

## Article

# Characterization and Development of Genomic SSRs in Pecan (*Carya illinoensis*)

Chengcai Zhang <sup>1</sup>, Xiaohua Yao <sup>1,\*</sup>, Huadong Ren <sup>1</sup>, Jun Chang <sup>1</sup>, Jun Wu <sup>2</sup>, Weizhong Shao <sup>3</sup> and Qing Fang <sup>3</sup>

<sup>1</sup> Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Fuyang District, Hangzhou 311400, China; c.c.zhang@caf.ac.cn (C.Z.); renhd@163.com (H.R.); ylchj163@163.com (J.C.)

<sup>2</sup> Jiande Forestry Bureau, Jiande, Hangzhou 311600, China; wujun\_jdlyj@163.com

<sup>3</sup> Forestry Center of Jiande, Jiande, Hangzhou 311600, China; shaoyunzhu@126.com (W.S.); fangqing\_jdlc@163.com (Q.F.)

\* Correspondence: yaoxh168@163.com; Tel.: +86-571-6331-0094

Received: 4 December 2019; Accepted: 2 January 2020; Published: 4 January 2020



**Abstract:** *Research Highlights:* The distribution of simple sequence repeat (SSR) motifs in two draft genomes of pecan was evaluated. Sixty-six SSR loci were validated by PCR amplification in pecan. Twenty-two new development markers can be used for genetic study in genus *Carya*. *Background and Objectives:* Pecan has good nutritional and health benefits and is an important crop worldwide. However, the genetic research in this species is insufficient. One of the main reasons for this is the lack of enough accurate, convenient, and economical molecular markers. Among different marker types, SSR loci are enormously useful in genetic studies. However, the number of SSRs in *C. illinoensis* (Wangenh.) K. Koch is limited. *Materials and Methods:* The distribution of SSR motifs in the pecan genome was analyzed. Then, the primers for each SSR were designed. To evaluate their availability, 74 SSR loci were randomly selected and amplified in pecan. Finally, 22 new SSRs and eight former ones were picked to evaluate the genetic diversity in 60 pecan genotypes and to determine their transferability in other *Carya* species. *Results:* 145,714 and 143,041 SSR motifs were obtained from two draft genomes of ‘87MX3-2’ and ‘Pawnee’, respectively. In total, 9145 candidate primers were obtained. Sixty-six (89.19%) primers amplified the target products. Among the 30 SSRs, 29 loci showed polymorphism in 60 pecan genotypes. The polymorphic information content (PIC) values ranged from 0.012 to 0.906. In total, 26, 25, and 22 SSRs can be used in *C. cathayensis* Sarg., *C. dabieshanensis* W. C. Cheng & R. H. Chang, and *C. hunanensis* W.C. Liu, respectively. Finally, the dendrogram of all individuals was constructed. The results agree with the geographic origin of the four species and the pedigree relationships between different pecan cultivars. *Conclusions:* The characterization of SSRs in the pecan genome and the new SSRs will promote the progress of genetic study and breeding in pecan, as well as other species of genus *Carya*.

**Keywords:** pecan; *Carya*; SSR; genetic diversity; DNA marker; molecular marker assisted selection

## 1. Introduction

Pecan (*Carya illinoensis* (Wangenh.) K. Koch), native to North America, is an important crop worldwide [1]. Pecan nuts are rich in unsaturated fatty acids, phenolics, and flavonoids and have good nutritional and health benefits [2–4]. The nut shell contains high levels of bioactive compounds, including tocopherols, phytosterols, total phenolics, and condensed tannins, and shows antioxidant, antimicrobial, and potential anticancer activity [5,6]. Meanwhile, the high oil content (>70% of the fresh weight) and high mono-unsaturated fatty acids content of the nut make pecan an excellent oil crop [4]. In addition, the biomass waste of the tree makes pecan a potential energy crop [7]. Previous studies

have reported a significant variation in phytochemical profiles between different pecan cultivars [8,9]. Therefore, the identity verification and genetic study of this multipurpose tree are important.

To date, several molecular marker types have been used in pecan, such as amplified fragment length polymorphism (AFLP) [10], random amplified polymorphic DNA (RAPD) [11], inter-simple sequence repeat (ISSR) [12], simple sequence repeat (SSR) [13], and single nucleotide polymorphism (SNP) [14]. Compared to the dominant markers (AFLP, RAPD and ISSR), SSRs are superior in terms of their high informative content, high allelic variation, and good reproducibility [15]. They have been extensively employed in genetic diversity, identity verification, kinship analysis, genetic mapping, and association analysis in various plant species [16,17]. In total, 28 SSRs (including 25 genome SSRs and three chloroplast SSRs) have been developed in pecan, and a set of 19 SSR loci have been used in genetic diversity studies [12,13,18,19]. Li et al. (2014) developed a set of SSRs based on transcriptome sequences of Chinese hickory (*C. cathayensis* Sarg.) and identified the cross-transferability to other species of *Carya*; however, the availability of these markers has not been fully evaluated in pecan [20]. Recently, 87,446 SNPs have been discovered in pecan individuals by genotyping by sequencing (GBS), and 17 SNPs were significantly associated with the flowering type [14]. This study significantly promoted the genetic study in genus *Carya*. However, considering the large information content, high convenience, and low cost of SSR, SNP is unlikely to replace it in the near future [14,16].

Pecan ( $2n = 32$ ) belongs to the Juglandaceae, with an estimated genome size of approximately 650 Mb [4]. It is a woody plant, which is highly heterozygous and has a lengthy generation time, large plant size, and long juvenile phase [21]. These characteristics inhibited the genetic research and cultivar improvement in this plant. The traditional breeding strategies in this plant are laborious, costly, and time consuming. Recently, molecular marker-assisted selection (MAS) has become a research hotspot in the crop breeding area that can utilize trait-linked markers to shorten the breeding period [22,23]. However, MAS study in pecan is less often carried out. Thereby, MAS in pecan has lagged behind other fruit trees and forestry species [24–27].

In recent years, transcriptome and genome data of pecan have been published [3,4,28,29], offering a good opportunity to discover and develop SSRs on a large scale. Compared to genic SSRs, genomic SSRs (gSSRs) are more abundant [30]. In the present study, the genome sequences of two pecan cultivars, including ‘87MX3-2’ and ‘Pawnee’, were downloaded from public databases. The distribution of SSR motifs in the two cultivars was evaluated. A large number of candidate primers for gSSRs were designed using bioinformatic tools. Then, a set of gSSRs was randomly selected, and evaluated in ‘Pawnee’ and ‘Mahan’. Finally, the gSSRs with a good amplifying effect were picked and then used to analyze the genetic diversity of different types of pecan germplasms in China, and to verify the cross-transferability to three species of the genus *Carya*. We believe that this study will increase the number of usable SSRs in pecan, and will facilitate genetic diversity study, cultivar identification, genetic mapping, and MAS research in pecan, as well as other species of *Carya*.

