR Identification and Marker Development

Micro Satellite (MISA; <http://pgrc.ipk-gatersleben.de/misa>) was used to detect, locate and identify SSRs. The minimum number of motifs used to select the SSRs was ten for mono-nucleotide repeats, and six for di-nucleotide motifs, five for tri-, tetra-, penta-, and hexa-nucleotide repeats. Primers were designed using Primer 3.0 [28] by default settings and selected according to the following criteria: primer lengths of 18–24 bases, GC content of 40–60%, annealing temperature of 56–62 °C, and the predicted PCR product sizes of 150–300 bp.

2.4. SSR Marker Validation, Transferability, and PCR Conditions

Subsequently, 192 SSR markers were randomly selected to test the transferability among the three *Dalbergia* species, and first amplified using three leaf samples (one each *Dalbergia* species) by PCR as described below. The PCR reactions were performed in 10 µL final volume, containing 5 µL water, 1 µL 10 × DNA polymerase buffer, 1.5 µL MgCl₂ (10 mM), 0.8 µL dNTPs (2.5 mM each), 0.3 µL of each primer at 5 µM, 0.1 µL Taq polymerase at 3 units/µL (TaqUBA), and 1 µL of genomic DNA (40–50 ng). Thermocyclers were programmed as follow: pre-denaturation at 94 °C for 4 min, followed by 35 cycles of 94 °C for 30 sec, appropriate annealing temperature for 30 sec, and 72 °C for 30 sec, and a final step at 72 °C for 5 min. The PCR products in *D. tonkinensis* and *D. cochinchinensis*, with clear, stable, and specific bands were considered as successful PCR amplifications. In *D. odorifera*, the clear, stable, and specific bands with approximately expected lengths of 100–350 bp (i.e., 50 nt longer or shorter than the predicted PCR product size) were considered as successful PCR amplifications. All the PCR reactions were repeated at least once, to confirm positive results and/or detect false negatives due to technical issues. Finally, a set of 19 SSR markers was randomly selected from the validated ones and 60 *Dalbergia*

leaf samples were used to confirm the polymorphism, and their diluted PCR products were separated by capillary electrophoresis.

2.5. Analysis of Marker Polymorphism and *Dalbergia* Genetic Relationship

Allele sizes of polymorphic SSR markers were genotyped using ABI3730 (Applied Biosystems, Foster, CA, USA) and Gene Mapper v4.0 (Applied Biosystems, Foster, CA, USA). POPGENE v1.3.2 software [29] was used to estimate the following statistics: the number of alleles (N_a), observed (H_o) and expected heterozygosities (H_e), and the number of allele frequencies, respectively. The polymorphism information content (PIC) was calculated for each locus [30], so did the percentage of polymorphic loci (PPBs) for each species. A pair-wise similarity matrix was constructed using the SM similarity coefficient (SSC), based on which, an un-weighted pair group method with arithmetic mean (UPGMA) tree was constructed to reveal the relationship among these 60 *Dalbergia* trees using NTSYS-pc software (version 2.1) [31]. Tree confidence was also analyzed using the NTSYS-pc software.

3. Results

3.1. Transcriptome Assembly and Annotation

To select different genotypes and achieve a comprehensive overview of *D. odorifera* for SSR identification, leaves from three *D. odorifera* trees (DO27, DO98, and DO100) were used for Illumina draft reads and sequence assembly. In total, more than 143 million raw reads were obtained via the deep paired-end sequencing (Table 2). Followed rigorous quality control, 138,516,418 clean Illumina reads were obtained, of which, 40,896,098 was for DO27, 43,885,716 for DO98, and 53,734,604 for DO100, with Q20 ranging from 97.46% to 98.07% and GC contents from 44.72% to 45.09%.

Table 2. Summary of *D. odorifera* base quality.

Sample	Raw Reads	Clean Reads	Clean Bases	Error (%)	Q20 (%)	Q30 (%)	GC (%)
DO27	42,267,758	40,896,098	6.13G	0.01	97.46	93.94	45.09
DO98	45,415,150	43,885,716	6.58G	0.01	98.07	95.26	45.01
DO100	55,666,026	53,734,604	8.06G	0.01	97.86	94.9	44.72

In total, 154,854 individual transcripts and 115,292 unigenes were obtained with a mean length of 881 bp (N_{50} = 1634 bp) and 676 bp (N_{50} = 1080 bp), respectively. The length of unigenes ranged from 201 to 12,117 bp with a total of 77,979,362 nucleotides. Among these unigenes, 63.95% (73,730) had lengths ranging from 201 to 500 bp, 18.76% (21,634) from 501 to 1000 bp, 10.48% (12,091) from 1001 to 2000 bp, and 7837 (6.80%) with lengths longer than 2000 bp (Figure 1).

After the functional annotation, 56,898 of all the 115,292 unigenes were successfully annotated in at least one of the seven databases, and 7508 were annotated in all the databases (Table S1). A total of 43,597 unigenes showed significant BLAST (basic local alignment search tool) hits in the Nr database and 99.86% (43,537) showed significant similarity to known proteins. Additionally, 36,710 (84.20%) presented similarity to known proteins in the Swiss-Prot database.

Based on Nr annotations, 33,219 unigenes assigned to gene ontology (GO) terms clustered into biological process, cellular component and molecular function categories, including 56 sub-categories (Figure 2a). In the biological process category, cellular process (18,131, 54.58%), metabolic process (17,343, 52.21%), and single-organism process (13,536, 40.76%) were the most abundant of the 24 sub-groups. Within the cellular component category, the cell (28.48%) and cell part (28.47%) components, both with unigenes more than 9400, were the most abundant terms of all the 22 sub-groups, whereas only a few unigenes were assigned to synapse part (4), synapse (4), symplast (2), extracellular matrix component (1) and nucleoid (1). Of the last 10 sub-groups in the molecular function category, the binding and catalytic activity were prominently represented, assigned with

17,712 and 14,669 unigenes, respectively, whereas the metallochaperone activity sub-group contained only eight unigenes.

