Pak. J. Bot., 56(6): DOI: http://dx.doi.org/10.30848/PJB2024-6(32)

EST-SSR MARKER DEVELOPMENT BASED ON DE NOVO TRANSCRIPTOMIC ASSEMBLY OF *AESCULUS CHINENSIS* BUNGE

YU LING LI[1,2θ], XIN MAO[3θ], TIAN XU SUN[4], CUI LAN LIU[1], YUAN SHUAI ZHANG[5], YUAN FU DONG[5] AND XIU HONG MAO[1*]

[1]*Shandong Provincial Key Laboratory of Forest Tree Genetic Improvement, Shandong Academy of Forestry, Jinan, 250014, Shandong, China*
[2]*Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou, 510520, Guangdong, China*
[3]*College of Forestry, Henan Agricultural University, Zhengzhou, 450000, Henan, China*
[4]*Shandong Institute of Territorial and Spatial Planning, Jinan, 250014, Shandong, China*
[5]*DaQingShan forest farm of Feixian, Linyi, 273400, Shandong, China*
**Corresponding author's email: xiuhongmao@shandong.cn*
[θ]*Contributed equally to this study*

Abstract

Aesculus chinensis Bunge, a deciduous tree in the genus *Aesculus* of the Hippocastanaceae family, has ornamental, economic, and medicinal values. But the scarcity of comprehensive genetic data hinders the utilization of this species, especially the EST-SSR markers in *A. chinensis* Bunge have not been reported previously. In this study, the transcriptomes of *A. chinensis* Bunge were assembled from novo, which enabled the identification of 33,365 unigenes with a mean length of 1348 bp. The sum of 25450 (76.27%) unigenes showed similarities with at least one database including Nr (71.39%), Swiss Prot (57.03%), GO (53.31%), KO (27.13%) and KOG (19.27%). Based on the transcriptome data collected for *A. chinensis* Bunge, 16,103 SSR loci were detected with a distribution density of one SSR per 2.79 kb. The most dinucleotide and trinucleotide repeats were AG/CT and AAG/CTT. Then 58 primer pairs were randomly selected for polymorphism validation and 31 primer pairs were polymorphic among eight *Aesculus* samples. There were 167 alleles detected, varying from 2 to 9 with a mean of 5.387 alleles per locus. According to the allelic variation, 7 primers were capable of distinguishing the 8 *Aesculus* samples from 6 species. These findings will be useful for the conservation, genetic diversity assessment and molecular identification of *Aesculus* resources.

Key words: *Aesculus chinensis* Bunge; Transcriptome sequencing; Gene annotation; EST-SSR markers; Molecular identification.

Introduction

A. chinensis Bunge, also known as Sala tree, belongs to the genus *Aesculus* of the Hippocastanaceae family. As deciduous trees, *A. chinensis* Bunge has a tall trunk, dark brown or grayish brown bark, leaflets ranging from five to seven and inflorescence rachis with puberulent, florets with 5-10 flowers, and subglobose seeds. *A. chinensis* Bunge grows wild only in the Qin Ling Mountains while being cultivated in southern Shaanxi, southern Hebei, and northern Henan of China (Yu, 1981). It was one of the four famous street trees around the world, which has many functions, such as ornament function, medicinal function, edible function, and so on. Seeds of the genus *Aesculus* contain aescin, which has a wide range of pharmacological activities, including anti-inflammation, inhibition of gastric acid secretion, anti-edema, scavenging of reactive oxygen species, and anti-liver injury (Liu & Zhou, 2010; Kim *et al*., 2017; Chen *et al*., 2019). Studies about molecular biology and the preservation of genetic resources in *A. chinensis* Bunge were very limited. Only the chloroplast genome of *A. chinensis* has been recently sequenced (Zhang *et al*., 2019; Liu *et al*., 2020).

Transcriptome Sequencing based on the Illumina HiSeq TM2000 has been extensively employed for transcriptome profiling, gene expression analysis, and the identification of functional genes due to its high accuracy, high speed, and low cost (Severin *et al*., 2010). Transcriptome sequencing can obtain the law of growth and metabolism of species under the condition of lack of gene resources and reveal the internal relationship between their biological characteristics and genes. It can also analyze and explore the gene structure and function on a deeper level, which is a cost-effective means of gene sequence research (Simon *et al*., 2009; Reddy *et al*., 2015). It is particularly useful for the analysis of non-model species containing large nuclear genomes as the assembly of the de novo transcriptome is functionally independent of the existing genome sequence (Wang *et al*., 2009). To date, transcriptome sequencing technology has been widely used in woody plants such as Liriodendron chinense (Yang *et al*., 2014), Hazelnut (Cheng *et al*., 2015), Tea (Guo *et al*., 2018), Blueberry (Qi *et al*., 2019), Date palm (Naganeeswaran *et al*., 2020), etc.

For non-model plants without reference genomes, transcriptome sequencing is an effective means to discover new genes and collect ESTs (expression sequence tags), which will make the development of molecular markers simpler (Wang *et al*., 2009; Grabherr *et al*., 2011). Molecular markers development via transcriptome sequencing has the advantages of being cheap, simple, and rapid, high polymorphism, and transferable among plant species, which can greatly improve working efficiency. Simple sequence repeat (SSR) is one of the most prevalent molecular markers, which is a subclass of tandem repeats consisting of 1-6 nucleotides, found in the genomes of all prokaryotes and eukaryotes (Taheri *et al*., 2014). SSRs are generally classified into EST-SSR (expression sequence tag-Simple sequence repeat) and genomic-SSR. EST-SSR markers are found in more conserved regions than markers derived from genomic sequences, so they are more transferable between species. EST-SSR have been applied in genetic diversity assessment, evolution, linkage map, variety identification, and marker-assisted selection

breeding because of their high polymorphism, repeatability, sensitivity, and easy detection (Liu *et al*., 2013, Zhou *et al*., 2016, Taheri *et al*., 2018; He *et al*., 2020).

The breeding programs in *A. chinensis* Bunge have been hindered due to the limited availability of genomic resources. In the current study, transcriptome sequencing and assemble were carried out on the leaves and inflorescences of *A. chinensis* Bunge and generated a set of SSR markers. The purpose of the study is to provide a theoretical basis for screening genes, developing SSR molecular markers, and further functional genome research.

Material and Methods

Materials and RNA extraction: The materials used in this experiment were the leaves and inflorescences of three *A. chinensis* Bunge plants, which were collected on the campus of Shandong University (117.060105E, 36.673287N, China) on April 10 of 2019. After sampling, the materials were quickly placed in liquid nitrogen and then transferred to a -80°C refrigerator. RNA degradation and contamination were monitored using 1% agarose gels.

The RNA purity was determined using a Nano Photometer® spectrophotometer (IMPLEN, CA, USA). To measure RNA integrity, the Agilent Bioanalyzer 2100 system's RNA Nano 6000 Assay Kit (Agilent Technologies, CA, USA) was utilized.

Transcriptome sequencing library preparation: RNA samples were prepared with 1.5 μg of RNA samples. Sequencing libraries were created using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) according to the manufacturer's instructions. Oligo (dT) magnetic beads were used to concentrate the mRNA + poly (A) tail, and fragmentation buffer was added to split the mRNA into short fragments. The first strand of cDNA was generated by the M-MuLV reverse transcriptase system using fragment mRNA as a template and random oligonucleotides as primers, while the second-strand cDNA was produced by the buffer, dNTPs, RNaseH, and DNA polymerase I. On the basis of amplified products, cDNAs were purified using the AMPure XP equipment, and library quality was assessed using the Agilent Bioanalyzer 2100 system.

Clustering and sequencing: According to the manufacturer's instructions, the TruSeq PE Cluster Kit v3-cBot-HS (Illumia) was used to cluster the index-coded sample data on a cBot Cluster Generation System. The library preparations were sequenced on an Illumina Hiseq TM 2000 platform after cluster creation, and paired-end reads were produced (Reuter *et al*., 2015).

Analysis of data

Quality control: Raw fast q format data were initially processed using in-house Perl programs. Clean data were acquired in this stage by eliminating reads including adaptor, reads containing ploy-N, and low quality reads from raw data. At the same time, the clean data's Q20, Q30, GC-content, and sequence duplication levels were evaluated (Grabherr *et al*., 2011). All downstream analysis relied on clean, high-quality data.

Transcriptome assembly: All libraries/samples' left files (read1 files) were combined into a single large left. fq file, and all libraries/samples' right files (read2 files) were combined into a single large right. fq file. Trinity v2.4.0 (https://github.com/trinityrnaseq/trinityrnaseq/issues/270), with min kmer cov set to 2 by default and all other settings set to default, was used to assemble the transcriptome based on the left. fq and right. fq.

Annotation of gene functions: The following databases have had gene functions annotated.: Nr (e-value = 1e-5); Pfam (Protein family; e-value = 0.01); Nt (e-value = 1e-5); KOG (Clusters of Orthologous Groups of proteins; e-value = 1e-5); Swiss-Prot (e-value = 1e-5); KO (KEGG Ortholog database; e-value = 1e-10); GO (Gene Ontology; e-value = 1e-6).

SSR identification and primer design: MISA was used to identify SSR loci in the transcriptome (Haas *et al*., 2013). The following repetition units were selected: Mono-nucleotides contain 10 repeat units, di-nucleotides contain 6, and tri-, tetra-, penta-, and hexa-nucleotides contain 5. Primer 3 (version 2.3.5 with default parameter values) was used to SSR primers design (Untergasser *et al*., 2012) and primer synthesis was carried out by Beijing Rui Bo Xing Ke Biotechnology Co., LTD (Beijing, China).