## 2. Materials and Methods

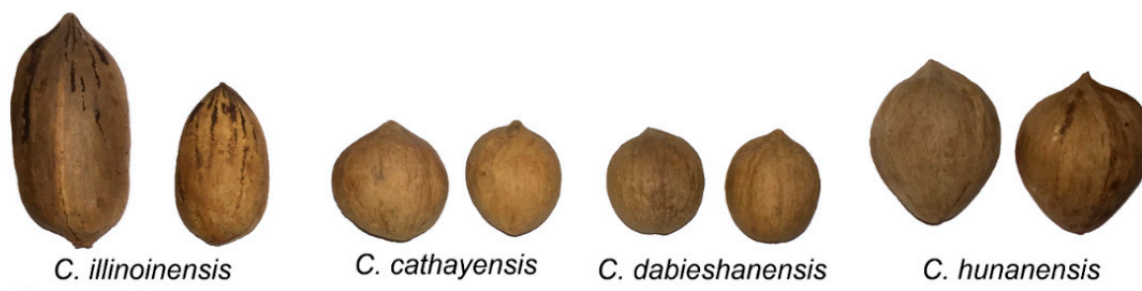
### 2.1. Plant Materials and DNA Extraction

In total, 80 plant samples were used in this study, including 21 pecan cultivars (20 cultivars introduced from USA), 14 excellent seedlings, 25 pecan seedling trees, 10 Chinese hickory strains, five *C. hunanensis* W. C. Cheng & R. H. Chang seedling trees, and five *C. dabieshanensis* W.C. Liu seedling trees (Table 1; Figure 1). Genomic DNA was extracted from buds or young leaves of each individual using a Plant DNA Extraction Kit (Tiangen, Beijing, China). All DNA was stored at  $-80^{\circ}\text{C}$  until use.

**Table 1.** The information about the plant samples.

No.	Name	Species <sup>1</sup>	Type <sup>2</sup>	Origin <sup>3</sup>	No.	Name	Species	Type	Origin	No.	Name	Species	Type	Origin
1	Osage	C. il	C	USA	28	J1	C. il	ES	Jinhua, ZJ	55	JLSX	C. il	SP	Jiande, ZJ
2	Pawnee	C. il	C	USA	29	J2	C. il	ES	Jinhua, ZJ	56	0071A	C. il	SP	Jiande, ZJ
3	McMillan	C. il	C	USA	30	1101	C. il	ES	Nanjing, JS	57	0072A	C. il	SP	Jiande, ZJ
4	Lakota	C. il	C	USA	31	1103	C. il	ES	Nanjing, JS	58	0073A	C. il	SP	Jiande, ZJ
5	Carter	C. il	C	USA	32	1104	C. il	ES	Nanjing, JS	59	SDHD1	C. il	SP	Jiande, ZJ
6	Colby	C. il	C	USA	33	1106	C. il	ES	Nanjing, JS	60	SDHDX1	C. il	SP	Jiande, ZJ
7	Stuart	C. il	C	USA	34	1201	C. il	ES	Nanjing, JS	61	JP-18	C. hu	SP	Jinping, GZ
8	Greenriver	C. il	C	USA	35	1202	C. il	ES	Nanjing, JS	62	JP-17	C. hu	SP	Jinping, GZ
9	Waco	C. il	C	USA	36	JL72	C. il	SP	Jiande, ZJ	63	JP-19	C. hu	SP	Jinping, GZ
10	Major	C. il	C	USA	37	JL72D	C. il	SP	Jiande, ZJ	64	JP-20	C. hu	SP	Jinping, GZ
11	Oconee	C. il	C	USA	38	JL1	C. il	SP	Jiande, ZJ	65	JP-18	C. hu	SP	Jinping, GZ
12	Syrup Mill	C. il	C	USA	39	JL2	C. il	SP	Jiande, ZJ	66	TTZ-15	C. da	SP	Luan, AH
13	Navaho	C. il	C	USA	40	JL3	C. il	SP	Jiande, ZJ	67	TTZ-21	C. da	SP	Luan, AH
14	Gloria Grande	C. il	C	USA	41	JL00011	C. il	SP	Jiande, ZJ	68	TTZ-20	C. da	SP	Luan, AH
15	Forkert	C. il	C	USA	42	SD1-2	C. il	SP	Jiande, ZJ	69	TTZ-22	C. da	SP	Luan, AH
16	Choctaw	C. il	C	USA	43	SJ2	C. il	SP	Jiande, ZJ	70	TTZ-11	C. da	SP	Luan, AH
17	Creek	C. il	C	USA	44	SJ1	C. il	SP	Jiande, ZJ	71	GL2	C. ca	S	Jinhua, ZJ
18	Mohawk	C. il	C	USA	45	YZXT	C. il	SP	Jiande, ZJ	72	XK24	C. ca	S	Jinhua, ZJ
19	Elliott	C. il	C	USA	46	YZXD	C. il	SP	Jiande, ZJ	73	DY1	C. ca	S	Jinhua, ZJ
20	Mahan	C. il	C	USA	47	YZXG1	C. il	SP	Jiande, ZJ	74	DY6	C. ca	S	Jinhua, ZJ
21	YLC21	C. il	C	Jiande, ZJ	48	YZXG2	C. il	SP	Jiande, ZJ	75	GL8	C. ca	S	Jinhua, ZJ
22	XH1	C. il	ES	Hangzhou, ZJ	49	JL4	C. il	SP	Jiande, ZJ	76	XK88	C. ca	S	Jinhua, ZJ
23	XH4	C. il	ES	Hangzhou, ZJ	50	CHK1	C. il	SP	Hangzhou, ZJ	77	DY5	C. ca	S	Jinhua, ZJ
24	XH5	C. il	ES	Hangzhou, ZJ	51	CHK2	C. il	SP	Hangzhou, ZJ	78	GL7	C. ca	S	Jinhua, ZJ
25	XH6	C. il	ES	Hangzhou, ZJ	52	CHK3	C. il	SP	Hangzhou, ZJ	79	XK89	C. ca	S	Jinhua, ZJ
26	XH14	C. il	ES	Hangzhou, ZJ	53	CHK4	C. il	SP	Hangzhou, ZJ	80	XK40	C. ca	S	Jinhua, ZJ
27	DFH	C. il	ES	Jinhua, ZJ	54	JLCD	C. il	SP	Jiande, ZJ					

**Note:** <sup>1</sup> C. il = *C. illinoensis*, C. ca = *C. cathayensis*, C. da = *C. dabieshanensis*, C. hu = *C. hunanensis*; <sup>2</sup> C = Cultivar, ES = Excellent seedling, S = Strain, SP = Seeding plant; <sup>3</sup> ZJ = Zhejiang province, GZ = Guizhou province, JS = Jiangsu province.



