Basic Characteristics of Flower Transcriptome Data and Derived Novel EST-SSR Markers of *Luculia yunnanensis*, an Endangered Species Endemic to Yunnan, Southwestern China

Yao Zhang [1,2], Xi Liu [3], Yuying Li [4], Xiongfang Liu [1,2], Hong Ma [1,5,*], Suping Qu [6,*] and Zhenghong Li [1,*]

1   Institute of Highland Forest Science, Chinese Academy of Forestry, Kunming 650224, China;
    zhangyao949@163.com (Y.Z.); liuxiongfang@caf.ac.cn (X.L.)
2   College of Forestry, Nanjing Forestry University, Nanjing 210037, China
3   School of Geography and Ecotourism, Southwest Forestry University, Kunming 650224, China;
    lx13320441881@163.com
4   College of Grassland Science, Shanxi Agricultural University, Jinzhong 030801, China; liyuy_ing@163.com
5   Key Laboratory of Breeding and Utilization of Resource Insects, National Forestry and Grassland
    Administration, Kunming 650224, China
6   Flower Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China
*   Correspondence: hortscience@163.com (H.M.); qsp@yaas.org.cn (S.Q.); lzh4949@163.com (Z.L.)

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: *Luculia yunnanensis* (Rubiaceae), an evergreen shrub or small tree, is endemic to China and confined to Nujiang Prefecture, Yunnan Province. This plant is of high ornamental value owing to its attractive pink flowers, sweet fragrance, and long flowering period. Due to the influence of climate change and human factors, the distribution range of *L. yunnanensis* has exhibited a significant shrinking trend, and it has become a vulnerable species that is in urgent need of conservation and rational utilization research. In this study, the flower transcriptome sequencing of *L. yunnanensis* was conducted using an Illumina HiSeq platform. We designed and developed a series of EST-SSR primers based on the flower transcriptome data of *L. yunnanensis*. The results showed that 98,389 unigenes were obtained from the *L. yunnanensis* flower transcriptome, all of which were aligned with sequences in public databases. Nr, Nt, Pfam, KOG/COG, Swiss-Prot, KEGG, and GO annotated 31,859, 13,853, 22,684, 10,947, 21,416, 9722, and 23,390 unigenes, respectively. The MISA (Microsatellite) tool was used to identify SSR loci from all unigenes, and a total of 15,384 SSRs were identified. Repeat motifs were given priority with mononucleotides, dinucleotides, and trinucleotides. The 81 primer pairs were synthesized randomly, of which 44 pairs showed effective amplification. A total of 17 primers showed stable amplification, and rich polymorphism was observed in 6 populations. We concluded via genetic diversity analysis that the average effective number of alleles (Ne), Shannon's information index (I), and polymorphism information content (PIC) were 1.925, 0.837, and 0.403, respectively. In conclusion, 17 EST-SSR primers can be used for subsequent population genetic diversity analysis and molecular-marker-assisted breeding, which is of great significance for formulating resource conservation and utilization strategies for *L. yunnanensis*.

Keywords: *Luculia yunnanensis*; transcriptome; EST-SSR markers; polymorphism; conservation genetics

1. Introduction

As an endangered species endemic to Yunnan Province, Southwestern China, *Luculia yunnanensis* is an evergreen shrub or small tree that belongs to the genus *Luculia*, of the family Rubiaceae. This species mainly grows on limestone mountains, secondary shrubby woodland, and open slopes at altitudes of between 1200 and 3200 m. It possesses high ornamental value due to its attractive plant shape, long flowering period, striking pink flowers, sweet fragrance, and the fact that it is evergreen. There is thus great value in its development and utilization.

Our previous field investigations of *L. yunnanensis* and its recorded distribution showed an obvious shrinking trend in its distribution range, which currently only comprises Lushui, Fugong, and Gongshan counties in Nujiang Prefecture, Yunnan Province. The distribution range has been reduced from seven counties in five prefectures or cities to three counties in one prefecture, and the administrative area of the distribution area is 10,465.48 km^2 (only 20% of its historical range). According to the IUCN threat level criteria, *L. yunnanensis* belongs to the vulnerable species category [1]; thus, there is an urgent need for relevant biological conservation research.

The assessment of genetic diversity is one of the most significant parts of the research on the conservation of rare and endangered plants [2]. Considered to be an important component that could be seen as a determination of the viability of a species, it may affect species' adaptive capacity and evolutionary potential, and is often regarded as an indirect warning of the potential extinction of endangered species [3,4]. A decrease in genetic diversity in a species may lead to the degradation of its adaptability to environmental changes, leading to a decline in its evolutionary ability [5,6]. Thus, only by clarifying the genetic diversity of threatened species and taking various corresponding factors into consideration can we propose feasible strategies regarding their protection and utilization.

DNA molecular markers are usually used to estimate the genetic diversity of plants, and a sufficient number of polymorphic primers is the prerequisite for molecular genetic diversity analysis. As one of the most commonly used molecular marker techniques, EST-SSR has the advantages of simple operation, good stability, high accuracy, rich variability, good repeatability, and interspecific transmissibility [7–10]. Compared with genomic SSR markers, EST-SSR markers have low development cost and can be applied to species without reference genomes, and the application of EST-SSR markers has greatly promoted the use of molecular marker techniques in plant research [10–12], as well as being widely used in fields such as plant genetics and breeding, conservation, and the development of germplasm [13–16]. To date, the development of EST-SSR primers for *L. yunnanensis* has not been reported. Zhou [17] made use of a modified biotin–streptavidin capture method to develop 13 pairs of SSR primers suitable for analyzing the genetic diversity of *Luculia*. These 13 pairs of primers can effectively distinguish between *Luculia pinceana* and *L. yunnanensis*. With the same method, Ma [18] developed 24 pairs of SSR primers, and 11 pairs of them were polymorphic in two populations of *L. yunnanensis*. However, this number of markers is far from meeting the needs of follow-up studies on *L. yunnanensis*. Therefore, in this study we designed a series of EST-SSR primers based on the flower transcriptome data of *L. yunnanensis*, and developed a set of polymorphic EST-SSR primers that can be used for subsequent population genetic diversity analysis and molecular-marker-assisted breeding, which are of great significance for formulating resource conservation and utilization strategies for *L. yunnanensis*.

2. Results and Discussion

2.1. Functional Annotation and Classification

The transcriptome sequencing was conducted using an Illumina HiSeq platform (Beijing Novogene Biological Information Technology Co., Ltd., Beijing, China); clean reads were used for de novo assembly using Trinity with the default parameters to obtain 140,042 transcripts, and TGICL version 2.1 software was used to cluster the transcripts into 98,389 unigenes. There were 18,521 unigenes (18.82% of the total unigenes) with a length of more than 1 Kb and high assembly integrity (Table 1). To learn more about the features and functions of these unigenes, we aligned them with the sequences of public databases, including Nr, Nt, Pfam, KOG/COG, Swiss-Prot, KEGG, and GO (see Table 2). It can be seen from the results that at least some of the unigenes were not annotated in all seven databases, which demonstrates that there are unknown unigene sequences. This is consistent with the fact that the annotation of the unigenes of *Glycyrrhiza uralensis* [19], *Luculia gratissima* [20], and *Prunella vulgaris* [21] did not reach 100%. This result may be caused by the length of some unigenes being so short that the annotation information is

incomplete, or by the specific unigenes of this species not being fully recognized and the relevant information not being included in the database. The specific reasons for this need to be further studied via sequence analysis and gene expression verification.