Amplification and validation of EST-SSR: Based on the obtained EST-SSRs, eight species of *Aesculus* were chosen for polymorphism tests. FastPure Plant DNA Isolation Mini Kit (Vazyme Biotech Co., Ltd.) was used to extract genomic DNA. The PCR reaction was performed out in a volume of 20 μL, containing 2 × Mix (10 μL), M13 forward primer (0.2 μL), reverse primer (0.3μL), forward primer (0.1μL), DNA (1μL) and add H_2O to a volume of 20 μL. PCR amplification was performed at 95°C for 5 min, followed by 20 denaturation cycles (95°C for 30 s, 55°C for 30 s, and 72°C for 30 s. then 95°C for 30 s, 52°C for 30 s, 72°C for 30 s) and a final extension for 10 min at 72°C. Capillary electrophoresis (ABI manufacturer) was used to confirm the PCR fragments. Capillary electrophoretograms were read and analyzed using Gene Marker V2.2.0 software (https://genem arker. soltw are.infor mer.com/2.2/). Using the GenALEX6.5 tool, the number of alleles (*Na*), the number of effective alleles (*Ne*), Observed heterozygosity (*Ho*), expected heterozygosity (*He*), and Shannon's information index (*I*) were calculated (Peakall & Smouse., 2012). The SIMQUAL program of NTSYS-pc Version 2.1 was used to estimate genetic similarity coefficients for 8 *Aesculus* species. The dendrogram was constructed with NTSYS-pc SHAN clustering program using the UPGMA algorithm.

Results

De novo assembly: A total of 758,298,28 raw reads were generated by the Illumina HiSeq TM 2000 sequencing of *A. chinensis* Bunge. To carry out the de novo assembly, 748,057,34 clean reads were used to generate 888,69 transcripts. The transcript lengths ranged from 301 to 18,146 bp, with the average length of 1550 bp and N50 of 2109 bp (Table 1). 64,576 (72.66%) transcripts length ranged from 301 to 2000 bp (Fig. 1). The sequencing result was accurate and can be used for subsequent transcriptome

assembly. The transcripts were assembled into 333,65 unigenes, with N50 length of 1959 bp and a mean length of 1348bp (Table 1), which were longer than the *Lycium barbarum* with N50 and the average length was 1579bp and 1049bp (Chen *et al*., 2017). Among the 333,65 unigenes, 26,118 (78.28%) were between 301bp and 2 kb in size, 7247 (21.72%) were over 2 kb in length (Fig. 1).

Gene functional annotation: The 23820, 19031, 17789, 17789, 9054, and 6430 unigenes were annotated according to Nr, Swiss-Prot, Pfam, GO, KO, and KOG databases, respectively (Table 2). Among them, unigenes had the most functional annotations in the Nr database (71.39%). In at least one database, 25450 (76.27%) unigenes had homologous matches, and 3989 (11.95%) unigenes were annotated in all seven public databases (Table 2).

By comparing and annotating with the Nr database, we can obtain the similarity between the gene sequence of *A. chinensis* Bunge and the gene sequences of related species. 23820 (71.39%) unigenes were assigned to Nr database in this study. The e-value distributions of high homologous sequences in Nr database had 48.20% (0-1e-100) with high homology (Fig. 2A). while the similarity distribution shows that 91.6% of the sequences with more than 60% similarity and 52.7% of the sequences with more than 80% comparability (Fig. 2B). Furthermore, the highest matches in the Nr database were *Citrus clementina* (18.8%), *Citrus sinensis* (17.3%), and *Citrus unshiu* (14.8%) (Fig. 2C). In addition, 42.1% of the unigenes belonged to other sequences, which may be different from other tree species or belonged to a sequence with a non-coding region.

6430 unigenes were assigned to KOG database, and 7164 functional annotations were obtained, which were divided into 25 categories. Posttranslational modification, protein turnover, and chaperones (876, 12.22%) group were maximal. followed by General function prediction only (833,11.62%), Translation, ribosomal structure, and biosynthesis (673, 9.39%) (Fig. 3). The smallest group was Cell motility (3), followed by extracellular structure (9). Through the functional annotation classification of KOG database. It was found that these unigenes were abundant in protein posttranslational modification, protein conversion, translation, ribosomal structure, and biosynthesis. In addition, General function prediction and unknown function accounted for 17.92% of the total, indicating that some genes whose functions are unclear and need to be further studied.

A total of 97,978 annotations of 17,789 sequenced genes were obtained in the GO database, which was divided into 3 categories and 56 subcategories (Fig. 4). The unigenes were classified into biological processes categories (46,486, 47.45%), followed by cellular components (29,402, 30.00%) and molecular functions (22,090, 22.55%). The greatest categories in the category of biological process were cellular process (10,409, 22.39%) and metabolic process (9518, 20.47%), followed by single organism process (7596, 16.34%), biological regulation (3760, 8.09%), regulation of biological process (3465, 7.45%) and localization (3071, 6.61%) (Fig. 4). In the cellular component, cells and cell parts accounted for 39.02%, the organelle and organelle part accounted for 20.34%, and the membrane and membrane part accounted for 20.53%. In the molecular function cluster, binding (10,375, 46.97%) and catalytic activities (8149, 36.89%) were the largest categories (Fig. 4).

Table 1. Summary of Illumina sequencing-based sequence assembly.

Category	Items	Value
Reads	Total raw reads	758,298,28
	Total clean reads	748,057,34
	Clean base	11.22Gb
	Error rate (%)	0.03
	Q20 percentage (%)	98.05
	Q30 percentage (%)	94.18
	GC percentage (%)	45.41
Transcripts	Minimum length	301
	Maximum length	18146
	Average length	1550
	Length of N50	2109
	Total nucleotides (nt)	137,750,895
Unigenes	Minimum length	301
	Maximum length	18146
	Average length	1348
	Length of N50	1959
	Total nucleotides (nt)	449,654,05

Note: Q20/Q30: represents the correct base identification rate of 99%/99.9%

N50: Arrange all of the sequences from long to short and sum the length of each in that order. when the added length equals 50% of the entire length of the sequence, the length of the current sequence is reached

Table 2. Unigene annotations in the public database.

Database	Number of unigenes	Percentage (%)
Nr	23820	71.39
NT	19747	59.18
Swiss-Prot	19031	57.03
PFAM	17789	53.31
GO	17789	53.31
KO	9054	27.13
KOG	6430	19.27
Annotated in all databases	3989	11.95
Annotated in at least one databases	25450	76.27
Without annotation genes	7915	23.73
Total unigenes	33365	100.00

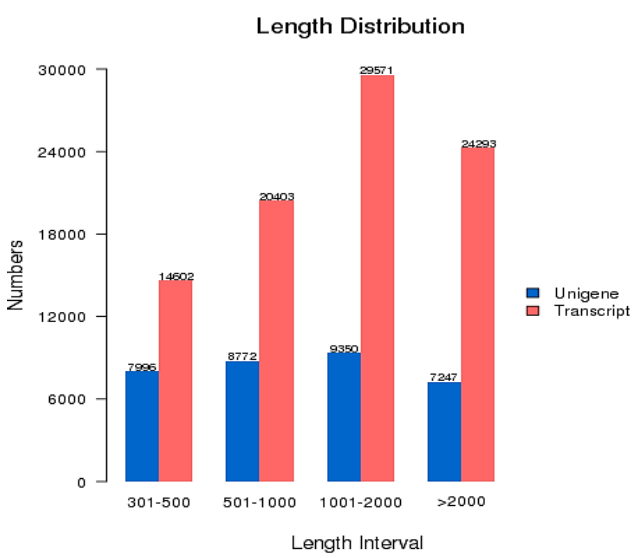


Fig. 1. Length distribution of Unigene and Transcripts.