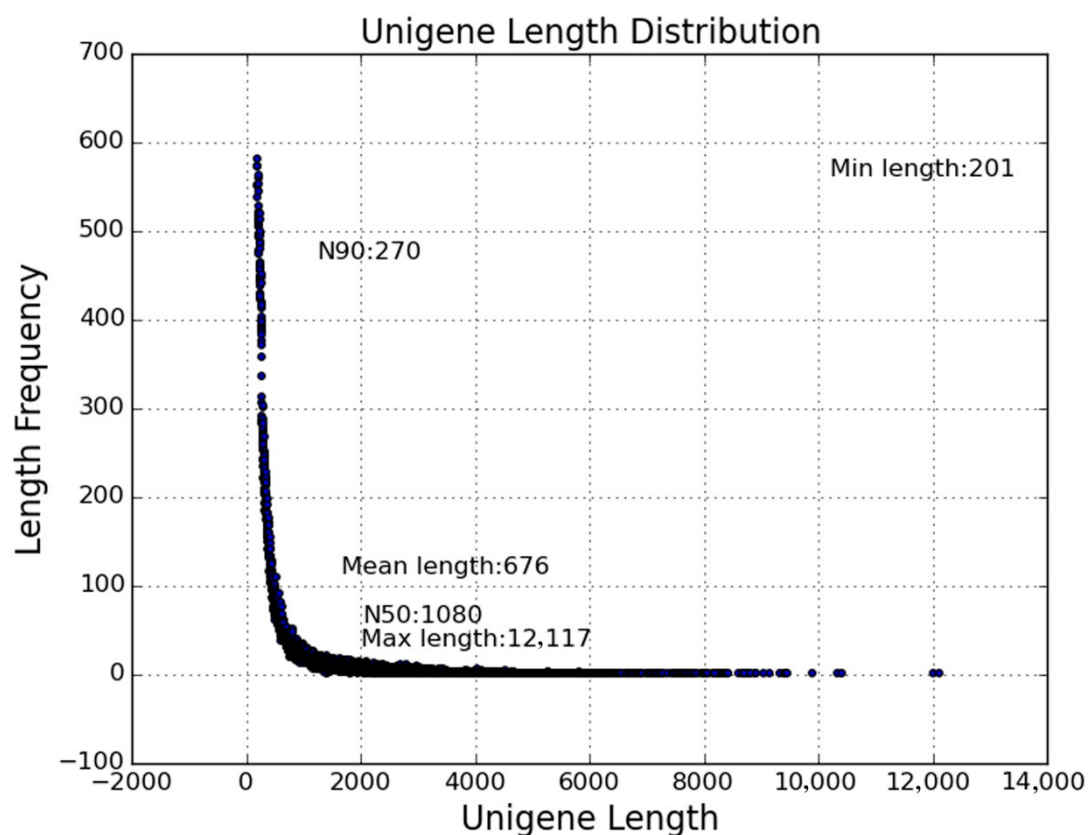


Figure 1. Unigenes length distribution in *D. odorifera* transcriptome. Horizontal and vertical axes show the size and number of unigenes, respectively.

According to the KOG (eukaryotic ortholog groups) functional classification, 17,860 assigned unigenes were classified into 26 functional groups (Figure 2b). The largest group was general function prediction with 2987 unigenes (16.72%), followed by posttranslational modification, protein turnover, chaperones (2352, 13.17%), and signal transduction mechanisms (2018, 11.30%).

To characterize the active biological pathways in *D. odorifera*, we performed a kyoto encyclopedia of genes and genomes analysis (KEGG). The classification showed 16,571 unigenes were annotated to 129 pathways, and categorized into 5 clades (cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems) and 19 sub-groups (Figure 2c). Of the five clades, cellular processes contained only one sub-group of transport and catabolism (877 unigenes), so did organismal systems with the sub-group of environmental adaptation (680 unigenes). Among the 19 sub-groups, carbohydrate metabolism (1775) and translation (1670) were the most represented pathways.

