forests

MDPI

Article

Assessing the Genetic Diversity and Genealogical Reconstruction of Cypress (*Cupressus funebris* Endl.) Breeding Parents Using SSR Markers

Hanbo Yang [1], Rui Zhang [1], Guoqing Jin [1,*], Zhongping Feng [2] and Zhichun Zhou [1]

[1] Institute of Subtropical Forestry, Chinese Academy of Forestry, Zhejiang Provincial Key Laboratory of Tree Breeding, Fuyang 311400, China; yanghanbo6@outlook.com (H.Y.); zhongli6@126.com (R.Z.); zhichun_zhou@yeah.net (Z.Z.)

[2] Laoshan Forest Farm of Chun'an, Chun'an 311700, China; zhongping_feng@163.com

* Correspondence: jin_guoqing@126.com; Tel.: +86-182-8136-6168

Academic Editor: Timothy A. Martin

Received: 21 April 2016; Accepted: 23 July 2016; Published: 28 July 2016

Abstract: To identify genetic diversity, genetic structure and the relationship among accessions, and further establish a core collection for the long-term breeding of cypress (*Cupressus funebris* Endl.), the genealogy of breeding parents was reconstructed using simple sequence repeat (SSR) molecular markers. Seventeen SSR markers were used to detect molecular polymorphisms among 290 cypress accessions from five provinces and 53 accessions with unknown origin in China. A total of 92 alleles (*Na*) were detected with 5.412 alleles per locus and an average polymorphism information content (*PIC*) of 0.593. The haplotype diversity (*H*) ranged from 0.021 to 0.832, with an average of 0.406. The number of alleles (*Na*) and the effective number of alleles (*Ne*) ranged from 4.294 to 5.176 and from 2.488 to 2.817 among five populations, respectively. The pairwise population matrix of Nei's genetic distance ranged from 0.008 to 0.023. Based on the results of unweighted pair group method average (UPGMA) cluster and population structure analyses, 343 breeding parents were divided into two major groups. Lower genetic differentiation coefficients and closer genetic relationships were observed among cypress breeding parents, suggesting that the genetic basis was narrow, and the genetic relationship was confused by frequent introduction and wide cultivation. Moreover, we reconstructed the genealogy between breeding parents and 30 accessions of breeding parents from an identified core collection. According to the present study, not only geographic origin but also the relationship of the individuals should be considered in future crossbreeding work.

Keywords: *Cupressus funebris* Endl.; SSR; genealogical reconstruction; genetic diversity; core collection

1. Introduction

Cupressus funebris Endl. (cypress) is an endemic conifer tree species in southern China and exists in a wide range of habitats ranging from relatively fertile sites to nutrient-deficient bogs up to an altitude of approximately 2000 m [1,2]. This species is beautiful in gross morphology and of extreme importance in horticulture ecology and economic profit (for example, wood and incense) [3]. The field survey shows that the long cultivation history of cypress is blurred in the boundary between wild and cultivated accessions [1]. The large proportion of artificial forests that have emerged in the distribution regions in southern China could complicate genealogy and further limit breeding work [2].

The genetic diversity and evolutionary relationship can be evaluated through morphological, cytological, biochemical and DNA molecular markers [4]. However, morphological analysis is labor-intensive and lacking in definition and objectives [5]. In addition, an analysis based on allozyme cannot be applied because of the limited polymorphisms of these variants [6]. The development of DNA-based markers has facilitated the monitoring and manipulation of genetic variation, and the

results have been less affected by environment [7]. The intraspecific or interspecific genealogies can also be accurately inferred through DNA markers, although the field observation data are unavailable [8–10]. In woody perennial species, high management costs and high vulnerability to environmental conditions exist in the conservation and utilization of breeding resources. Moreover, the population structure in most cultivated woody perennial species differs from that of annual species [11]. The advantages of developing core collections could resolve these problems [12], and DNA markers can also be used to successfully construct core collections in crops and woody species [11], such as rice, olive, soybean, tea, etc. [13–16].

Simple sequence repeat (SSR), a type of microsatellite marker, is particularly useful for spectrum reconstruction and genetic diversity and phylogenic analyses of plants because of the high polymorphism and codominance of these markers [17]. In *Cupressus* species, including *Cupressus sempervirens* L., *Cupressus sempervirens* var. *horizontalis*, *Cupressus chengiana* S. Y. Hu, *Cupressus duclouxiana* Hickel, *Cupressus gigantea* Cheng et L. K. Fu and *Cupressus funebris* Endl. [2,6,18,19], previous studies have been conducted using SSR markers to examine genetic diversity and population structure. Shahroodian et al. used SSR markers to determine the genetic variation of *C.sempervirens* var. *horizontalis* at three sites, which provided 81.41% of total variance in this species [19]. Bagnoli et al. provided an interpretation of the history of *Cupressus* distribution, characterized by ancient eastern populations (Turkey and Greek islands) and depauperate populations in the central Mediterranean age using SSR markers [18]. Overall, these previous studies confirm that SSR markers are a powerful tool for elucidating genetic diversity and revealing the phylogenetic relationships of *Cupressus* species. The lack of information for the basic population genetics of cypress represents a serious impediment for establishing genetic management plans. The objectives of this study were to: (i) examine the level of genetic diversity; (ii) reconstruct the genealogy; and (iii) establish a core collection of cypress. This study will enable the calculation of genetic diversity within various species of cypress and establish a core collection that has not been previously reported.

2. Materials and Methods

2.1. Plant Materials

The breeding parents of cypress were derived from five provinces of China (50 accessions from Sichuan Province, 35 accessions from Guizhou Province, 53 accessions from Hubei Province, 86 accessions from Zhejiang Province, and 66 accessions from Chongqing Province). Moreover, additional accessions (53) were initially acquired from other collections around these provinces, and their original geographic origins and populations are unknown. All 343 breeding parents were cultivated at Laoshan Forestry Farm of Chun'an (29°32′ N, 119°03′ E) in 1981, 2010, 2011, 2012 and 2014 (Table 1). Young needles were collected from 343 individual breeding parents, immediately frozen in liquid nitrogen and subsequently stored at −80 °C prior to DNA isolation.

Table 1. The origin areas of cypress breeding parents.

Origin	Numbers	Latitude	Longitude	Collection Time
unknown	53	-	-	1981
Chongqing Province, China	66	28°88′ N	106°68′ E	2012
Zhejiang Province, China	86	30°26′ N	120°22′ E	2011
Hubei Province, China	53	32°57′ N	111°19′ E	2014
Guizhou Province, China	35	26°47′ N	107°62′ E	2014
Sichuan Province, China	50	30°37′ N	102°90′ E	2010
Total	343			

2.2. DNA Isolation and PCR Amplification

DNA was extracted from young leaves using a standard CTAB protocol [20]. The DNA concentration was measured using a NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA), and the samples were diluted with TE buffer to 15 ng/μL for PCR amplification.