Table 1. Length distribution after assembly in the flower transcriptome of *L. yunnanensis*.

Transcript Length Interval	200–500 bp	500–1 kbp	1 k–2 kbp	>2 kbp	Total
Number of transcripts	71,181 (50.83%)	25,812 (18.43%)	23,641 (16.88%)	19,408 (13.86%)	140,042
Number of unigenes	62,121 (63.14%)	17,747 (18.04%)	10,870 (11.05%)	7651 (7.78%)	98,389

Table 2. The statistics of the unigenes' annotation in the flower transcriptome of *L. yunnanensis*.

Annotation	Number of Annotated Unigenes	Percentage of Annotated Unigenes (%)	Percentage of Total Unigenes (%)
Annotated in Nr	31,859	32.38	22.75
Annotated in Nt	13,853	14.07	9.89
Annotated in KO	9722	9.88	6.94
Annotated in Swiss-Prot	21,416	21.76	15.29
Annotated in Pfam	22,684	23.05	16.20
Annotated in GO	23,390	23.77	16.70
Annotated in KOG	10,947	11.12	7.82
Annotated in all databases	4273	4.34	3.05
Annotated in at least one database	36,497	37.09	26.06
Total annotated unigenes	98,389	100	70.26

According to the species classification from the result of BLAST with Nr (Figure 1A), *Coffea canephora* (Rubiaceae, *Coffea*) had the highest matching rate, followed by *Vitis vinifera* (Vitaceae, *Vitis*), with the next highest being *Nicotiana tomentosiformis* (Solanaceae, *Nicotiana*). As can be seen from Figure 1B, 38.5% of the unigenes can be fully matched to Nr (the smaller the E-value, the higher the degree of matching and the similarity), and the degree of matching in Nr is relatively high. Figure 1C shows that 11.0% of the unigenes had more than 95% similarity, and 54% of the unigenes had 80–95% similarity (the higher the similarity, the higher the confidence).

All unigenes were assessed for GO assignments based on the Nr annotations. The unigenes were categorized into biological process (BP), cellular component (CC), and molecular function (MF) to reveal the gene function classification (Figure 2). Within these functional groups, "metabolic process" was the dominant group among biological processes (282,588), which contained 144,606 unigenes, and second was "cellular process" (62,514). Among cellular components (83,109), "cell part" (24,619) contained the largest number of unigenes, followed by "organelle" (19,336). With a total of only 57,203 unigenes in the molecular function category, the two largest groups were "binding" and "catalytic activity", which included 41,251 and 11,311 unigenes, respectively. This result was similar to those for *Lycium barbarum* [22], *Rhododendron fortune* [23], and *Elaeagnus mollis* Diels [24], which showed that metabolic processes and cellular processes contained the largest numbers of unigenes in all subcategories. The genes successfully annotated by KOG were classified into 26 KOG groups, as shown in Figure 3. The dominant group was general function prediction (1936), followed by the posttranslational modification, protein turnover, and chaperones (1459), while the smallest group was cell motility (only 4). The KOG classification results of the three species mentioned above are still consistent with the results of our study. In addition, the group of unknown functions contained 443 unigenes, accounting for 3.6% of the total annotation information in the *L. yunnanensis* flower transcriptome. Studies on *Castanea henryi* (Skan) Rehd. et Wils. [25], *Pinus yunnanensis* Franch. [26], *Rhododendron longipedicellatum* Lei Cai et Y. P. Ma [27], and *Phyllanthus emblica* [28] have produced

similar results. We speculate that this may be caused by insufficient annotation information. By comparing the KEGG database, the unigenes can be classified into 22 KEGG pathways according to the signaling pathways involved. Of these pathways, the most represented was translation (957), followed by carbohydrate metabolism (918) and signal transduction (910) (Figure 4).
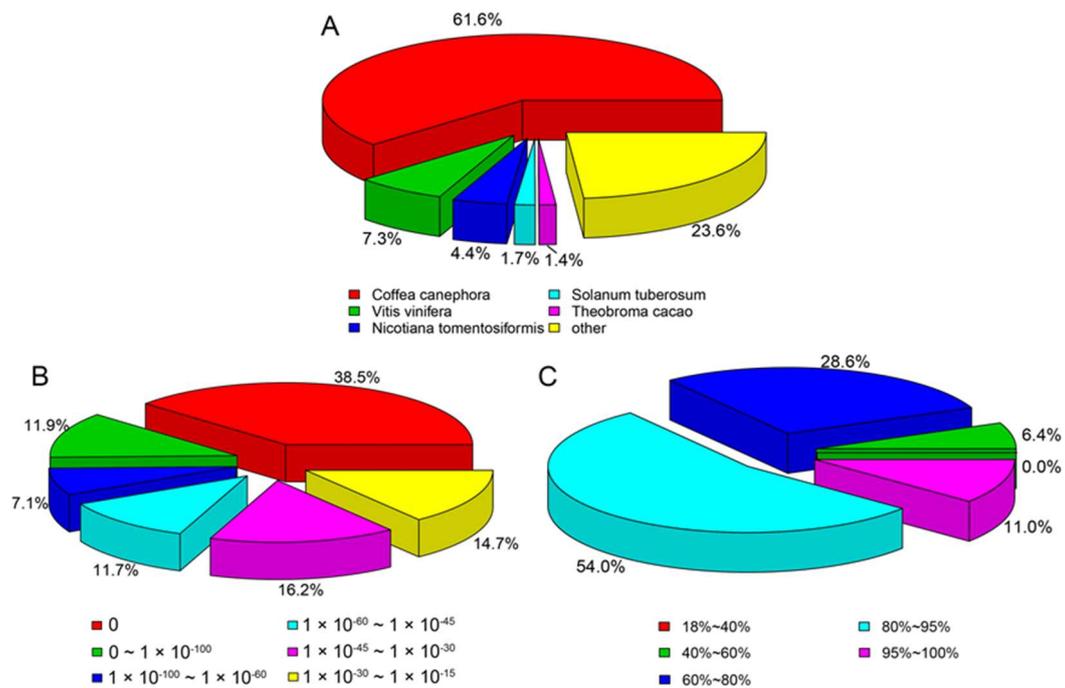