**Figure 1.** The nuts of four *Carya* species. The scale is one centimeter.

## 2.2. gSSR Mining and Primer Design

The genomic sequences of ‘87MX3-2’ (<http://www.hagsc.org/pecan>) and ‘Pawnee’ ([ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001\\_101000/100571/](ftp://parrot.genomics.cn/gigadb/pub/10.5524/100001_101000/100571/)) were downloaded from public databases, respectively. The Microsatellite identification tool (MISA, <http://pgrc.ipk-gatersleben.de/misa/misa.html>) was used to search SSR motifs in the genome of each cultivar (Thiel et al., 2003), with repeats with a minimum of 6, 5, 5, 5, and 5 for di-, tri-, tetra-, penta-, and hexanucleotides, respectively. For compound SSRs, the maximum interruption between two SSRs was set as 100 bases. Primers for the SSR loci were designed using p3\_in.pl, primer3\_core, and p3\_out.pl programs under default parameters (Untergasser et al., 2012). Then, the low-quality primer pairs were subsequently removed, such as those featuring the following: (1) the primers containing unknown bases; (2) a 3′ end of primers with three consecutive identical bases; (3) a  $T_m$  value difference between forward and reverse primers of more than 1 °C; (4) a GC content of each primer >55% or <35%; (5) the ratio of GC content between forward and reverse primers of >1.2 or <0.8.

## 2.3. Amplification and Validation of gSSRs

A total of 74 gSSR loci were randomly selected. Genomic DNA from ‘Pawnee’ and ‘Mahan’ were selected to amplify target products and validate the usability of these primers. PCR reactions were performed in 16 µL reaction volumes, containing 7.5 µL of 2 × Tsingke Master mix (Beijing Tsingke Biological Technology Co., Ltd., Beijing, China), 1 µL of each primer (10 pmol), and 1 µL of cDNA and 4.5 µL ddH<sub>2</sub>O. The PCR procedure was 94 °C for 5 min, then 35 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s, with a final extension at 72 °C for 5 min. PCR products were electrophoresed using 1% agarose under 300 V for 12 min. The primers which amplify the expected product size were selected and labeled with HEX, ROX, FAM, or TRAMA. The PCR amplifications were performed again, and the products were separated using capillary electrophoresis on an ABI 3730 sequencer (Applied Biosystems, Foster City, CA, USA). Finally, the primers that gave clearly distinguishable peaks were used for the genetic study of the pecan population.

## 2.4. Genetic Diversity Analysis of Pecan and Transferability of gSSRs to Other Species

Twenty-two validated gSSRs and eight SSRs from Grauk et al. (2003) were chosen to amplify genomic DNA from 80 genotypes [13]. PCR products from each individual were separated using capillary electrophoresis. The genotype of each SSR locus was analyzed by Gene Mapper 4.1 software (Applied Biosystems, Foster City, CA, USA). Several genetic parameters, including observed allele number ( $N_a$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), and polymorphic information content (PIC), were calculated by using the Popgen 32 program (University of Alberta and Center for International Forestry Research, Canada). A phylogenetic dendrogram was constructed by using the unweighted pair group method with arithmetic average (UPGMA).

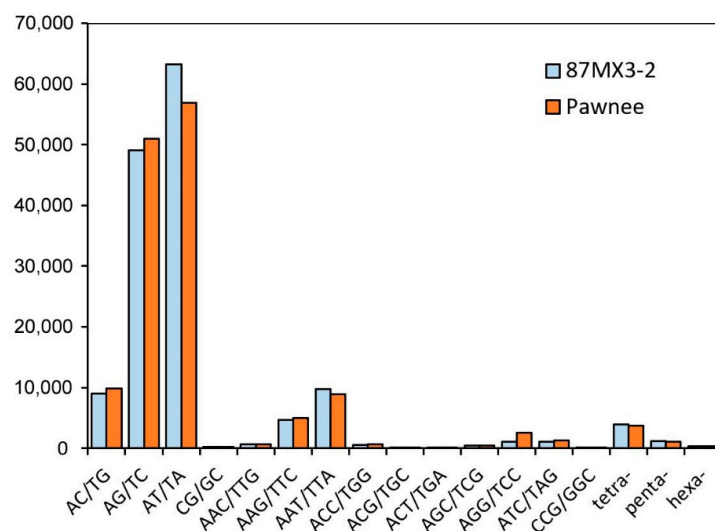
### 3. Results

#### 3.1. Characteristics of gSSRs

The genomic sequences of ‘87MX3-2’ and ‘Pawnee’ were downloaded and analyzed using the MISA program. The results showed that 48,009 (38.54%) out of 124,560 scaffolds (531.7 Mb) of ‘87MX3-2’ were found to contain SSR motifs (Table 2). In total, 145,714 gSSRs were confirmed (Table S1). So, the frequency of gSSR in ‘87MX3-2’ was 1 gSSR per 3.65 kb. Among different types of SSR motifs, the dinucleotide (121,626, 83.47%) was the most predominant, followed by trinucleotide (18,593, 12.76%) (Figure 2). For the dinucleotide repeat motif, AT/TA (53.60%) was the main type, while CG/GC (0.21%) was the least frequent. In total, 143,041 gSSRs were identified in 4955 (11.47% out of 43,183 examined sequences) sequences in the genome of ‘Pawnee’ (651.2 Mb; Table S1). The frequency of gSSR is 1 gSSR per 4.55 kb. Dinucleotide repeat motifs (118,045, 82.53%) formed the main category, followed by trinucleotide (19,818, 13.85%). The most and the least present types of dinucleotide repeat motif were AT/TA (48.22%) and CG/GC (0.22%), respectively.