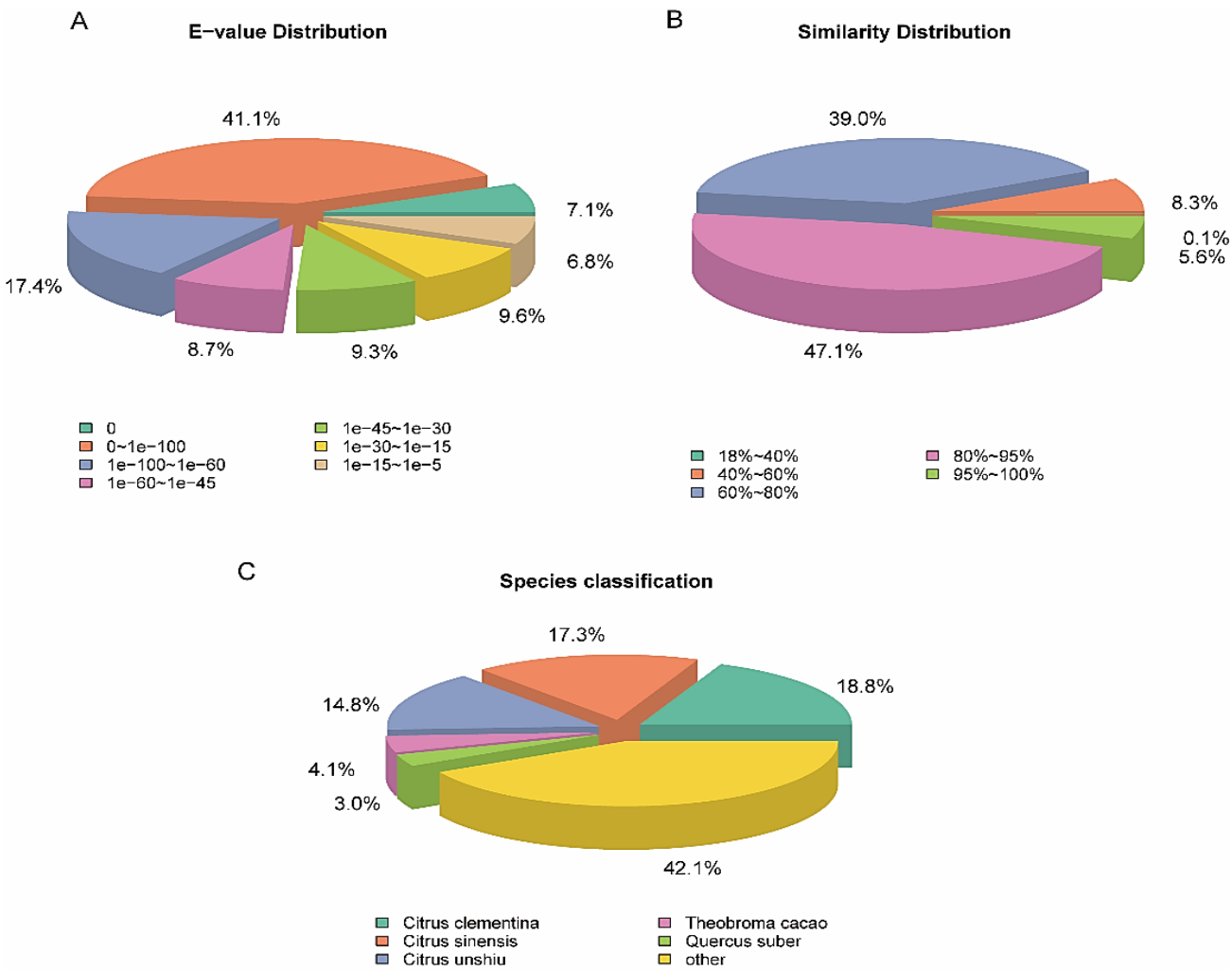
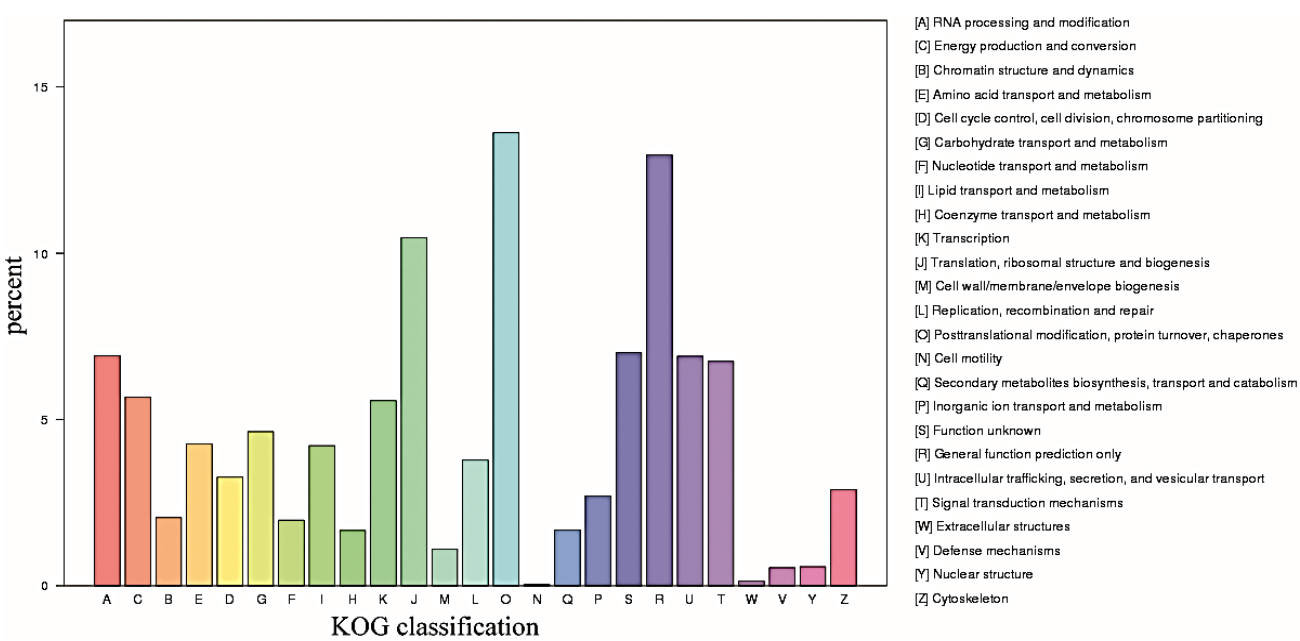


Fig. 2. Unigene annotations in the Nr database.
(A). The E-value distribution of unigene's BLAST x hits against the Nr database. (B). Similarity distribution of the BLAST x hits for unigene against the Nr database. (C). Species classification of the top BLAST x hits for unigene against the Nr database.



Functional Classification by KOG and GO
Fig. 3. Functional classification of Unigene in KOG databa

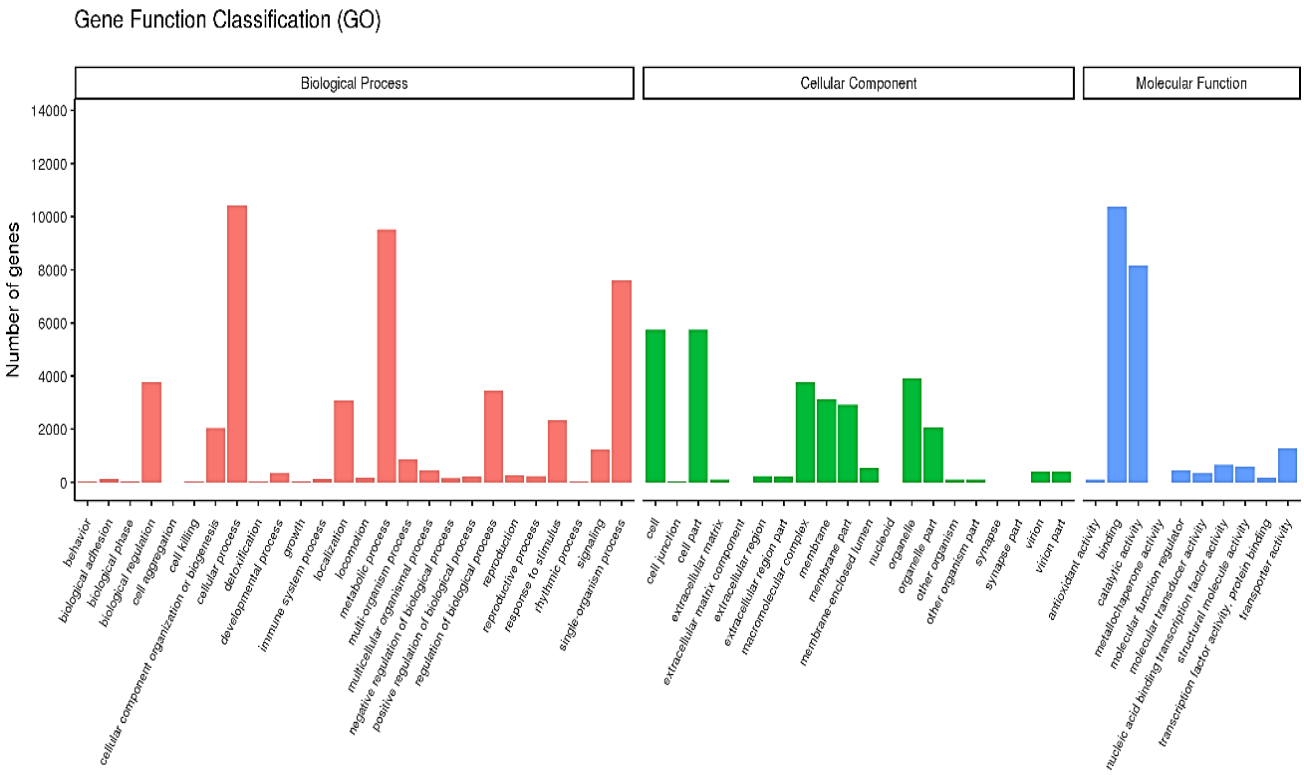


Fig. 4. Functional classification of Unigene in GO database.