The sequences of seven primer pairs for cypress (Cupressaceae) were obtained from Li et al. [21], nine primer pairs for *Cupressus chengiana* S. Y. Hu (Cupressaceae) were obtained from Xu et al. [22] and one primer pair for *Cupressus sempervirens* L. (Cupressaceae) was obtained from Sebastiani et al. [23] (Table S1). The PCR reaction included 15 ng DNA, 10× PCR buffer (Takara), 0.3 mM dNTPs, 0.8 U Taq DNA polymerase (Takara), and 0.3 μM primers. The PCR reactions were performed on a Takara PCR Thermal Cycler Dice® Touch (Takara, Dalian, China) under the following conditions: denaturation at 94 °C for 5 min, 36 amplification cycles at 94 °C for 40 s, 45–60 °C for 40 s and 72 °C for 80 s, with a final elongation step at 72 °C for 7 min. The amplified products were separated on 10% polyacrylamide gel and visualized through silver staining. According to the band migration, the alleles were numbered as A, B, C, D . . . based on size.

2.3. Statistical Analysis of Genetic Diversity

Cervus version 2.0 [24] was used to calculate the polymorphism information content (*PIC*) and detect the presence of null alleles. The observed and expected heterozygosity (*Ho* and *He*), haplotype diversity (*H*), Shannon's information index (*I*), inbreeding coefficient (*Fis*), fixation index (*Fst*), the observed number of alleles (*Na*), and effective number of alleles (*Ne*) were estimated, and the Hardy–Weinberg equilibrium was tested in GenAlEx 6.5 [25]. The probability of identity (*PI*) was computed according to Paetkau et al. [26], $PI = 2(\sum p_i^2)^2 - \sum p_i^4$, where P_i is the frequency of the ith allele, using GenAlEx 6.5. The genetic diversity parameters of each population were also calculated, separately. The average value of 17 SSR markers served as the parameter for the genetic diversity of the population. To validate the genetic variation among and within populations, a hierarchical analysis of molecular variance (AMOVA) using ARLEQUIN v.3.1 was performed [27].

2.4. Clustering Analysis

The collected data were organized in a square matrix in which the codes "0" and "1" were used for the absence and presence of alleles, respectively (the code for missing data was 9). The genetic distance between breeding parents was subsequently calculated using the DICE coefficient [28] through the SimQual Procedure of NTSYSpc 2.0 [29]. A dendrogram was constructed using unweighted pair group method average (UPGMA) clustering and drawn using NTSYSpc 2.0 [29].

2.5. Population Structure

The software program Structure 2.0 [30] was utilized to infer the population structure and assign individuals to populations based on SSR genotypes. The population structure using model-based clustering through a Bayesian approach was used to identify clusters based on Hardy–Weinberg equilibrium and linkage equilibrium. Multiple runs of population structure were performed, and *K* (the number of populations) ranged from 1 to 20. The burn-in time and replication number were set to 10,000 for each run, and each run was replicated ten times. An ad hoc statistic ΔK based on the rate of change in the log probability of data between successive K values, according to Evanno et al. [31], was calculated using the Structure Harvester v.0.9.93 website [32], and this value was used to estimate the most likely number of clusters (K). The software CLUMPP v.1.1.2 [33] was used to account for label switching and compute the symmetric similarity coefficient between pairs of runs using the Greedy algorithm to identify potential distinct modes among the results of independent replicate runs and compute the average individual *Q* values over replicated runs showing a similar mode. The output of CLUMPP v.1.1.2 was used directly as the input for the program DISTRUCT for cluster visualization [34]. For this approach, the ΔK method was used in adjudication for the best *K*, and the individuals with a proportional membership $Q > 0.8$ in the primary population were considered in subsequent analyses. To examine the effect of geographic distance on genetic structure, correlations between the pairwise genetic distances among five breeding parent populations were calculated according to the latitude and longitude of each site, and these distances were evaluated using the Mantel test implemented through GenAlEx 6.5.

2.6. Construction of Genetic Core Collection

Patterns of linkage disequilibrium (LD) were investigated among SSR loci. The squared correlation of allele frequencies R^2 [35] was used to test the LD between pairs of SSR markers with 10^5 permutations using the software package TASSEL version 2.0.1 (http://www.maizegenetics.net/). To construct the genetic core collection, we used the "M" (maximization) method, according to Schoen et al. [36], implemented through Core Finder software [37]. The M strategy uses a specific combination of accessions to maximize the number of observed alleles at each marker locus and the cofinder through iterative procedures to select samples with the highest allelic diversity. The efficiency of the strategy was assessed after comparing the parameter of genetic diversity using the cofinder to select the parameters of genetic diversity in initial collections.

3. Results

3.1. Subsection

3.1.1. SSR Markers

The total number of alleles in breeding parents (290 cultivars from five provinces in China) was 92 with a mean number of alleles per locus of 5.412, ranging between a minimum of 2 (CF02) and a maximum of 10 different alleles (CF11, Table 2). Only 69 alleles were detected in the 53 accessions of unknown origin. The mean effective number of alleles (*N*e) was 2.698 (range from 1.021 to 6.298). The highest *PI* value (0.96) was at CF02 and the lowest *PI* (0.04) was at CF01. The cumulative *PI* was 2.37×10^{-11}, indicating the unlikelihood of two randomly selected individuals having the same SSR profile. The haplotype diversity (*H*) ranged between 0.021 and 0.832, with an average of 0.406. In general, allele frequencies were unevenly (either low or high) distributed within the investigated loci and most were present at low frequencies, as approximately 55.43% alleles were present at frequencies of less than 0.10 (Figure S1). The percentage of heterozygotes per marker detected for the breeding parents in the present study ranged from 0.7% for marker CF02 to 96.2% for marker CYP101. The unbiased expected heterozygosity (*uH*e) ranged from 0.021 (CF02) to 0.843 (CF01) with an average of 0.528. The Shannon's information index (*I*) ranged from 0.058 (CF02) to 1.977 (CF01) with an average of 1.035. The *Fis* value was negative for 290 cultivars (−0.037), indicating a slight excess of heterozygosity. The *PIC* values for the 17 markers ranged from 0.370 (CYP84) to 0.802 (CF04), with a mean value of 0.593.

3.1.2. Genetic Relationship Analysis

The unweighted pair group method average (UPGMA) tree clustered the majority of the breeding parents into two major groups at a DICE coefficient of 0.65 (Figure S2). Group 1 was the largest cluster with two subgroups (pink and blue). One subgroup (pink) contained 93 breeding parents, comprising breeding parents (54.8%) from accessions with unknown origin (51), 13.2% (12) breeding parents from Sichuan, 5.4% (5) breeding parents from Guizhou, 18.3% (17) breeding parents from Zhejiang, 3.2% (3) breeding parents from Chongqing, and 5.4% (5) breeding parents from Hubei. Another subgroup (blue) comprised 32, 17, 23, 10 and 15 breeding parents from Chongqing, Zhejiang, Hubei, Guizhou and Sichuan, respectively. Group 2 (light blue) predominantly (150) comprised 31 (47.0%) breeding parents from Chongqing, 50 (58.1%) breeding parents from Zhejiang, 24 (45.3%) breeding parents from Hubei, 20 (57.1%) breeding parents from Guizhou, 23 (46.0%) breeding parents from Sichuan, and only two accessions from groups with unknown origin. Three accessions (Hubei 44, Zjejiang 21 and Zhejiang 49) were not grouped into separate clusters and had a relatively far genetic relationship with the other accessions in the two major groups.