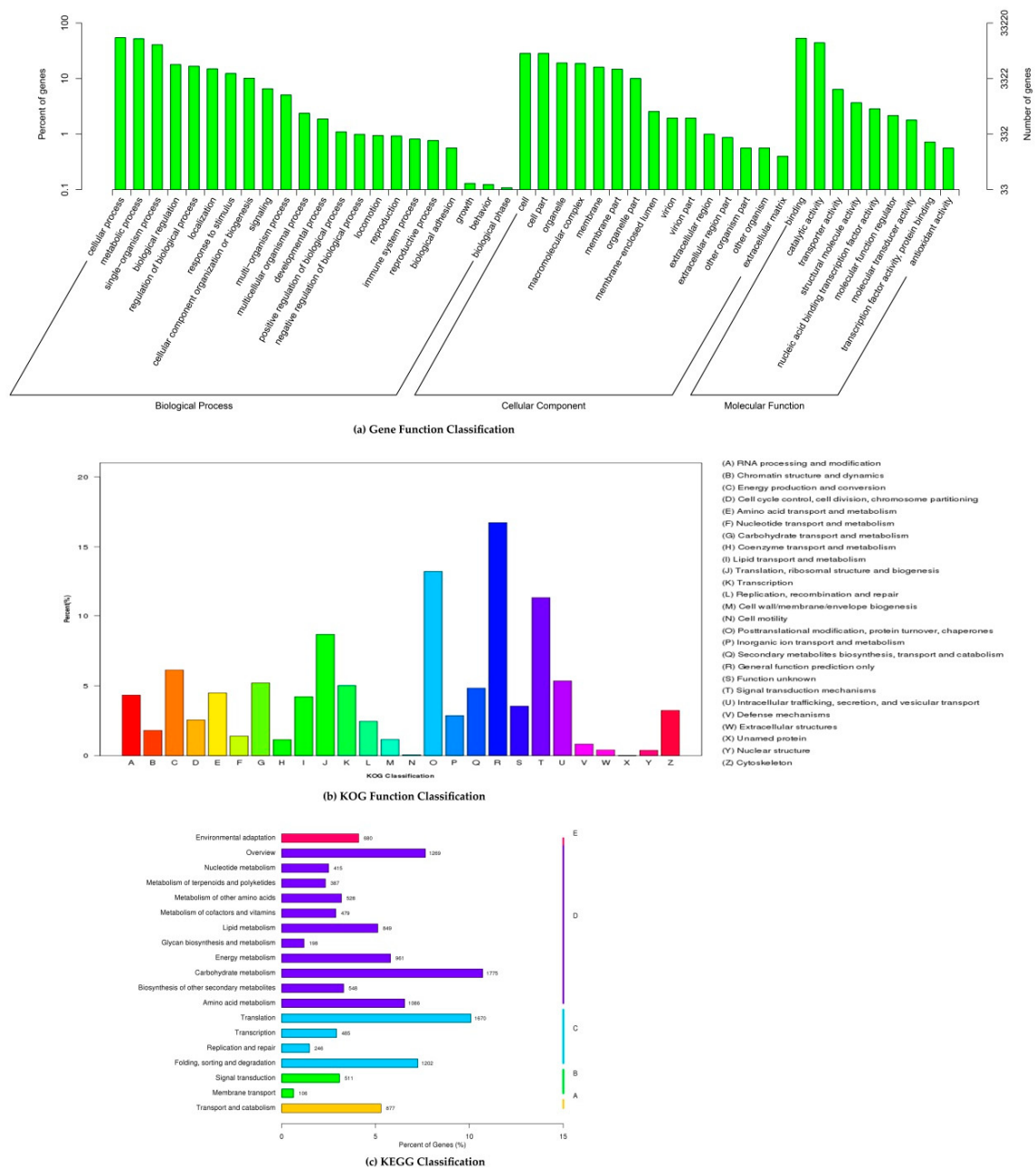


Figure 2. Functional classifications of *D. odorifera* unigenes. **(a)** Gene ontology (GO) classification of assembled unigenes. The left *y*-axis indicates the percentage of a specific category of genes in each main category. The right *y*-axis represents the amount of unigenes in a category; **(b)** Eukaryotic ortholog groups (KOG) classification of assembled unigenes. The *y*-axis indicates the amount of unigenes in a specific functional cluster. The *x*-axis shows function class; **(c)** Kyoto encyclopedia of genes and genomes (KEGG) classification of assembled unigenes. The *y*-axis indicates the KEGG pathway, and the *x*-axis is the ratio of the number of unigenes.

3.2. Frequency and Distribution of SSRs in *D. odorifera* Transcriptome

From the 115,292 unigenes (77,979,362 bp), 35,774 potential SSRs were identified and distributed in 26,880 unigenes. Of which, 24.7% (6629) contained more than one SSRs (Table S2). The frequency of SSRs in the *D. odorifera* transcriptome was 31.0% with a distribution density of 469.72 SSRs per Mb.

The most common type of repeat motifs was mono-nucleotide (21,623, 60.44%), followed by di- (7612, 21.28%) and tri-nucleotide (6112, 17.09%) (Table 3). The three predominant motif types represented 98.81% in all, whereas only 40 and 14 SSRs showed in penta- and hexa-nucleotide repeat motifs, respectively. The average repeat sequence length of di-, tri-, tetra-, penta-, and hexa-nucleotides was 12.75, 15.78, 16.82, 20.49, 25.75, and 38.57 bp, with the tandem repeats number of 6–24, 5–12, 5–14, 5–9, 5–8, and 5–10, respectively. In total, up to 85 different repeat motifs were identified in this research. And the mono-nucleotide (A/T)_n and (T/A)_n repeat motifs showed the highest number among all the SSRs, amounting to 20,927 (58.49%). Of the dinucleotide repeat motifs, (AG/CT)_n was the most frequent (52.21%, 3974), whereas only 4.5% (16) for (CG/CG)_n. Additionally, (AAG/CTT)_n (1419) and (AAT/ATT)_n (1116) were the majorities represented in trinucleotide repeat motifs.

3.3. Development and Transferability of Polymorphic SSR Markers

A set of 192 SSRs was randomly selected, and their primers were designed to test specificity of amplification for three leaf samples (one sample from each *Dalbergia* species). Of these designed primers, 88 (45.8%) amplified successfully in *D. odorifera*, followed by 66 (34.4%) in *D. tonkinensis*, and 83 (43.2%) in *D. cochinchinensis*. Moreover, 63 (32.8%) produced clear amplicons across the three *Dalbergia* species. Details on these SSRs, including ID of the template DNA sequence carrying the SSR, SSR type, repeat motif, position in template sequence, primer sequence, annealing temperature, and expected amplicon length (for developing alternative primers if desired) are presented in Table S3.

Ultimately, 19 of these 66 successfully amplified SSR markers were randomly selected for polymorphic validation and all these selected markers showed polymorphism among the 60 *Dalbergia* individuals (Table 4). In total, 112 alleles represented across these individuals. The allele number of per locus ranged from 3 to 13 with an average of 5.89. Observed (*Ho*) and expected heterozygosity (*He*) was in a range of 0.05–0.65 and 0.44–0.79, respectively. The polymorphic information content (*PIC*) varied from 0.38 (S01) to 0.75 (S26) with the mean 0.59. Moreover, 78.9% (15) of these markers were highly polymorphic with the *PIC* values higher than 0.5.

Table 3. Details of the different SSRs (simple sequence repeats) distribute in *D. odorifera*.

Repeat Motif (Sum)	No. of Repeats									Total
	5	6	7	8	9	10	11	12	>12	
Mono-nucleotide (21623)										
A/T	-	-	-	-	-	6114	3554	2773	8486	20,927
C/G	-	-	-	-	-	6114	3554	2773	380	696
Di-nucleotide (7612)										
AG/CT	-	883	733	757	903	585	112	1	-	3974
AT/AT	-	486	345	359	417	287	53		-	1947
AC/GT	-	528	359	285	242	191	66	4	-	1675
CG/CG	-	11	4	1					-	16
Tri-nucleotide (6112)										
AAG/CTT	733	469	212	4	-	1	-	-	-	1419
AAT/ATT	558	379	172	6	1	-	-	-	-	1116
AAC/GTT	472	263	128	3	-	-	-	1	-	867
ATC/ATG	413	209	100	4	-	-	-	-	1	727
Others	1173	547	245	16	-	1	-	1	-	1983
Tetra-nucleotide (373)										
AAAT/ATTT	94	4	-	-	1	-	-	-	-	99
AAAG/CTTT	50	4	-	-	-	-	-	-	-	54
AAGG/CCTT	27	1	-	-	-	-	-	-	-	28
AATG/ATTC	18	3	-	-	-	-	-	-	-	21
Others	143	27	-	1	-	-	-	-	-	171
Penta-nucleotide (40)										
AAAAG/CTTTT	3	-	-	-	-	-	-	-	-	3
AAACC/GGTTT	3	-	-	-	-	-	-	-	-	3
AACAC/GTGTT	3	-	-	-	-	-	-	-	-	3
ACAGC/CTGTG	3	-	-	-	-	-	-	-	-	3
Others	24	3	-	1	-	-	-	-	-	28
Hexa-nucleotide (14)										
ACAGCC/CTGTGG	-	-	-	-	-	1	-	-	-	1
ACCCTG/AGGGTC	-	-	-	-	-	1	-	-	-	1
ACCTCC/AGGTGG	-	-	-	-	1	-	-	-	-	1
Others	5	6	-	-	-	-	-	-	-	11
Total	3722	3823	2298	1437	1565	7293	3891	2878	8867	35,774
%	10.4	10.69	6.42	4.02	4.37	20.39	10.88	8.04	24.79	100

Table 4. Characteristics of 19 SSR (simple sequence repeat) markers developed from *D. odorifera* transcriptome and summary statistics of polymorphism transferred across three *Dalbergia* species.

