Development of EST-SSR Markers and their Use in Assessing Genetic Diversity in Chinese Fir Infusion Populations

Shuangyun Yang[1], Liming Bian[1],*, Zhiqiang Chen[1,2], Renhua Zheng[3], Shunde Su[3], Long Zhang[4], Xueyan Zheng[4], Yousery A. El-Kassaby[5], Xie Zhang[6]

[1] State Key Laboratory of Tree Genetics and Breeding, Co-Innovation Center for Sustainable Forestry in Southern China, Key Laboratory of Forest Genetics and Biotechnology of Ministry of Education, Nanjing Forestry University, Nanjing 210037, People's Republic of China
[2] Department of Forest Genetics and Plant Physiology, Umeå Plant Science Centre, Swedish University of Agricultural Sciences, Umeå 901-83, Sweden
[3] Key Laboratory of Timber Forest Breeding and Cultivation for Mountainous Areas in Southern China, Fujian Academy of Forestry Science, Fuzhou 350012, People's Republic of China
[4] Yangkou State-owned Forest Farm, Nanping 353211, People's Republic of China
[5] Department of Forest and Conservation Sciences, Faculty of Forestry, The University of British Columbia, Vancouver B.C. V6T 1Z4, Canada
[6] Hunan Academy of Forestry, Changsha 410004, People's Republic of China
*Correspondence: lmbian@njfu.edu.cn

Abstract

In advanced-generation tree breeding program, infusion populations are often used as an effective method to expand and maintain genetic diversity. To analyze the genetic diversity and population structure of six Chinese fir (*Cunninghamia lanceolata*) geographical populations, a total of 20 expressed sequence tag-derived simple sequence repeat markers pairs (EST-SSR) were developed and applied for genetic diversity analysis. The evaluated populations exhibited moderate genetic diversity with the following parameters: number of alleles (N_a: 4.850), effective number of alleles (N_e: 2.920), information index (*I*: 0.958), observed heterozygosity (H_o: 0.319), expected heterozygosity (H_e: 0.481), unbiased heterozygosity (uH_e: 0.496), and fixation index (*F*: 0.321). The analysis of molecular variance (AMOVA) suggested that only 9.42 % of genetic variation existed among populations, whereas the majority (90.58 %) resided within populations. Cluster analysis showed one population (Sichuan Dechang) as a separate taxon, likely due to its geographical isolation. The present study demonstrated the effectiveness of the developed EST - SSR in analyzing genetic diversity for population.

Keywords: Chinese fir; EST-SSR; Infusion populations; Population structure; Genetic diversity
Introduction

Introduction

Cunninghamia lanceolata (Lamb.) Hook., a tree species belonging to the genus *Cunninghamia* R. Br. in the Taxodiaceae family (Cao et al. 2022; Jing et al. 2023; Xiang et al. 2022), is an endemic timber species in China, primarily distributed in the Yangtze River basin and south of the Qinling Mountains (Huang et al. 2022) with diverse geographical variation. According to the Ninth National Forest Inventory, Chinese fir plantations covered 9.9 million hectares with a substantial volume of 755 million m^3, accounting for around a quarter of the dominant species in the country's plantation forests (Ji et al. 2022; Xu et al. 2022). Since the 1950s, genetic improvement of Chinese fir has been an active research endeavor in China. At present, the species is in its fourth breeding cycle, each with different breeding objectives. The first cycle focused on fast-growing and productive traits, while the fourth cycle aimed at producing fast-growing, high-quality families or clones of strong resistance to pest and insect. These objectives resulted in significant genetic gains in growth and volume. To ensure breeding objectives continually meet future climate change and market needs, it is crucial to introduce new germplasms and enrich existing breeding sources in a multi-objective breeding program (Kang 2019).The new germplasm is usually called an infusion population, which is commonly derived from superior trees selected from the original base population or the introduction of superior genetic resources from outside the breeding zone. In advanced-generation breeding, the infusion population's genetic material is typically utilized to improve specific traits or to enhance the genetic diversity within the advanced-generation breeding population (White et al. 2007).

Currently, the genetic diversity evaluation of Chinese fir germplasm relies on phenotypic traits, supported by molecular markers (Li et al. 2022). Traditional genetic diversity analysis based on phenotypic data mainly relied on morphological indicators, which are time-consuming and affected by

DOI:10.2478/sg-2025-0015
edited by the Thünen Institute of Forest Genetics

environmental factors. Molecular markers, such as SSR (simple sequence repeat) and SNP (single nucleotide polymorphisms), have been extensively applied in population genetic diversity studies, driven by the rapid development and continuous improvement of molecular biology techniques. SNP markers offer superior features for large-scale population analyses, including higher genomic density, greater reproducibility, compatibility with contemporary genomic methodologies, and an increased ability to detect subtle variations among individuals and populations. In contrast, SSR markers are less expensive to develop, remain widely used for specific applications, and have been standardized for variety identification and large-scale resource evaluation in species lacking high-quality genomes (Liu et al. 2022; Shi et al. 2022; Uddin et al. 2023). Among them, SSR markers, including g-SSR (genome-based development) and EST-SSR (expressed sequence tags-based), are commonly used. EST-SSR has the advantages of easy development and genotyping and higher efficiency compared to g-SSR markers. (Shi et al. 2022).

EST-SSR polymorphic primers developed in Chinese fir, are used for "plus trees" population, that is phenotypically superior individuals selected for breeding programs, collected from Anhui, Jiangxi, Sichuan, and Yunnan provinces as breeding material. By integrating existing markers with transcriptome sequencing, we developed novel EST-SSR polymorphic primers of Chinese fir. These markers were subsequently used to assess genetic diversity at the molecular level across several geographically distinct populations.

Materials and Methods

Plant materials and DNA extraction

During the period from November 2018 to December 2019, 112 germplasm resources of Chinese fir were collected from six areas (AHHS: Anhui Huangshan; JXGZ: Jiangxi Ganzhou; SCDC: Sichuan Dechang; SCYA: Sichuan Yaan; YNTC: Yunnan Tengchong; YNLL: Yunnan Longling), with SCDC representing the high-altitude area of Dechang in Sichuan province (Fig. 1, Table 1). The selected "plus tree" was older than 30 years and met the necessary phenotypic criteria (Xu et al. 2011). Scion materials were grafted onto rootstock planted at the state-owned forest farm, Yangkou, Fujian Province. In June 2022, 8-10 young needles with no noticeable disease symptoms or pests were collected from each grafted clone. Collected leaves were stored in -80°C until future use.

Genomic DNA of Chinese fir was extracted with the Tiangen kit (DP-320-02), and 1 μL DNA solution was obtained. Then, concentration and purity of the DNA were analyzed with the Thermo NanoDrop 2000 Ultra spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Furthermore, using 3% agarose gel electrophoresis, DNA quality was identified. The final extracted DNA was preserved at -80 °C.