Figure 1. Characterization of assembled *L. yunnanensis* unigenes using the Nr database: (**A**) Species distribution for the assembled unigenes. (**B**) E-value distribution for the assembled unigenes. (**C**) Similarity distribution for the assembled unigenes.
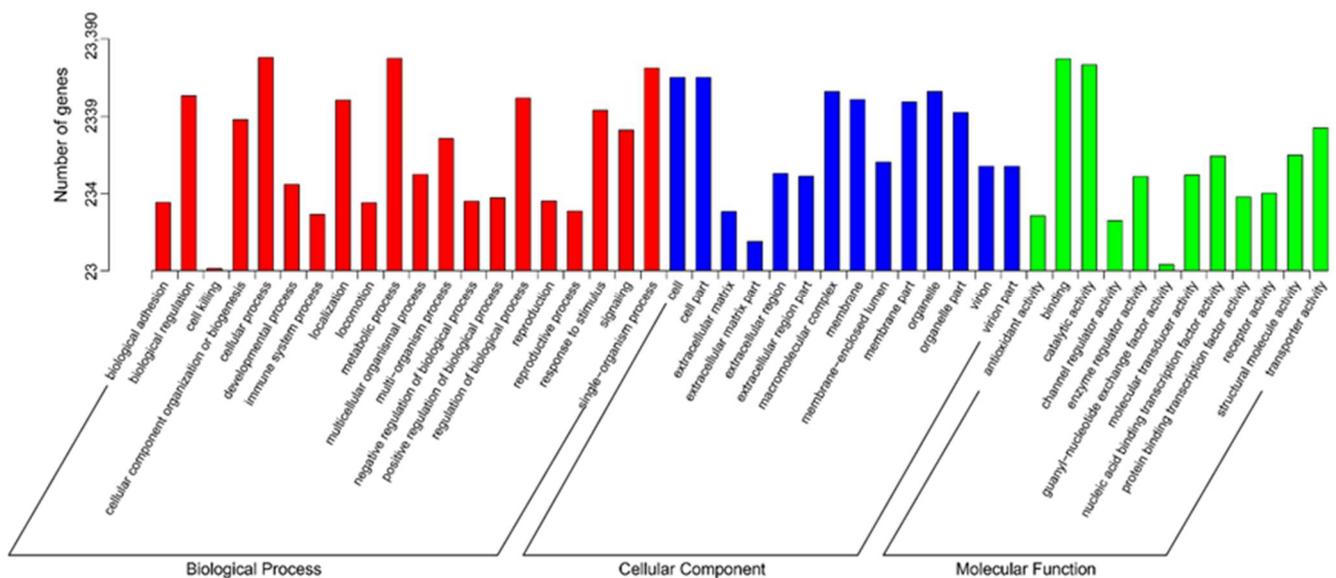

Figure 2. GO classification of assembled unigenes of the *L. yunnanensis* flower transcriptome. The *x*-axis indicates the subgroups in GO annotation; the *y*-axis indicates the number of genes in each category.
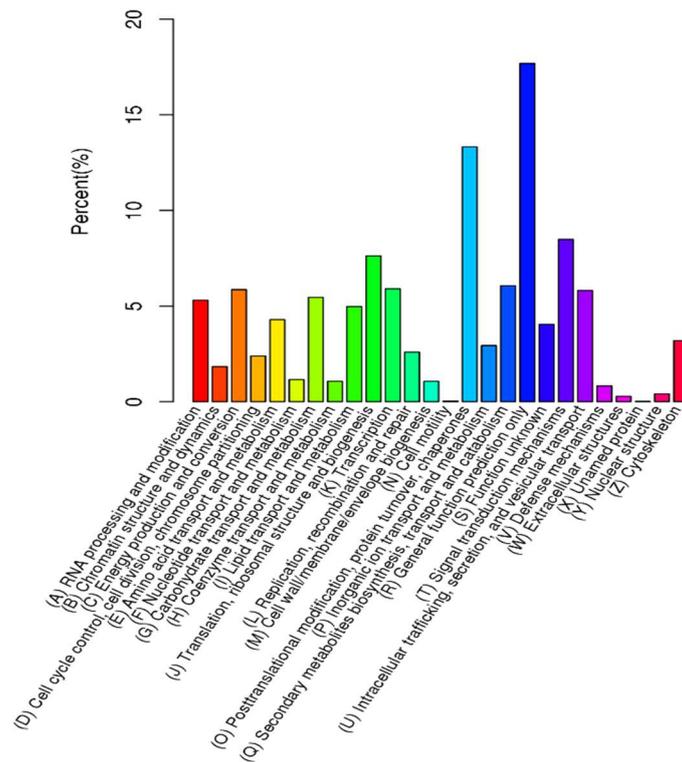
Figure 3. KOG classification of assembled unigenes of the *L. yunnanensis* flower transcriptome. The *x*-axis indicates the 26 groups in KOG annotation; the *y*-axis indicates the percentage of the number of genes in each group relative to the total number of annotated genes.

Figure 4. KEGG metabolic pathways of assembled unigenes of the *L. yunnanensis* flower transcriptome: (**A**) cellular processes; (**B**) environmental information processing; (**C**) genetic information processing; (**D**) metabolism; (**E**) organismal systems. The *x*-axis indicates the number of genes in each metabolic pathway and the ratios of the number of genes to total number of annotated genes; the *y*-axis indicates the names of the KEGG metabolic pathways.

The unigenes were roughly divided into 3 functional categories and 56 subcategories according to GO function, among which metabolic process, cellular process, cell, cell part, organelle, binding, and catalytic activity were highly enriched. In other words, the expression levels of genes related to cellular activity, metabolic activity, and catalytic activity was high, indicating that *L. yunnanensis* has strong metabolic capacity. We hypothesized that this may be related to the continuous cell proliferation of the meristem during flower development and the vigorous metabolic activities in the flower organs of *L. yunnanensis*. Xia et al. [29] analyzed the flower transcriptome of *Camellia sinensis*, and their results and hypotheses were similar to ours. KEGG functional annotation analysis showed that unigenes could be grouped into 5 categories, among which the pathways related to metabolism and genetic information processing accounted for the highest proportion, and the number of genes related to metabolism-related pathways was the largest. This was further proof that there was strong metabolic activity in *L. yunnanensis* during this period. Additionally, the metabolic pathways of amino acid metabolism, carbohydrate metabolism, lipid metabolism, biosynthesis of other secondary metabolites, and environmental adaptation were involved in the KEGG functional annotation analysis. These data provide a molecular basis for further research on resistance mechanisms, and allow us to explore the genes related to the flowering regulation and environmental adaptability of *L. yunnanensis*.