**Table 2.** Information of simple sequence repeats (SSRs) identified in the pecan genome.

Features	Values	
	‘87MX3-2’	‘Pawnee’
Total number of sequences examined	124,560	43,183
Total size of examined sequences (bp)	531,741,495	651,240,950
Total number of identified SSRs	145,714	143,041
Number of SSR containing sequences	48,009	4955
Number of sequences containing more than 1 SSR	27,045	1961
Number of SSRs present in compound formation	21,060	17,816



**Figure 2.** The motif distribution pattern of pecan genomic simple sequence repeats (gSSRs) (considering sequence complementary).

#### 3.2. Primer Design and Validation

A total of 60,018 primer pairs for 20,006 gSSRs in 7143 sequences were designed by using Primer 3 software (Whitehead Institute, Cambridge, MA, USA). Then, 9145 (15.24%) candidate primer pairs to 5002 gSSRs (25.00%) were obtained after the deletion of low-quality ones (Table S2). To validate this, a set of 74 primer pairs were randomly picked and synthesized. All of them were distributed in 74 contigs of ‘87MX3-2’ genome sequences. These loci were distributed in 70 scaffolds of ‘Pawnee’ genome sequences: Ciz034 and Ciz035 in scaffold72459, Ciz037 and Ciz058 in scaffold69125, Ciz039 and Ciz044

in scaffold101014, and Ciz065 and Ciz074 in scaffold89570. To validate these markers, the 74 primers were amplified in ‘Pawnee’ and ‘Mahan’. After electrophoresis, 66 primer pairs (89.19%) yielded the expected bands (Table S3). The size of amplicons ranged from 111 bp (Ciz028) to 280 bp (Ciz006). Six loci (Ciz018, Ciz032, Ciz042, Ciz063, Ciz064, and Ciz065) had no PCR product observed. Two loci (Ciz012 and Ciz026) obtained a nontarget product.

### 3.3. Genetic Diversity Analysis and Cross-Species Transferability of gSSRs

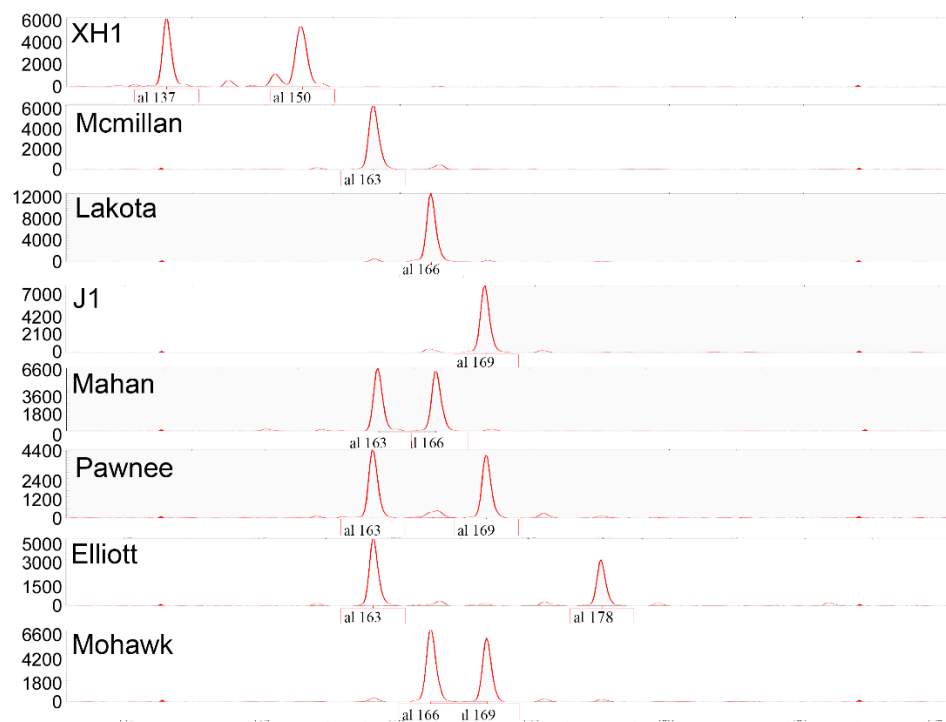
Twenty-two randomly selected gSSRs from 66 validated loci and eight SSRs from a previous study were employed in this section (Table 3) [13]. In total, 60 pecan genotypes and 20 individuals from three other species were used to perform genetic diversity analysis and cross-transferability, respectively. The PCR products were separated using capillary electrophoresis (Figure 3). Among the 22 newly developed gSSR loci, 21 (95.45%, except Ciz036) exhibited polymorphism in the examined population (Table 3). For all loci, the observed allele number ( $N_a$ ) ranged from 1 (Ciz036) to 18 (Ciz031). A total of 221 alleles were detected in pecan, and the mean number of alleles per locus was 7.369. The observed heterozygosity ranged from 0.00 (Ciz036) to 1 (Ciz047) with a mean value of 0.439. In addition, the minimum and the maximum of PIC values were 0.00 (Ciz036) and 0.893 (Ciz031), with 23 loci showing high PIC values ( $PIC > 0.5$ ), and the mean value was 0.547. Twenty-six (86.67%) gSSRs could yield the target product in hickory, while 25 (83.33%) and 22 (73.33%) loci could be used in *C. dabieshanensis* and *C. hunanensis*, respectively (Table 4).