Development of primers for EST-SSR markers in Chinese fir

Two methods were used to develop EST-SSR marker primers for Chinese fir: (1) For the reported polymorphic primers, polyacrylamide gel and capillary electrophoresis were adopted for screening the polymorphic primers; (2) Polymorphic primers were developed on the basis of transcriptome sequencing.

Screening of existing SSR primers

First, 91 pairs of primers with acceptable good amplification effect were selected from 270 pairs of primers published for Chinese fir. DNA of one single tree from each population was selected as the template, and the 91 primer pairs were used for amplification. Their effectiveness and polymorphism were screened with polyacrylamide gel electrophoresis and capillary electrophoresis, respectively.

Development of SSR primers based on transcriptome sequencing

The transcriptome was obtained by sequencing on the Illumina NovaSeq 6000 platform. In June 2022, needles of one individual in Jiangxi Ganzhou population (JXGZ) were collected and sent to Beijing Qingke Biotechnology Co., Ltd. for transcriptome sequencing.

MISA (Scott et al. 2000) was applied to identify the simple repeats in transcriptome sequences and to count the different SSR types and their distribution. Parameters of the minimum number of single, two, three, four, five, and six nucleotide repeats were 10, 7, 6, 5, 5, and 4, respectively (Supplementary Table 1).

Batch design was performed with Primer 6 (Rozen and Skaletsky 2000), with 110 primers pairs (the expected amplification product length was 100-500bp) were randomly selected and synthesized by Tongyong Biosystems (Anhui) Co., Ltd. Primer screening method is the same as above.

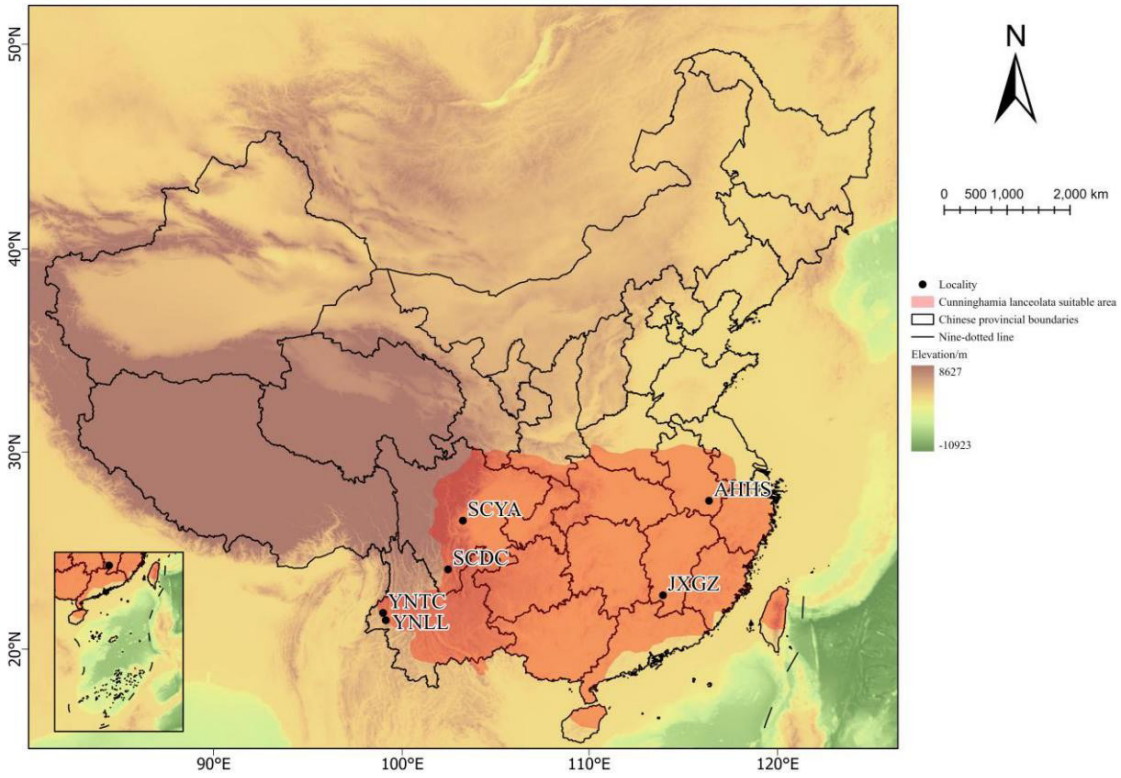

Fig. 1
Geographic locations of six Chinese fir populations.

Table 1
Sources and identities of six Chinese fir populations.

Population[1]	Identity	Population size
AHHS	1~9	9
JXGZ	10~42	33
SCDC	43~53	11
SCYA	54~68	15
YNTC	69~94	26
YNLL	95~112	18

[1]AHHS: Anhui Huangshan; JXGZ: Jiangxi Ganzhou; SCDC: Sichuan Dechang; SCYA: Sichuan Yaan; YNTC: Yunnan Tengchong; YNLL: Yunnan Longling.

Data analysis

(1) Genetic diversity analysis: DataFormater (Fan et al. 2016) was adopted for converting the different software data formats. GenAlEx 6.5 (Peakall and Smouse 2012) was used for calculating the number of alleles (N_a), effective number of alleles (N_e), information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased heterozygosity (uH_e), and fixation index (F) for each geographical population. PowerMarker V3.25 (Liu and Muse 2005) was utilized to determine the polymorphism information content (PIC). To equalize each population's sample size, a population with a larger sample size (e.g., Jiangxi Ganzhou) was randomly selected each time, and the sampling was repeated 10 times to calculate the genetic diversity parameters and estimate their mean value.
(2) Molecular ANOVAs: To assess the genetic variation between and within populations, AMOVA was carried out on genetic variation with Arlequin V3.5.2.2 (Excoffier and Lischer 2010).
(3) Cluster analysis: Popgen 32 (Zang et al. 2021) was employed to generate a genetic similarity matrix between the six Chinese fir geographical populations, while UPGMA in MEGA 11 (Hall 2013) was used for systematic clustering analysis.
(4) Population structure analysis: To explore the Chinese fir geographical population structure, the Bayesian clustering method of Structure V2.3.4 (Pritchard et al. 2000) was applied. The number of groups K was defined as 1–6, the length of the burn-in periods at the beginning of the MCMC was defined as 10,000, and the MCMC after the length of burn-in periods was defined as 100,000. The cluster value was defined by the ΔK method, with the maximum ΔK value being calculated by the Structure Harvester.
(5) Principal coordinate analysis: Using GenAlEx V6.5, the principal coordinate analysis (PCoA) of the 112 individuals was calculated.

Results

Eight pairs of polymorphic primers were screened from the available reported primers.