2.2. Identification of SSRs

The MISA (Microsatellite) tool was used to identify SSR loci in the transcriptome data, and a total of 15,384 SSRs were identified from 98,389 unigenes; the SSR frequency in the transcriptome was 15.63%, and the mean distance of the SSRs in the unigenes was 6.39 kb. It is well known that the distribution frequency of SSR loci varies greatly between different species, which may be related to the genome size of species, the organ and period of the sampled plant used for transcriptome sequencing, the SSR development methods, and the screening criteria [30]. Species with higher frequencies include, for instance, hybrid *Cymbidium* (58.64%) [31], *Morus alba* L. (45.56%) [32], and *Bougainvillea glabra* (44.91%) [33]. Species with lower frequencies include *Paeonia suffruticosa* (6.4%) [34], *Pinus elliottii* (4.80%) [35], and *Chrysanthemum morifolium* (2.84%) [36]. Kumpatal [37] found that the distribution frequency of SSR loci in dicotyledons ranged from 2.65% to 16.82%, with an average of 9.73%. The distribution frequency of SSR loci in the flower transcriptome of *L. yunnanensis* is consistent with the distribution frequency range of SSR loci in dicotyledons, and higher than the average. Therefore, the flower transcriptome data of *L. yunnanensis* had abundant SSR loci, and could be used for the development of SSR markers.

There were abundant SSR types in the flower transcriptome of *L. yunnanensis*, with the distribution of mononucleotide to hexanucleotide repeat types, but the proportion of different repeat types was significantly different. The number of mononucleotide repeat motifs found to be the most frequent was 8889, accounting for 60.29% of the total repeat motifs. Second, dinucleotide and trinucleotide repeat motifs were 3340 and 2309, accounting for 22.65% and 15.66%, respectively, of the total repeat motifs. The tetranucleotide, pentanucleotide, and hexanucleotide repeat motifs accounted for only 158, 23, and 25, respectively. In this study, the number of short-repeat motifs (mononucleotide, dinucleotide, and trinucleotide) was more than that of long-repeat motifs, which is consistent with the distribution in most species, and supports the view that long-repeat motifs have high variability [38]. As shown in Table 3, the mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs with the highest frequency were A/T, AG/CT, AAG/CTT, AAAG/CTTT, CCTTC/GAAGG, and GAGAC/GTCTC, respectively. The frequency of mononucleotide repeats was the highest. This result was similar to those for *Ricinus communis* [39], *Glycyrrhiza uralensis* Fisch [40], *Arachis hypogaea* L. [41], and *Elaeagnus mollis* Diels [24], which showed that the A/T repeat motif accounted for the largest proportion of SSRs. The proportion of 12–21 bp repeats in SSRs under different repeat types and repeat times was more than 95%, among which the proportion of 5 times and trinucleotide repeat motifs was the largest (22.25%), followed by that of 6 times and dinucleotide repeat motifs (16.48%) (Table 4).

Table 3. Frequency of different repeat motifs in the SSRs of the *L. yunnanensis* flower transcriptome.

Repeat Types	No.	Frequency (%)	Maximum Repeat Motif (Number and %)
Mononucleotide repeat (p1)	8889	60.29	A/T (8816, 99.18%)
Dinucleotide repeat (p2)	3340	22.65	AG/CT (2174, 65.09%)
Trinucleotide repeat (p3)	2309	15.66	AAG/CTT (428, 18.54%)
Tetranucleotide repeat (p4)	158	1.07	AAAG/CTTT (22, 13.92%)
Pentanucleotide repeat (p5)	23	0.16	CCTTC/GAAGG (2, 8.70%) GAGAC/GTCTC (2, 8.70%)
Hexanucleotide repeat (p6)	25	0.16	CCTTC/GAAGG (2, 8.70%) GAGAC/GTCTC (2, 8.70%)

Table 4. Number of SSRs of different repeat motifs on different repeat times in the flower transcriptome of *L. yunnanensis*.

Repeat Types	Repeat Times						
	5	6	7	8	9	10	>10
Dinucleotide repeat	—	965	688	558	591	417	121
Trinucleotide repeat	1303	604	374	23	5	—	—
Tetranucleotide repeat	132	25	—	—	—	1	—
Pentanucleotide repeat	14	6	2	—	1	—	—
Hexanucleotide repeat	7	9	5	4	—	—	—

The polymorphism level of SSR markers is an important basis for evaluating their availability, and the length of SSRs is an important factor affecting their polymorphism. The results showed that the polymorphism was high when the length of SSRs was more than or equal to 20 bp, medium when the length was 12–19 bp, and extremely low when the length was less than 12 bp [42]. Therefore, more than 95% SSRs were had a moderate-to-high level of polymorphism in this study. In addition, lower-grade SSR motifs (i.e., mononucleotide, dinucleotide, and trinucleotide) tend to produce polymorphism more easily than higher-order SSR motifs [43,44]. Due to the large number of mononucleotides, dinucleotides, and trinucleotides in this study, it can be predicted that SSRs obtained from the *L. yunnanensis* flower transcriptome have a high potential for polymorphism, and this leads to a high application value in the study of its molecular markers.

2.3. SSR Primer Screening and Verification

A total of 9195 primer pairs were designed using Primer 3.0, and 81 primer pairs were randomly selected to amplify across the genomic DNA template of *L. yunnanensis*. Of the 81 primer pairs, 44 of them successfully amplified fragments from *L. yunnanensis* genomic DNA, with a success rate of 54.32%. Therefore, the primers developed and designed in this study have a certain applicability to the population of *L. yunnanensis*. According to the preliminary screening results, 10 individuals were randomly selected from each population (a total of 60 individuals from 6 populations) to further verify the polymorphism of the 44 primer pairs that could amplify bands. Among them, 17 primer pairs were found to be polymorphic in *L. yunnanensis*, producing clear amplicons of the expected size, which were suitable for the detection of SSR loci by capillary electrophoresis. The specific information of 17 EST-SSR primers is shown in Table 5. A dendrogram clustered the 6 populations of *L. yunnanensis* into 2 clusters (Figure 5). One cluster (I) included 2 populations (GDX and GCG) of Gongshan County and one population (FMM) of Fugong County. The other cluster (II) included 2 populations (LCM and LLO) of Lushui County and one population (FPK) of Fugong County. Geographically, FMM is close to Gongshan County and FPK is close to Lushui County, which is consistent with the results of cluster analysis.

Table 5. The information of 17 EST-SSR primers of *L. yunnanensis*.