Locus	ID	Repeat Motif	Forward Primer(5'-3')	Reverse Primer(3'-5')	Predicted Product Size (bp)	Product Size (bp)	SS Position	Tm (°C)	Size ¹	Na ²	Ho ³	He ⁴	PIC ⁵
S01	c102105_g1	(ATA)6	AGTCCCGCCACAAAATCAT	CTGGTCAGTCATCCCCAC	259	225–258	Unknown	60	60	8	0.45	0.79	0.75
S02	c11754_g1	(AAG)6	GGTCCCTGACTCACTGAAGC	CAACCTCTCTCTGCAGAACCA	269	254–290	5'UTR	60	60	7	0.18	0.53	0.48
S03	c1204_g1	(ATA)6	GCACGTGGTCAAAGCAATCA	ATGAGCCCTTCTGCACTTC	267	254–266	Unknown	60	60	3	0.23	0.67	0.59
S04	c25868_g1	(GAT)6	GCTGTGGAGTCACGTTCTCA	TCCCCACAGAATCACAAGCC	277	272–293	3'UTR	60	60	6	0.42	0.60	0.54
S07	c29390_g1	(CCT)6	GCCAATGACATAATGGGCGG	TGCAGAGAGTCAGGAGCTCT	226	226–250	CDS	60	60	4	0.33	0.62	0.55
S08	c31361_g1	(TGT)6	GGAAGAGAAATGGAGGGTAGCT	TGCCAGACAACCAGAATGCT	245	299–320	Unknown	60	59	6	0.27	0.74	0.69
S09	c33497_g1	(CAT)6	ACCCTCCTCCTCCACCTTTT	ACCGGCTTCAGTGATTGGTT	228	224–254	5'UTR	60	60	6	0.45	0.77	0.73
S10	c40172_g1	(TGC)7	CACGTACCCAACCGTCAAGA	TCCGACGACCACCTAATCCT	273	246–294	5'UTR	60	42	5	0.52	0.54	0.50
S11	c40452_g1	(ATC)6	AAAAAGCGAGGACTACGGCA	TGGAGAAGCAGTGCTCGTTT	229	218–227	Unknown	60	60	3	0.18	0.64	0.56
S12	c34672_g2	(GAT)7	GGTGAACAAGCTGGAGTGGA	AAGCCCAGCATCTAAACCCC	270	259–277	CDS	60	60	4	0.37	0.44	0.40
S21	c56000_g1	(TCCC)6	GAGCCTTGAGTTACCTCCC	TTGGGTGTGAGATTGAGGGC	248	230–250	5'UTR	60	39	6	0.65	0.75	0.70
S22	c53146_g1	(TC)10	CCACCGATCTTAACCTCCGG	ACTACAAGTGCGTGTGACCC	255	230–282	Unknown	60	60	11	0.43	0.65	0.63
S23	c56684_g2	(TC)10	TGGCGTTGACTTCCAGCATT	GAGCAGTGTGAGCATGATGC	277	242–284	3'UTR	60	60	7	0.05	0.50	0.41
S24	c59001_g1	(CTA)7	GCTGCAAAATGCCAGTGCTTA	CGCTGTTGTGTCAGTGATTGG	234	219–268	Unknown	60	60	8	0.25	0.77	0.74
S26	c60831_g5	(GTT)7	CCAATCCCACCAGTGAGGAG	GCAGCACCTCTGAGACAAGT	244	223–262	CDS	60	60	4	0.05	0.48	0.38
S27	c49315_g1	(TAC)7	GAACCTTTCTCTTGCGCCT	CCTATGAAGCGGTGTCATGC	265	260–272	3'UTR	60	60	4	0.25	0.73	0.67
S28	c63495_g1	(TAC)7	ACAGCATTTGTGTTGTGCA	CAGCTGCGCTCTCATTCTA	249	201–249	Unknown	60	60	3	0.12	0.62	0.54
S29	c57231_g1	(TAT)7	TCCCCGTTCCTCTCTCTCAG	GGACTGTACATGGCACTCA	152	141–174	5'UTR	60	60	6	0.22	0.76	0.71
S30	c48304_g3	(TGT)7	TGCCTTGATCCGCTGAGATC	TCCCAAAATCGATGCAAAGCA	250	240–258	5'UTR	59	60	6	0.35	0.65	0.58
Mean										5.89	0.30	0.64	0.59

¹ Number of sampled individuals with expected band, ² number of alleles, ³ observed heterozygosity, ⁴ expected heterozygosity, ⁵ polymorphic information content.

3.4. SSR Polymorphism and Phylogenetic Analysis of the Three *Dalbergia* Species

The nineteen SSR markers showed distinctive allelic patterns among *D. odorifera*, *D. tonkinensis*, and *D. cochinchinensis* (Table 5). Among the three *Dalbergia* species, *D. odorifera* presented the largest number of alleles (56) with 2–5 alleles/locus, followed by *D. cochinchinensis* (54, 0–7 alleles/locus), and *D. tonkinensis* (47, 1–5 alleles/locus). The number of polymorphic markers varied among target species, all the 19 (100%) SSR markers represented polymorphic in *D. odorifera*, while 17 (89.5%) represented in *D. tonkinensis*, and 11 (57.9%) in *D. cochinchinensis*, respectively. The mean observed (H_o) and mean expected heterozygosity (H_e) was 0.34 and 0.40 in *D. odorifera*, 0.27 and 0.32 in *D. tonkinensis*, and 0.29 and 0.33 in *D. cochinchinensis*, respectively. Among these polymorphic markers, the PIC ranged from 0.05 (S26) to 0.62 (S21) in *D. odorifera*, 0.05 (S02, S23, S24, and S26) to 0.72 (S29) in *D. tonkinensis*, and 0.05 (S26) to 0.73 (S22) in *D. cochinchinensis*, with the mean of 0.34, 0.31, and 0.43, respectively. Moreover, locus S09, S21, and S24 were highly polymorphic in *D. odorifera* with the PIC values higher than 0.5, so was locus S21 and S29 in *D. tonkinensis*, and locus S04, S07, S12, S22, and S24 in *D. cochinchinensis*. Additionally, locus S11 only showed polymorphism in *D. odorifera*. Subsequently, the 19 newly developed SSR markers were used to assess the genetic relationships of these 60 *Dalbergia* trees. The UPGMA dendrogram revealed that: the three clearly divided clusters were associated with the three *Dalbergia* species based on genetic similarity coefficients (SSM, $r = 0.999$) (Figure 3).