The initial screening of 91 primer pairs for polymorphism using polyacrylamide gel electrophoresis revealed that 21 primer pairs were polymorphic and had clear amplified bands. However, interpreting these bands was challenging due to their proximity in size (Supplementary Fig. 1, a). While polyacrylamide gel electrophoresis theoretically can differentiate between 1-bp differences, it often comes with significant errors. To overcome this limitation, capillary electrophoresis, which provides more accurate and intuitive results, was employed to further screen the 21 polymorphic primer pairs for the selection of core primers (Supplementary Fig. 1, b). Finally, eight pairs of polymorphic primers were identified, based on criteria such as high amplified band polymorphism, good reproducibility, and absence of shadow peaks (Table 2).

Table 2
Information on 8 pairs of polymorphic EST-SSR primers from Chinese fir.

IDX	NO.	Forward and Reverse primers	Primer sequence(5′-3′)	Fragment length (bp)	Repeat unit	GC (%)	Annealing temperature (°C)
1	P7	F	AGCTCCACTCCGCAAATAGT	151	$(TGAC)_3$	50.0	52.4
	P8	R	CTTGAACTTCAGCCATGGAA			45.0	50.4
2	P13	F	GGCAATGCTTACTGTGGAGA	144	$(TGG)_7$	50.0	52.4
	P14	R	TACAGTCACTGCCTGCTCCT			55.0	54.4
3	P55	F	TACCATCCAAACCCACATCC	316	$(GCAGAG)_3$	50.0	52.4
	P56	R	AACAGTTATGGCCCGTCAAG			50.0	52.4
4	P75	F	ACCCAGATTGCCTCAT	430	$(TGC)_8$	52.6	51.5
	P76	R	CAGTTACAGTAGTGCTTCC			52.6	51.5
5	P101	F	TGTTCTTCATGCCAGCATCT	163	$(GAG)_6$	50.4	50.4
	P102	R	CACCACCTCCTCCTCCATAC			56.5	56.5
6	P105	F	AGAATGGGCAGTCGCTAATC	135	$(TCA)_5$	50.0	52.4
	P106	R	GGGAAATCAGATACGGAGGA			50.0	52.4
7	P127	F	TCCTTTCCTTGTAGCCCATT	154	$(CTTC)_3$	45.0	50.4
	P128	R	CTAAAGCCAGGAGGAAGCTC			55.0	54.4
8	P153	F	AGATAGTTCTCCCAGGCTATCCAAGAT	207-289	$(AG)_{33}$	44.4	59.1
	P154	R	CCACCTACATACAACATAAGCGACCAA			44.4	59.1

Transcriptome data assembly

Sequencing of the Chinese fir transcriptome yielded a total of 42,411,732 raw reads. A total of 41,640,586 clean reads remained after removing adaptors and low - quality reads (including those with >10 % of N removed and those with over 50 % of bases with quality Q ≤ 10). Among these, Q30 high - quality sequences occupied 95.4 %, and the GC content accounted for 43.66 % of the total bases. The base error rate was 0.1 %, suggesting high quality of the Chinese fir data obtained through Illumina NovaSeq 6000 sequencing, which is suitable for subsequent bioinformatics analysis.

By assembling with Trinity software and removing duplicate spliced sequences, a total of 198,389 unigenes were acquired, with a total length of 100,725,460 bp. The minimum splice length was 88 bp, the maximum splice length was 538,899 bp, and the average length was 507.7 bp. The lengths of N75, N50, and N25 were 304, 550, and 1653 bp, respectively. Supplementary Table 2 demonstrates a decreasing trend in the number of unigene sequences with increasing sequence length, indicating a successful sequencing result and assembly of the Chinese fir.

Distribution characteristics and number of each repeat type of transcriptome

The diversity of Chinese fir transcriptome SSR repeat types is depicted in Supplementary Table 1, for the number and distribution ratios. Trinucleotide SSR repeat types were the most abundant (289), comprising 49.5 % of the total EST-SSR loci. Following closely were single and hexanucleotide repeat types, accounting for 24.3 and 21.1 % of the total SSR, respectively. In contrast, tetranucleotide and pentanucleotide repeat types were least frequent, both representing a mere 0.3 % of the total (Supplementary Table 1). When considering base complementarity, the Chinese fir transcriptome exhibited a total of 152 motif types across the six nucleotide repeat types (Supplementary Table 3). The most frequent repeating motif was A/T, constituting 22.60 % of the total EST-SSR. Among the dinucleotide repeats, AG/CT and GA/TC accounted for 1.88 and 1.20 % of the total EST-SSR, respectively. The trinucleotide repeat GAA/TTC was the most common, representing 6.16% of the total EST-SSR. Within the hexanucleotide repeats, the most frequent repeating motif was GAGGAA/TTCCTC and accounted for 1.03 % of the total number of EST-SSR, and was the prevalent motif among the six-nucleotide repeats. Tetranucleotide and pentanucleotide repeats had fewer SSR repeat motifs (Supplementary Table 3).

Transcriptome SSR primer design and screening

Using Primer 6 software, we designed and obtained a total of 523 primer pairs. From these, we randomly selected 110 pairs for synthesis and performed PCR amplification using DNA templates from the six Chinese fir geographical populations. For the primary selection, we used polyacrylamide gel electrophoresis to assess the validity and polymorphism of these 110 primer pairs. As a result, 18 primer pairs were found to be valid and polymorphic (Supplementary Fig. 2, a). Next, we focused on the 18 primer pairs using the initial screening and then used capillary electrophoresis. Finally, we chose 12 primer pairs for further analysis (Supplementary Fig. 2, b) and Table 3).

Transcriptome SSR primer design and screening

Using Primer 6 software, we designed and obtained a total of 523 primer pairs. From these, we randomly selected 110 pairs for synthesis and performed PCR amplification using DNA templates from the six Chinese fir geographical populations. For the primary selection, we used polyacrylamide gel electrophoresis to assess the validity and polymorphism of these 110 primer pairs. As a result, 18 primer pairs were found to be valid and polymorphic (Supplementary Fig. 2, a). Next, we focused on the 18 primer pairs using the initial screening and then used capillary electrophoresis. Finally, we chose 12 primer pairs for further analysis (Supplementary Fig. 2, a) and Table 3).

Table 3
Information on SSR primer pairs for 12 polymorphic pairs of Chinese fir.