Table 2. Genetic diversity for 17 microsatellite loci analyzed in breeding parents.

	Locus	*Na*	*Ne*	*PIC*	*H*	*I*	*He*	*Ho*	*uHe*	*Fis*	*PI*
Breeding parents from the five provinces (*n* = 290)	CF01 [a]	9	6.298	0.689	0.832	1.977	0.841	0.407	0.843	0.513 ***	0.04
	CF02 [a]	2	1.021	0.682	0.021	0.058	0.021	0.007	0.021	0.702 ***	0.96
	CF03	9	4.191	0.64	0.682	1.685	0.761	0.719	0.763	0.048 [ns]	0.09
	CF04	8	4.679	0.802	0.693	1.740	0.786	0.781	0.788	−0.014 [ns]	0.07
	CF09	7	3.947	0.772	0.618	1.544	0.747	0.736	0.748	0.002 [ns]	0.11
	CF11	10	3.007	0.722	0.614	1.506	0.667	0.866	0.669	−0.293 **	0.14
	CF12 [a]	4	1.479	0.59	0.308	0.639	0.324	0.194	0.324	0.369 ***	0.48
	CUC1	3	1.575	0.572	0.275	0.608	0.365	0.431	0.366	−0.202 **	0.46
	CUC2	4	1.285	0.393	0.210	0.478	0.222	0.236	0.222	−0.073 [ns]	0.62
	CUC3 [a]	4	1.843	0.424	0.246	0.707	0.457	0.664	0.458	−0.474 ***	0.39
	CUC4	6	2.732	0.547	0.382	1.235	0.634	0.853	0.635	−0.363 **	0.19
	CUC6 [a]	6	1.428	0.48	0.292	0.636	0.300	0.172	0.300	0.414 ***	0.51
	CUC7 [a]	4	2.064	0.575	0.149	0.784	0.515	0.885	0.516	−0.730 ***	0.35
	CUC8	5	2.995	0.61	0.609	1.207	0.666	0.591	0.667	0.118 [ns]	0.18
	CYP52	4	3.554	0.739	0.527	1.322	0.719	0.951	0.720	−0.321 **	0.13
	CYP84 [a]	4	1.688	0.37	0.392	0.701	0.407	0.192	0.408	0.530 ***	0.41
	CYP101 [a]	3	2.075	0.471	0.057	0.774	0.518	0.962	0.519	−0.856 ***	0.35
	Mean	5.412	2.698	0.593	0.406	1.035	0.527	0.568	0.528	−0.037	-
	SD	0.582	0.350	0.133	0.046	0.129	0.056	0.076	0.056	0.108	-
	Total	92	45.859	-	-	-	-	-	-	-	2.37×10^{-11}
Breeding parents with unknown origin (*n* = 53)	CF01 [a]	7	4.886	0.674	0.788	1.731	0.795	0.286	0.805	0.641 ***	0.07
	CF02 [a]	1	1.000	0.65	0.000	0.000	0.000	0.000	0.000	-	1.00
	CF03 [a]	5	2.136	0.594	0.501	0.955	0.532	0.368	0.546	0.307 ***	0.29
	CF04	8	4.723	0.786	0.699	1.770	0.788	0.913	0.806	−0.158 **	0.07
	CF09	3	2.504	0.733	0.382	0.982	0.601	0.778	0.609	−0.295 **	0.24
	CF11	6	2.658	0.667	0.583	1.292	0.624	0.642	0.630	−0.029 [ns]	0.17
	CF12 [a]	2	1.328	0.565	0.205	0.413	0.247	0.288	0.249	−0.169 **	0.60
	CUC1	3	1.286	0.496	0.214	0.457	0.223	0.132	0.225	0.407 ***	0.62
	CUC2 [a]	4	2.509	0.304	0.391	1.075	0.601	0.943	0.607	−0.569 ***	0.23
	CUC3 [a]	3	2.091	0.514	0.168	0.797	0.522	0.925	0.527	−0.772 ***	0.34
	CUC4	6	2.301	0.641	0.560	1.135	0.565	0.160	0.571	0.717 ***	0.23
	CUC6 [a]	5	2.445	0.584	0.247	1.041	0.591	0.925	0.597	−0.564 ***	0.25
	CUC7 [a]	2	2.000	0.428	0.000	0.693	0.500	1.000	0.505	−1.000 ***	0.38
	CUC8	5	2.351	0.516	0.499	0.981	0.575	0.528	0.580	0.081 [ns]	0.27
	CYP52 [a]	4	3.756	0.731	0.577	1.352	0.734	0.962	0.741	−0.310 ***	0.12
	CYP84 [a]	2	1.168	0.336	0.144	0.274	0.144	0.031	0.146	0.783 ***	0.74
	CYP101 [a]	3	2.038	0.419	0.019	0.740	0.509	1.000	0.514	−0.964 ***	0.36
	Mean	4.059	2.422	0.567	0.352	0.923	0.503	0.581	0.509	−0.118	-
	SD	0.473	0.270	0.14	0.045	0.116	0.054	0.091	0.055	0.140	-
	Total	69	41.180	-	-	-	-	-	-	-	3.29×10^{-10}

Na, number of different alleles; *Ne*, effective number of alleles; *H*, haplotype diversity index; *I*, Shannon's information index; *Ho*, observed heterozygosity; *He*, expected heterozygosity; *uHe*, unbiased expected heterozygosity; *PIC*, polymorphic information content; *PI*, probability of identity; *Fis*, inbreeding coefficient; SD, standard deviation; [a], Loci with null alleles and deviating from Hardy–Weinberg equilibrium with significant $p < 0.001$; [ns], not significant; **, $p < 0.01$; ***, $p < 0.001$.

3.1.3. Population Diversity and Structure

The population genetic parameters used in the present study are summarized in Table 3. The highest degree of genetic diversity was observed in the Zhejiang population (*Ne* = 2.817, *I* = 1.055, and *He* = 0.538), and the genetic diversity was lowest in the Chongqing population (*Ne* = 2.488, *I* = 0.943, and *He* = 0.495). The haplotype diversity (*H*) ranged from 0.365 (Guizhou population) to 0.418 (Zhejiang population). All results showed that there was no significant difference among the five populations in the level of genetic diversity, and these populations had a close genetic relationship. The *Fis* values were negative in all five populations, indicating a slight excess of heterozygosity among the five populations. Five population-specific alleles (alleles detected in only one population) produced from 5 of the 17 SSR markers were identified (Figure 1). The five population-specific alleles were distributed in the Zhejiang, Chongqing and Hubei populations, and the allele frequencies ranged from 0.7% to 17.0%. The Nei's genetic distance ranged from 0.008 (Hubei vs. Chongqing) to 0.023 (Guizhou vs. Hubei) (Table 4). The greatest geographic distance (1925.9 km) was between Zhejiang and Sichuan; however, this pairing did not reveal the largest genetic distance (0.023). The Mantel test (Figure 2) showed that genetic distance was not correlated with geographic distance (r = 0.02 and p = 0.530). A hierarchical analysis of molecular variance (AMOVA) indicated that the proportion of genetic differentiation within the five populations accounted for most of the molecular variance (99.668%), and only 0.332% of the molecular variance was accounted for in the variation among populations (Table 5). The *Fst* value between populations was also low, and the greatest *Fst* values were observed between populations from Guizhou, Hubei, Chongqing and Zhejiang.