**Table 3.** The genetic characteristics of 30 SSR loci in 60 pecan genotypes.

NO. <sup>1</sup>	Loci	Fluorescence Dye	$N_a$ <sup>2</sup>	$H_o$ <sup>3</sup>	$H_e$ <sup>4</sup>	PIC <sup>5</sup>	Size (bp)
1	Ciz003	ROX	10	0.695	0.741	0.735	137–178
2	Ciz009	ROX	5	0.525	0.517	0.513	157–163
3	Ciz011	ROX	2	0.068	0.097	0.096	202–204
4	Ciz022	FAM	5	0.650	0.630	0.625	223–237
5	Ciz031	FAM	18	0.542	0.901	0.893	237–257
6	Ciz036	TRAMA	1	0.000	0.000	0.000	264
7	Ciz038	HEX	3	0.300	0.551	0.546	213–216
8	Ciz039	ROX	4	0.200	0.390	0.387	198–202
9	Ciz040	FAM	5	0.367	0.475	0.471	219–227
10	Ciz043	TRAMA	14	0.712	0.800	0.793	260–278
11	Ciz045	FAM	6	0.224	0.549	0.544	182–253
12	Ciz046	FAM	12	0.237	0.755	0.748	250–262
13	Ciz047	FAM	5	1.000	0.646	0.641	236–241
14	Ciz050	FAM	3	0.070	0.102	0.101	243–261
15	Ciz052	ROX	9	0.517	0.601	0.596	203–223
16	Ciz055	TRAMA	16	0.690	0.899	0.891	249–268
17	Ciz058	FAM	13	0.550	0.807	0.801	199–247
18	Ciz059	TRAMA	7	0.633	0.708	0.702	232–265
19	Ciz070	TRAMA	12	0.695	0.824	0.817	265–280
20	Ciz071	TRAMA	5	0.817	0.751	0.744	269–273
21	Ciz073	TRAMA	5	0.390	0.512	0.507	273–285
22	Ciz074	ROX	3	0.339	0.333	0.330	180–186
23	PM-CA10	HEX	2	0.051	0.050	0.050	92–113
24	PM-CIN13	HEX	6	0.441	0.510	0.506	102–116
25	PM-CIN20	ROX	2	0.017	0.017	0.016	139–142
26	PM-CIN22	HEX	5	0.117	0.590	0.585	96–103
27	PM-CIN27	HEX	12	0.800	0.775	0.768	62–92
28	PM-CIN4	HEX	16	0.783	0.796	0.789	90–143
29	PM-GA38	HEX	9	0.627	0.602	0.596	81–100
30	PM-GA41	HEX	6	0.105	0.637	0.631	69–90
-	Mean	-	7.367	0.439	0.552	0.547	-
-	St. Dev	-	4.752	0.283	0.266	0.264	-

**Note:** <sup>1</sup> The SSR loci of 23–30 from a previous study [13]; <sup>2</sup>  $N_a$  = observed allele number; <sup>3</sup>  $H_o$  = observed heterozygosity; <sup>4</sup>  $H_e$  = expected heterozygosity; and <sup>5</sup> PIC = polymorphic information content.





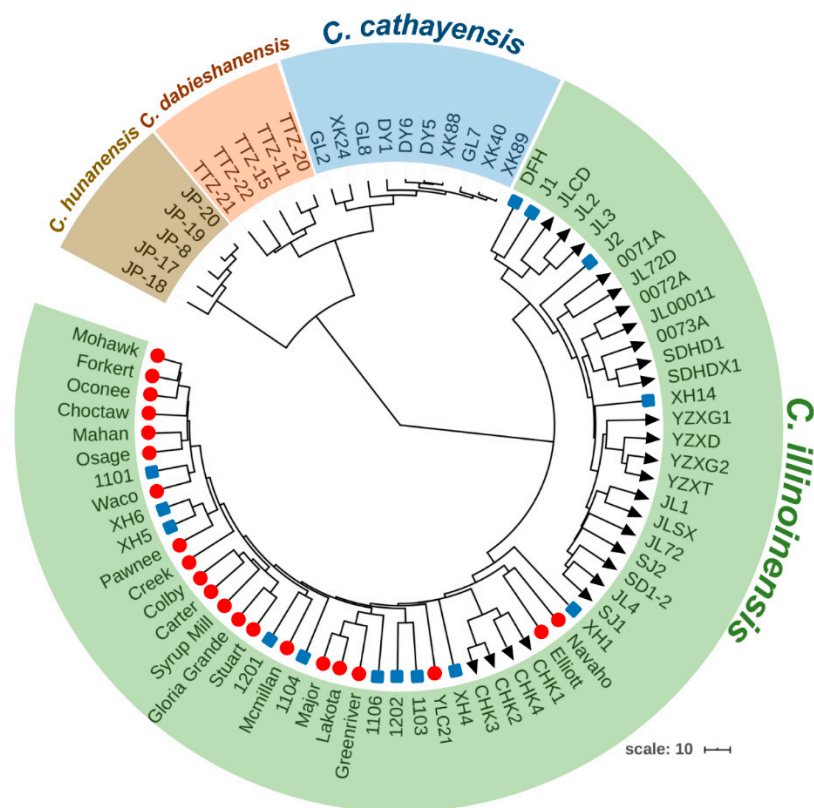
**Figure 3.** The capillary electrophoresis patterns of primer Ciz003 in eight pecan germplasms.

**Table 4.** The cross-transferability of 30 SSR loci in three species.

Locus	<i>C. cathayensis</i> (n = 10)		<i>C. dabieshanensis</i> (n = 5)		<i>C. hunanensis</i> (n = 5)	
	Na	Size (bp)	Na	Size (bp)	Na	Size (bp)
Ciz003	-	-	-	-	-	-
Ciz009	1	167	1	167	2	167–176
Ciz011	1	203	1	203	1	203
Ciz022	2	228–230	1	228	1	230
Ciz031	2	249–255	3	252–255	1	253
Ciz036	1	264	1	264	1	264
Ciz038	2	216–218	1	218	-	-
Ciz039	1	199	3	195–202	2	196–197
Ciz040	1	222	1	222	2	216–222
Ciz043	1	260	1	260	2	260–261
Ciz045	6	205–227	4	205–224	1	224
Ciz046	-	-	-	-	-	-
Ciz047	1	238	2	238–241	2	238–241
Ciz050	4	239–254	2	230–240	2	222–240
Ciz052	3	251–255	1	254	1	254
Ciz055	1	263	1	263	3	260–264
Ciz058	-	-	3	233–241	-	-
Ciz059	2	266–267	2	264–267	1	264
Ciz070	-	-	-	-	-	-
Ciz071	2	255–266	2	255–266	2	255–266
Ciz073	4	266–270	1	270	2	267–270
Ciz074	3	175–180	3	177–183	-	-
PM-CA10	1	134	3	116–136	-	-
PM-CIN13	3	104–111	-	-	-	-
PM-CIN20	1	139	1	139	1	139
PM-CIN22	1	95	1	95	1	94
PM-CIN27	2	62–66	-	-	1	68
PM-CIN4	3	113–118	2	113–114	2	105–118
PM-GA38	2	80–90	2	80–82	2	82–92
PM-GA41	2	95–101	1	101	1	99

### 3.4. Population Structure and Cluster Analysis

The 80 individuals were mainly divided into two clusters (Figure 4). Cluster I contained all samples originating from China, including 10 hickory, five *C. dabieshanensis*, and five *C. hunanensis*. According to three different species, cluster I was divided into three subclasses. The genetic distance between hickory and *C. dabieshanensis* was relatively close. The samples of pecan were clustered together and showed two subclasses. Interestingly, all the cultivars and most excellent strains (10 out of 14) were classified together, while nearly all the seedling trees (21 out of 25) and four excellent strains were clustered into another subgroup. Nine cultivars ('Mohawk', 'Forkert', 'Oconee', 'Choctaw', 'Mahan', 'Osage', 'Waco', 'Pawnee', and 'Creek') and three excellent strains (1101, XH5, and XH6) appeared to have a short genetic distance. In addition, 'Colby', 'Carter', 'Syrup Mill', 'Gloria Grande', and 'Stuart'; 'Major', 'Lakota', and 'Greenriver'; and 'Elliott' and 'Navaho' were clustered together, respectively (Figure 4).