IDX	NO.	Forward and Reverse primers	Primer sequence(5′-3′)	Fragment length (bp)	Repeat unit	GC (%)	Annealing temperature (°C)
1	P189	F	TTGTTTGAGAGGCTGGGGTG	136	$(GTTGGT)_4$	55.0	54.4
	P190	R	GGAGCTGCTGTGCGATTCTA			55.0	54.4
2	P213	F	GCGACTTTGCACCTTCCATC	136	$(T)_{10}$	55.0	54.4
	P214	R	TCAGCTCCTCCAGAAACCCT			55.0	54.4
3	P215	F	CCTAGGGTCAGAGGCAGAGG	149	$(CCATCT)_4$	65.0	58.6
	P216	R	TGTAAGTGCCGCGATCAGAA			50.0	52.4
4	P243	F	TTCTGGTTCCCGTTACGGTG	111	$(CACCTG)_4$	55.0	54.4
	P244	R	CATCTCCTCCAGCTGCCTTC			60.0	54.4
5	P265	F	ACATGCTTGTTGACCTCCTCA	142	$(A)_{11}$	47.6	52.4
	P266	R	CTCCTCTGGGCGAACTTCTC			60.0	56.5
6	P299	F	GTAGCCATGGAAAGCCCTGT	172	$(TTCATA)_5$	55.0	54.4
	P300	R	GAGACGGCAAAAGACCAGGA			55.0	54.4
7	P301	F	CAGCAGCAGCAACAACAACA	170	$(TCA)_9$	50.0	52.4
	P302	R	TTGAGACAGCCGGATTGGAC			55.0	54.4
8	P305	F	ATGTGTCGGAGTTGGGGTTC	114	$(GAA)_7$	55.0	54.4
	P306	R	AAGTGAAAGAGAAGGGCGGG			55.0	54.4
9	P351	F	AAAAATCCCTGGCGGTGGAT	127	$(CTT)_6$	50.0	52.4
	P352	R	AACAAACCCCCGTCACTTGT			50.0	52.4
10	P375	F	CCCTGGTCATCATTCAGCCC	169	$(GAG)_6$	60.0	56.5
	P376	R	AGCCTGCCCAAAATCAAAACC			47.6	52.4
11	P379	F	CAGCCGCAAAGAAAGAAGCT	198	$(GGT)_7$	50.0	52.4
	P380	R	GCTTCGTAGCCATGTTTTGGATT			43.5	53.7
12	P383	F	AGATTTCACTGGAGGTGGCG	196	$(GCA)_8$	55.0	54.4
	P384	R	CAAGCACGCATAGCCAAGTG			55.0	54.4

Genetic diversity

By employing 20 pairs of EST-SSR primers to amplify 112 individuals (Table 4), a total of 208 alleles were identified, with N_a variation ranging from 2.0 to 63.0 at different loci and a mean value of 10.4. N_e ranged from 1.041 to 15.719, with a mean of 2.920. *I* ranged from 0.074 to 2.839, with an average of 0.945. H_o and H_e at the species level were 0.319 and 0.481, respectively. The uH_e ranged from 0.033 to 0.959, with a mean value of 0.496. The mean value of *F* is 0.336. The average PIC value was 0.508. Among the loci, one locus (P305P306) had a PIC value < 0.25, indicating low polymorphism. Ten loci with 0.25 ≤ PIC ≤ 0.5 indicated moderate polymorphism, while nine loci had a PIC value > 0.5, indicating high polymorphism. These findings demonstrate the effectiveness of these markers in assessing the genetic diversity of Chinese fir.

Table 5 reveals variations in genetic diversity levels among the six Chinese fir geographical populations. The Yunnan Tengchong population (YNTC) exhibited the highest genetic diversity, whereas the Sichuan Dechang population (SCDC) displayed the lowest genetic diversity.

Table 4
Genetic diversity parameters for 20 SSR loci.

NO.	Locus	N_a	N_e	I	H_o	H_e	uH_e	F	PIC
1	P7P8	3	2.166	0.846	0.000	0.514	0.531	1.000	0.509
2	P13P14	5	2.085	0.815	0.109	0.498	0.515	0.773	0.468
3	P55P56	23	3.258	1.397	0.212	0.627	0.647	0.632	0.714
4	P75P76	6	1.500	0.541	0.338	0.299	0.310	-0.010	0.267
5	P101P102	5	1.602	0.652	0.415	0.367	0.379	-0.073	0.433
6	P105P106	5	2.018	0.781	0.311	0.455	0.470	0.426	0.520
7	P127P128	2	1.586	0.151	0.293	0.344	0.355	0.235	0.335
8	P153P154	63	15.719	2.839	0.569	0.928	0.959	0.386	0.973
9	P189P190	4	2.100	0.798	0.563	0.492	0.507	-0.140	0.492
10	P213P214	9	2.366	1.047	0.538	0.573	0.593	0.067	0.560
11	P215P216	4	1.809	0.720	0.453	0.405	0.419	-0.042	0.386
12	P243P244	4	1.965	0.726	0.393	0.467	0.482	0.252	0.407
13	P265P266	10	3.420	1.341	0.454	0.688	0.711	0.362	0.750
14	P299P300	8	1.537	0.647	0.264	0.326	0.338	0.098	0.331
15	P301P302	7	2.186	0.967	0.217	0.479	0.494	0.451	0.584
16	P305P306	4	1.041	0.074	0.010	0.033	0.033	0.692	0.060
17	P351P352	6	1.441	0.476	0.157	0.256	0.263	0.372	0.296
18	P375P376	22	5.272	1.787	0.548	0.739	0.763	0.276	0.860
19	P379P380	8	1.808	0.808	0.282	0.428	0.442	0.338	0.478
20	P383P384	10	3.526	1.381	0.250	0.692	0.715	0.626	0.737
Mean		10.4	2.920	0.319	0.319	0.481	0.496	0.336	0.508

NO.: Number of the SSR; N_a: Number of alleles; N_e: Effective number of alleles; I: Information index; H_o: Observed heterozygosity; H_e: Expected heterozygosity; uH_e: Unbiased heterozygosity; F: Fixation index; PIC: Polymorphism information content.

Table 5
Genetic parameters of the six Chinese fir geographical populations.

NO.	Population	N_a	N_e	I	H_o	H_e	uH_e	F
1	AHHS	3.500	2.397	0.855	0.289	0.465	0.492	0.383
2	JXGZ	6.650	3.411	1.098	0.347	0.523	0.531	0.302
3	SCDC	3.050	2.070	0.673	0.225	0.368	0.386	0.420
4	SCYA	4.100	2.783	0.915	0.340	0.480	0.496	0.240
5	YNTC	6.200	3.560	1.115	0.356	0.531	0.541	0.264
6	YNLL	5.600	3.300	1.091	0.356	0.517	0.532	0.319
Mean		4.850	2.920	0.958	0.319	0.481	0.496	0.321

NO.: Number; N_a: Number of alleles; N_e: Effective number of alleles; I: Information index; H_o: Observed heterozygosity; H_e: Expected heterozygosity; uH_e: Unbiased expected heterozygosity; F: Fixation index.

Population genetic relationships and genetic structure

AMOVA analysis (Table 6) revealed that within- and among-population molecular variation accounted for 90.58 % and 9.42 % of the total genetic variation, respectively.

Cluster analysis using UPGMA was conducted on the six Chinese fir geographical populations (Fig. 2). At an Euclidean distance of 2.82, the populations were categorized into four groups. The first group comprised the Sichuan Dechang (SCDC), Sichuan Yaan (SCYA), and Anhui Huangshan (AHHS) populations, with increasing distances between the groups indicating a gradual decrease in kinship. The second group consisted of the Yunnan Longling (YNLL) population. The third group included the Yunnan Tengchong (YNTC) population. The fourth group encompassed the Jiangxi Ganzhou (JXGZ) population, which exhibited the greatest distance from the first group and a more distant relationship.