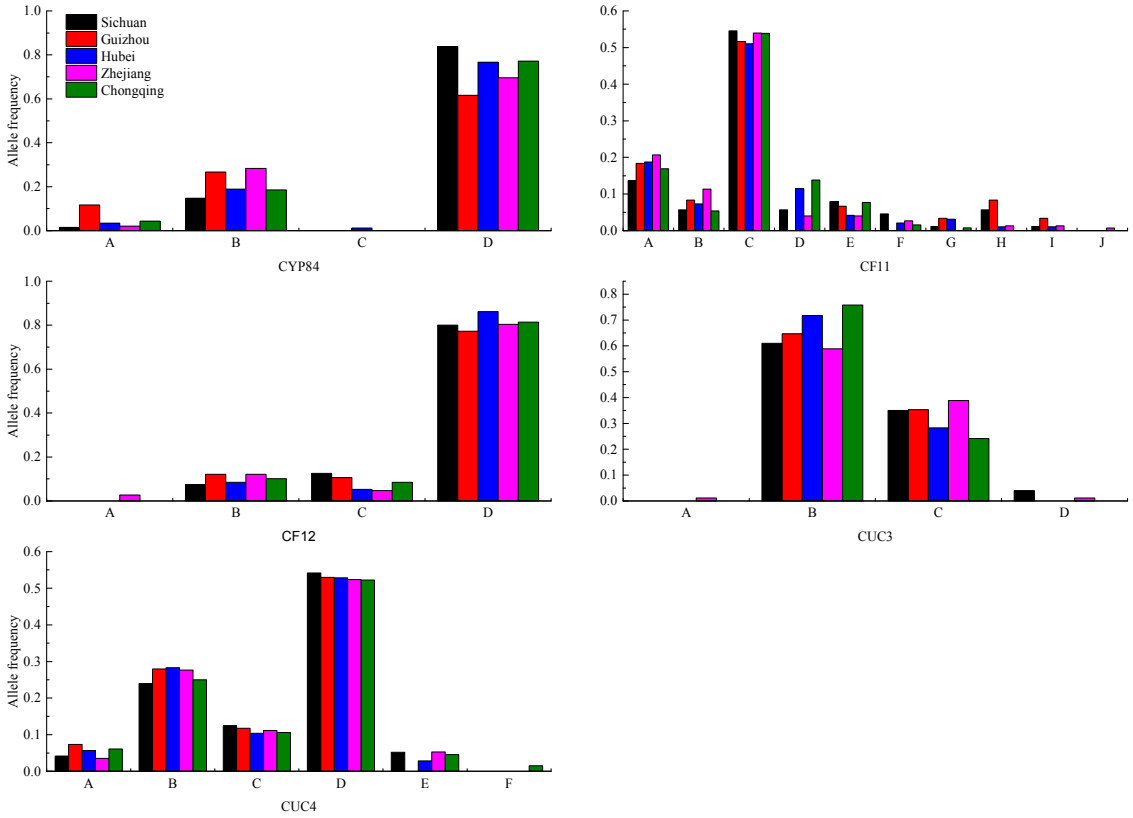

Figure 1. Five private alleles of 17 microsatellite loci in five populations.

Table 3. Polymorphic information among populations. [1]

Population	*Na*	*Ne*	*H*	*I*	*Ho*	*He*	*uHe*	*Fis*	*Na* Freq. ⩾ 5%	No. Private Alleles
Sichuan	4.765	2.674	0.415	1.035	0.562	0.531	0.538	−0.025	3.412	0.000
	(0.546)	(0.323)	(0.045)	(0.126)	(0.070)	(0.054)	(0.054)	(0.097)	(0.438)	(0.000)
Guizhou	4.294	2.628	0.400	1.004	0.553	0.525	0.534	−0.067	3.353	0.000
	(0.567)	(0.317)	(0.047)	(0.129)	(0.079)	(0.056)	(0.057)	(0.115)	(0.420)	(0.000)
Hubei	4.882	2.552	0.385	0.998	0.588	0.511	0.517	−0.070	3.235	0.059
	(0.562)	(0.302)	(0.045)	(0.126)	(0.079)	(0.056)	(0.056)	(0.112)	(0.369)	(0.059)
Zhejiang	5.176	2.817	0.418	1.055	0.570	0.538	0.542	−0.058	3.176	0.176
	(0.577)	(0.391)	(0.046)	(0.131)	(0.076)	(0.055)	(0.056)	(0.103)	(0.413)	(0.095)
Chongqing	4.529	2.488	0.365	0.943	0.557	0.495	0.500	−0.095	2.941	0.059
	(0.595)	(0.307)	(0.046)	(0.127)	(0.084)	(0.059)	(0.059)	(0.110)	(0.337)	(0.059)

[1] Brackets for standard deviation (SD); *Na*, number of different alleles; *Ne*, effective number of alleles; *H*, haplotype diversity index; *I*, Shannon's information index; *Ho*, observed heterozygosity; *He*, expected heterozygosity; *uHe*, unbiased expected heterozygosity; *Fis*, inbreeding coefficient; *Na* freq. ⩾ 5%, No. of different alleles with a frequency ⩾ 5%.

Table 4. Nei's standard genetic distances (lower diagonal) and pairwise *Fst* values (upper diagonal) between populations.

Population	Sichuan	Guizhou	Hubei	Zhejiang	Chongqing
Sichuan		0.008	0.005	0.005	0.008
Guizhou	0.017		0.009	0.006	0.008
Hubei	0.010	0.023		0.007	0.004
Zhejiang	0.011	0.017	0.017		0.009
Chongqing	0.013	0.017	0.008	0.020	

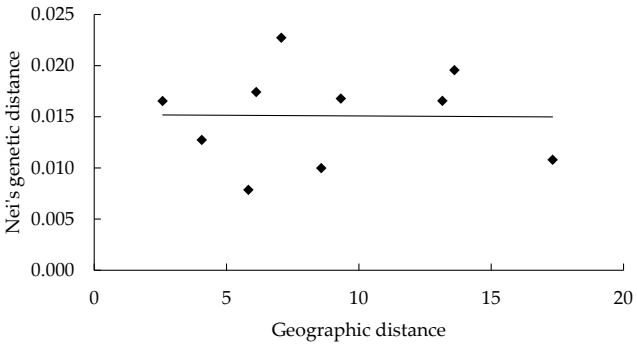

Figure 2. Correlation between genetic distance and geographic distance for cypress populations.

Table 5. Analysis of molecular variance (AMOVA) within and among populations.

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage Variation (%)
Among populations	4	22.589	0.015	0.332
Within populations	575	2226.637	4.473	99.668
Total	579	2249.226	4.488	

These results showed a clear maximum ΔK at K = 2 (Figure 3), in which all individuals were classified into two different clusters (Figure 4). Approximately 90% of the individuals belonged to each genetic cluster, indicating strong ancestry values (Q value) with an average greater than 0.9. Regarding genetic cluster 1 (green), including the Guizhou (44.7%), Hubei (46%), Zhejiang (56.7%), Chongqing (41.6%), and Sichuan (39.6%) populations and an unknown (only 7.8%) group, only one individual (0.7%) showed ancestry values (Q value) < 0.6. Most of the breeding parents in the unknown group (92.2%) belonged to genetic cluster 2 (red), with ancestry values > 0.8. The average distance (expected heterozygosity) between the individuals in clusters 1 and 2 was 0.5028 and 0.5214, respectively.