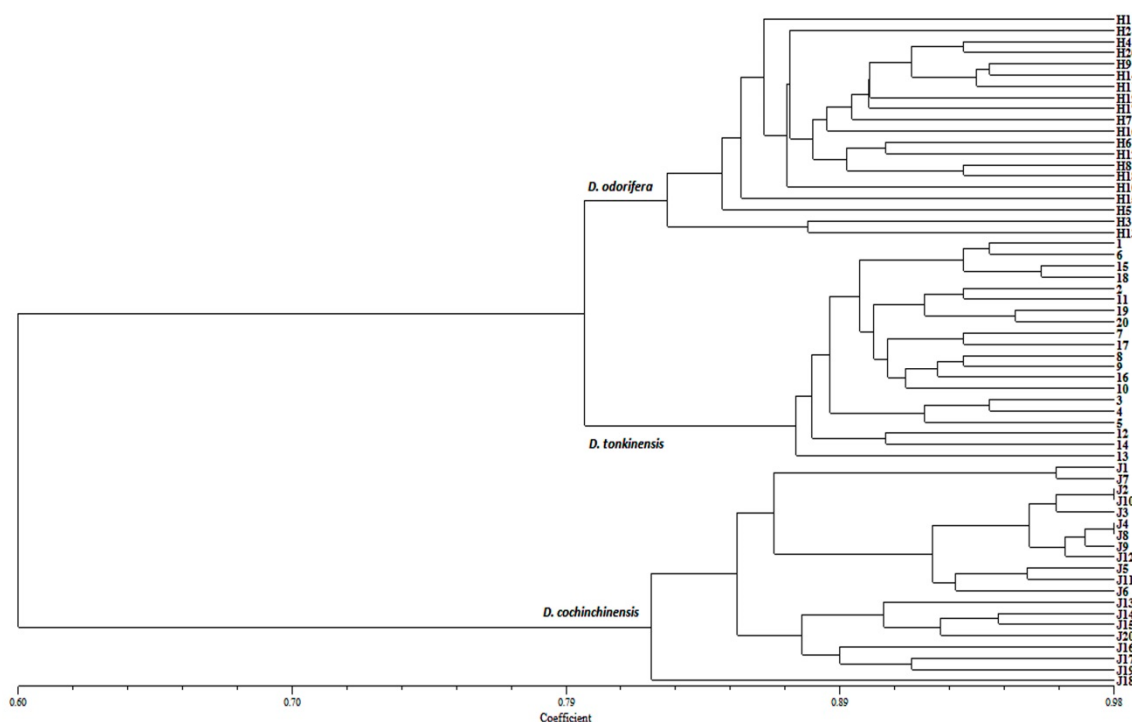


Figure 3. The UPGMA (un-weighted pair group method with arithmetic mean) cluster diagram of 60 *Dalbergia* trees based on the 19 newly developed SSR markers.

Table 5. Polymorphism of 19 SSR markers in *D. odorifera*, *D. tonkinensis*, and *D. cochinchinensis*.

Locus	<i>D. odorifera</i>					<i>D. tonkinensis</i>					<i>D. cochinchinensis</i>				
	Size ¹	Na ²	Ho ³	He ⁴	PIC ⁵	Size ¹	Na ²	Ho ³	He ⁴	PIC ⁵	Size ¹	Na ²	Ho ³	He ⁴	PIC ⁵
S01	20	3	0.50	0.54	0.46	20	2	0.40	0.43	0.33	20	6	0.45	0.53	0.49
S02	20	2	0.00	0.10	0.09	20	2	0.05	0.05	0.05	20	5	0.5	0.49	0.45
S03	20	2	0.45	0.50	0.37	20	2	0.25	0.51	0.37	20	1	\	\	\
S04	20	4	0.35	0.50	0.44	20	2	0.15	0.22	0.19	20	5	0.75	0.67	0.6
S07	20	2	0.55	0.50	0.37	20	1	\	\	\	20	4	0.45	0.56	0.5
S08	20	4	0.35	0.35	0.33	19	3	0.47	0.51	0.44	20	1	\	\	\
S09	20	4	0.50	0.63	0.56	20	3	0.35	0.56	0.44	20	2	0.5	0.47	0.35
S10	20	3	0.50	0.45	0.40	20	3	0.60	0.52	0.45	2	1	\	\	\
S11	20	2	0.55	0.48	0.36	20	1	\	\	\	20	1	\	\	\
S12	20	2	0.20	0.18	0.16	20	2	0.30	0.26	0.22	20	3	0.6	0.66	0.57
S21	19	5	0.65	0.68	0.62	20	4	0.65	0.75	0.68	\	\	\	\	\
S22	20	5	0.35	0.39	0.36	20	4	0.15	0.15	0.14	20	6	0.8	0.78	0.73
S23	20	2	0.00	0.10	0.09	20	2	0.05	0.05	0.05	20	4	0.1	0.15	0.14
S24	20	4	0.50	0.63	0.55	20	2	0.05	0.05	0.05	20	7	0.2	0.68	0.61
S26	20	2	0.05	0.05	0.05	20	2	0.05	0.05	0.05	20	2	0.05	0.05	0.05
S27	20	2	0.05	0.14	0.13	20	3	0.70	0.55	0.44	20	1	\	\	\
S28	20	2	0.25	0.45	0.34	20	2	0.10	0.33	0.27	20	1	\	\	\
S29	20	2	0.30	0.47	0.35	20	5	0.35	0.78	0.72	20	1	\	\	\
S30	20	4	0.40	0.44	0.39	20	2	0.40	0.38	0.30	20	3	0.25	0.23	0.21
Mean	20	2.95	0.34	0.40	0.34	20	2.47	0.27	0.32	0.31	18	3.00	0.29	0.33	0.43

¹ Number of sampled individuals with expected band, ² number of alleles, ³ observed heterozygosity, ⁴ expected heterozygosity, ⁵ polymorphic information content.

4. Discussion

4.1. Transcriptome Sequencing, De Novo Assembly, and Annotation for *D. odorifera*

Currently, scarce genome information has been published for a few *Dalbergia* species [32–34], and to date, there is no de novo assembly data for the *D. odorifera* yet. Therefore, high-quality transcriptome resources, particularly those covering full-length sequences and have been filtered for redundancy, remain essential for *D. odorifera*. In this research, we carried out the transcriptome for three leaf-tissues from non-cloned individuals and generated approximately 14 million clean reads assembled into 115,292 unigenes, which are valuable resources for functional genomic studies of *Dalbergia* species prior to sequencing of the *D. odorifera* genome. The N50 and average length of generated unigenes were 1080 bp and 676 bp, respectively. These results are comparable to the results from recently published tree and legume transcriptome (leaf tissue) studies, such as *Camellia sinensis* (L.) Kuntze (867 bp, 601 bp) [35], *Morus atropurpurea* Roxb. (1219 bp, 793 bp) [36], *Millettia pinnata* Forst. (699 bp, 534 bp) [37], *Quercus kerrii* Craib (1166 bp, 720 bp), and *Quercus austrocochinchinensis* Hickel & A. Camus (1335 bp, 782 bp) [38].