To examine the population genetic structure of Chinese fir, STRUCTURE analysis was employed. The largest ΔK value (Fig. 3, a) was observed when K=4, indicating that the 112 germplasm resources could be classified into four distinct groups. These groups contained 27, 10, 8, and 67 germplasm resources of Chinese fir, respectively (Fig. 3, b). Although some genetic overlap occurred among groups, notable differences were still clear. The green gene pool exclusively comprised the Sichuan Dechang population, which is geographically isolated in the complex terrain of Liangshan Yi Autonomous Prefecture, Sichuan. Consequently, tracing the Chinese fir population in this region is relatively straightforward. In contrast, the genetic components of other gene pools exhibited greater mixing.

PCoA (Fig. 4) corroborated the STRUCTURE classification results, with the Sichuan Dechang population forming a distinct cluster, likely due to its geographical isolation. Given the less pronounced geographical isolation of the other populations, gene flow among Chinese fir populations may be frequent, leading to relatively mixed genetic components.

Table 6
Molecular variance analysis of the 6 Chinese fir geographical populations.

Source of variation	d. f.	Sum of squares	Variance components	Percentage of variation
Among populations	5	120.606	0.531	9.42
Within populations	218	1113.219	5.107	90.58
Total	223	1233.826	5.637	100
F_{ST}=0.09				

d.f. represents degree of freedom.

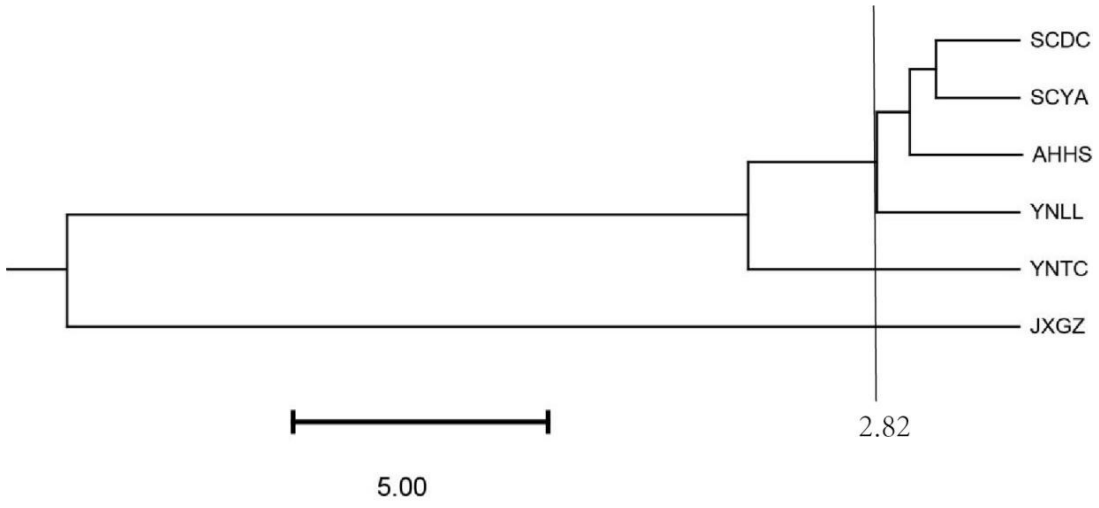

Fig. 2
UPGMA clustering trees for the six Chinese fir geographical populations.

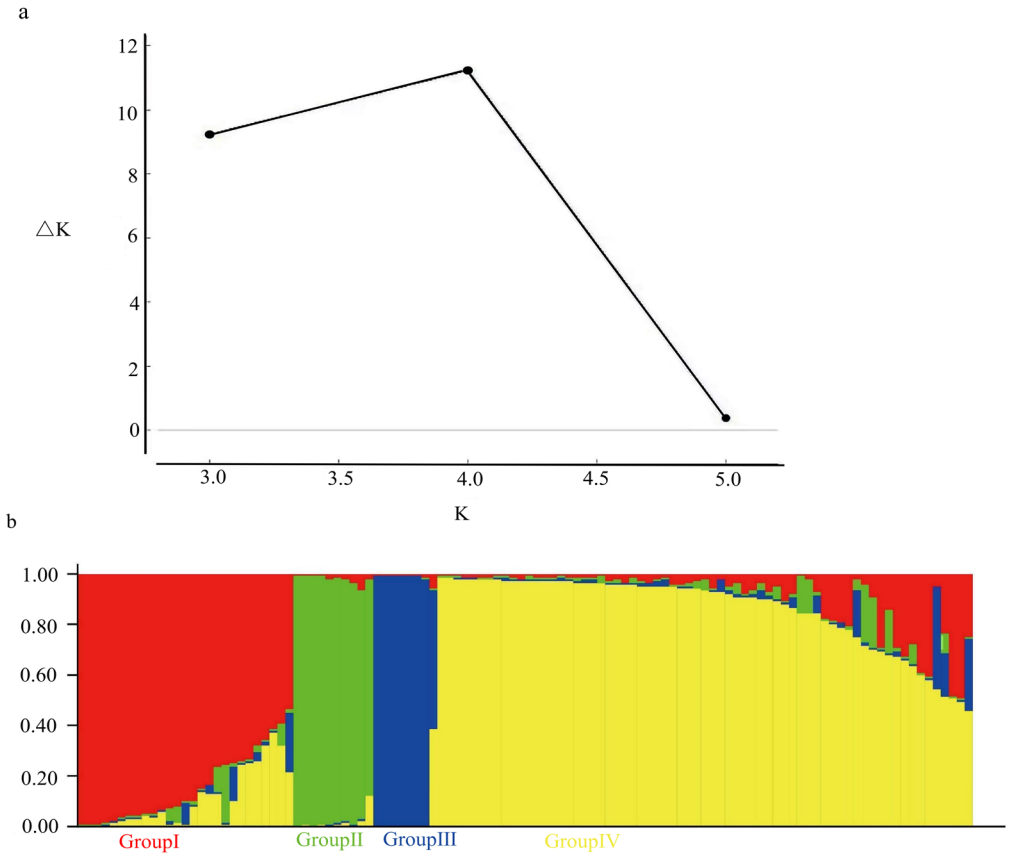

Fig. 3
Population structure of six Chinese fir geographical populations based on STRUCTRUR analysis. a) Estimation of the likelihood of clusters (*k*) for the most appropriate subpopulations (*ΔK*) and b) population structure of 112 in *k* = 4 clusters.