Primer ID	Primer Sequence (5′-3′)	Repeat Motif	*Ta* (°C)	Size (bp)	Na	Ne	*I*	PIC
N3	CAAATTCGCGCACCAAAACG GCTAGAGAGAAAAGGGGCCG	(TGGCGT)5	52	261–279	4	2.538	1.061	0.542
N6	ACTGCGTACCTCTCCCTCTT TCTCTCTCTCTCGGACGGAC	(TAATT)5	58	212–228	5	1.933	0.890	0.426
N9	GACCCCAAGTTGGCTGATCA AGGGCACTTCTGTCATTTCGA	(TGTTAC)8	60	127–187	6	2.212	0.923	0.446
N10	CTGGTGCACGAGGATTGAGT GAAGAGTGCCATGGAAACTGC	(TCAATT)7	60	194–224	2	1.427	0.476	0.255
Z13	CCTCCCATAGCAGCAGCAAT AGTAGTATTAATAATGGCTGGAGGT	(ACC)5	54	117–132	5	1.699	0.833	0.382
N22	CGCTTCTGTGTTCGAAACCA CAAAGCTTCCCGTCAACAGC	(ACAAC)7	60	165–180	4	2.575	1.091	0.549
N24	CCCACCGAGCAATACCCAAA ACCTTCTCTGTACTCTGCCT	(GAAA)5	56	268–280	4	1.344	0.555	0.245
Z31	GCAATCCTACTCGTGCTGGT AGCCAAGACTCGGCAGAAAA	(GGC)6	54	228–240	5	2.583	1.239	0.580
Z32	TGCACTCCATAAAAGAAGAAAACACA TGCAGTAACTTCGTGCCCTT	(TATT)5	53	114–122	3	1.405	0.544	0.264
Z33	CCCAACCCACCACACAAGT AGAGAGGAGGATCGAGGACG	(TCT)6	54	255–270	5	1.749	0.887	0.404
Z36	TCGGGTCCTAGGGCTTTCTT GGCCCTCCTTGAGCATTGAT	(CTTT)5	54	211–219	3	1.766	0.768	0.390
Z38	ACCCAAGGAACTCTGTCTCT ACACTTTCGTCGTCCTTAGGT	(AAT)6	53	109–118	4	1.522	0.684	0.321
N41	GCCAGAAGGATAGCTTTCGC GGTTTGTGGTGGTTTTTGGGA	(TCCT)5	53	197–221	6	1.298	0.547	0.223
Z48	AGGAAGGGCTTGTTTTTAAGGT GAGCCAATGACGATCCAGCT	(AG)8	52	215–225	6	1.724	0.914	0.400
Z50	TCTGCTGCATCCAATGTACTGT CCTGCCATAGGTGCCCATTT	(GT)8	54	144–154	5	2.690	1.205	0.576
Z54	AGTAAGTGGGTGGAGGTGGT AGGGGCTGATTCTCTAGCGA	(TTGA)5	59	212–216	2	1.910	0.670	0.363
X70	AGCTGGAAACTAAAGGTGGAGG CTCAGTCTGTCAGGCCTGTG	(ATC)5	58.5	244–250	3	2.355	0.944	0.493

Na = number of different alleles; Ne = number of effective alleles; *I* = Shannon's information index; PIC = polymorphism information content; *Ta* = annealing temperature.

The SSR marker diversity index was estimated using POPGENE 1.32; the effective number of alleles (Ne) and Shannon's information index (I) per locus varied from 1.298 to 2.690 (average of 1.925), and from 0.476 to 1.239 (average of 0.837), respectively. Using CERVUS 3.0 (Field Genetics, London, UK), we calculated that the allelic polymorphism information content (PIC) for each SSR locus ranged from 0.223 to 0.580, with an average of 0.403. Botstein [45] put forward a criterion that primers are highly polymorphic when PIC > 0.5, moderately polymorphic when 0.5 > PIC > 0.25, and low-grade polymorphic when PIC < 0.25. Most of the 17 EST-SSR primer pairs were moderately or highly polymorphic, indicating that these primers could be used for genetic diversity analysis and fingerprint construction of *L. yunnanensis*.

Today, plant diversity is gradually decreasing on a global scale, leading to a related decline in ecosystem service function [46]. Many rare species, such as *L. yunnanensis*, have seldom been studied, and lack protection due to a lack of social awareness. Thus, the numbers and distribution ranges of the species are shrinking. Using transcriptome sequencing technology to study the genetic diversity of *L. yunnanensis* can quickly and efficiently develop mass EST-SSR markers. The application of these markers is conducive to quickly supplementing the molecular biological information of this species, and provides a theoretical basis and scientific evidence for research on the genetic resources, conservation,

and utilization of *L. yunnanensis*. In addition, due to the high ornamental value of *L. yunnanensis*, we can refer to Ding's [47] study on *Chrysanthemum*, in which the development of SSR markers for genes related to the regulation of flower shape and color were studied, and these SSR markers were applied to the classification of more than 100 *Chrysanthemum* germplasm resources with different flower colors and shapes. In the same way, we could also screen the SSR loci of important ornamental traits such as the flower shape and color of *L. yunnanensis*, which could lay the foundation for breeding new varieties of *L. yunnanensis*.
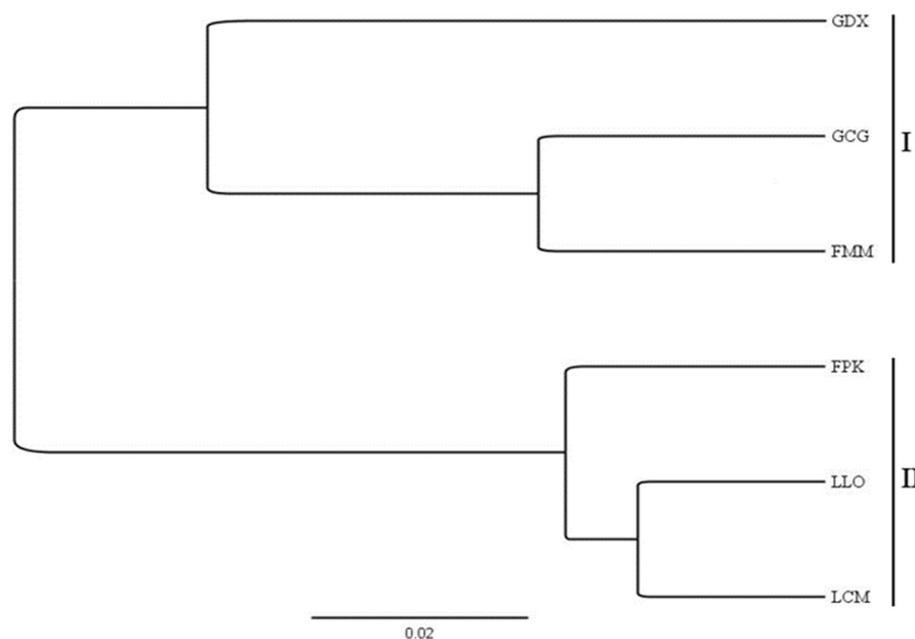

Figure 5. Cluster analysis of 6 *L. yunnanensis* populations based on 17 EST-SSR markers.

3. Materials and Methods

3.1. Plant Materials

We picked the experimental materials from Nujiang Lisu Autonomous Prefecture, Yunnan Province. Inflorescences of *L. yunnanensis* were collected, bottled, and brought back to Kunming for subsequent use. The blooming flowers of five individuals collected within LLO (Table 6) were frozen immediately in liquid nitrogen and stored at −80 °C for use as transcriptome sequencing materials. The plant samples were mixed, and RNA was extracted using an RNA extraction kit from TransGen (Beijing, China). Transcriptomic sequencing was performed at Beijing Novogene Biological Information Technology Co., Ltd., Beijing, China. Based on the results of previous investigations into the resource status of *L. yunnanensis*, 6 populations were selected as primer polymorphism detection materials in regions with a dense distribution and adequate quantity, and where sampling was convenient and relatively representative. These were GDX, GCG, FMM, FPK, LLO, and LCM (Table 6).