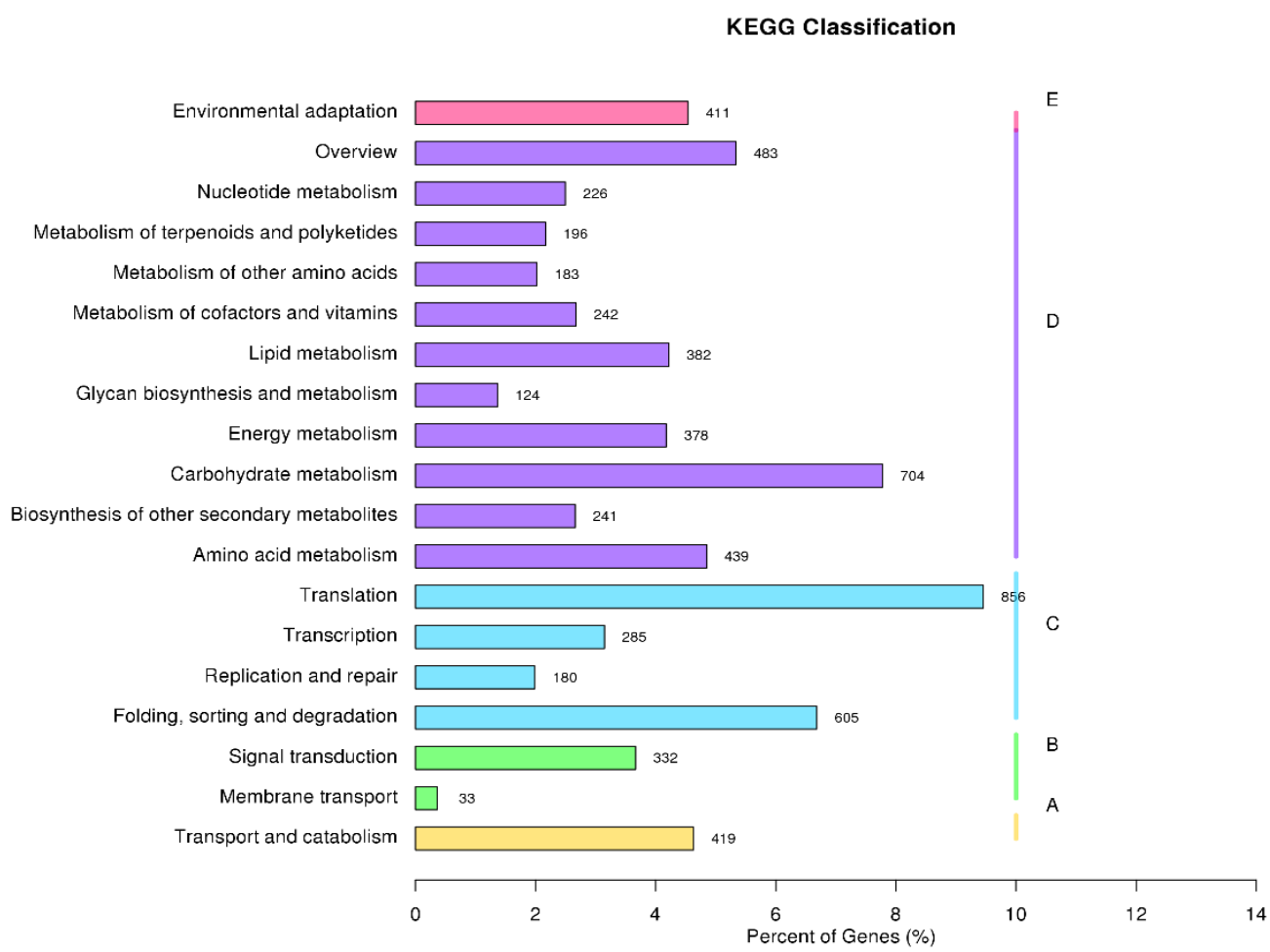


Fig. 5. Functional classification of Unigene in KEGG database.
A: Cellular Processes, B: Environmental Information Processing, C: Genetic Information Processing, D: Metabolism, E: Organismal Systems.

Table 3. The top 10 KEGG pathways for the assembled unigenes.

KEGG pathway	Pathway ID	Gene number	Function
Ribosome	ko03010	428	Genetic Information Processing
Plant-pathogen interaction	ko04626	358	Organismal Systems
Carbon metabolism	ko01200	277	Metabolism
Plant hormone signal transduction	ko04075	261	Environmental Information Processing
Biosynthesis of amino acids	ko01230	242	Metabolism
Protein processing in the endoplasmic reticulum	ko04141	213	Genetic Information Processing
Starch and sucrose metabolism	ko00500	195	Metabolism
Spliceosome	ko03040	194	Genetic Information Processing
Endocytosis	ko04144	191	Cellular Processes
RNA transport	ko03013	187	Genetic Information Processing

Table 4. Summary of SSR (simple sequence repeat).

Statistical item	Number	Percentage (%)
The total number of sequences examined	33365	
The total size of examined sequences (bp)	44965405	
Total number of identified SSRs	16103	
Number of sequences containing more than 1 SSR	3452	
Number of SSRs present in compound formation	1465	
Mono-nucleotide	7217	44.82
Di-nucleotide	4707	29.23
Tri-nucleotide	3715	23.07
Tetra-nucleotide	213	1.32
Penta-nucleotide	85	0.53
Hexa-nucleotide	166	1.03

Functional classification by KEGG: The KEGG pathway database is a knowledge base that enables for the systematic examination of gene activities in specific organisms. In the study, 6719 significant matching in the database were divided into 5 major groups, including 129 KEGG pathways. The largest category was Metabolism (3598;53.55%), followed by Genetic information processing (1926; 28.66%) (Fig. 5). It indicated that active metabolic processes were arising in the early growth of leaves and flowers. Metabolism include 11 sub-categories, and the largest sub-category was Carbohydrate metabolism (19.57%), followed by Overview (13.42%) and Amino acid metabolism (12.20%). Genetic information processing was divided into four categories, the largest category was Translation (44.44%), followed by Folding, classification, and degradation (31.41%) (Fig. 5). Among the 129 KEGG pathways, Ribosomes contained the greatest amount of unigenes (428) (Table 3). These data afford a solid foundation for the discovering the genes responsible for certain metabolic processes.

Frequency and distribution of SSRs in the unigenes: By searching the unigenes SSR loci of 33,365 (44,965,405 bp) -length *A. chinensis* Bunge, A total of 16,103 SSR loci were found. The SSR frequency was 48%, and the distribution density of one SSR per 2.79 kb. SSR repeat types were distributed from single nucleoside to hexanucleotide, of which the largest repeat type was single nucleotide, followed by dinucleotide and trinucleotide, and their numbers were 7217, 4707, and 3715, respectively, The sum of three accounting for 97.12% of the total SSR type (Table 4). Further analysis showed that the main type of single nucleotide was A/T, accounting for 99.78%. The most dinucleotide repeats were AG/CT (58.49%), followed by AT/AT (34.65%). There are 10 repeat types in trinucleotides, among which the number of AAG/CTT (29.37%) was the largest. The main tetranucleotides including ACAT/ATGT, AAAT/ATTT, AAAG/CTTT, AAAC/GTTT (Fig. 6A).

In addition, Among the single nucleotides, the most repeats were 9-12 (5038, 69.81%), followed by 13-16 (1411,19.55%). Dinucleotides were mainly 5-8 repeats (2884, 61.27%), followed by 9-12 repeats (1114, 23.67%). Trinucleotides were mainly composed of 5-8 repeat units (3529, 94.99%) (Table 5; Fig. 6B). The others (tetra-/ Penta-/ Hexa-) were also mainly 5-8 repeats. Overall, the main repeat units of the SSR motif (from single nucleoside to hexanucleotide) were 5-8 and 9-12 repeats, and the sum of them accounts for 81.93% of the total (Table 5). According to the location of the SSR locus in the gene, SSR can be divided into three categories, located in 3′UTR (untranslated region), 5′UTR, and CDS (Coding DNA sequence). 1290 SSRs are located in CDS, and the trinucleotides were 1088, accounting for 84.34%. 7138 SSRs were located in non-coding regions, of which 4203 are 5'UTR and 2935 are 3'UTR (Fig. 6C). The number of dinucleotides and trinucleotides in 5′UTR was significantly higher than that in 3'UTR.

Table 5. SSR motif distribution in the transcriptome of *A. chinensis* Bunge.

Repeat number	Motif type						Total	Ratio (%)
	Mono-	Di-	Tri-	Tetra-	Penta-	Hexa-		
5-8	-	2884	3529	205	84	148	6850	42.54
9-12	5038	1114	167	7	-	17	6343	39.39
13-16	1411	453	11	1	1	1	1878	11.66
17-20	497	183	6	-	-	-	686	4.26
21-24	165	47	2	-	-	-	214	1.33
25-28	59	13	-	-	-	-	72	0.45
29-32	18	6	-	-	-	-	24	0.15
>32	29	7	-	-	-	-	36	0.22
total	7217	4707	3715	213	85	166	16103	
Proportion%	44.82	29.23	23.07	1.32	0.53	1.03		100.00