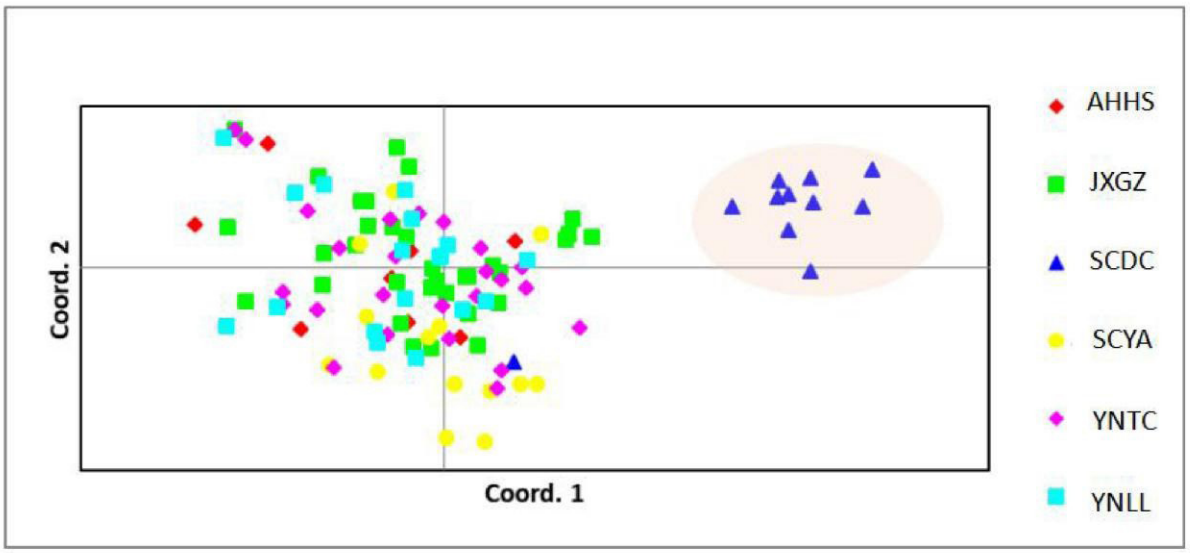

Fig. 4
Population structure of 112 Chinese fir individuals based on principal component analysis. AHHS: Anhui Huangshan; JXGZ: Jiangxi Ganzhou; SCDC: Sichuan Dechang; SCYA: Sichuan Yaan; YNTC: Yunnan Tengchong; YNLL: Yunnan Longling.

Discussion

Screening for SSR molecular markers

SSR primers are conserved and stable, making them generally applicable across relatively closely related species. Therefore, screening SSR primers from closely related species is also a more efficient method. However, the degree of generality varies depending on the species' own characteristics (Chistiakov et al. 2006). *Taiwania cryptomerioides* is closely related to Chinese fir. Zhang et al. (2013) developed EST-SSR markers for both species and found 10 pairs of polymorphic primers, with only one pair specific to *T. cryptomerioides*. This suggests a genetic relationship between the two species, but these polymorphic loci may be conserved genes. Consequently, the development of EST-SSR markers in Chinese fir still relies on its own markers (Zhang et al. 2013). From 270 pairs of published EST-SSR primers for Chinese fir, eight pairs of valid and polymorphic primers were selected, yielding a success rate of 2.96 %. However, due to the limited number of primers for final screening from published Chinese fir EST-SSR primers, transcriptome sequencing was used for additional primer development for secondary supplementation.

Transcriptome sequencing is a fast, efficient, and accurate method for acquiring data, with wide coverage. It plays a crucial role in gene expression analysis, new gene discovery, functional gene localization, and molecular marker development (Gao et al. 2022; Uddin et al. 2022). The length of assembled sequences reflects the transcriptome quality acquired from Illumina sequencing platform (Jia et al. 2016). Compared to other gymnosperms using the same platform, Chinese fir had longer average unigenes (507.7 bp). Although lower than *Pinus dabeshanensis* (Xiang et al. 2015)(1074 bp) and *Platycladus orientalis* (Zhou et al. 2019) (908.4 bp), it was higher than *Pinus pinaster* (Cañas et al. 2015) (495 bp), indicating high sequence assembly quality. These sequences provide valuable information for molecular marker development and genetic diversity analysis of Chinese fir (Canales et al. 2014; Gao et al. 2022; He et al. 2015; Liu et al. 2017). Trinucleotide repeats were the most abundant in the Chinese fir transcriptome, confirming previous reports on other plant species such as *Pinus pinaster*, *Pinus tabuliformis*, and *Pinus dabeshanensis* (Canales et al. 2014; Lawson and Zhang 2006; Niu et al. 2013; Wang et al. 2010; Xiang et al. 2015). Among trinucleotide repeats, GGA/TTC was the most common, serving as a reference for SSR primer screening.

Genetic diversity of geographical populations

Genetic diversity means the biological variation that occurs within and among populations (Nasiri et al. 2023; Yan et al. 2022; Zhu et al. 2023). Heterozygosity is a crucial measure of population genetic diversity (Jing et al. 2023). In this study, observed (H_o) and expected (H_e) heterozygosity for Chinese fir geographical population were 0.319 and 0.481, respectively, which are consistent with previous analyses of Chinese fir genetic diversity that used EST-SSR markers. Hu et al. (2016) reported an average heterozygosity of 0.384 based on 11 Chinese fir families. Comparatively, *Taxodium distichum* (L.) Rich. (Duan et al. 2020), another Taxodiaceae species, exhibited Ho and He values of 0.381 and 0.245, respectively. Similarly, *Metasequoia glyptostroboides* (Jin et al. 2015), also a Taxodiaceae species, showed an H_o of 0.376 and H_e of 0.452. These findings suggest that the genetic diversity of Chinese fir is moderate compared to other Taxodiaceae species, indicating that a relatively larger rich collection of Chinese fir is needed to capture adequate genetic diversity. However, in order to ensure sufficient material support for the development and utilization of this resource and enhance the preservation of Chinese fir germplasm diversity, it is necessary to expand resource collection and analysis to more regions.

Population genetic structure

Population genetic structure means the spatial and temporal patterns of genetic diversity within and among populations. It is influenced by factors including habitat fragmentation and breeding systems. FST, an important indicator of population genetic structure (Hamrick and Godt 1996), can be adopted for evaluating the degree of genetic differentiation among populations. Additionally, in our study, we also conducted AMOVA analysis which showed that the Chinese fir geographical populations exhibited a moderate level of genetic differentiation (FST = 0.09).The main source of genetic variation was found within populations, which aligns with the genetic pattern observed in other species like *Ginkgo biloba* (Zhou et al. 2020) and *Pinus massoniana* (Yang et al. 2021).These species also exhibit high genetic variation within populations and low genetic variation among populations (Lu et al. 2022; Vieira Ldo et al. 2014; Yang et al. 2021; Zhou et al. 2020). Based on these findings, it is important to focus on the selection and improvement of plus tree individuals in Chinese fir breeding.

This study, through multiple population genetics methods, reveals that the Dechang, Sichuan (SCDC) population exhibits significant genetic uniqueness. Both STRUCTURE analysis and principal coordinate analysis (PCoA) results consistently indicate that the SCDC population is independent of all other populations: it is identified as an independent genetic cluster (K) in STRUCTURE and forms an isolated point group far from other populations in PCoA. This phenomenon is attributed to its unique allele frequencies, which are likely the result of strong natural selection pressure exerted by the special local climate in the Hengduan Mountains, leading to rapid genetic differentiation in recent times rather than due to ancient isolation events (White et al. 2007).

