forests

MDPI

Article

Construction of a Core Collection of Germplasms from Chinese Fir Seed Orchards

Hanbin Wu [1,2], Aiguo Duan [1,2], Xihan Wang [3], Zhiyun Chen [3], Xie Zhang [4], Guiping He [5] and Jianguo Zhang [1,2,*]

[1] State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China
[2] Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China
[3] State-Owned Forestry Farm of Weimin, Shaowu 354006, China
[4] Institute of Forestry, Hunan Academy of Forestry, Changsha 410004, China
[5] Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China
* Correspondence: zhangjg@caf.ac.cn

Abstract: Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook) is one of the most important tree species for afforestation in China. First-, second-, and third-generation seed orchards of Chinese fir have been established successively, and rich germplasms have been accumulated in the process of genetic improvement. It is necessary to build a core collection of germplasms from Chinese fir seed orchards. In this work, we constructed core collections representing the genetic diversity of Chinese fir seed orchard resources based on SSR data. A total of 607 seed orchard materials from three generations were used to determine the best sampling method and intensity by comparing and analyzing nine methods for constructing core collections. Core Hunter's multi-strategy optimizes allele coverage and the distance criterion under a 30% sampling intensity (weight: A–NE, 0.7; CV, 0.3 and E–NE, 0.5; CV, 0.5), which is superior to other strategies and was selected as the best method. The two core collections (A–NE&CV73, E–NE&CV55) constructed contained all the alleles of the whole collection and effectively limited the homology in the core collections; each core collection contained 182 accessions. Our findings could contribute greatly towards improving the management of genetic resources in Chinese fir seed orchards and provide elite materials for future studies.

Keywords: Chinese fir; core collection; SSR; seed orchard

check for updates

Citation: Wu, H.; Duan, A.; Wang, X.; Chen, Z.; Zhang, X.; He, G.; Zhang, J. Construction of a Core Collection of Germplasms from Chinese Fir Seed Orchards. *Forests* **2023**, *14*, 305. https://doi.org/10.3390/f14020305

Academic Editor: Witold Wachowiak

Received: 5 January 2023
Revised: 29 January 2023
Accepted: 31 January 2023
Published: 3 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Plant germplasms are the important material basis of breeding work and the key to determining the breeding effect. Germplasm collection ensures the long-term protection of genetic resources and provides a rich genetic basis for plant genetic improvement and variety breeding [1,2]. The ultimate purpose of mastering and studying germplasms is to use them, mainly for creating new crops and breeding new varieties. Of course, germplasm work can also be used for theoretical research on crop origin, evolution, classification, and other aspects. All countries in the world attach great importance to the investigation, collection, evaluation, preservation, and utilization of germplasms [1]. However, the management and use of large germplasms is associated with a high economic cost to complete daily tasks, such as preservation, regeneration, reproduction, recording, and evaluation [2–4]. Therefore, the concept of a core set was introduced to help make these decisions. It is defined as a subset of a complete set, which can best represent the diversity of the entire set with minimum redundancy [5,6]. Core germplasm collections provide a new method for the utilization and protection of germplasms, and play a number of roles in the management and use of genetic resources.

Since the introduction of the concept of the core collection, a large number of documents on the theory and practice of core collection have been accumulated. A core

collection is established based on geographical origin and morphological characteristics. Due to problems related to provenance and environmental interaction, phenotypic diversity is not enough to represent the genetic diversity of germplasms [2,7–10]. Therefore, molecular markers are considered to be able to capture the genetic diversity of germplasms at the DNA sequence level [11]. A core collection has been developed using SSR and SNP markers in many species, such as apple (Malus × domestica Borkh.) [10], strawberry (*Fragaria* × *ananassa*) [12], olive (*Olea europaea* L.) [13], walnut (*Juglans regia* L.) [14], *schima superba* [15], etc. At present, different bioinformatics tools have been proposed to build core germplasm collections, such as Coreset in Powermaker v 3.25 [16], Power Core [17], and Core Hunter [18,19].

Chinese fir, or *Cunninghamia lanceolata* (Lamb.) Hook, is a unique economic and afforestation tree species in China, which is widely distributed in the southern subtropical region. Since the 1960s, provenance tests have been carried out on Chinese fir in the entire distribution area, collecting and preserving a large number of excellent provenance resources [20–22]. A large number of seed orchard resources with excellent genetic quality were selected for a study of multi-generation genetic improvement of Chinese fir [23] based on understanding the rules of geographic variation and the excellent provenance materials of Chinese fir [20,22]. In addition, a series of breeding studies [24–27] have accumulated rich germplasms. The construction of core collections represents important and necessary work of the Chinese fir breeding plan. However, the direction and method of construction have mainly been considered from the perspective of geographical origin. With the advancement of Chinese fir breeding generations, excellent varieties of Chinese fir have been introduced and cultivated among provinces and regions, and they have often been used for hybrid seed production. However, there are duplicate and redundant materials in the germplasm collections of seed orchards, which may not only hinder the effective protection of germplasms, but also impede the effectiveness of the evaluation and use of germplasms. Consequently, it is necessary to build a core collection of seed quality of Chinese fir seed orchards.