For gene annotation, it was not surprising that more than 60% annotated unigenes was top-hit to legume species (Figure S1), particularly *Glycine max* (L.) Merr. (7679) [39] and *Glycine soja* Sieb. et Zucc. (7428) [40], the species with the earliest complete genome sequences. This may be because they all belong to the family Fabaceae and share a high level of genomes sequence similarity. Similar observations have been reported on other transcriptome studies [41,42]. It is thought that KEGG provides a basic platform for systematic analysis of gene function in terms of gene product networks [43]. Our KEGG analysis revealed 16,571 unigenes involved in 129 pathways. In addition, 33,219 unigenes were assigned to GO assignments, and 17,860 were functional classified according to the KOG analysis. All these annotations and classifications are useful resources for investigating biological function-specific unigenes in *D. odorifera*.

4.2. SSR Prediction, Validation, and Application

Molecular markers have been widely and effectively applied in plants for genetic diversity and association analysis [8–12]. Prior to this study, however, there were only six RAPD (random amplified polymorphic DNA) markers and 25 SRAP (sequence-related amplified polymorphism) markers for the genotyping of *D. odorifera* germplasm [7,13], and no co-dominant SSR markers have been specifically developed for this endangered tree species. Therefore, the development of a known polymorphic SSR marker database for *D. odorifera* is of great use for diversity studies and breeding programs, as well as being transferable to closely *Dalbergia* species. In the present study, we identified 35,774 putative SSRs from the transcriptome dataset, which is of great differences in numbers, dominant repeats, and motif types reported from other legume species using Illumina sequencing [41,42,44]. This may be attributable to inconsistencies in genome structure or composition, dataset size, search method, or criteria.

The effectiveness and success of SSRs rely considerably on the quality of the markers, the accuracy of the genotyping data, and the plant materials used [45]. Therefore, the next step in building a working marker set for genetic improvement efforts was validation of the identified SSRs. Of these 192 primer pairs, 32.8% (63) yielded clear bands across the 60 *Dalbergia* trees. The success rate is comparable for SSR markers developed from the transcriptome of *Robinia* (25%) [46], *Allium* (31%) [47], and *Osmanthus* (35%) [14,48]. Furthermore, all of these 19 markers exhibited polymorphism among these 60 *Dalbergia* trees, with a polymorphic information content (PIC) range of 0.38–0.75, which were comparable to, or substantially higher than the PIC values of cross-species EST-SSR markers reported for *Melilotus* (0.10–0.87) [48], *Robinia* (0.03–0.76) [46], and *Casuarina* (0.26–0.62) [19].

Highly polymorphic and stable SSR markers were important resources for genetic relationship analysis. In the present research, 78.9% (15) of these markers were highly polymorphic, with PIC values higher than 0.50, indicating those alleles were found in more than 50% of these germplasms. Deciding

whether an SSR marker is useful depends upon the scientific issue since *PIC* values may be influenced by many factors, such as sampling schemes, the number of SSRs, and types of SSR motif repeats [45]. Highly polymorphic SSR markers are useful for genetic diversity but using only the most polymorphic markers would bias the overall genetic diversity [15], especially in conservation studies [49]. Therefore, the entire newly developed SSR markers were used to evaluate the genetic relationships of these 60 sampled *Dalbergia* trees. Based on genetic similarity coefficients (SSM, $r = 0.999$), three major clusters according to the three *Dalbergia* species were revealed (Figure 3), indicating these transferable SSR markers are advantageous to explore the relationships among *Dalbergia* species and could be available for assisting genetic research and breeding in the future.

5. Conclusions

This study represents the first attempt to obtain the transcriptome information and mine SSR markers of *D. odorifera* using RNA-seq, providing a large and well-assembled transcript data set. These abundant resources are an initial step towards understanding the genetic backgrounds of this endangered species and may serve as a basic reference for other *Dalbergia* species. Moreover, a novel set of SSR markers was successfully developed and transferred to other *Dalbergia* species. These results are useful for evaluating the genetic diversity, population genetic structure, and marker-assisted breeding applications in the future.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4907/10/2/98/s1>, Figure S1: Species classification, Table S1: Summary for the annotation of *D. odorifera* unigenes, Table S2: Summary of SSRs identified from the transcriptome, Table S3 Details of 63 transferable SSR markers.

Author Contributions: D.-P.X. and Z.H. conceived and designed the research, revised the manuscript; Z.-J.Y. and N.-N.Z. collected the *Dalbergia* materials, analyzed geographic information, and performed parts of experiments; X.-J.L. investigated and provided technical support for the statistics, and F.-M.L. performed the experiments, analyzed the data, and wrote the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (31500537), the Natural Science Foundation of Guangxi Province (2016GXNSFBA380089), and the Central Non-Profit Research Institution of Chinese Academy of Forestry (Grant No. CAFYBB2017ZX001-4).

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Institute of Botany, Chinese Academy of Sciences. *Flora of China*, 1st ed.; Science Press: Beijing, China, 1994; Volume 40, p. 98.
2. So, T.; Dell, B. Conservation and utilization of threatened hardwood species through reforestation—An example of (*Kruz.*) craib and Pierre in Cambodia. *Pac. Conserv Biol.* **2010**, *16*, 101–116. [CrossRef]
3. *Dalbergia cochinchinensis*. The IUCN Red List of Threatened Species 1998:e.T32625A9719096. Available online: <http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32625A9719096.en> (accessed on 9 December 2018).
4. *Dalbergia odorifera*. The IUCN Red List of Threatened Species 1998:e.T32398A9698077. Available online: <http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32398A9698077.en> (accessed on 9 December 2018).
5. *Dalbergia tonkinensis*. The IUCN Red List of Threatened Species 1998:e.T32819A9732061. Available online: <http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32819A9732061.en> (accessed on 9 December 2018).
6. Hartvig, I.; So, T.; Changtragoon, S.; Tran, H.T.; Bouamanivong, S.; Theilade, I.; Kjaer, E.D.; Nielsen, L.R. Population genetic structure of the endemic rosewoods *Dalbergia cochinchinensis* and *D. oliveri* at a regional scale reflects the Indochinese landscape and life-history traits. *Ecol. Evol.* **2018**, *8*, 530–545. [CrossRef] [PubMed]
7. Yang, Q.X.; Feng, J.D.; Wei, J.H.; Li, R.T.; He, M.J. Genetic diversity of China's endangered medicinal plant *Dalbergia odorifera*. *World Sci. Technol. Mod. Tradit. Chin. Med. Mater. Med.* **2007**, *9*, 73–79. Available online: <http://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFD2007&filename=SJKX200702016&v=MTgxNzJGeUhuVmJ2TE5pZkFkckc0SHRiTjZOUVZb1I4ZVgxTHV4WVM3RGxVDNvVHJXTTFGckNVUkxLZVpIZHE=> (accessed on 20 June 2018).