Infusion populations

To maintain the original breeding objectives while introducing infusion populations for potentially gaining genetic improvements on certain traits, it is crucial to thoroughly assess the traits of the introduced materials through genetic evaluation before incorporating them into the new breeding cycle. Infusion populations, in the context of a genetic improvement project, refer to the new accessions and individuals that were introduced from other breeding populations or external

populations (White et al. 2007). The purpose of introducing these external individuals/populations is to maintain genetic diversity in advanced-generation breeding populations (Kang 2019). In order to avoid loss of genetic gain after the introduction of external individuals/populations, a comprehensive genetic evaluation of the gain on certain traits of introduced materials should be conducted before their integration into the new breeding cycle (White et al. 2007). The average heritability values for tree height, diameter at breast height (DBH), and wood density in the second-generation breeding population of Chinese fir were 0.37, 0.33, and 0.17, respectively (Bian et al. 2014), resulting in a significant increase in average realistic genetic gain. The breeding goal for Chinese fir has evolved from fast growth and high yield in the initial round to fast growth, high quality, and strong biotic resistance in the fourth round (Jing et al. 2023). In this study, the molecular evaluation of infusion populations of Chinese fir aids in the preliminary selection of infusion populations that align with the breeding objectives of the fourth round. After genetic evaluation, infusion populations may be introduced to enhance the target breeding traits and preserve the genetic diversity of advanced-generation breeding populations.

Data Availability Statement: The original transcriptome data which can support the present study will be provided by the author unreservedly. The repository is NCBI; the login number is PRJNA1012109.

Author Contributions: SY, LB, YAE and JS were responsible for conception and design of the study. SY, RZ, SS, LZ, XYZ and XZ collected the preliminary data. SY and LB carried out the statistical analysis. SY wrote the manuscript. LB, ZC and YAE contributed to the writing and revisions. All authors contributed to the article and approved.

Funding: The current work was supported by the National Key Research and Development Program of China (2022YFD2200201), the National Natural Science Foundation of China (32171818), Fujian Province Science and Technology Research Funding on the Fourth Tree Breeding Program of Chinese fir (ZMGG-0801), Forestry Science and Technology Promotion Demonstration Project (2022XT17) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Conflict of Interest: The authors claim that they show no conflict of interest.

Acknowledgements: We wish to express our appreciation to Lejin Fang (Huangshan University), Bin Ou (Jiangxi Environmental Engineering Vocational College), Zhen Huang (Sichuan Academy of Forestry), Linyuan Fan (Yunnan Provincial Forest Seedling Work Station) for helping us collect germplasm resources.

References

Bian L, Shi J, Zheng R, Chen J, Wu HX (2014) Genetic parameters and genotype–environment interactions of Chinese fir (*Cunninghamia lanceolata*) in Fujian Province. Can J Forest Res 44: 582-592. https://doi.org/10.1139/cjfr-2013-0427

Canales J, Bautista R, Label P et al. (2014) De novo assembly of maritime pine transcriptome: implications for forest breeding and biotechnology. Plant Biotechnol J 12: 286-299. https://doi.org/10.1111/pbi.12136

Cañas RA, Feito I, Fuente-Maqueda JF, Ávila C, Majada J, Cánovas FM (2015) Transcriptome-wide analysis supports environmental adaptations of two *Pinus pinaster* populations from contrasting habitats. BMC Genomics 16: 909. https://doi.org/10.1186/s12864-015-2177-x

Cao S, Deng H, Zhao Y et al. (2022) Metabolite Profiling and Transcriptome Analysis Unveil the Mechanisms of Red-Heart Chinese Fir [*Cunninghamia lanceolata* (Lamb.) Hook] Heartwood Coloration. Front Plant Sci 13: 854716. https://doi.org/10.3389/fpls.2022.854716

Chistiakov DA, Hellemans B, Volckaert FAM (2006) Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. Aquaculture 255: 1-29. https://doi.org/10.1016/j.aquaculture.2005.11.031

Duan H, Xu J, Wang Z (2020) Construction of SSR fingerprint and analysis on genetic relationship of Taxodium species and their hybrid progenies T. 'Zhongshanshan' series cultivars. J Plant Resour Environ 29: 11-18.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10: 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fan WQ, Gai HM, Sun X, Yang AG, Zhang ZF, Ren M (2016) DataFormater, a software for SSR data formatting to develop population genetics analysis. Mol Plant Breed 14: 265-270. https://doi.org/10.5376/mpb.cn.2016.14.0005

Gao C, Ren L, Wang M, Wang Z, Fu N, Wang H, Shi J (2022) Full-Length Transcriptome Sequencing-Based Analysis of *Pinus sylvestris* var. mongolica in Response to Sirex noctilio Venom. Insects 13. https://doi.org/10.3390/insects13040338

Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol 30: 1229-1235. https://doi.org/10.1093/molbev/mst012

Hamrick JL, Godt MW (1996) Effects of life history traits on genetic diversity in plant species. Philosoph Trans R Soc London. Ser B Biol Sci 351: 1291-1298. https://doi.org/10.1098/rstb.1996.0112

He B, Gu Y, Xu M, Wang J, Cao F, Xu LA (2015) Transcriptome analysis of Ginkgo biloba kernels. Front Plant Sci 6: 819. https://doi.org/10.3389/fpls.2015.00819

Huang R, Zeng W, Deng H, Hu D, Wang R, Zheng H (2022) Inbreeding in Chinese Fir: Insight into the Rare Self-Fertilizing Event from a Genetic View. Genes 13. https://doi.org/10.3390/genes13112105

Ji Y, Zhu L, Hao Z et al. (2022) Exploring the *Cunninghamia lanceolata* (Lamb.) Hook Genome by BAC Sequencing. Front Bioeng Biotechnol 10: 854130. https://doi.org/10.3389/fbioe.2022.854130

Jia X, Deng Y, Sun X, Liang L, Su J (2016) De novo assembly of the transcriptome of *Neottopteris nidus* using Illumina paired-end sequencing and development of EST-SSR markers. Mol Breeding 36: 1-12. https://doi.org/10.1007/s11032-016-0519-2

Jin Y, Bi Q, Guan W, Mao JF (2015) Development of 23 novel polymorphic EST-SSR markers for the endangered relict conifer *Metasequoia glyptostroboides*. Appl Plant Sci 3. https://doi.org/10.3732/apps.1500038

Jing Y, Bian L, Zhang X et al. (2023) Genetic diversity and structure of the 4(th) cycle breeding population of Chinese fir (*Cunninghamia lanceolata* (lamb.) hook). Front Plant Sci 14: 1106615. https://doi.org/10.3389/fpls.2023.1106615