**Figure 4.** The dendrogram of 80 germplasms constructed by the unweighted pair group method with arithmetic average (UPGMA). The plant samples included 60 *C. illinoensis*, 10 *C. cathayensis*, five *C. dabieshanensis*, and five *C. hunanensis*. The positions of cultivars, strains, and seedling plants of pecan are indicated by red circles, blue squares, and black triangles.

## 4. Discussion

Recently, next and third-generation sequencing technologies have been employed to investigate molecular mechanisms of key traits and construct a reference genome in pecan [3,4,12,28]. Transcriptomes and draft genomes offered an opportunity to identify SSR sites on a large scale. In the present study, two draft genomes of '87MX3-2' and 'Pawnee' were independently analyzed to mine genomic SSR loci [14,28]. The frequencies of the occurrence of gSSRs were 1/3.65 kb and 1/4.55 kb in '87MX3-2' and 'Pawnee', respectively. The gSSR densities between two cultivars showed slight differences. The genomic SSR frequency was significantly higher than genic SSR (1/6.10 kb; 6860 SSRs, including di-, tri-, tetra-, penta-, and hexanucleotides, in 41,858,722 bp sequences) in



pecan [29], which confirmed that the gSSR is more abundant than transcriptomic SSRs in this plant. The gSSR density is higher than bamboo (1 SSR/16 kb) [31], and lower than walnut (1 SSR/2.3 kb) and Chinese jujube (1 SSR/1.17 kb) [32,33]. Among dinucleotide repeat motifs, AT/TA was the most abundant type, followed by AG/TC (Figure 2). This result was similar to gSSRs in Chinese Jujube and cumin [15,34]. Sixty-six (89.19%) out of 74 primers were successfully validated by using PCR in two cultivars (Table S3). The success rate was similar to *Cyamopsis tetragonoloba* (86.4%) [35], higher than *Camellia sinensis* (32.00%) and *Exbucklandia tonkinensis* (76.26%) [30,36].

The genus *Carya* consists of 17 species [1], of which pecan, hickory, *C. hunanensis*, and *C. dabieshanensis* have significant economic values in China [3]. However, the molecular markers in these species are far from sufficient. In this study, 66 gSSRs were validated in ‘Pawnee’ and ‘Mahan’, and 21 newly developed loci and eight previous SSRs showed genetic polymorphisms in different germplasms (Table 3). Among them, 23 loci showed high PIC values (>0.5). Therefore, these gSSR loci exhibited high polymorphisms in different germplasms and could be used in genetic diversity study, genetic map construction, Quantitative trait loci (QTL) mapping, as well as cultivar identification in pecan. For the cross-transferability of gSSR loci, 86.67% (26/30) of gSSRs can be used in hickory, which was higher than that in a previous study (63.02%) [20]. Moreover, the transferability rates were 83.33% (25/30) and 73.33% (22/30) between pecan and two other *Carya* species of *C. dabieshanensis* and *C. hunanensis*, respectively (Table 4). Similarly, high transferability rates also have been reported in other species, such as *Pistacia vera* L. [37], *Casuarina* L. ex Adans [38], and *Melilotus* [39].

SSR is a useful tool for evolutionary studies and pedigree relationships evaluation [16]. As is shown in Figure 4, the species of *C. hunanensis*, *C. dabieshanensis*, and *C. cathayensis* were clustered together, and the last two species showed a closer relationship. These observations are consistent with the results of Zhang et al. (2013), who reported that the *Carya* species from eastern Asian and eastern North American were classified into two different groups, respectively [40]. For pecan, ‘Mohawk’, ‘Forkert’, ‘Oconee’, ‘Choctaw’, ‘Mahan’, ‘Osage’, ‘Waco’, ‘Pawnee’, and ‘Creek’ exhibited a wide range of similarities. Similar results were also reported by previous studies [12,13]. ‘Mohawk’ and ‘Choctaw’ originated from a controlled cross of ‘Success’ × ‘Mahan’ [14,41] (Table S4), and ‘Mohawk’ was the parent of ‘Pawnee’ and ‘Creek’. ‘Forkert’ was selected from offspring of ‘Success’ × ‘Schley’ [14,41]. ‘Schley’ was an ancestor of ‘Oconee’ and ‘Waco’ [41]. Therefore, these individuals have pedigree relationships with each other, and the group patterns agreed with this fact. Incidentally, ‘1101’, ‘XH5’, and ‘XH6’ were also contained in this subgroup, which implies that these excellent seedlings might be progenies of these cultivars. Presumably, ‘Stuart’ might be the parent of ‘Gloria Grande’ [41]. Here, ‘Gloria Grande’ and ‘Stuart’ showed a higher similarity, which provided good support for this speculation. In addition, ‘Lakota’ originated from ‘Mahan’ × ‘Major’, and ‘Greenriver’ was selected from the same woods as ‘Major’. Therefore, these three cultivars were joined together. Pecan was introduced into China in the 1900s. Currently, a large number of seedling trees is spread through several provinces in China. However, the collection and genetic study of these germplasms is still insufficient. As shown in Figure 4, a relatively long genetic distance exists between cultivars and most seedling trees in China. Therefore, the hybridization between the two types of germplasms might create more genetic variation, which might be useful for cross breeding in the future.

## 5. Conclusions

Pecan is an important multipurpose tree worldwide. However, the progress of genetic study in pecan is limited by the lack of a sufficient number of accurate, convenient, and economical molecular markers. In the present study, the distribution of SSR motifs was evaluated in the pecan genome. The gSSRs densities were 1/4.55 kb (‘Pawnee’) to 1/3.65 kb (‘87MX3-2’). In total, 60,018 primer pairs for 20,006 gSSRs were designed, and a set of 66 gSSRs were successfully validated. Thirty gSSRs were employed to analyze the genetic diversity and progeny relationship between pecan genotypes and their cross-transferability to other *Carya* species. In total, 29 SSR loci showed polymorphism, and 26, 25, and 22 gSSRs can be used in hickory, *C. dabieshanensis*, and *C. hunanensis*, respectively. We believe

that the characterization of SSRs in the pecan genome and the new gSSRs reported here will promote the progress of genetic study and MAS research in pecan, as well as other species of the genus *Carya*.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1999-4907/11/1/61/s1>, Table S1: The frequency of identified SSR motifs, Table S2: The information of primers, Table S3: The information of 66 primer pairs used in the present study, Table S4: The pedigree information of 21 cultivars.