8. Sanabam, R.; Singh, N.S.; Sahoo, D.; Devi, H.S. Genetic relationship of rough lemon landraces and under-utilised citrus genotypes from North-East India revealed by SSR and RAPD markers. *Trees* **2018**, *32*, 1043–1059. [CrossRef]
9. Vu Thi Thu, H. Genetic diversity among endangered rare *Dalbergia cochinchinensis* (Fabaceae) genotypes in Vietnam revealed by random amplified polymorphic DNA (RAPD) and inter simple sequence repeats (ISSR) markers. *Afr. J. Biotechnol.* **2012**, *11*. [CrossRef]
10. Mohammad-Panah, N.; Shabaniyan, N.; Khadivi, A.; Rahmani, M.-S.; Emami, A. Genetic structure of gall oak (*Quercus infectoria*) characterized by nuclear and chloroplast SSR markers. *Tree Genet. Genomes* **2017**, *13*. [CrossRef]
11. Chhajer, S.; Jukanti, A.K.; Bhatt, R.K.; Kalia, R.K. Genetic diversity studies in endangered desert teak [*Tecomella undulata* (Sm) Seem] using arbitrary (RAPD), semi-arbitrary (ISSR) and sequence based (nuclear rDNA) markers. *Trees* **2018**, *32*, 1083–1101. [CrossRef]
12. Nybom, H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* **2004**, *13*, 1143–1155. [CrossRef]
13. Yang, Y.; Meng, H.; Wu, Y.; Chen, B.; Gan, B.C. Primer Screening of SRAP Molecular Marker in *Dalbergia odorifera*. *Acta Agric. Jiangxi* **2011**, *23*, 29–31. [CrossRef]
14. Alexander, L.W.; Thammina, C.S.; Kramer, M. Cross-transferability of SSR markers in *Osmanthus*. *Genet. Resour. Crop Evol.* **2018**, *65*, 125–136. [CrossRef]
15. Liu, S.; An, Y.; Li, F.; Li, S.; Liu, L.; Zhou, Q.; Zhao, S.; Wei, C. Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). *Mol. Breed.* **2018**, *38*, 59. [CrossRef]
16. Göl, Ş.; Göktay, M.; Allmer, J.; Doğanlar, S.; Frary, A. Newly developed SSR markers reveal genetic diversity and geographical clustering in spinach (*Spinacia oleracea*). *Mol. Genet. Genom.* **2017**, *292*, 847–855. [CrossRef] [PubMed]
17. Onoue, N.; Kobayashi, S.; Kono, A.; Sato, A. SSR-based molecular profiling of 237 persimmon (*Diospyros kaki* Thunb.) germplasms using an ASTRINGENCY-linked marker. *Tree Genet. Genomes* **2018**, *14*, 28. [CrossRef]
18. Taheri, S.; Lee Abdullah, T.; Yusop, M.; Hanafi, M.; Sahebi, M.; Azizi, P.; Shamshiri, R. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. *Molecules* **2018**, *23*, 399. [CrossRef] [PubMed]
19. Li, N.; Zheng, Y.-Q.; Ding, H.-M.; Li, H.-P.; Peng, H.-Z.; Jiang, B.; Li, H.-B. Development and validation of SSR markers based on transcriptome sequencing of *Casuarina equisetifolia*. *Trees* **2017**, *32*, 41–49. [CrossRef]
20. Dervishi, A.; Jakše, J.; Ismaili, H.; Javornik, B.; Štajner, N. Comparative assessment of genetic diversity in Albanian olive (*Olea europaea* L.) using SSRs from anonymous and transcribed genomic regions. *Tree Genet. Genomes* **2018**, *14*. [CrossRef]
21. Dong, M.; Wang, Z.; He, Q.; Zhao, J.; Fan, Z.; Zhang, J. Development of EST-SSR markers in *Larix principis-rupprechtii* Mayr and evaluation of their polymorphism and cross-species amplification. *Trees* **2018**, *32*, 1559–1571. [CrossRef]
22. Zhai, S.H.; Yin, G.S.; Yang, X.H. Population Genetics of the Endangered and Wild Edible Plant *Ottelia acuminata* in Southwestern China Using Novel SSR Markers. *Biochem. Genet.* **2018**, *56*, 235–254. [CrossRef]
23. Vu, D.-D.; Bui, T.T.-X.; Nguyen, M.-D.; Shah, S.N.M.; Vu, D.-G.; Zhang, Y.; Nguyen, M.-T.; Huang, X.-H. Genetic diversity and conservation of two threatened dipterocarps (*Dipterocarpaceae*) in southeast Vietnam. *J. For. Res.* **2018**, *1*–9. [CrossRef]
24. Yan, L.-P.; Liu, C.-L.; Wu, D.-J.; Li, L.; Shu, J.; Sun, C.; Xia, Y.; Zhao, L.-J. De novo transcriptome analysis of *Fraxinus velutina* using Illumina platform and development of EST-SSR markers. *Biol. Plant.* **2017**, *61*, 210–218. [CrossRef]
25. Chen, L.-Y.; Cao, Y.-N.; Yuan, N.; Nakamura, K.; Wang, G.-M.; Qiu, Y.-X. Characterization of transcriptome and development of novel EST-SSR makers based on next-generation sequencing technology in *Neolitsea sericea* (Lauraceae) endemic to East Asian land-bridge islands. *Mol. Breed.* **2015**, *35*, 187. [CrossRef]
26. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644. Available online: <https://www.nature.com/articles/nbt.1883> (accessed on 28 June 2018). [CrossRef] [PubMed]

27. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. Available online: <https://www.nature.com/articles/nprot.2013.084> (accessed on 28 June 2018). [CrossRef] [PubMed]
28. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [CrossRef] [PubMed]
29. Popgene. version 1.32. the User-Friendly Shareware for Population Genetic Analysis molecular Biology and Biotechnology Center, University of Alberta Edmonton. Available online: <http://www.ualberta.ca/~jfyeh> (accessed on 23 November 2017).
30. Powell, W.; Morgante, M.; Andre, C.; Hanafey, M.; Vogel, J.; Tingey, S.; Rafalski, A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* **1996**, *2*, 225–238. [CrossRef]
31. Rohlf, F.J. NTSYS-pc: Microcomputer programs for numerical taxonomy and multivariate analysis. *Am. Stat.* **1987**, *41*, 330. Available online: <https://www.jstor.org/stable/2684761> (accessed on 28 June 2018). [CrossRef]
32. Moritsuka, E.; Chhang, P.; Tagane, S.; Toyama, H.; Sokh, H.; Yahara, T.; Tachida, H. Genetic variation and population structure of a threatened timber tree *Dalbergia cochinchinensis* in Cambodia. *Tree Genet. Genomes* **2017**, *13*. [CrossRef]
33. Wariss, H.M.; Yi, T.-S.; Wang, H.; Zhang, R. Characterization of the complete chloroplast genome of *Dalbergia odorifera* (Leguminosae), a rare and critically endangered legume endemic to China. *Conserv. Genet. Resour.* **2018**, *10*, 527–530. [CrossRef]
34. Deng, C.-Y.; Xin, G.-L.; Zhang, J.-Q.; Zhao, D.-M. Characterization of the complete chloroplast genome of *Dalbergia hainanensis* (Leguminosae), a vulnerably endangered legume endemic to China. *Conserv. Genet. Resour.* **2018**. [CrossRef]
35. Li, N.N.; Yue, C.; Cao, H.L.; Qian, W.J.; Hao, X.Y.; Wang, Y.C.; Wang, L.; Ding, C.Q.; Wang, X.C.; Yang, Y.J. Transcriptome sequencing dissection of the mechanisms underlying differential cold sensitivity in young and mature leaves of the tea plant (*Camellia sinensis*). *J. Plant Physiol.* **2018**, *224–225*, 144–155. [CrossRef] [PubMed]
36. Dai, F.; Tang, C.; Wang, Z.; Luo, G.; He, L.; Yao, L. De novo assembly, gene annotation, and marker development of mulberry (*Morus atropurpurea*) transcriptome. *Tree Genet. Genomes* **2015**, *11*, 26. [CrossRef]
37. Huang, J.; Lu, X.; Yan, H.; Chen, S.; Zhang, W.; Huang, R.; Zheng, Y. Transcriptome characterization and sequencing-based identification of salt-responsive genes in *Millettia pinnata*, a semi-mangrove plant. *DNA Res.* **2012**, *19*, 195–207. [CrossRef] [PubMed]
38. An, M.; Deng, M.; Zheng, S.-S.; Song, Y.-G. De novo transcriptome assembly and development of SSR markers of oaks *Quercus austrocochinchinensis* and *Q. kerrii* (Fagaceae). *Tree Genet. Genomes* **2016**, *12*. [CrossRef]
39. Sato, S.; Nakamura, Y.; Kaneko, T.; Asamizu, E.; Kato, T.; Nakao, M.; Sasamoto, S.; Watanabe, A.; Ono, A.; Kawashima, K.; et al. Genome structure of the legume, *Lotus japonicus*. *DNA Res.* **2008**, *15*, 227–239. [CrossRef] [PubMed]
40. Kim, M.Y.; Lee, S.; Van, K.; Kim, T.H.; Jeong, S.C.; Choi, I.Y.; Kim, D.S.; Lee, Y.S.; Park, D.; Ma, J.; et al. Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 22032–22037. [CrossRef] [PubMed]
41. Torales, S.L.; Rivarola, M.; Pomponio, M.F.; Gonzalez, S.; Acuña, C.V.; Fernández, P.; Lauenstein, D.L.; Verga, A.R.; Hopp, H.E.; Paniego, N.B.; et al. De novo assembly and characterization of leaf transcriptome for the development of functional molecular markers of the extremophile multipurpose tree species *Prosopis alba*. *BMC Genom.* **2013**, *14*, 705. [CrossRef]
42. Huang, J.; Guo, X.; Hao, X.; Zhang, W.; Chen, S.; Huang, R.; Gresshoff, P.M.; Zheng, Y. De novo sequencing and characterization of seed transcriptome of the tree legume *Millettia pinnata* for gene discovery and SSR marker development. *Mol. Breed.* **2016**, *36*, 75. [CrossRef]
43. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40*, D109–D114. [CrossRef]
44. Sathyanarayana, N.; Pittala, R.K.; Tripathi, P.K.; Chopra, R.; Singh, H.R.; Belamkar, V.; Bhardwaj, P.K.; Doyle, J.J.; Egan, A.N. Transcriptomic resources for the medicinal legume *Mucuna pruriens*: De novo transcriptome assembly, annotation, identification and validation of EST-SSR markers. *BMC Genom.* **2017**, *18*, 409. [CrossRef]

45. Liu, S.; Liu, H.; Wu, A.; Hou, Y.; An, Y.; Wei, C. Construction of fingerprinting for tea plant (*Camellia sinensis*) accessions using new genomic SSR markers. *Mol. Breed.* **2017**, *37*, 93. [CrossRef]
46. Guo, Q.; Wang, J.-X.; Su, L.-Z.; Lv, W.; Sun, Y.-H.; Li, Y. Development and evaluation of a novel set of EST-SSR markers based on transcriptome sequences of Black Locust (*Robinia pseudoacacia* L.). *Genes* **2017**, *8*, 177. Available online: <https://www.mdpi.com/2073-4425/8/7/177> (accessed on 28 June 2018). [CrossRef] [PubMed]
47. Barboza, K.; Beretta, V.; Kozub, P.C.; Salinas, C.; Morgenfeld, M.M.; Galmarini, C.R.; Cavagnaro, P.F. Microsatellite analysis and marker development in garlic: Distribution in EST sequence, genetic diversity analysis, and marker transferability across Alliaceae. *Mol. Genet. Genom.* **2018**, *293*, 1091–1106. [CrossRef]
48. Yan, Z.; Wu, F.; Luo, K.; Zhao, Y.; Yan, Q.; Zhang, Y.; Wang, Y.; Zhang, J. Cross-species transferability of EST-SSR markers developed from the transcriptome of *Melilotus* and their application to population genetics research. *Sci. Rep.* **2017**, *7*, 17959. [CrossRef] [PubMed]
49. Väli, Ü.; Einarsson, A.; Waits, L.; Ellegren, H. To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Mol. Ecol.* **2008**, *17*, 3808–3817. [CrossRef] [PubMed]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).