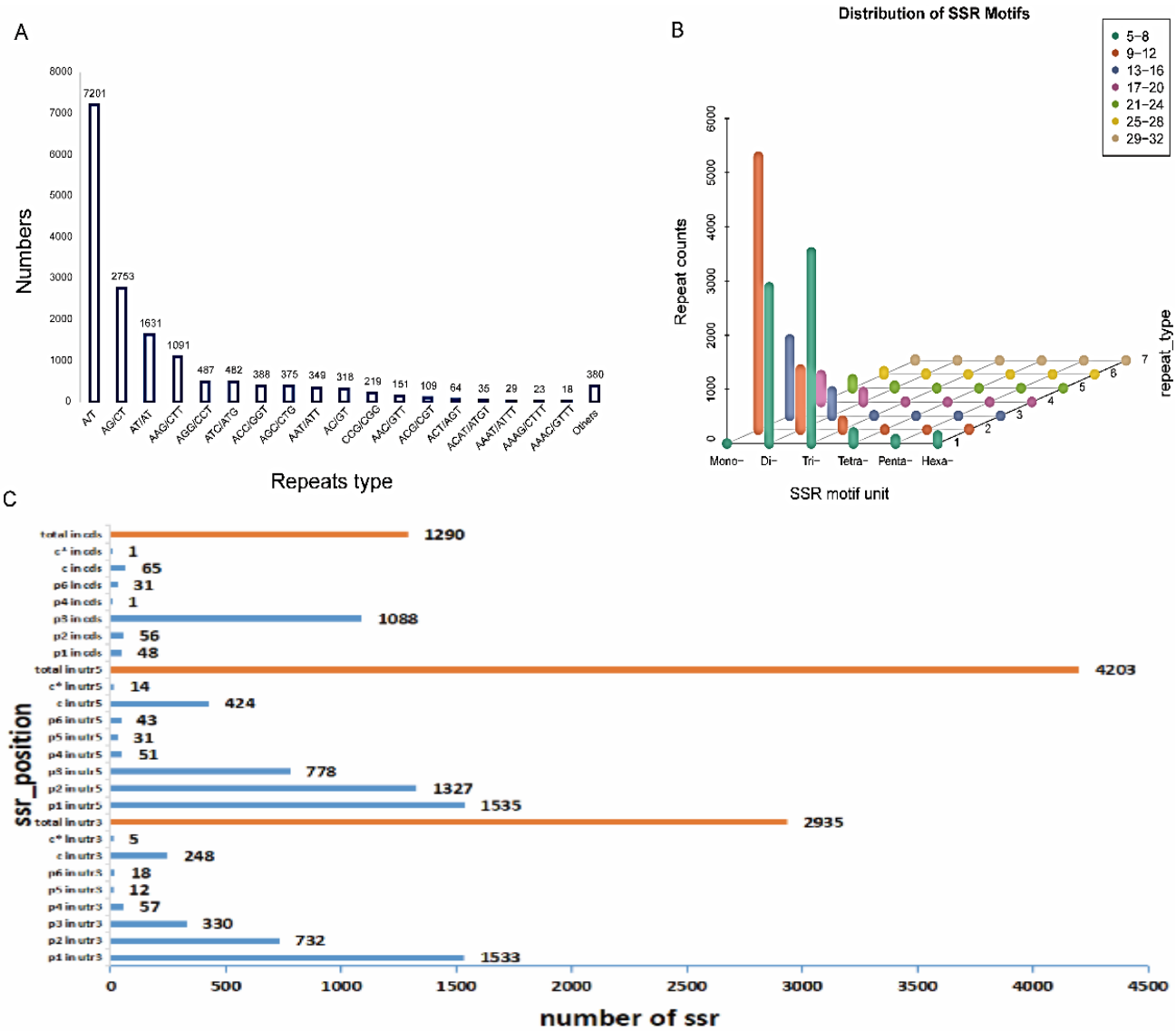


Fig. 6. (A): SSR repetition type distribution, (B): Distribution of SSR motif in the transcriptome, (C): Quantity distribution of SSR locations in *A. chinensis* Bunge. Note: P1-P6: Mono-, Di-, Tri-, Tetra-, Penta-, Hexa- nucleotide; c and c*: Complex and repetition type.

EST-SSR validation and fingerprint construction: According to the designed 5842 EST-SSR primers, 58 primer pairs were chosen at random and successfully utilized for validation via PCR amplification. 48 of the 58 primer pairs were highly successfully amplified by *Aesculus* genomic DNA, with an amplification efficiency rate of 82.7%. The remaining 10 primer pairs were unable to amplify PCR products at various annealing temperatures. Eight *Aesculus* samples from six species were used as PCR templates, 31 of the 48 primer pairs were polymorphic, 1 primer was recognized as monomorphic, and the polymorphism frequency was 64.6%.

The 31 polymorphic SSR primers were utilized to evaluate the eight *Aesculus* samples. Capillary electrophoresis was used to identify locus polymorphisms (Fig. 7). As presented in Table 6, 167 alleles were detected and the number of alleles per locus varied from 2 to 9, with an average of 5.387 alleles. The Number of effective alleles (*Ne*) varied from 1.438 to 6.095 with an average of 3.628, The Shannon's information index (*I*) varied from 0.483 to 1.960, with an average of 1.380. The Observed heterozygosity (*Ho*) varied from 0.000 to 0.875, with an average of 0.476. and Expected heterozygosity (*He*) varied from 0.305 to 0.844 with an average of 0.680. Except for P6 and P28 primers, the values of Expected heterozygosity (*He*) were all higher than 0.5 (Table 6), suggesting that genetic diversity was abundant in these EST-SSR markers.

According to the allelic variation of each locus, fingerprints of 8 *Aesculus* samples from six species were constructed (Table 7). Among the 31 pairs of polymorphic primers, 7 pairs of primers can distinguish 8 *Aesculus* samples completely alone. Then the genetic distance was estimated using the NTSYS software and the UPGMA dendrogram showed that the 8 *Aesculus* samples could be separated into 2 groups at the coefficient value of 0.54 (Fig. 8). Group I comprises two *A. chinensis* species and the other four species belong to group II. Among group II, The *A.*×*carnea* 'Neill Red' was closest to *A. hippocastanum* 'Baumannii', and the *A. pavia* L., was closest to *A. flava* Sol (Fig. 8).

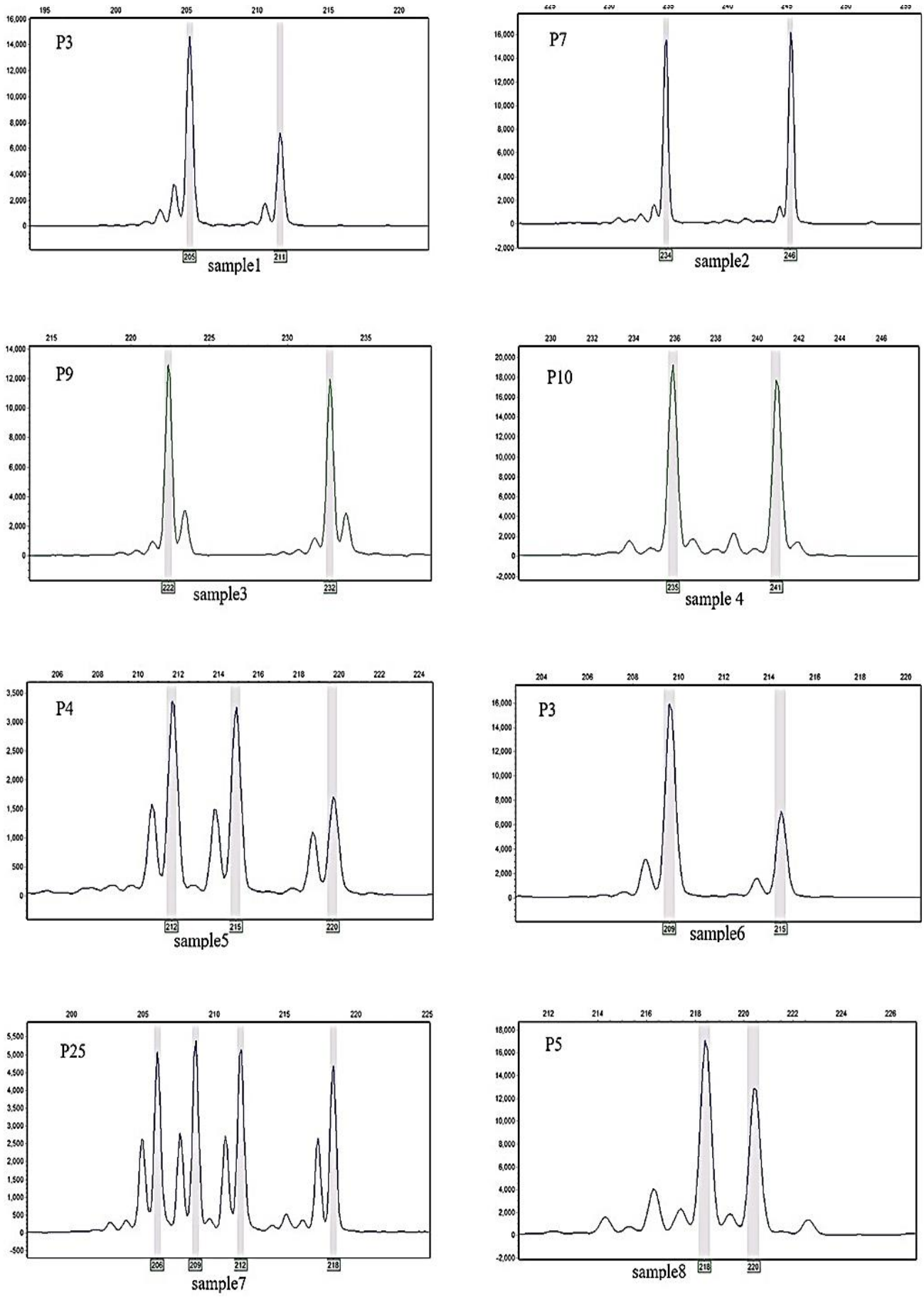


Fig. 7. Diagram of partial capillary electrophoresis of eight *Aesculus* samples. Sample1-8 represent clone S1, S123, S124, S50, S78, S8, S80, S98 respectively.

Table 6. The 31 EST-SSR markers in *A. chinensis* Bunge's characteristics.