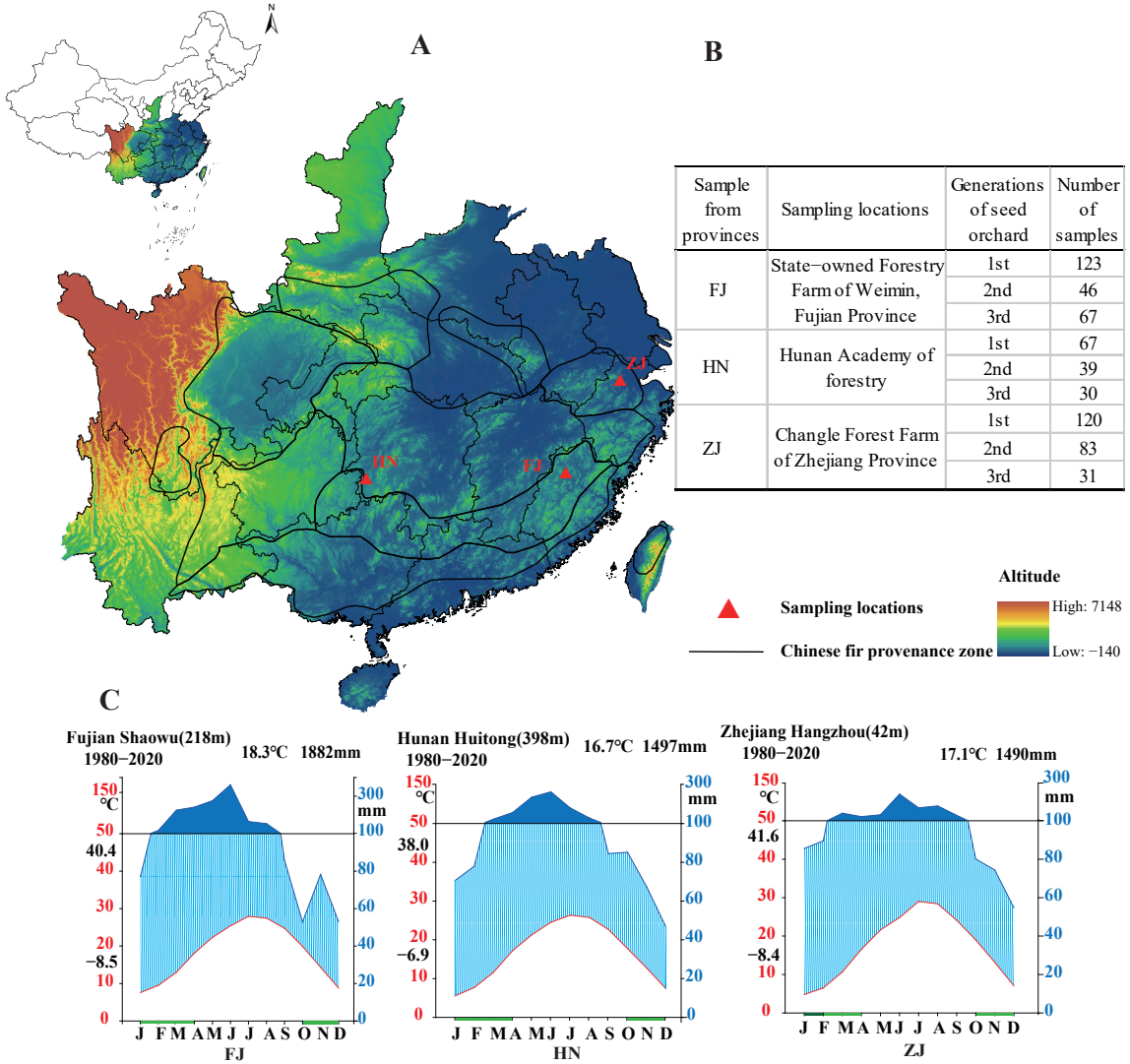
Given this, the main purpose of this study was to develop a core set that can represent the genetic diversity of Chinese fir seed orchard germplasms. For this purpose, we used a series of different methods to construct the core set based on 21 pairs of SSR primer markers, and we compared and analyzed genetic diversity parameters of the subset to select the final core set. Finally, we also studied the correlation between the core set and the whole germplasm. Our results improved important materials for the development of Chinese fir breeding resources.