3.2. De Novo Assembly, Functional Annotation, and Classification

Transcriptome sequencing was carried using an Illumina HiSeq 4000 platform (Beijing Novogene Biological Information Technology Co., Ltd., Beijing, China); clean reads were used for de novo assembly using Trinity (version: trinityrnaseq-r20131110) with the default parameters to obtain transcripts, and TGICL version 2.1 software was used to cluster the transcripts into unigenes [48,49]. For functional prediction and classification, all unigenes were compared against public databases, including NCBI non-redundant protein sequences (Nr), Nt, Protein Family (Pfam), Eukaryotic Orthologous Groups of Proteins (KOG/COG), Swiss-Prot, the KEGG Orthology database (KO), and Gene Ontology (GO). BLAST [50] was used for unigene annotation in the Nt, Nr, COG, and Swiss-Prot databases. The

KEGG Automatic Annotation Server [51] was used for KEGG annotation of unigenes. Blast2GO [52] and Nr annotation results were used for the GO annotation.

Table 6. The location and habitat of the 6 *L. yunnanensis* populations.

Code	GDX	GCG	FMM	FPK	LLO	LCM
Location	Dulongjiang township, Gongshan County	Cikai town, Gongshan County	Maji township, Fugong County	Pihe township, Fugong County	Luobenzhuo township, Lushui County	Chenggan township, Lushui County
Altitude (m)	1324–1788	1325–1778	1354–1455	1738–1752	2029–2153	1664–1736
Longitude and latitude (°)	98.3188 −98.3466 /27.6915 −27.9160	98.6088 −98.8334 /27.5254 −27.7472	98.8388 −98.8453 /27.3632 −27.3651	98.9102 −98.9106 /26.4799 −27.4801	98.8116 −98.8172 /26.4780 −26.4812	98.8253 −98.8312 /26.3540 −26.3563

3.3. Detection of EST-SSR Markers and Designing of Primers

SSR loci in the transcriptome were identified using the microsatellite identification tool MISA (http://misaweb.ipk-gatersleben.de/) (accessed on 3 December 2020) [53]. The parameters of identifying the mono-, di-, tri-, tetra-, penta-, and hexanucleotide motifs were set to minimum numbers of 10, 6, 5, 5, 5, and 5 repeat units, respectively. Primer pairs were designed using Primer 3.0 software [54], according to the following parameters: the length range of the optimal primer sizes was 18–23 bp, the optimal annealing temperature was from 55–65 °C, the PCR product size range was 80–300 bp, and the GC content was 40–60%.

3.4. DNA Extraction and Quality Check

Total genomic DNA was extracted from the silica-dried leaves of *L. yunnanensis* with a plant genomic DNA extraction kit containing RNase A (TransGen, Beijing, China), following the manufacturer's protocol. The integrity and quality of the genomic DNA was checked using 0.8% agarose gel electrophoresis. The DNA purity and concentration were determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA), and samples were stored at −20 °C until used for SSR amplification.

3.5. EST-SSR Amplification and Data Analysis

The 81 pairs of randomly selected SSR primers were used for preliminary screening, and 10 individuals per population were used to test the primers amplified with clear bands in order to further determine the polymorphism of the SSR primers. PCR amplification was carried out in a 10 μL reaction mixture containing 10–30 ng of genomic DNA, 0.15 mmol/L of each dNTP, 1 μL of 10×Taq buffer, 0.6 mmol/L each of reverse and forward primer, and 1 unit of Taq DNA polymerase (TransGen, Beijing, China). The PCR amplification was carried out using a Bio-Rad C1000 Touch ™ Thermal Cycler (Bio-Rad Laboratories, CA, USA), as follows: 5 min of initial denaturation at 94 °C, followed by 30 cycles of denaturation for 30 s at 94 °C, a temperature gradient for annealing from 50 °C to 60 °C for 30 s, and extension at 72 °C for 1 min, with a final 10 min at 72 °C for final extension. The PCR products were resolved by electrophoresis in 1.5% agarose gels to determine whether amplification was successful. PCR fluorescent tagging and capillary electrophoresis were performed to further screen polymorphisms. The 5′ end of each forward primer was labeled with FAM or HEX fluorescent dyes (Thermo Fisher Scientific, Wilmington, DE, USA). The PCR amplifications were performed using the PCR conditions mentioned above. The multiplex DNA products labeled with the above-mentioned fluorescent dyes were analyzed on an ABI 3730xl DNA Analyzer with a GeneScan 500 LIZ Size Standard (Thermo Fisher Scientific, Wilmington, DE, USA), and allele sizes were assessed using GeneMapper 4.1 (Thermo Fisher Scientific, Wilmington, DE, USA). The parameters of genetic diversity—including the number of alleles (Na), effective number of alleles (Ne), and Shannon's information index (I)—of these SSR loci were estimated using POPGENE 1.32 [55], and the

allelic polymorphism information content (PIC) for each SSR locus was calculated using CERVUS 3.0 (Field Genetics, London, UK) [56]. Cluster analysis of 6 populations on the basis of shared allele distance (DAS) was performed using POPULATIONS 1.2.30 [57], with the unweighted pair group method with arithmetic mean (UPGMA) method. The clustering tree was visualized and edited using Interactive Tree of Life (iTOL) version 3 [58]. The primers with good repeatability, clear bands, and high polymorphism were selected as candidate primers for molecular genetic diversity research on *L. yunnanensis*.

4. Conclusions

The *L. yunnanensis* transcriptome was obtained using an Illumina HiSeq platform. In total, 98,389 unigenes were generated, of which 98,389 unigenes were aligned with the sequences of public databases. Nr, Nt, Pfam, KOG/COG, Swiss-Prot, KEGG, and GO annotated 31,859, 13,853, 22,684, 10,947, 21,416, 9722, and 23,390 unigenes, respectively. EST-SSR specific primer pairs were successfully designed. In addition, MISA was used to identify SSR loci from all unigenes, and a total of 15,384 SSRs was identified. The repeat motif was given priority with mononucleotides, dinucleotides, and trinucleotides. Furthermore, the 81 primer pairs were synthesized randomly, of which 44 pairs showed effective amplification, 17 primers showed stable amplification, and rich polymorphism was observed in 6 populations. This sequence resource and these potential EST-SSR markers can be used for subsequent population genetic diversity analysis and marker-assisted breeding, which is of great significance for formulating resource conservation and utilization strategies for *L. yunnanensis*.