primer	Forward primer (5'-3')	Reverse primer (5'-3')	Tm (°C)	SSRs	*Na*	*Ne*	*I*	*Ho*	*He*
P 1	AGGCTCTTTCTGAGGTTGGC	AAGAGCCTGTTCACCACTGG	59.9	(AC)6	5	3.200	1.332	0.625	0.688
P 2	AGAGTTGGGGAGGATGCTCT	ACGTATGCACCCGAGACTTC	59.9	(TA)8	5	2.783	1.230	0.625	0.641
P 3	CCAACTCGGCCATCTTGACT	GGAAGAAATGGCGCATGCTT	59.9	(TA)6	7	4.923	1.750	0.750	0.797
P 4	CACCACAAGCCTTCCATCCT	CCATCCATGTGCCTGACTGT	60.0	(CCT)5	6	3.765	1.511	0.875	0.734
P 5	TCTGCTTTGACCTCGCCATT	GCTGTCTCCTTTGACGGTGA	60.0	(TA)14	7	6.095	1.862	0.375	0.836
P 6	AAATGCTCACTCTCTCCGGC	AGACGGCACAGATGGTTTGT	60.0	(GC)6	2	1.438	0.483	0.375	0.305
P 7	CCTCCTCCTCCATCTGCTCT	GCACCCACCTGTCAACCTTA	60.0	(TC)7	5	2.000	1.037	0.250	0.500
P 8	GGCAAGAGGGTGCTTGAGAT	AAGGATGGCCACAGCTCTTC	60.0	(TG)6	5	3.556	1.423	0.375	0.719
P 9	TGGTTGCTCTTTCCAGCCAT	TTTGGCCCTTCATTGGAGCA	60.0	(GAAGT)5	6	3.765	1.511	0.625	0.734
P 10	ACCTCGTGGATTTGGCAGAG	CGCTGTCAACGACTCCATCT	60.1	(TC)11	6	3.556	1.488	0.500	0.719
P 11	ATTGGGCTCGGTCTTCTTGG	AGCCACCCGAGTCTATGACT	60.0	(AT)7	5	3.657	1.424	0.250	0.727
P 12	CCCTGGAGGCACAAACTGAT	AGTCCACATCTCACTGCTGC	60.0	(AG)14	8	4.923	1.841	0.750	0.797
P 13	CTTCCGCTTGCAGCATTTGT	CCGGGATTACATGGACGGTT	59.9	(TCT)6	3	2.133	0.900	0.000	0.531
P 14	TGGTGGCATGCCTTCTTCTT	AGTTGCTGCTGCTGTTGTTG	59.9	(CTT)6	3	2.462	0.974	0.250	0.594
P 15	CTTCCGATGTGCCTTCCAGT	ATTCATCGCTACCGCTACCG	60.0	(GGAGAG)5	4	2.844	1.163	0.125	0.648
P16	CCAGGCGCATAGTCATGACT	TGGCATGCACTTGACAGGAT	59.9	(AT)6	4	3.122	1.223	0.375	0.680
P 17	ATTCCGCACAACTCACCAGT	GCAGGAAACAGAGCACTTGC	60.0	(CT)9	8	6.095	1.927	0.625	0.836
P 18	TCTGTCTACCAAGCGGCAAG	ATCGTCCCCATCAGAGGTCA	60.0	(CT)8	6	4.000	1.548	0.625	0.750
P 19	CACCGCACCCACTTACTCTT	AGCAACTGCCCAAGATCCTC	60.0	(TTTA)5	5	3.368	1.369	0.375	0.703
P 20	TACTGGTCTGAGGCAGCTCT	ACCTTGCCCAGATTACCTGC	60.0	(TGC)5	2	2.000	0.693	0.000	0.500
P 21	ACATCCAGACAGAAGCCAGC	TGCTGCAGGAATAACAGGCA	60.0	(CT)7	6	4.129	1.581	0.500	0.758
P 22	GTGGCCCATACCAGTAGAGC	GCTTTTGGGTGCTGTTGAGG	59.9	(AG)13	8	6.400	1.960	0.625	0.844
P 23	GTCTGTCCGTCCGCGATTAT	CGCAAACTTGTCGACGGATC	59.9	(GA)8	7	5.333	1.787	0.625	0.813
P 24	CGGTCCCATTCAACTCACGA	TTTCACGATCTCCGGCGTAG	60.0	(TC)8	8	5.818	1.906	0.625	0.828
P 25	AGCTTGCTTTCTCGAGCCAT	ATGAGAGTGCGATCGAGCTG	60.0	(AGG)6	5	2.977	1.300	0.375	0.664
P 26	TGGTGGCCTTCAACTGGATC	CAATCCAATGCCGAGGGAGT	60.0	(GGC)5	4	2.246	1.041	0.625	0.555
P 27	GCAGCATCTCCAGCCTTAGT	AATTTGGTTGCAAGCGCCTT	59.8	(GA)13	9	5.333	1.923	0.875	0.813
P 28	TCCCACCCTCCTCTTCTCTG	AGAGCCACTGTACAACCTGC	60.0	(CT)6	3	1.684	0.736	0.375	0.406
P 29	AGGCAAGTACAGCAAGCCAT	TTTCTCCAGTCAGTCGTGGC	60.0	(TA)10	5	3.122	1.354	0.125	0.680
P 30	AGCCAGAGAGAGAGGAGGTG	CTCCCACACTCCTTGTTGCT	60.0	(TC)10	3	2.462	0.974	0.500	0.594
P 31	TACAACGCACCTACACCACC	ATGAGCAATTCCAGCAGCCT	60.0	(TC)13	7	3.282	1.527	0.750	0.695
mean					5.387	3.628	1.380	0.476	0.680

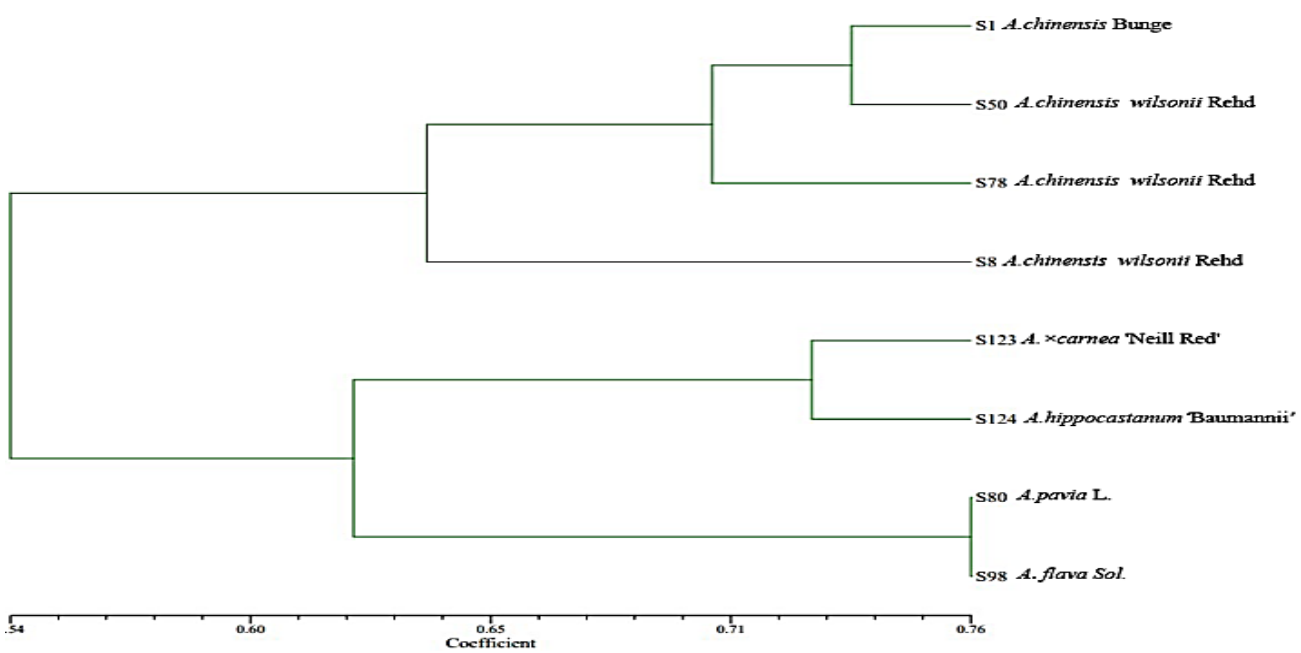


Fig. 8. The phylogenetic relationship of the selected *Aesculus* species constructed by 31EST-SSR markers. The NTSYS2.1 software under the UPGMA algorithm was employed to construct the phylogenetic trees.

Table 7. Fingerprints map of *Aesculus*.

Clones	Primers						
	P3	P5	P17	P23	P27	P28	P29
S1 (*A. chinensis* Bunge)	205/211	238/242	280/282	280/282	371/389	220	262
S123 (*Aesculus*×*carnea* 'Neill Red')	217/219	208	274/280	276/288	379/395	222	270
S124 (*A. hippocastanum* 'Baumannii')	201/215	218/220	274	288	379/393	246	270/272
S50 (*A. chinensis wilsonii* Rehd)	205	238	284/290	282	379/381	220/224	264
S78 (*A. chinensis wilsonii* Rehd)	205/209	220	278	282/284	383/391	220/228	268/272
S8 (*A. chinensis wilsonii* Rehd)	209	226	288	286/300	387	224/228	274/284
S80 (*A. pavia* L.)	209/217	208/216	278/292	276/280	351	226	260
S98 (*A. flava* Sol.)	209/215	216	278/280	276	381	220/222	260/274

Discussion

Transcriptome characteristics of *A. chinensis* Bunge: Transcriptome assembly quality was assessed by several criteria, including N50, average length, and examination of the RNA-Seq (Ribonucleic acid-sequencing) read (Al-Qurainy *et al*., 2019). The N50 was used to assess transcriptome assemblies, in which a high value corresponds to a high quality. In this research, a total of 33,365 unigenes with N50 (1959 bp) and a mean length of 1348 bp were obtained, which was longer than previously reported such as *A. chinensis* (N50=1807 bp, mean length = 999.86 bp) (Wei *et al*., 2019). *Platanus acerifolia* (N50 = 1579 bp, mean length= 1015.15 bp) (Li *et al*., 2019). *Pigeonpea* (N50=1393 bp, mean length = 396 bp) (Nigam *et al*., 2017). *Sodom apple* (N50 = 1733 bp, mean length =858.83 bp) (Muriira *et al*., 2015). The Q20 of 98.05% with GC of 45.41% reflects a high-quality sequencing run too. This indicated sequencing data and transcriptome assembly for *A. chinensis* Bunge had high quality in the present study. By comparing the homology with the Nr protein database, *A. chinensis* Bunge has higher homology with *Citrus clementina* and *Citrus sinensis*, which was consistent with Wei's conclusion (Wei *et al*., 2019). Maybe there are fewer protein sequences of *Aesculus* in the public database while there are more protein sequences related to *Citrus*. The transcriptome data will offer a solid basis for further gene mining and breeding using marker-assisted selection for *A. chinensis* Bunge.