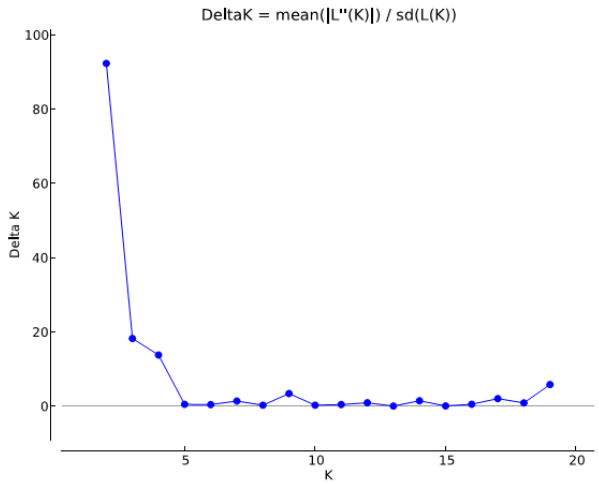

Figure 3. Estimates of the rate of the slope of the log probability curve (ΔK) plotted against K.

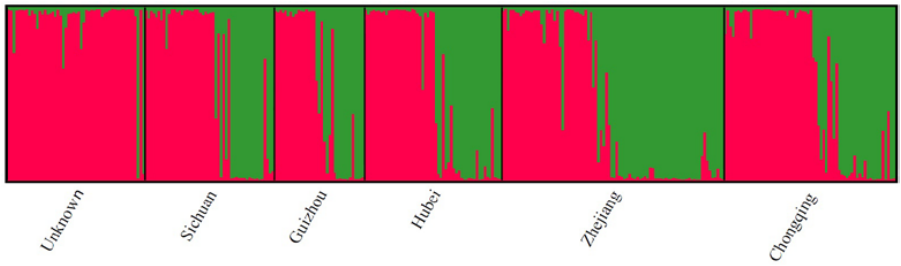

Figure 4. Assignment of 343 breeding parents to two populations using Structure version 2.0.

3.1.4. Construction of Genetic Core Collection

Figure 5 shows SSRs with significant linkage equilibrium ($R^2 < 0.3$; $p < 0.01$), and no SSRs showed linkage disequilibrium (LD). A core collection containing 30 genotypes retaining all 93 alleles identified from breeding parents was defined. The core collection contained six (11.3%) breeding parents from Hubei, seven (14.0%) breeding parents from Sichuan, seven (8.1%) breeding parents from Zhejiang, three (8.6%) breeding parents from Guizhou, three (4.5%) breeding parents from Chongqing, and four (7.5%) breeding parents from an unknown population. The observed and expected heterozygosity (*Ho* and *He*) values calculated for the core collection were close to those values for the overall genotypes (Table 6). Regarding the allele frequency, the chi-square test showed that the core collection had a similar allelic distribution for all 17 SSR markers ($p > 0.05$). Moreover, the frequency of alleles between this core collection and the overall breeding parents was highly correlated: not only were the same alleles detected but these alleles were also represented at a nearly identical frequency ($r = 0.98$, Figure 6).

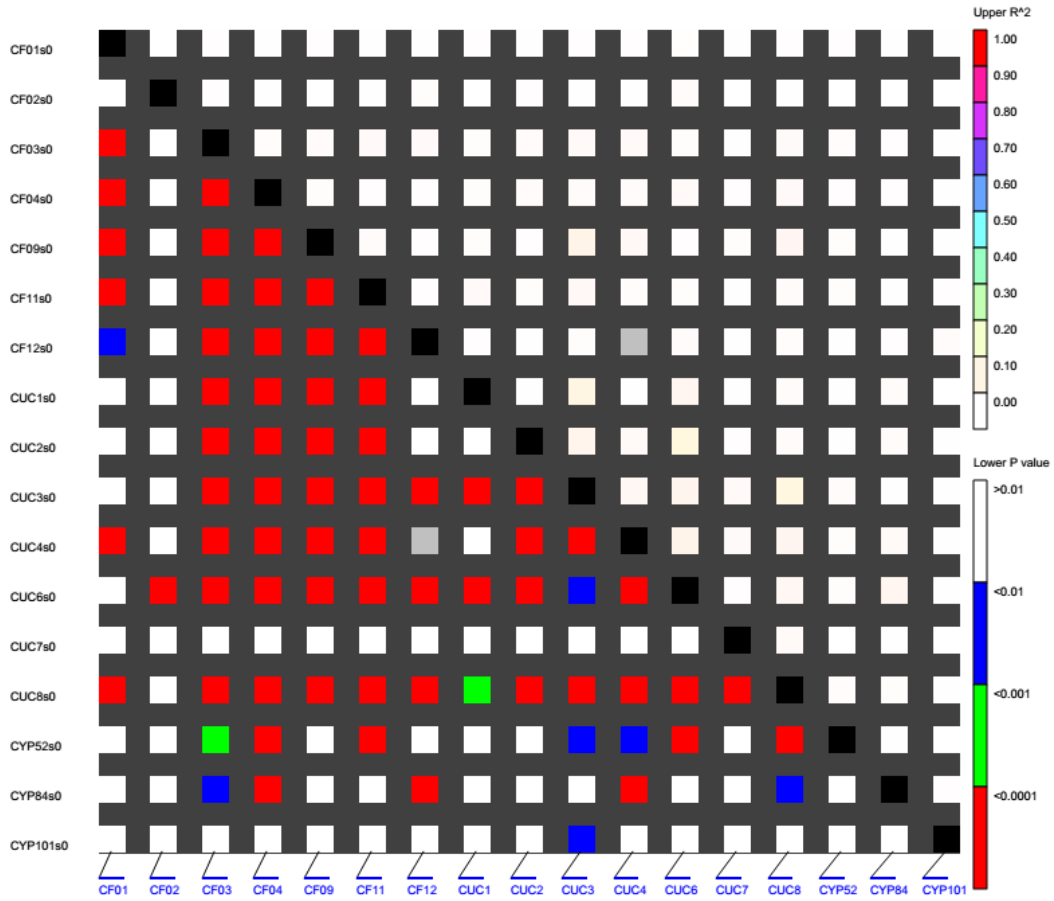

Figure 5. Pairwise linkage disequilibrium (LD) (R^2) between simple sequence repeats (SSRs) in cypress.

Table 6. SSR diversity within the core collection compared with the overall breeding parents.

Population	Size	*Na*	*Ne*	*I*	*Ho*	*He*	Correlation of Allele Frequency
Overall breeding parents	343	93	2.7	1.06	0.57	0.54	0.96
Core collection	30	93	3.0	1.16	0.57	0.57	

Na, number of different alleles; *Ne*, effective number of alleles; *I*, Shannon's information index; *Ho*, observed heterozygosity; *He*, expected heterozygosity.