Author Contributions: Z.L., H.M. and Y.Z. conceived and designed the research; X.L. (Xi Liu), X.L. (Xiongfang Liu) and Y.L. performed the experiments; Y.Z. and Y.L. analyzed the data and wrote the paper; Y.Z., H.M. and S.Q. revised the manuscript; Z.L. supervised the project. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

References

1. Li, Y.Y. Studies on Conservation Genetics of Chinese Endemic Species *Luculia yunnanensis* Hu. Doctoral Dissertation, Chinese Academy of Forestry, Beijing, China, 2017.
2. Allendorf, F.; Ryman, N. The role of genetics in population viability analysis. In *Population Viability Analysis*; Beissinger, S.R., McCullough, D.R., Eds.; The University of Chicago Press: Chicago, IL, USA, 2000; pp. 50–85.
3. Holsinger, K.E.; Gottlieb, L.D. Conservation of rare and endangered plants: Principles and prospects. In *Genetics and Conservation of Rare Plants*; Falk, D.A., Holsinger, K.E., Eds.; Oxford University Press: New York, NY, USA, 1991; pp. 195–208.
4. Lande, R. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* **1993**, *142*, 911–927. [CrossRef] [PubMed]
5. Soule, M.; Simberloff, D. What do genetics and ecology tell us about the design of nature reserves? *Biol. Conserv.* **1986**, *35*, 19–40. [CrossRef]
6. Kaljund, K.; Jaaska, V. No loss of genetic diversity in small and isolated populations of *Medicago sativa* subsp. *falcata*. *Biochem. Syst. Ecol.* **2010**, *38*, 510–520. [CrossRef]
7. Varshney, R.K.; Graner, A.; Sorrells, M.E. Genetic microsatellitemarkers in plants: Features and applications. *Trends Biotechnol.* **2005**, *23*, 48–55. [CrossRef] [PubMed]
8. Kalia, R.K.; Rai, M.K.; Kalia, S.; Singh, R.; Dhawan, A.K. Microsatellite markers: An overview of the recent progress in plants. *Euphytica* **2011**, *177*, 309–334. [CrossRef]
9. Liu, G.; Xie, Y.J.; Zhang, D.Q.; Chen, H.P. Analysis of SSR loci and development of SSR primers in Eucalyptus. *J. For. Res.* **2018**, *29*, 273–282. [CrossRef]
10. Liu, Y.; Fang, X.M.; Tang, T.; Wang, Y.D.; Wu, Y.H.; Luo, J.Y.; Wu, H.T.; Wang, Y.Q.; Zhang, J.; Ruan, R.W.; et al. Inflorescence transcriptome sequencing and development of new EST-SSR markers in common buckwheat (*Fagopyrum esculentum*). *Plants* **2022**, *11*, 742. [CrossRef]

11. Ellis, J.R.; Burke, J.M. EST-SSRs as a resource for population genetic analyses. *Heredity* **2007**, *99*, 125–132. [CrossRef]

12. Yang, X.; Yang, N.; Yuan, Q.; Zhao, G.J.; Li, G.L.; Jia, L.M.; Chen, Z. Development and application of EST-SSR molecular markers in *Pinus bungeana*. *J. Beijing For. Univ.* **2021**, *43*, 1–11.

13. Xiang, X.Y.; Zhang, Z.X.; Wang, Z.G.; Zhang, X.P.; Wu, G.L. Transcriptome sequencing and development of EST-SSR markers in *Pinus dabeshanensis*, an endangered conifer endemic to China. *Mol. Breed.* **2015**, *35*, 1–10. [CrossRef]

14. Li, X.Y.; Guo, Q.; Dong, L.; Sun, Y.H.; Niu, D.S.; Liu, J.P.; Wang, H.S.; Li, Y. EST-SSR analysis of genetic diversity of *Robinia pseudoacacia* clones in Jixian county, Shanxi province of northern China. *J. Beijing For. Univ.* **2019**, *41*, 39–48.

15. Dehkordi, M.K.; Beigzadeh, T.; Sorkheh, K. Novel in silico EST-SSR markers and bioinformatic approaches to detect genetic variation among peach (*Prunus persica* L.) germplasm. *J. For. Res.* **2020**, *31*, 1359–1370. [CrossRef]

16. Wani, M.S.; Sharma, V.; Gupta, R.C.; Munshi, A.H. Development and characterization of SSR markers in Himalayan species Betula utilis. *J. For. Res.* **2020**, *31*, 1453–1460. [CrossRef]

17. Zhou, X.J.; Dong, Y.; Zhao, J.J.; Huang, L.; Ren, X.P.; Chen, Y.N.; Huang, S.M.; Liao, B.S.; Lei, Y.; Yan, L.Y.; et al. Genomic survey sequencing for development and validation of single-locus SSR markers in peanut (*Arachis hypogaea* L.). *BMC Genom.* **2016**, *17*, 420. [CrossRef]

18. Ma, H.; Wang, L.; Wan, Y.M.; Li, H.Z.; Li, Z.H.; Liu, X.X.; Liang, N.; Li, W.J. A set of novel microsatellite markers developed for *Luculia yunnanensis* (Rubiaceae), an endangered plant endemic to Yunnan, China. *Int. J. Mol. Sci.* **2012**, *13*, 534–539. [CrossRef] [PubMed]

19. Ramilowski, J.A.; Sawai, S.; Seki, H.; Mochida, K.; Yoshida, T.; Sakurai, T.; Muranaka, T.; Saito, K.; Daub, C.O. *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals. *Plant Cell Physiol.* **2013**, *54*, 697–710. [CrossRef]

20. Wan, Y.M. Mechanism on photoperiodic control of flowering in *Luculia gratissima* 'Xiangfei'. Doctoral Dissertation, Chinese Academy of Forestry, Beijing, China, 2019.

21. Zhu, Y.H.; Zhang, M.J.; Li, L.; Zhao, L.; Dong, C.M. Transcriptome analysis of *Prunella vulgaris* and identification of putative genes involved in second metabolism biosynthesis. *Zhong Cao Yao* **2019**, *50*, 1220–1226.

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

23. Xu, Q.W.; Lou, X.Z.; Yang, B.; Lin, E.P.; Tong, Z.K. Transcriptome sequencing and analysis of *Rhododendron fortune*. *J. Zhejiang AF Univ.* **2019**, *36*, 1190–1198.

24. Liu, Y.L.; Li, S.Q.; Wang, Y.Y.; Liu, P.Y.; Han, W.J. *De novo* assembly of the seed transcriptome and search for potential EST-SSR markers for an endangered, economically important tree species: *Elaeagnus mollis* Diels. *J. For. Res.* **2020**, *31*, 69–77. [CrossRef]

25. Zhang, L.; Fan, X.M.; Lin, Q.; Xiang, H.; Wang, W.; Yuan, D.Y. Transcriptome Analysis for developing kernel and expression analysis of starch and sucrose metabolism-related genes in *Castanea henryi*. *J. Plant Genet. Resour.* **2015**, *16*, 603–611.

26. Cai, N.H.; Deng, L.L.; Xu, Y.L.; Xu, Y.; Zhou, L.; Wang, D.W.; Tian, B.; He, C.Z.; Duan, A.A. Transcriptome analysis for *Pinus yunnanensis* based on high throuput sequencing. *Bull. Bot. Res.* **2016**, *36*, 75–83.