2. Materials and Methods

2.1. Plant Material and Microsatellite Genotyping

The experimental materials were purchased from clonal seed orchards in Fujian, Zhejiang, and Hunan provinces in China. Fujian and Hunan provinces are high-yield production areas of Chinese fir, and Zhejiang province is a marginal production area. A total of 607 germplasms were collected from nine seed orchards of three generations, including 223 in Fujian (FJ), 137 in Hunan (HN), and 236 in Zhejiang (ZJ) (Figure 1, Table S1).

We used CTAB plant genome DNA extraction kits to extract the total DNA of all Chinese fir germplasms, and 21 pairs of well-amplified SSR primers (15 pairs of dinucleotides and 6 pairs of trinucleotides) (Table S2) were used for genotyping [28,29]. The reaction volume of the PCR reaction system was 15 µL, including 50 ng DNA template, 0.6 µL 10 uM F primer (including fluorescent primer), 0.6 µL 10 um R primer, and 12.8 µL 1.1 × Golden Star T6 Super PCR Mix (TsingKe, Beijing, China). PCR cycle parameters followed those outlined by Wen and Li et al. [28,29]. The amplification products were subjected to capillary electrophoresis (Liz 500 as the internal size standard) on the ABI3730 sequencer (Applied Biosystems, Life Technologies, Carlsbad, CA, USA). Finally, the electrophoresis results were read using Genemapper 4.0 software (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) and analyzed to obtain each genotype.



Sample from provinces	Sampling locations	Generations of seed orchard	Number of samples
FJ	State−owned Forestry Farm of Weimin, Fujian Province	1st	123
		2nd	46
		3rd	67
HN	Hunan Academy of forestry	1st	67
		2nd	39
		3rd	30
ZJ	Changle Forest Farm of Zhejiang Province	1st	120
		2nd	83
		3rd	31

Figure 1. Summary of experimental materials. (**A**,**B**) Locations and sampling information of Chinese fir seed orchards; (**C**) Walter and Lieth climatic diagram of study sites.

2.2. Construction of Core Collections

We used nine methods to build core collections: (1) We used the "maximum length subtree" (MLS) function in Darwin 6.0.14 [30] to identify a subset, minimizing the redundancy between samples, and limiting the loss of diversity where possible (diversity was represented by the constructed tree here). (2) We used the advanced M strategy proposed by Kim et al. [17] to build a subset, which was implemented in Power Core v1.0 [17]. (3) Based on the A–NE standard (average distance between each access and the nearest entry), we created a type 1 core collection (A–NE) to comprehensively understand the genetic diversity of the whole collection [31]. (4) Based on the E–NE standard (average distance between each entry and the nearest neighbor entry), we determined a type 2 core collection (E–NE), which was intended to represent the extremes of the entire set [32]. (5, 6, 7) We optimized the allele coverage (O_CV), expected heterozygosity (O_He), and Shannon diversity index (O_Shannon). (8) We ran a simulated annealing algorithm based on allele maximization (SANA). (9) We ran a simulated annealing algorithm based on maximum genetic diversity (SAGD).

Methods (3, 4, 5, 6, 7) were implemented in Core Hunter 3.0 [19]. Methods (7, 8) were executed in the CoreSet of PowerMarker v3.25 [16]. Then, we used 10%, 20%, 30%, 40%, and 50% sampling intensities to observe the difference in the selected materials to explore the best sampling intensity. To evaluate the performance of each construction method, we randomly selected an equal proportion of accessions as a control (R).

2.3. Indicators for Evaluating the Core Collections

We used the following genetic diversity parameters to evaluate the created core subset: number of observed alleles (Na) and effective alleles (Ne); observed (Ho) and expected (He) heterozygosity; Shannon diversity index (I); and fixation index (F), $F = (He - Ho)/He$. The above parameters were estimated using GenALEx 6.5.1 software [33]. Homogeneity refers to the proportion of first-degree relatives in the population (P = individuals with first-degree relationships within populations/the total number of populations). The TrioML estimator (Wang, 2007) in COANCESTRY 1.0.1.10 software [34] was used to estimate the co-ancestry (r) between individuals. The boundary of the first-degree relationship was referred to as the standard by Wu et al. [35]. If $r \geq 0.4511$, it was considered to be a first-degree relationship.