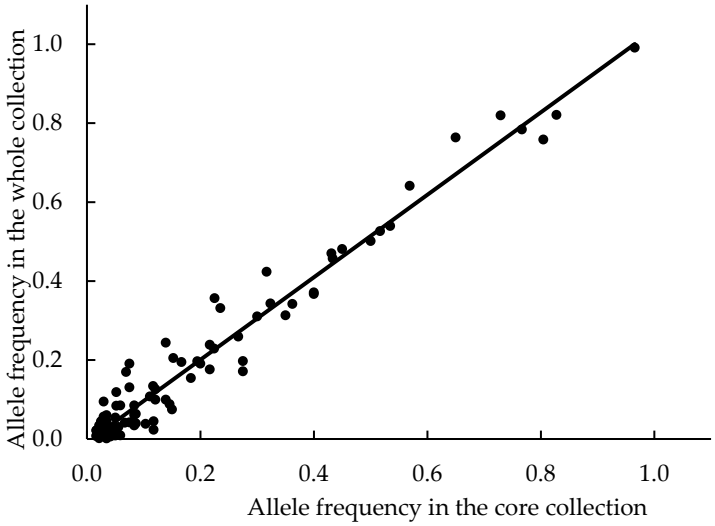

Figure 6. Correlation between the allele frequencies of 93 SSR alleles in the core collection (30 individuals) and the breeding parents (343 individuals).

4. Discussion

Most conifers have high genetic diversity and low differentiation among populations [38]. The present study represents the first large-scale analysis using molecular markers to assess genetic diversity and the relationships among cultivated cypress breeding parents throughout China. In the present study, the cypress breeding parents had low genetic diversity (H = 0.406), and low haplotype diversity (H) in each geographic population, reflecting a long history of cuttage cultivation or asexual reproduction. The positive values for *Fis* reflect the presence of null alleles [39]. In the present study, eight loci were detected as containing null alleles using the Cervus software; for CYP84, the previously reported null alleles [6] significantly ($p < 0.001$) deviated from Hardy–Weinberg equilibrium. However, the average *Fis* (−0.037) was close to zero, suggesting that the cypress breeding parents in the present study were comparable to a random mating unit [40]. Allelic diversity (*Na*, *Ne*), expected heterozygosity (*He*), and Shannon's information index (I) are three important and commonly used parameters to study genetic diversity in populations [41,42]. The allelic variation (*Na*), expected heterozygosity (*He*) and Shannon's information index (I) among the breeding parents examined in the present study (*Ne* = 2.698, *He* = 0.527, and I = 1.035) were lower than those values observed in wild populations (*Ne* = 3.42, *He* = 0.706, and I = 1.333) and comparable to putatively cultivated populations (*Ne* = 3.23, *He* = 0.620, and I = 1.149) [2]. These findings also supported the conclusion that the average genetic diversity indices of wild populations were higher than those indices of putatively cultivated populations [2]. Compared with other *Cupressus* species, the Shannon's information index (I = 1.035) was comparable to that observed in *C. duclouxiana* Hickel (I = 1.169) and *C. gigantea* Cheng et L. K. Fu (I = 1.022) [2]. However, a reduced set of *C. sempervirens* var. *Horizontalis* (I = 0.410) [43] displayed a lower I value, while a higher I value (1.377) was observed in *C. chengiana* S. Y. Hu [2]. This low I value likely reflects a less diverse genetic background within these breeding parents. Reflecting the long history of exploitation and cultivation of cypress, cultivated or mosaic populations are

widespread [44]. Although these five cypress populations might experience different demographic histories, the observed level of microsatellite genetic diversity was similar, and Nei's genetic distance was low (Tables 3 and 4). These results showed that the genetic relationship of breeding parents from different geographic origins was close. These results are consistent with a long history of cultivation and increased selection pressure by humans for cypress. In populations grouped according to geographical origin, an extremely low percentage of variation (0.332%) and reduced F_{ST} values could reflect geographic factors, as expected in a woody perennial breeding species [45]. Five alleles were identified as population-private alleles. Some of these alleles differed in size from the most frequent population alleles and likely originated through a direct mutation, indicating that some individuals were particular to the geographic population [40].

The UPGMA tree clustered the majority of breeding parents into two major groups. The breeding parents with unknown origin were clustered closely together with short branches between accessions, and these breeding parents were clustered into a single subgroup (pink), suggesting that they could be a single male or female parent population in crossbreeding work. The clustering analysis showed that the genetic basis of these varieties was relatively narrow, and the genetic diversity of these varieties was not completely related to geographic distribution. In addition, the three breeding parents were not clustered into two major groups, and they had a higher genetic distance with other breeding parents, implying that rare spontaneous somatic mutations likely occurred in long-cultivated trees propagated through grafting [46]. A Bayesian clustering approach probabilistically assigns individuals to a single population or to multiple populations when their genotype indicates admixture [30]. The ΔK statistics provided the highest value at $K = 2$ ($\Delta K = 82.59$). As the ad hoc statistic ΔK preferentially detects the uppermost level of structure [47,48], $K = 2$ was considered as the starting point for the first ground clustering in all 343 individuals. Structure analysis identified two subgroups among the breeding parents. This analysis is consistent with the results of previous UPGMA clustering analyses, showing that the genetic subdivision of clusters is not consistent with geographic origin. This phenomenon in the breeding parents of cypress was also observed in other annual and perennial plants, such as sorghum and poplar [49,50], suggesting that there was no obvious differentiation in the population structure of cypress from different geographic origins. The relationship and geographic origin of the parents play an important role in maintaining a high level of genetic diversity and the generation of heterotic progenies through crossbreeding [51,52]. The analysis based on the Mantel test (Figure 2) showed that the genetic distance was not correlated with the geographic distance ($r = 0.02$ and $p = 0.530$), suggesting that the geographic distance is not the principal factor influencing genetic differentiation among the breeding parents of cypress. Therefore, based on the results of the present study and previous experiences in crossbreeding, both distant geography and evolutionary relationship should be considered when selecting the crossbreeding parents of cypress. Moreover, we understood the genetic relationship between individuals from 343 breeding parents and five populations with different geographic origins through SSR markers, and in-depth research on the genetic diversity and population structure of breeding parents. This information would be helpful to shorten the period in crossbreeding and accelerate the utilization of cypress breeding materials. Based on the above findings concerning the genetic diversity of cypress breeding parents, we should strengthen the protective measures for natural populations of cypress to maintain a high level of genetic diversity, collect more breeding materials of cypress to broaden the genetic background, and lay a foundation for long-term breeding goals.

The advanced "M" (maximization) strategy employed in the present study, however, is considered the most appropriate approach for selecting the most diverse alleles and eliminating redundancy [53]. A core collection of 30 genotypes sufficient to retain all 93 alleles identified from the analyzed breeding parent genotypes was defined (Table 6, Figure 6). This core collection thus guarantees the preservation of all alleles, which is critical for maintaining the genetic diversity of a population and is likely involved in plant adaption to environmental shift [54,55]. Overall, the 30 genotypes selected for the collection constitute a representative sample of the diversity retained from the breeding parents. Furthermore,

the 30 genotypes for core collection should be extensively evaluated for morphology traits to identify accessions with desirable characteristics in wood improvement research and genomic studies.