27. Li, T.Q.; Liu, X.F.; Wan, Y.M.; Li, Z.H.; Qi, G.H.; Li, Y.Y.; Liu, X.X.; He, R.; Ma, Y.; Ma, H. Transcriptome analysis for *Rhododendron longipedicellatum* (plant species with extremely small populations) based on high throughput sequencing. *Bull. Bot. Res.* **2017**, *37*, 825–834.

28. Liu, X.F.; Li, T.Q.; Li, Z.H.; Wan, Y.M.; Liu, X.X.; Zhang, X.; An, J.; Ma, H. Transcriptome analysis for *Phyllanthus emblica* distributed in dry-hot valleys in Yunnan, China. *Fore. Res.* **2018**, *31*, 1–8.

29. Xia, L.F.; Zhu, X.Z.; Liang, M.Z.; Ma, W.; Song, W.X.; Tian, Y.P.; Zhou, M.; Chen, L.B. Transcriptome analysis of flower from tea plant [*Camellia sinensis* (L.) O. Kuntz] via RNA-Seq technology. *Southwest China J. Agric. Sci.* **2016**, *29*, 2058–2062.

30. Biswas, M.K.; Chai, L.J.; Mayer, C.; Xu, Q.; Guo, W.W.; Deng, X.X. Exploiting BAC−end sequences for the mining, characterization and utility of new short sequences repeat (SSR) markers in *Citrus*. *Mol. Biol. Rep.* **2012**, *39*, 5373–5386. [CrossRef] [PubMed]

31. Zhong, H.Q.; Lin, R.Y.; Lin, B.; Wu, J.S.; Fan, R.H.; Ye, X.X. Analysis on SSR information in transcriptome and development of EST-SSR markers for hybrid cymbidium. *Chin. J. Cell. Biol.* **2020**, *42*, 286–295.

32. Zhao, T.; Chang, S.X.; Leng, Q.Y.; Xu, S.S.; Yin, J.M.; Niu, J.H. Development of SSR molecular markers based on transcriptome sequencing of *Bougainvillea*. *Mol. Plant Breed.* **2019**, *17*, 4331–4341.

33. Wang, H.; Xie, Y.; Sun, Z.C.; Gao, Y.J.; Zhang, D.H.; Song, Y.X.; Gao, Y.X. Analysis on SSR locis in mulberry fruit transcriptome. *Mol. Plant Breed.* **2020**, *18*, 208–215.

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

35. Yi, M.; Zhang, L.; Lei, L.; Cheng, Z.S.; Sun, S.W.; Lai, M. Analysis of SSR information in transcriptome and development of EST-SSR molecular markers in *Pinus elliottii* Engelm. *J. Nanjing. For. Univ.* **2020**, *44*, 75–83.

36. Feng, S.; He, R.; Lu, J.; Jiang, M.; Shen, X.; Yan, J.; Wang, Z.; Wang, H. Development of SSR markers and assessment of genetic diversity in medicinal *Chrysanthemum morifolium* cultivars. *Front. Genet.* **2016**, *7*, 113. [CrossRef] [PubMed]

37. Kumpatal, S.P.; Mukhopadhyay, S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* **2005**, *48*, 985–998. [CrossRef] [PubMed]

38. Li, Q.; Li, B.; Guo, S.X. SSR information in transcriptome of *Dendrobium nobile*. *China J. Chin. Mater. Med.* **2017**, *42*, 63–69.

39. Zou, M.L.; Xia, Z.Q.; Chen, J.R.; Wang, W.Q. Analysis of SSR information and annotation in EST resource of Caster Bean (*Ricinus communis*). *Chin. J. Trop. Crop.* **2012**, *33*, 2138–2143.

40. Liu, Y.L.; Zhang, P.F.; Song, M.L.; Hou, J.L.; Mei, Q.; Wang, W.Q.; Liu, C.S. Transcriptome analysis and development of SSR molecular markers in *Glycyrrhiza uralensis* Fisch. *PLoS ONE* **2015**, *10*, e0143017. [CrossRef] [PubMed]

41. Zhou, W.; Wang, H.; Li, D.Z.; Yang, J.B.; Zhou, W. Isolation and characterization of 13 microsatellite loci from *Luculia pinceana* (Rubiaceae), a typical distylous species. *Hortscience* **2010**, *45*, 840–841. [CrossRef]

42. Temnykh, S.; De Clerck, G.; Lukashova, A.; Lipovich, L.; Cartinhour, S.; Mc Couch, S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* **2001**, *11*, 1441–1452. [CrossRef]

43. Dreisigacker, S.; Zhang, P.; Warburton, M.L.; Van Ginkel, M.; Hoisington, D.; Bohn, M.; Melchinger, A.E. SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different megaenvironments. *Crop. Sci.* **2004**, *44*, 381–388. [CrossRef]

44. Liu, S.S.; Qiao, Z.Q.; Zeng, H.J.; Li, Y.X.; Cai, N.; Liu, X.M.; Wang, X.M. Analysis on SSR loci in transcriptome and development of EST-SSR molecular markers in *Lonicera macranthoides*. *Mol. Plant Breed.* **2021**, *19*, 3015–3021.

45. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331. [PubMed]

46. Li, D.Z.; Yang, X.Y.; Hugh, W.P. Problems and challenges in germplasm preservation. *Bull. Chin. Acad. Sci.* **2010**, *25*, 533–540.

47. Ding, H.X. Development and Application of SSR Markers in *Chrysanthemum* Based on Full-Length Transcriptome Sequence. Master Dissertation, Henan University, Kaifeng, China, 2020.

48. Pertea, G.; Huang, X.Q.; Liang, F.; Antonescu, V.; Sultana, R.; Karamycheva, S.; Lee, Y.D.; White, J.; Cheung, F.; Parvizi, B.; et al. TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics* **2003**, *19*, 651–652. [CrossRef]

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

50. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]

51. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic. Acids. Res.* **2007**, *35*, W182–W185. [CrossRef]

52. Götz, S.; García-Gómez, J.M.; Terol, J.; Williams, T.D.; Nagaraj, S.H.; Nueda, M.J.; Robles, M.; Talón, M.; Dopazo, J.; Conesa, A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic. Acids. Res.* **2008**, *36*, 3420–3435. [CrossRef]

53. Beier, S.; Thiel, T.; Munch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [CrossRef]

54. Rozen, S.; Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **2000**, *132*, 365–386.

55. Yeh, F.C.; Yang, R.C.; Boyle, T.; Ye, Z.H.; Mao, J.X. *POPGENE, the User-Friendly Shareware for Population Genetic Analysis*; Molecular Biology and Biotechnology Centre, University of Alberta: Edmonton, AB, Canada, 1997; Volume 10, pp. 295–301.

56. Kalinowski, S.T.; Taper, M.L.; Marshall, T.C. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* **2007**, *16*, 1099–1106. [CrossRef]

57. Populations 1.2.30: Population Genetic Software (Individuals or Populations Distances, Phylogenetic Trees). Available online: http://bioinformatics.org/~{}tryphon/populations/ (accessed on 13 April 2022).

58. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **2016**, *44*, W242–W245. [CrossRef] [PubMed]