Transcriptome annotation of *A. chinensis* Bunge: Seven public databases were successfully annotated with the assembled unigenes, at least one database was annotated to 25,450 unigenes (76.27%). However, 23.73% of unigenes were unannotated in any database, it may be that these genes of the genus *Aesculus* have only short non-coding regions or fewer gene annotations in the public database. By comparing with KOG database, 6430 unigenes were divided into 25 categories, of which most categories were connected to Posttranslational modification, protein conversion, and molecular chaperone (12.22%), followed by General functional prediction only (11.62%) and Translation, ribosomal structure and biosynthesis (9.39%), which were greatly similar to *Apocynum venetum* L (Yuan *et al*., 2020). A total of 97,978 annotations were obtained from the GO database, which was assigned to 3 categories (biological process, cellular components, and molecular function). In the biological process, the largest group was the cellular process (22.39%), followed by the metabolic process (20.47%) and the single organism process (16.34%). Coincidentally, the main group was cellular process (22.04%), metabolic process (20.99%), and single organism process of the biological process in *Apocynum venetum L* (Yuan *et al*., 2020) too. The most abundant categories were binding and catalytic activity in molecular function, which was similar to *Apocynum venetum* L., and *Platanus acerifolia* (Yuan *et al*., 2020; Li *et al*., 2019). The metabolic pathway was the largest category in the KEGG database, among which the largest three groups were carbohydrate metabolism, overview, and amino acid metabolism with a proportion of 19.57%, 13.42%, and 12.20%, respectively, which were similar to *Apocynum venetum* L., (Yuan *et al*., 2020), *Platanus acerifolia* (Li *et al*., 2019) and *guar* (Tanwar *et al*., 2017). These transcriptome descriptions will be useful for studying specific physiological processes, gene structure and function, and metabolic pathways in *A. chinensis* Bunge.

SSR characterization of *A. chinensis* Bunge: SSR molecular markers are codominant, specific, multi-allelic, highly polymorphic, and can be transferred between species within genera (Varshney *et al*., 2005). To date, there has been no reported development of EST-SSR in *A. chinensis* Bunge. In the research, 16,103 SSR loci were found and the largest number of SSR repeat types was single nucleotides, followed by dinucleotides and trinucleotides. The density of SSR distribution in *A. chinensis* Bunge unigenes was 1/2.79kb. which was much higher than those obtained in *R. rex* (1/5.65kb) (Zhang *et al*., 2017) and *tree peony* (1/9.24kb) (Wu *et al*., 2014), but lower than the frequency of *Stephanandra incisa* (1/1.60kb) (Zhang *et al*., 2021). Among the SSR repeat types, The main dinucleotide types were AG/CT (58.49%), AT/AT (34.65%), and trinucleotides were mainly AAG/CTT, consistent with *Apocynum venetum* L. (Yuan *et al*., 2020). Meanwhile, the most types of SSR are AG/CT (15.4%), AAG/CTT (9.8%), and AT/AT (6.5%) in *Pistacia vera* L. (Karcı *et al*., 2020), and AG/CT is also dominant in *Jerusalem artichoke* (32.9%) (Yang *et al*., 2019), AAG/CTT was also the most (12.89%) of trinucleotide in *Chelidonium majus* L. (Pourmazaheri *et al*., 2019). The above results indicate

that dinucleotide (AG/CT) is the most common and trinucleotide (AAG/CTT) is the most common in plants (Wei *et al*., 2011; Liu *et al*., 2015). Further analysis, 1290 SSRs (trinucleotide was 1088, 84.34%) were located in CDS. Similarly, trinucleotide (444, 68.7%) and dinucleotide (242, 49.4%) repeats were preferentially present in the CDS of *Blackgram* (Raizada & Souframanien, 2019). 7138 SSRs were identified in the non-coding region, and the number of di- and tri-nucleotides in 5'UTR is significantly larger than that in 3'UTR. Interestingly, in *Blackgram*, the 5'UTRs contained more trinucleotides than 3'UTRs too (Raizada & Souframanien, 2019; Souframanien & Reddy, 2015). The enormous number of SSRs can substantially benefit large-scale genotyping investigations such as genetic variety assessment and association mapping for critical characteristics in *A. chinensis* Bunge.

Validation of EST-SSR Makers and molecular identification: 58 pairs of primers were selected for PCR validation, and 48 (82.7%) primers resulted in clear bands. The PCR amplification efficiency was greater than *Boea clarkeana* (81.08%) (Wang *et al*., 2017), *Hevea brasiliensis* (75%) (Triwitayakorn *et al*., 2011), and *R.rex* (50%) (Zhang *et al*., 2017). But lower than the rate reported for *Haloxylon ammodendron* (85%) (Long *et al*., 2014), *Caragana korshinskii* Kom (93.7%) (Long *et al*., 2015), and *Lycium barbarum* (88%) (Chen *et al*., 2017). which imply the SSR molecules markers have moderate cross-transferability in *Aesculus*. The polymorphism ratio among eight *Aesculus* samples from six species was 64.6%, and this was much higher than that in *R. rex* (36%) (Zhang *et al*., 2017) and *Stephanandra incisa* (29%) (Zhang *et al*., 2021). Suggesting these primers can be used in molecular-assisted breeding.

Then 167 alleles were identified among the eight *Aesculus* species using the 31 EST-SSR polymorphic makers. The alleles varied from 2 to 9 per locus with a mean of 5.387 alleles. The Observed heterozygosity (*Ho*) averaged 0.476, Expected heterozygosity (*He*) averaged 0.680, which were higher than that of EST-SSRs in *Apocynum venetum* L. (*Ho* = 0.336, *He* = 0.342) (Yuan *et al*., 2020). *Stephanandra incisa* (*He* = 0.434) (Zhang *et al*., 2021), *Pistachiovera* L. (*Ho* = 0.39, *He* = 0.65) (Karcı *et al*., 2020). *Torreya grandis* (*He*=0.432) (Zeng *et al*., 2018). The Shannon's information index (*I*) can also reflect molecules marker diversity, ranging from 0.483 to 1.960, with an average of 1.380. This finding indicated a high level of informativeness within these EST-SSR markers and may be used in a variety of genetic analyses among species and evolutionary adaptation at the species level.

It showed that 7 pairs of primers had a 100% transferability rate and were capable of discriminating the 8 samples of *Aesculus* plants alone, the molecular identification rate was 100%. This indicates that the 7 pairs of primers can be widely used in other *Aesculus* species. Similarly, primers mined for *Black Locust* (Guo *et al*., 2017) and *Gleditsia sinensis* (Lin *et al*., 2017) can be successfully transferred to other *Robinia* and *Gleditsia* genus species, respectively. All of the above studies illustrated the high transferability of EST-SSR primers in sibling species.

Conclusion

To the best of our knowledge, this is the first research in which *de novo* transcriptomes sequencing of *A. chinensis* Bunge was used to develop SSR markers. 7 of 31 polymorphic SSR markers could completely distinguish the 8 species of *Aesculus* plants alone, and the molecular identification rate was 100%. The findings not only to give a valuable resource for genetic improvement in *A. chinensis* Bunge, as well as potent molecular tool for genetic linkage study in other *Aesculus* plants.

Acknowledgment

We are really thankful to Professor Fang-fang Ma of the Forestry College of Shandong Agricultural University for her Revision of the language in the paper. The Agricultural Improved Seed Project of Shandong Province, China, provided financial assistance for this work through grant 2020LZGC011-3.

References

Al-Qurainy, F., A. Alshameri, A.R. Gaafar, S. Khan, M. Nadeem and A.A. Alameri. 2019. Comprehensive stress-based de novo transcriptome assembly and annotation of Guar (*Cyamopsis tetragonoloba* (L.) Taub.): an important industrial and forage crop. *Int. J. Genomics*, 2019: 7295859.

Chen, C.J., S.G. Wang, J.Y. Chen, X.L. Liu, M.C. Zhang, X. Wang and W.H. Xu. 2019. Escin suppresses HMGB1-induced overexpression of aquaporin-1and increased permeability in endothelial cells. *FEBS Open Biol.*, 9: 891-900.

Chen, C.L., M.L. Xu, C.P. Wang, G.X. Qiao, W.W. Wang, Z.Y. Tan, T.T. Wu and Z.S. Zhang. 2017. Characterization of the *Lycium barbarum* fruit transcriptome and development of EST-SSR markers. *PLoS One*, 12: e0187738.

Cheng, Y.Q., J.F. Liu, H.D. Zhang, J. Wang, Y.X. Zhao and W.T. Geng. 2015. Transcriptome analysis and gene expression profiling of abortive and developing ovules during fruit development in hazelnut. *PLoS One*, 10: e0122072.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury and Q. Zeng. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, 29: 644-652.

Guo, Q., J.X. Wang, L.Z. Su, W. Lv, Y.H. Sun and Y. Li. 2017. Development and evaluation of a novel set of EST-SSR markers based on transcriptome sequences of Black Locust (*Robinia pseudoacacia* L.). *Genes*, 8: 177.

Guo, Y.Q., X.J. Chang, C. Zhu, S.T. Zhang, X.Z. Li and H.F. Fu. 2018. De novo transcriptome combined with spectrophotometry and gas chromatography-mass spectrometer (GC-MS) reveals differentially expressed genes during accumulation of secondary metabolites in purple-leaf tea (*Camellia sinensis* cv Hongyafoshou). *J. Hort. Sci. Biotechnol.*, 94: 349-367.

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

He, D., J. R. Zhang, X.F. Zhang, S.L. He, D.B. Xie, Y. Liu, C.M. Li, Z. Wang and Y.P. Liu. 2020. Development of SSR markers in *Paeonia* based on de novo transcriptomic assemblies. *PLoS One*, 15: e0227794.

Karcı, H., A. Paizila, H. Topçu, E. Ilikçioğlu and S. Kafkas. 2020. Transcriptome sequencing and development of novel genic SSR markers from *Pistacia vera* L. *Front Genet.*, 11: 1021.

Kim, S.E., T.H. Kim, S.A. Park, W.T. Kim, Y.W. Park and J.S. Ahn. 2017. Efficacy of horse chestnut leaf extract ALH-L1005 as a matrix metalloproteinase inhibitor in ligature-induced periodontitis in canine model. *J. Vet Sci.*, 18: 245-251.