5. Conclusions

We hereby reported the first molecular characterization of a collection of 343 breeding parents belonging to five provinces and an unknown population using a set of 17 polymorphic SSR markers. We clarified the genetic relationships between breeding parents through clustering analysis. The pattern detection methods of neighbor joining and Bayesian models similarly clustered these breeding parents. The breeding parents in unknown populations were clustered into one subgroup, whereas the breeding parents from the five populations were clustered into two groups. There was no obvious geographic context of the genetic relationship and diversity in the breeding parents. A manageable core collection with 30 genotypes was selected to maintain the highest variability in the lowest number of plants. This collection represents the genetic variability of the breeding parents of cypress, as all 93 SSR alleles detected were retained among the 343 breeding parents.

Acknowledgments: This study was funded through grants from the Forestry Industry Research Special Foundation for public welfare projects, China (grant number 201304103-1) and Tree Breeding Research Project of Zhejiang Province, China.

Author Contributions: Hanbo Yang and Guoqing Jin conceived and designed the experiments; Hanbo Yang performed the experiments; Hanbo Yang and Rui Zhang analyzed the data; Zhongping Feng and Zhichun Zhou contributed reagents/materials/analysis tools; and Hanbo Yang drafted the manuscript.

Conflicts of Interest: The authors declare that there are no financial and personal relationships with other people or organizations that could inappropriately influence this work, there are no professional or other personal interests of any nature or in any product, service and/or company, which could influence the position presented herein or affect the review of the manuscript.

References

1. Farjon, A. *A Monograph of Cupressaceae and Sciadopitys*; Royal Botanic Gardens, Kew: London, UK, 2005.
2. Xu, L.; Xu, H.; Li, Z.; Shang, H.; Adams, R.P.; Mao, K. Genetic diversity and conservation implications of four cupressus species in china as revealed by microsatellite markers. *Biochem. Genet.* **2014**, *52*, 181–202.
3. Fu, L.G.; Yu, Y.F.; Farjon, A. Cupressaceae. In *Flora of China*; Raven, P.H., Wu, C.Y., Eds.; Science Press: Beijing, China; Missouri Botanical Garden Press: St. Louis, MO, USA, 1999; pp. 62–65.
4. Bretting, P.; Wildrlechner, M. Genetic markers and plant genetic resource management. *Plant Breed. Rev.* **2010**, *13*, 11–86.
5. Kearsey, M.J.; Farquhar, A.G.L. Qtl analysis in plants; where are we now? *Heredity* **1998**, *80*, 137–142. [CrossRef] [PubMed]
6. Vermigli, M. Genetic Varability in Italian Populations of *Cupressus sempervirens* L. Assessed by SSR and Rapd Markers. Ph.D. Thesis, University of Verona, Verona, Italy, 2005.
7. Parker, P.G.; Snow, A.A.; Schug, M.D.; Booton, G.C.; Fuerst, P.A. What molecules can tell us about populations: Choosing andusing a molecular marker. *Ecology* **1998**, *79*, 361–382. [CrossRef]
8. Hai-Fei, W.; Xu-Xiao, Z.; Jian-Ping, G.; Tao, Y.; Xue-Lian, S.; Yu, M.; Robert, R. Genetic diversity and relationship of global faba bean (*Vicia faba* L.) germplasm revealed by issr markers. *Tag. Theor. Appl. Genet. Theor. Angew. Genet.* **2012**, *124*, 789–797.
9. Khadivi-Khub, A.; Zamani, Z.; Fattahi, R.; Wünsch, A. Genetic variation in wild Prunus L. Subgen. Cerasus germplasm from iran characterized by nuclear and chloroplast ssr markers. *Trees* **2014**, *28*, 471–485. [CrossRef]
10. Nimnual, A.; Romkaew, J.; Chukeatirote, E.; Nilthong, S. Evaluation of genetic relationship among some important Japanese and Thai soybean varieties using AFLP analysis. *Aust. J. Crop Sci* **2014**, *8*, 481–485.
11. Escribano, P.; Viruel, M.A.; Hormaza, J.I. Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*annona cherimola*, annonaceae), an underutilised subtropical fruit tree species. *Ann. Appl. Biol.* **2008**, *153*, 25–32. [CrossRef]

12. Wang, J.C.; Hu, J.; Xu, H.M.; Zhang, S. A strategy on constructing core collections by least distance stepwise sampling. *Tag. Theo. Appl. Genet. Theor. Angew. Genet.* **2007**, *115*, 1–8. [CrossRef] [PubMed]
13. Zhang, H.; Zhang, D.; Wang, M.; Sun, J.; Qi, Y.; Li, J.; Wei, X.; Han, L.; Qiu, Z.; Tang, S. A core collection and mini core collection of *Oryza sativa* L. In China. *Tag. Theor. Appl. Genet. Theor. Angew. Genet.* **2011**, *122*, 49–61. [CrossRef] [PubMed]
14. Belaj, A.; Atienza, S.G.; Urdíroz, N.M.; Rosa, R.D.L.; Satovic, Z.; Martín, A.; Kilian, A.; Trujillo, I.; Valpuesta, V. Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genet. Genomes* **2012**, *8*, 365–378. [CrossRef]
15. Guo, Y.; Li, Y.; Hong, H.; Qiu, L.J. Establishment of the integrated applied core collection and its comparison with mini core collection in soybean (*Glycine max*). *Crop J.* **2005**, *44*, 174–180. [CrossRef]
16. Taniguchi, F.; Kimura, K.; Saba, T.; Ogino, A.; Yamaguchi, S.; Tanaka, J. Worldwide core collections of tea (*Camellia Sinensis*) based on genome-wide simple sequence repeat (SSR) markers. In Proceedings of the International Plant and Animal Genome Conference XXII, San Diego, CA, USA, 10–15 January 2014.
17. Kejun, L.; Major, G.; Spencer, M.; Smith, J.S.; Ed, B.; John, D. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* **2003**, *165*, 2117–2128.
18. Bagnoli, F.; Vendramin, G.G.; Buonamici, A.; Doulis, A.G.; González-Martínez, S.C.; La Porta, N.; Magri, D.; Raddi, P.; Sebastiani, F.; Fineschi, S. Is cupressus sempervirens native in italy? An answer from genetic and palaeobotanical data. *Mol. Ecol.* **2009**, *18*, 2276–2286. [CrossRef] [PubMed]
19. Shahroodian, S.H.; Azadfar, D.; Soltanloo, H.; Ramezanpour, S.S. Genetic variability in natural iranian populations of *cupressus sempervirens* var. Horizontalis in Caspian sea coastward assessed by SSR markers. *Plant Omics J.* **2011**, *4*, 19–24.
20. Webb, D.M.; Knapp, S.J. DNA extraction from a previously recalcitrant plant genus. *Plant Mol. Biol. Rep.* **1990**, *8*, 180–185. [CrossRef]
21. Li, Z.; Xu, H.; Zhao, G. Isolation and characterization of polymorphic microsatellite loci primers for *Cupressus funebris* (cupressaceae). *Conserv. Genet. Resour.* **2012**, *5*, 307–309. [CrossRef]
22. Xu, H.; Shi, D.; Wang, J.; Xu, T.; Wu, Y. Isolation and characterization of polymorphic microsatellite markers in *Cupressus chenggiana* S. Y. Hu (cupressaceae). *Conserv. Genet.* **2008**, *9*, 1023–1026. [CrossRef]
23. Sebastiani, F.; Buonamici, A.; Fineschi, S.; Racchi, M.L.; Raddi, P.; Vendramin, G.G. Novel polymorphic nuclear microsatellites in *Cupressus sempervirens* L. *Mol. Ecol. Notes* **2005**, *5*, 393–394. [CrossRef]
24. Marshall, T.C.; Slate, J.; Kruuk, L.E.; Pemberton, J.M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **1998**, *7*, 639–655. [CrossRef] [PubMed]
25. Peakall, R.; Smouse, P.E. Genalex 6.5. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef] [PubMed]
26. Paetkau, D.; Calvert, W.; Stirling, I.; Strobeck, C. Microsatellite analysis of population structure in Canadian polar bears. *Mol. Ecol.* **1995**, *4*, 347–354. [CrossRef] [PubMed]
27. Excoffier, L.; Laval, G.; Schneider, S. *Arlequin Version 3.01. An Integrated Software Package for Population Genetics Data Analysis*; Computational and Molecular Population Genetic Lab, Institute of Zoology, University of Berne: Bern, Swizerland, 2006.
28. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. [CrossRef]
29. Rohlf, F.J. *NTSYS-pc Numerical Taxonomy and Multivariate Analysis System (Version 1.80)*; State University of New York: New York, NY, USA, 1994.
30. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *7*, 574–578.
31. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef] [PubMed]
32. Earl, D.A. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [CrossRef]
33. Jakobsson, M.; Rosenberg, N.A. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **2007**, *23*, 1801–1806. [CrossRef] [PubMed]
34. Rosenberg, N.A. DISTRUCT: A program for the graphical display of population structure. *Mol. Ecol. Notes* **2004**, *4*, 137–138. [CrossRef]