Kang X (2019) Thoughts on tree breeding strategies. J Beijing Forestry Univ 41: 15-22. https://doi.org/10.12171/j.1000-1522.20190412

Lawson MJ, Zhang L (2006) Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. Genome Biol 7: R14. https://doi.org/10.1186/gb-2006-7-2-r14

Li N, Yang Y, Xu F et al. (2022) Genetic Diversity and Population Structure Analysis of *Castanopsis hystrix* and Construction of a Core Collection Using Phenotypic Traits and Molecular Markers. Genes 13. https://doi.org/10.3390/genes13122383

Liu H, Yan Z, Xu H, Li C, Fan Q, Liao W, Liao B (2017) Development and characterization of EST-SSR markers via transcriptome sequencing in *Brainea insignis* (Aspleniaceae s.l.). Appl Plant Sci 5. https://doi.org/10.3732/apps.1700067

Liu J, Gao S, Xu Y et al. (2022) Genetic Diversity Analysis of Sapindus in China and Extraction of a Core Germplasm Collection Using EST-SSR Markers. Front Plant Sci 13: 857993. https://doi.org/10.3389/fpls.2022.857993

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128-2129. https://doi.org/10.1093/bioinformatics/bti282

Lu D, Huang H, Wang A, Zhang G (2022) Genetic Evaluation of *Juniperus sabina* L. (Cupressaceae) in Arid and Semi-Arid Regions of China Based on SSR Markers. Forests 13: 231. https://doi.org/10.3390/f13020231

Nasiri M, Yousefzadeh H, Shirvany A, Etemad V, Espahbodi K, Amirchakhmaghi N, Rajora OP (2023) Effects of fifty years of shelterwood harvesting on genetic diversity and population structure of Oriental beech (*Fagus orientalis* L.) in the relict Hyrcanian forest. Forest Ecol Manage 529: 120623. https://doi.org/10.1016/j.foreco.2022.120623

Niu SH, Li ZX, Yuan HW, Chen XY, Li Y, Li W (2013) Transcriptome characterisation of *Pinus tabuliformis* and evolution of genes in the *Pinus phylogeny*. BMC Genomics 14: 263. https://doi.org/10.1186/1471-2164-14-263

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28: 2537-2539. https://doi.org/10.1093/bioinformatics/bts460

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959. https://doi.org/10.1093/genetics/155.2.945

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132: 365-386. https://doi.org/10.1385/1-59259-192-2:365

Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. Theor Appl Genet 100: 723-726. https://doi.org/10.1007/s001220051344

Shi Z, Zhao W, Li Z, Kang D, Ai P, Ding H, Wang Z (2022) Development and validation of SSR markers related to flower color based on full-length transcriptome sequencing in Chrysanthemum. Sci Rep 12: 22310. https://doi.org/10.1038/s41598-022-26664-3

Uddin N, Muhammad N, Ali SS, Ullah R, Bari A, Hussain H, Zhu D (2023) Characterization of the Genetic Variability within *Ziziphus nummularia* Genotypes by Phenotypic Traits and SSR Markers with Special Reference to Geographic Distribution. Genes 14. https://doi.org/10.3390/genes14010155

Uddin S, Munir MZ, Gull S et al. (2022) Transcriptome Profiling Reveals Role of MicroRNAs and Their Targeted Genes during Adventitious Root Formation in Dark-Pretreated Micro-Shoot Cuttings of Tetraploid *Robinia pseudoacacia* L. Genes 13. https://doi.org/10.3390/genes13030441

Vieira Ldo N, Faoro H, Rogalski M et al. (2014) The complete chloroplast genome sequence of *Podocarpus lambertii*: genome structure, evolutionary aspects, gene content and SSR detection. PLoS One 9: e90618. https://doi.org/10.1371/journal.pone.0090618

Wang Z, Fang B, Chen J et al. (2010) De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (Ipomoea batatas). BMC Genomics 11: 726. https://doi.org/10.1186/1471-2164-11-726

White TL, Adams WT, Neale DB (2007) Forest genetics. CABI, Britain, p 261.

Xiang W, Xu L, Lei P et al. (2022) Rotation age extension synergistically increases ecosystem carbon storage and timber production of Chinese fir plantations in southern China. J Environ Manage 317: 115426. https://doi.org/10.1016/j.jenvman.2022.115426

Xiang X, Zhang Z, Wang Z, Zhang X, Wu G (2015) Transcriptome sequencing and development of EST-SSR markers in *Pinus dabeshanensis*, an endangered conifer endemic to China. Mol Breeding 35: 1-10. https://doi.org/10.1007/s11032-015-0351-0

Xu Q, Yin Y, Xu Z, Zhang X, Rong J (2011) The direction selection of superior clones of *Cunninghamia lanceolata* in hunan province. Hunan Forestry Sci Technol 38: 46-49. https://doi.org/10.3969/j.issn.1003-5710.2011.06.014

Xu R, Liu J, Wang LY, Yan Y, Ma XQ, Li M (2022) Analysis of root and leaf functional traits and C, N, P stoichiometry of *Cunninghamia lanceolate* from different provenances. Acta Ecol 42: 6311-6319.

Yan P, Xie Z, Feng K, Qiu X, Zhang L, Zhang H (2022) Genetic diversity analysis and fingerprint construction of Korean pine (*Pinus koraiensis*) clonal seed orchard. Front Plant Sci 13: 1079571. https://doi.org/10.3389/fpls.2022.1079571

Yang H, Liu Q, Zhang R, Zhou Z (2021) Genetic diversity of second generation-parental germplasm of masson pine revealed by SSR markers and establishment of a core germplasm collection. Scand J Forest Res 36: 524-531. https://doi.org/10.1080/02827581.2021.1981432

Zang M, Li X, Fang Y (2021) Genetic diversity analysis among natural populations of *Quercus fabri* based on SSR markers. J Nanjing Forestry Univ (Nat Sci Ed) 45: 63-69.

Zhang S, Huang H, Lin E, Tong Z (2013) Development and application of EST-SSR markers for *Cunninghamia lanceolata* and *Taiwania cryptomerioides*. Sci Silvae Sin 49: 173-180.

Zhou Q, Mu K, Ni Z, Liu X, Li Y, Xu L (2020) Analysis of genetic diversity of ancient Ginkgo populations using SSR markers. Ind Crops Prod 145: 111942. https://doi.org/10.1016/j.indcrop.2019.111942

Zhou W, Chen Q, Wang XB, Hughes TO, Liu JJ, Zhang X (2019) De novo assembly of the *Platycladus orientalis* (L.) Franco transcriptome provides insight into the development and pollination mechanism of female cone based on RNA-Seq data. Sci Rep 9: 10191. https://doi.org/10.1038/s41598-019-46696-6

Zhu X, Zou R, Tang J, Deng L, Wei X (2023) Genetic diversity variation during the natural regeneration of *Vatica guangxiensis*, an endangered tree species with extremely small populations. Glob Ecol Conserv 42: e02400. https://doi.org/10.1016/j.gecco.2023.e02400