**Author Contributions:** C.Z.: conceptualization, methodology, writing—original draft preparation. X.Y.: conceptualization, supervision. H.R.: data curation, writing—review and editing. J.C.: validation. J.W., W.S., and Q.F.: resources. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China No. 31800575, and the Fundamental Research Funds of CAF No. CAFYBB2018SY013.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Grauke, L.J.; Wood, B.W.; Harris, M.K. Crop Vulnerability: *Carya*. *J. Am. Soc. Hortic. Sci.* **2016**, *51*, 653–663. [CrossRef]
2. Gong, Y.; Pegg, R.B. Separation of ellagitannin-rich phenolics from U.S. pecans and Chinese hickory nuts using fused-core HPLC columns and their characterization. *J. Agric. Food Chem.* **2017**, *65*, 5810–5820. [CrossRef] [PubMed]
3. Zhang, C.; Yao, X.; Ren, H.; Chang, J.; Wang, K. RNA-Seq reveals flavonoid biosynthesis-related genes in pecan (*Carya illinoensis*) kernels. *J. Agric. Food Chem.* **2019**, *67*, 148–158. [CrossRef] [PubMed]
4. Huang, Y.; Xiao, L.; Zhang, Z.; Zhang, R.; Wang, Z.; Huang, C.; Huang, R.; Luan, Y.; Fan, T.; Wang, J.; et al. The genomes of pecan and Chinese hickory provide insights into *Carya* evolution and nut nutrition. *GigaScience* **2019**, *8*, giz036. [CrossRef]
5. Hilbig, J.; Policarpi, P.D.B.; Grinevicius, V.M.A.d.S.; Mota, N.S.R.S.; Toaldo, I.M.; Luiz, M.T.B.; Pedrosa, R.C.; Block, J.M. Aqueous extract from pecan nut [*Carya illinoensis* (Wangenh) C. Koch] shell show activity against breast cancer cell line MCF-7 and Ehrlich ascites tumor in Balb-C mice. *J. Ethnopharmacol.* **2018**, *211*, 256–266. [CrossRef]
6. Prado, A.C.P.d.; Manion, B.A.; Seetharaman, K.; Deschamps, F.C.; Barrera Arellano, D.; Block, J.M. Relationship between antioxidant properties and chemical composition of the oil and the shell of pecan nuts [*Carya illinoensis* (Wangenh) C. Koch]. *Ind. Crops Prod.* **2013**, *45*, 64–73. [CrossRef]
7. Qin, L.; Qian, H.; He, Y. Microbial lipid production from enzymatic hydrolysate of pecan nutshell pretreated by combined pretreatment. *Appl. Biochem. Biotechnol.* **2017**, *183*, 1336–1350. [CrossRef]
8. Villarreal-Lozoya, J.E.; Lombardini, L.; Cisneros-Zevallos, L. Phytochemical constituents and antioxidant capacity of different pecan [*Carya illinoensis* (Wangenh.) K. Koch] cultivars. *Food Chem.* **2007**, *102*, 1241–1249. [CrossRef]
9. Zhang, C.; Yao, X.; Ren, H.; Wang, K.; Chang, J. Isolation and characterization of three chalcone synthase genes in pecan (*Carya illinoensis*). *Biomolecules* **2019**, *9*, 236. [CrossRef]
10. Vendrame, W.A.; Kochert, G.; Wetzstein, H.Y. AFLP analysis of variation in pecan somatic embryos. *Plant Cell Rep.* **1999**, *18*, 853–857. [CrossRef]
11. Beedanagari, S.R.; Dove, S.K.; Wood, B.W.; Conner, P.J. A first linkage map of pecan cultivars based on RAPD and AFLP markers. *Theor. Appl. Genet.* **2005**, *110*, 1127–1137. [CrossRef] [PubMed]
12. Jia, X.D.; Wang, T.; Zhai, M.; Li, Y.R.; Guo, Z.R. Genetic diversity and identification of Chinese-grown pecan using ISSR and SSR markers. *Molecules* **2011**, *16*, 10078–10092. [CrossRef] [PubMed]
13. Grauke, L.J.; Iqbal, M.J.; Reddy, A.S.; Thompson, T.E. Developing microsatellite DNA markers in pecan. *J. Am. Soc. Hortic. Sci.* **2003**, *128*, 374–380. [CrossRef]
14. Bentley, N.; Grauke, L.J.; Klein, P. Genotyping by sequencing (GBS) and SNP marker analysis of diverse accessions of pecan (*Carya illinoensis*). *Tree Genet. Genom.* **2019**, *15*, 8. [CrossRef]
15. Bharti, R.; Kumar, S.; Parekh, M.J. Development of genomic simple sequence repeat (gSSR) markers in cumin and their application in diversity analyses and cross-transferability. *Ind. Crops Prod.* **2018**, *111*, 158–164. [CrossRef]
16. Vieira, M.L.; Santini, L.; Diniz, A.L.; Munhoz Cde, F. Microsatellite markers: What they mean and why they are so useful. *Genet. Mol. Biol.* **2016**, *39*, 312–328. [CrossRef]