35. Hill, W.G.; Robertson, A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **1968**, *38*, 226–231. [CrossRef] [PubMed]
36. Schoen, D.; Brown, A. Maximizing genetic diversity in core collections of wild crop relatives of crop species. In *Core Collections of Plant Genetic Resources*; Hodgkin, T., Brown, A.H.D., van Hintum, T.J.L., Morales, E.A.V., Eds.; John Wiley and Sons Ltd.: Chicester, UK, 1995; pp. 253–259.
37. Cipriani, G.; Spadotto, A.; Jurman, I.; Gaspero, G.D.; Crespan, M.; Meneghetti, S.; Frare, E.; Vignani, R.; Cresti, M.; Morgante, M. The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theor. Appl. Genet.* **2010**, *121*, 1569–1585. [CrossRef] [PubMed]
38. Hamrick, J.L.; Godt, M.J.W.; Sherman-Broyles, S.L. Factors influencing levels of genetic diversity in woody plant species. *New For.* **1992**, *6*, 95–124. [CrossRef]
39. Bruford, M.W.; Ciofi, C.; Funk, S.M. Characteristics of microsatellites. In *Molecular Tools for Screening Biodiversity: Plants and Animals*; Karp, A., Isaac, P.G., Ingram, D.S., Eds.; Chapman & Hall: London, UK, 1998; pp. 202–205.
40. Wei, L.; Luca, D.; Paolo, D.F.; Roberta, P.; Silviero, S.; Stefano, T. Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. *Plant Mol. Bio. Rep.* **2014**, *33*, 458–473.
41. Hamilton, M. *Population Genetics*; Wiley-Blackwell: New York, NY, USA, 2009.
42. Freeland, J.; Kirk, H.; Petersen, S. *Molecualr Ecology*, 2nd ed.; John Wiley & Sons: Chichester, UK, 2011.
43. Korori, A.; Azadfar, D.; Shirvany, A.; Valipour, K.; Matinizadeh, M. Assessment of genetic variability in long-lived *Cupressus sempervirens* var. *Horizontalis* using SSR markers. *Plant Gene Trait* **2012**, *3*, 43–49.
44. Zheng, W.; Fu, L. Cupressus linn. In *Editorial Committee of Flora Reipublicae Popularis Sinicae Flora Republicae Popularis Sinicae Tomus 7: Gymnospermae*; Science Press: Beijing, China, 1978; pp. 328–336.
45. Hamrick, J.L.; Godt, M.J.W. Allozyme diversity in plant species. In *Plant Population Genetics, Breeding and Genetic Resources*; Brown, A.H.D., Clegg, M.T., Kahler, A.L., Weir, B.S., Eds.; Sinauer: Sunderland, MA, USA, 1989; pp. 43–63.
46. Weissinger, A. Technologies for germplasm conservation ex situ. In *The Preservation and Valuation of Biological Resources*; Orians, G.H., Brown, G.M., Kunin, W.E., Swierbinski, J.E., Eds.; University of Washington Press: Seattle, WA, USA, 1990; pp. 3–31.
47. Lia, V.V.; Poggio, L.; Confalonieri, V.A. Microsatellite variation in maize landraces from northwestern argentina: Genetic diversity, population structure and racial affiliations. *Tag. Theor. Appl. Genet. Theor. Angew. Genet.* **2009**, *119*, 1053–1067. [CrossRef] [PubMed]
48. Emanuelli, F.; Lorenzi, S.; Grzeskowiak, L.; Catalano, V.; Stefanini, M.; Troggio, M.; Myles, S.; Martinez-Zapater, J.M.; Zyprian, E.; Moreira, F.M. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol.* **2013**, *13*, 1–17. [CrossRef] [PubMed]
49. Du, Q.; Wang, B.; Wei, Z.; Zhang, D.; Li, B. Genetic diversity and population structure of Chinese white poplar (*Populus tomentosa*) revealed by SSR markers. *J. Hered.* **2012**, *103*, 853–862. [CrossRef] [PubMed]
50. Adugna, A. Analysis of in situ diversity and population structure in Ethiopian cultivated *Sorghum bicolor* (L.) landraces using phenotypic traits and SSR markers. *Springerplus* **2014**, *3*, 212. [CrossRef] [PubMed]
51. Jena, K.K. The species of the genus oryza and transfer of useful genes from wild species into cultivated rice, *O. sativa*. *Breed. Sci.* **2010**, *60*, 518–523. [CrossRef]
52. Tripp, E.A.; Siti, F.; Iain, D.; Mcdade, L.A. Origin of African *Physacanthus* (Acanthaceae) via wide hybridization. *PLoS ONE* **2013**, *8*, 268–277. [CrossRef] [PubMed]
53. Franco, J.; Crossa, J.; Warburton, M.; Taba, S. Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. *Crop Sci* **2006**, *46*, 854–864. [CrossRef]
54. Richter, T.; Soltis, P.; Soltis, D. Genetic variation within and among populations of the narrow endemic, *Delphinium viridescens* (Ranunculaceae). *Am. J. Bot.* **1994**, *81*, 1070–1076. [CrossRef]
55. Bengtsson, B.O.; Weibull, P.; Ghatnekar, L. The loss of alleles by sampling: A study of the common outbreeding grass *Festuca Ovina* over three geographic scales. *Hereditas* **1995**, *122*, 221–238. [CrossRef]