Li, F.Q., C.Y. Wu, M.Z. Gao, M.M. Jiao, C. Qu, A. Gonzalez-Uriarte and C. Luo. 2019. Transcriptome sequencing, molecular markers, and transcription factor discovery of *Platanus acerifolia* in the presence of Corythucha ciliata. *Sci. Data*, 6: 128.

Lin, F.R., J.L. Xing, Y.Q. Meng, P. Huang and Y.Q. Zheng. 2017. Development and assessment of EST-SSR for *Gleditsia sinensis* Lam. based on transcriptome sequences. *J. Plant Genet. Resour.*, 18: 148-154.

Liu, L.J. and H.H. Zhou. 2010. Pharmacological effects of aescin and its clinical application. *Prog. Mod. Biomed*., 5: 957-960.

Liu, Y.L., P.F. Zhang, M.L. Song, J.L. Hou, M. Qing, W.Q. Wang and C.S. Liu. 2015. Transcriptome analysis and development of SSR molecular markers in *Glycyrrhiza uralensis Fisch*. *PLoS One*, 10: e0143017.

Liu, Z.P., T.L. Chen, L.C. Ma, Z.G. Zhao, P.X. Zhao, Z.B. Nan and Y.R. Wang. 2013. Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. *PLoS One*, 8: e83549.

Liu, Z., J.J. Zhang, Y.X. Zhou, Y.F. Liu, Z.G. Hu, G.H. Zheng and Z.H. Shi. 2020. The complete chloroplast genome of *Aesculus chinensis* var. wilsonii. *Mitochondrial DNA B Resour.*, 5: 2547-2549.

Long, Y., Y.Y. Wang, S.S. Wu, J. Wang, X.J. Tian and X.W. Pei. 2015. De novo assembly of transcriptome sequencing in *Caragana korshinskii* Kom. and characterization of EST-SSR markers. *PLoS One*, 10: e0115805.

Long, Y., J.W. Zhang, X.J. Tian, S.S. Wu, Q. Zhang, J.P. Zhang, Z.H. Dang and X.W. Pei. 2014. De novo assembly of the desert tree *Haloxylon ammodendron* (C. A. Mey.) based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. *BMC Genom.*, 15: 1111.

Muriira, N.G., W. Xu, A. Muchugi, J.C. Xu and A.Z. Liu. 2015. De novo sequencing and assembly analysis of transcriptome in the Sodom apple (*Calotropis gigantea*). *BMC Genom.*, 16: 723.

Naganeeswaran, S., T.P. Fayas and M.K. Rajesh. 2020. Dataset of transcriptome assembly of date palm embryogenic calli and functional annotation. *Data Brief*, 31: 105760.

Nigam, D., S. Saxena, G. Ramakrishna, A. Singh, N.K. Singh and K. Gaikwad. 2017. De novo assembly and characterization of *Cajanus scarabaeoides* (L.) thouars transcriptome by paired-end sequencing. *Front. Mol. Biosci.*, 4, 48: 1-4.

Peakall, R. and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, 28: 2537-2539.

Pourmazaheri, H., A. Soorni, B.B. Kohnerouz, N.K. Dehaghi, E. Kalantar, M. Omidi and M.R. Naghavi. 2019. Comparative analysis of the root and leaf transcriptomes in *Chelidonium majus* L. *PLoS One*, 14: e0215165.

Qi, X.P., E.L. Ogden, M.K. Ehlenfeldt and L.J. Rowland, 2019. Dataset of de novo assembly and functional annotation of the transcriptome of blueberry (*Vaccinium* spp.). *Data Brief*, 25: 104390.

Raizada, A. and J. Souframanien. 2019. Transcriptome sequencing, de novo assembly, characterisation of wild accession of blackgram (*Vigna mungo* var. silvestris) as a rich resource for development of molecular markers and validation of SNPs by high resolution melting (HRM) analysis. *BMC Plant Biol.*, 19: 358.

Reddy, B., A.K. Patel, K.M. Singh, D.B. Patil, P.V. Parikh and D.N. Kelawala. 2015. De novo assembly and transcriptome characterization of canine retina using high-throughput sequencing. *Genet. Res. Int*., 2015: 638679.

Reuter, J.A., D.V. Spacek and M.P. Snyder. 2015. High-throughput sequencing technologies. *Mol. Cell.*, 58: 586-597.

Severin, A.J., J.L. Woody, Y.T. Bolon, B. Joseph, B.W. Diers and A.D. Farmer. 2010. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.*, 10: 160.

Simon, S.A., J.X. Zhai, R.S. Nandety, K.P. McCormick, J. Zeng and D. Mejia. 2009. Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.*, 60: 305-333.

Souframanien, J. and K.S. Reddy. 2015. De novo assembly, characterization of immature seed transcriptome and development of genic-SSR markers in blackgram [*Vigna mungo* (L.) Hepper]. *PLoS One*, 10: e0128748.

Taheri, S., T.L. Abdullah, Z. Ahmad and N.A.P. Abdullah. 2014. Effect of acute gamma irradiation on *Curcuma alismatifolia* varieties and detection of DNA polymorphism through SSR marker. *BioMed. Res. Int.*, 2014: 631813.

Taheri, S., T.L. Abdullah, M.R. Yusop, M.M. Hanafi, M. Sahebi and P. Azizi. 2018. Mining and development of novel SSR markers using next generation sequencing (NGS) data in plants. *Molecules*, 23: 399.

Tanwar, U.K., V. Pruthi and G.S. Randhawa. 2017. RNA-Seq of Guar (*Cyamopsis tetragonoloba*, L. Taub.) leaves: de novo transcriptome assembly, functional annotation and development of genomic resources. *Front Plant Sci.*, 8: 91.

Triwitayakorn, K., P. Chatkulkawin, S. Kanjanawattanawong, S. Sraphet, T. Yoocha and D. Sangsrakru. 2011. Transcriptome sequencing of *Hevea brasiliensis* for development of microsatellite markers and construction of a genetic linkage map. *DNA Res.*, 18: 471-482.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm and S.G. Rozen. 2012. Primer3-new capabilities and interfaces. *Nucl. Acids Res.*, 40: e115.

Varshney, R.K., R. Sigmund, A. Börner, V. Korzun, N. Stein and M.E. Sorrells. 2005. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci.*, 168: 195-202.

Wang, Y., K. Liu, D. Bi, S.B. Zhou and J.W. Shao. 2017. Characterization of the transcriptome and EST-SSR development in *Boea clarkeana*, a desiccation-tolerant plant endemic to China. *Peer J*, 5: e3422.

Wang, Z., M. Gerstein and M. Snyder. 2009. RNA-SEQ: A revolutionary tool for transcriptomics. *Nat. Rev. Gene*, 10: 57-63.

Wei, W.L., X.Q. Qi, L.H. Wang, Y.X. Zhang, W. Hua, D.H. Li, H.X. Lv and X.R. Zhang. 2011. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genom.*, 12: 451.

Wei, Y.D., C. Xiong, T.Y. Zhang, H. Gao, Q.G. Yin and H. Yao. 2019. Synthesis of triterpenoid saponins from *Aesculus chinensis* based on transcriptome data. *China J. Chin. Mater. Med.*, 44: 1135-1144.

Wu, J., C.F. Cai, F.Y. Cheng, H.L. Cui and H. Zhou. 2014. Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. *Mol. Breed.*, 34: 1853-1866.

Yang, S., X. Sun, X. Jiang, L. Wang, J. Tian and L. Li. 2019. Characterization of the Tibet plateau Jerusalem artichoke (*Helianthus tuberosus* L.) transcriptome by de novo assembly to discover genes associated with fructan synthesis and SSR analysis. *Hereditas*, 156, 9: 1-13.

Yang, Y., M. Xu, Q.F. Luo, J. Wang and H.G. Li. 2014. De novo transcriptome analysis of *Liriodendron chinense* petals and leaves by Illumina sequencing. *Gene*, 534: 155-162.

Yu, D.J. 1981. *Flora Reipublicae Popularis Sinicae*. Vol: 46. Science Publishing House. P.R. China

Yuan, N., M.M. Li and C.L. Jia. 2020. De novo transcriptome assembly and population genetic analyses of an important coastal shrub, *Apocynum venetum* L. *BMC Plant Biol.*, 20(1): 408: 1-15.

Zeng, J., J. Chen, Y.X. Kou and Y.J. Wang. 2018. Application of EST-SSR markers developed from the transcriptome of *Torreya grandis* (Taxaceae), a threatened nut-yielding conifer tree. *Peer J*, 6: e5606.

Zhang, C.P., Z.L. Wu, X.Q. Jiang, W. Li, Y.Z. Lu and K.L. Wang. 2021. De novo transcriptomic analysis and identification of EST-SSR markers in *Stephanandra incisa*. *Sci. Rep.*, 11(1): 1059: 1-10.

Zhang, Y., X. Zhang, Y.H. Wang and S.K. Shen. 2017. De novo assembly of transcriptome and development of novel EST-SSR markers in *Rhododendron rex* Lévl. through Illumina sequencing. *Front. Plant Sci.*, 8: 1664.

Zhang, Z.Y., Y. Chen, X.B. Jiang, P. Zhu, L. Li, Y.L. Zeng and T.M. Tang. 2019. The complete chloroplast genome of *Aesculus chinensis*. *Mitochondrial DNA Part B*, 4: 1955-1956.

Zhou, Q., D. Luo, L. Ma, W. Xie, Y. Wang, Y.R. Wang and Z.P. Liu. 2016. Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus* L.) using Illumina sequencing. *Sci. Rep.*, 6: 20549.

(Received for publication 4 August 2023)