17. Zhong, Y.; Yang, A.; Li, Z.; Zhang, H.; Liu, L.; Wu, Z.; Li, Y.; Liu, T.; Xu, M.; Yu, F. Genetic diversity and population genetic structure of *Cinnamomum camphora* in south China revealed by EST-SSR markers. *Forests* **2019**, *10*, 1019. [\[CrossRef\]](#)
18. Grauke, L.J.; Mendoza-Herrera, M.A.; Binzel, M.L. Plastid microsatellite markers in *Carya*. *Acta Hortic.* **2010**, *859*, 237–246. [\[CrossRef\]](#)
19. Grauke, L.J.; Mendoza-Herrera, M.A.; Miller, A.J.; Wood, B.W. Geographic patterns of genetic variation in native pecans. *Tree Genet. Genom.* **2011**, *7*, 917. [\[CrossRef\]](#)
20. Li, J.; Zeng, Y.; Shen, D.; Xia, G.; Huang, Y.; Huang, Y.; Chang, J.; Huang, J.; Wang, Z. Development of SSR markers in hickory (*Carya cathayensis* Sarg.) and their transferability to other species of *Carya*. *Curr. Genom.* **2014**, *15*, 357–379. [\[CrossRef\]](#)
21. Conner, P.J. Roadblocks and hindrances to the development of molecular tools in pecan (*Carya illinoensis*): A breeder's perspective. *Acta Hortic.* **2015**, *1070*, 83–87. [\[CrossRef\]](#)
22. Wang, L.; Wang, B.; Du, Q.; Chen, J.; Tian, J.; Yang, X.; Zhang, D. Allelic variation in PtoPsbW associated with photosynthesis, growth, and wood properties in *Populus tomentosa*. *Mol. Genet. Genom.* **2016**, *292*, 1–15. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Zhang, M.; Zhou, C.; Song, Z.; Weng, Q.; Li, M.; Ji, H.; Mo, X.; Huang, H.; Lu, W.; Luo, J. The first identification of genomic loci in plants associated with resistance to galling insects: A case study in *Eucalyptus* L'Hér. (Myrtaceae). *Sci. Rep.* **2018**, *8*, 2319. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Baisan, J.; Vidalis, A.; Zhou, L.; Chen, Z.Q.; Li, Z.; Sillanpää, M.J.; Bernhardsson, C.; Scofield, D.; Forsberg, N.; Grahn, T.; et al. Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. *Plant J.* **2019**, *100*, 83–100. [\[CrossRef\]](#)
25. Cao, K.; Zhou, Z.; Wang, Q.; Guo, J.; Zhao, P.; Zhu, G.; Fang, W.; Chen, C.; Wang, X.; Wang, X.; et al. Genome-wide association study of 12 agronomic traits in peach. *Nat. Commun.* **2016**, *7*, 13246. [\[CrossRef\]](#)
26. Müller, B.S.F.; de Almeida Filho, J.E.; Lima, B.M.; Garcia, C.C.; Missiaggia, A.; Aguiar, A.M.; Takahashi, E.; Kirst, M.; Gezan, S.A.; Silva-Junior, O.B.; et al. Independent and Joint-GWAS for growth traits in *Eucalyptus* by assembling genome-wide data for 3373 individuals across four breeding populations. *New Phytol.* **2019**, *221*, 818–833. [\[CrossRef\]](#)
27. Xu, L.Y.; Wang, L.Y.; Wei, K.; Tan, L.Q.; Su, J.J.; Cheng, H. High-density SNP linkage map construction and QTL mapping for flavonoid-related traits in a tea plant (*Camellia sinensis*) using 2b-RAD sequencing. *BMC Genom.* **2018**, *19*, 955. [\[CrossRef\]](#)
28. Jenkins, J.; Wilson, B.; Grimwood, J.; Schmutz, J.; Grauke, L.J. Towards a reference pecan genome sequence. *Acta Hortic.* **2015**, *1070*, 101–108. [\[CrossRef\]](#)
29. Jia, Z.; Wang, G.; Xuan, J.; Zhang, J.; Zhai, M.; Jia, X.; Guo, Z.; Li, M. Comparative transcriptome analysis of pecan female and male inflorescences. *Russ. J. Plant Physiol.* **2018**, *65*, 186–196. [\[CrossRef\]](#)
30. 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. [\[CrossRef\]](#)
31. Zhao, H.; Yang, L.; Peng, Z.; Sun, H.; Yue, X.; Lou, Y.; Dong, L.; Wang, L.; Gao, Z. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. *Sci. Rep.* **2015**, *5*, 8018. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Wu, J.; Gu, Y.Q.; Hu, Y.; You, F.M.; Dandekar, A.M.; Leslie, C.A.; Aradhya, M.; Dvorak, J.; Luo, M.C. Characterizing the walnut genome through analyses of BAC end sequences. *Plant Mol. Biol.* **2012**, *78*, 95–107. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Fu, P.C.; Zhang, Y.Z.; Ya, H.Y.; Gao, Q.B. Characterization of SSR genomic abundance and identification of SSR markers for population genetics in Chinese jujube (*Ziziphus jujuba* Mill.). *PeerJ* **2016**, *4*, e1735. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Xiao, J.; Zhao, J.; Liu, M.; Liu, P.; Dai, L.; Zhao, Z. Genome-wide characterization of simple sequence repeat (SSR) loci in Chinese jujube and jujube SSR primer transferability. *PLoS ONE* **2015**, *10*, e0127812. [\[CrossRef\]](#)
35. Tribhuvan, K.U.; SV, A.M.; Sharma, P.; Das, A.; Kumar, K.; Tyagi, A.; Solanke, A.U.; Sharma, R.; Jadhav, P.V. Identification of genomic SSRs in cluster bean (*Cyamopsis tetragonoloba*) and demonstration of their utility in genetic diversity analysis. *Ind. Crops Prod.* **2019**, *133*, 221–231. [\[CrossRef\]](#)

36. Huang, C.; Yin, Q.; Khadka, D.; Meng, K.; Fan, Q.; Chen, S.; Liao, W. Identification and development of microsatellite (SSRs) makers of *Exbucklandia* (HAMAMELIDACEAE) by high-throughput sequencing. *Mol. Biol. Rep.* **2019**, *46*, 3381–3386. [[CrossRef](#)]
37. Zaloğlu, S.; Kafkas, S.; Doğan, Y.; Güney, M. Development and characterization of SSR markers from pistachio (*Pistacia vera* L.) and their transferability to eight *Pistacia* species. *Sci. Hortic.* **2015**, *189*, 94–103. [[CrossRef](#)]
38. Xu, X.; Zhou, C.; Zhang, Y.; Zhang, W.; Gan, X.; Zhang, H.; Guo, Y.; Gan, S. A novel set of 223 EST-SSR markers in *Casuarina* L. ex Adans.: Polymorphisms, cross-species transferability, and utility for commercial clone genotyping. *Tree Genet. Genomes* **2018**, *14*, 30. [[CrossRef](#)]
39. 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](#)]
40. Zhang, J.B.; Li, R.Q.; Xiang, X.G.; Manchester, S.R.; Lin, L.; Wang, W.; Wen, J.; Chen, Z.D. Integrated fossil and molecular data reveal the biogeographic diversification of the eastern asian-eastern north american disjunct hickory genus (*Carya* Nutt.). *PLoS ONE* **2013**, *8*, e70449. [[CrossRef](#)]
41. Conner, P.J.; Wood, B.W. Identification of pecan cultivars and their genetic relatedness as determined by randomly amplified polymorphic DNA analysis. *J. Am. Soc. Hortic. Sci.* **2001**, *126*, 474–480. [[CrossRef](#)]



© 2020 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/>).