3. Results

3.1. Comparison of Different Methods for Constructing Core Collections

Core collections of 607 germplasms in all of the studied Chinese fir seed orchards were constructed based on the M strategy and the distance strategy. According to a series of assumed subset sizes, the performance of each sampling strategy of the subsets was evaluated. Therefore, five subsets were developed, including 10%, 20%, 30%, 40%, and 50% sampling intensities. Table 1 reports the results of each genetic parameter of the constructed core collection and the whole seed orchard collection. Under a strong sampling intensity, the core collections constructed using MLS, E–NE, A–NE, O_He, SAGD, and R were significantly different from the whole collection. At a 30% sampling intensity, P was 0 for MLS and E–NE, and at a 40% sampling intensity, P was 0 for A–NE. However, Na decreased and F increased when the selection intensity increased. O_CV and Power Core covered all Na at a 20% sampling strength, but the P of O_CV showed an upward trend as the selection intensity decreased. The methods O_He and O_Shannon were better than the whole collection in terms of their respective objectives. With the increase in sampling intensity, He and I gradually increased, and P and Na gradually decreased. The P and Na of the SANA and SAGD methods decreased with an increase in sampling intensity.

Table 1. Parameters of different methods for constructing core collections. P, homogeneity (P = individuals with first-degree relationships within populations/the total number of populations); Na, observed number of alleles; Ne, effective number of alleles; I, Shannon diversity index; Ho, observed heterozygosity; He, expected heterozygosity; uHe, unbiased expected heterozygosity; F, fixation index. "*" and "**", statistically significant difference, Dunnett's test ("*", $p < 0.05$; "**", $p < 0.01$).

Method	Sampling Intensity	P	Na	Ne	I	Ho	He	uHe	F
Whole collection	100%	0.563	13.619	4.282	1.497	0.553	0.657	0.657	0.149
MLS	10%	0	9.619 *	4.851	1.637	0.531	0.723	0.729	0.269 *
	20%	0	10.810	4.807	1.610	0.535	0.706	0.709	0.238
	30%	0	11.476	4.677	1.584	0.536	0.693	0.695	0.222
	40%	0.017	12.095	4.578	1.565	0.537	0.684	0.685	0.209
	50%	0.026	12.571	4.505	1.547	0.539	0.676	0.677	0.196
Power Core	20%	0.180	13.619	4.830	1.625	0.552	0.685	0.688	0.191
E–NE	10%	0	9.333 *	4.483	1.558	0.470	0.694	0.700	0.320 **
	20%	0	10.857	4.639	1.576	0.507	0.691	0.694	0.259 *
	30%	0	11.476	4.545	1.564	0.520	0.686	0.688	0.234
	40%	0.008	12.048	4.505	1.556	0.524	0.679	0.681	0.221
	50%	0.023	12.286	4.489	1.537	0.531	0.671	0.672	0.202
A–NE	10%	0	9.333 *	4.478	1.563	0.476	0.699	0.704	0.317 *
	20%	0	10.857	4.676	1.584	0.510	0.693	0.696	0.260 *
	30%	0	11.429	4.525	1.566	0.522	0.687	0.689	0.233
	40%	0	12.000	4.499	1.553	0.523	0.679	0.680	0.223
	50%	0.03	12.381	4.476	1.539	0.530	0.672	0.673	0.205
O_CV	10%	0	12.667	4.579	1.598	0.560	0.672	0.677	0.163
	20%	0.149	13.619	4.586	1.562	0.563	0.666	0.669	0.142
	30%	0.247	13.619	4.501	1.541	0.560	0.667	0.669	0.151
	40%	0.314	13.619	4.270	1.509	0.556	0.658	0.660	0.146
	50%	0.366	13.619	4.275	1.504	0.553	0.659	0.66	0.149
O_He	10%	0.067	9.571 *	4.996	1.668	0.606	0.738	0.745	0.178
	20%	0.322	10.810	4.988	1.653	0.607	0.724	0.727	0.156
	30%	0.302	11.619	4.861	1.630	0.597	0.714	0.716	0.157
	40%	0.393	11.905	4.776	1.609	0.596	0.706	0.708	0.150
	50%	0.419	12.190	4.670	1.587	0.588	0.698	0.699	0.149
O_Shannon	10%	0.067	10.857	5.382	1.704	0.597	0.723	0.729	0.175
	20%	0.132	11.952	5.191	1.678	0.587	0.714	0.717	0.174
	30%	0.275	12.667	5.050	1.650	0.589	0.705	0.707	0.159
	40%	0.347	12.952	4.911	1.626	0.584	0.697	0.699	0.158
	50%	0.393	13.048	4.816	1.603	0.58	0.691	0.692	0.152
SANA	10%	0.033	10.810	4.665	1.538	0.542	0.664	0.670	0.169
	20%	0.116	11.952	4.508	1.530	0.549	0.660	0.663	0.154
	30%	0.258	12.095	4.367	1.517	0.554	0.663	0.665	0.158
	40%	0.289	12.667	4.316	1.506	0.554	0.659	0.660	0.150
	50%	0.373	13.048	4.326	1.503	0.551	0.656	0.657	0.149
SAGD	10%	0.131	9.476 *	4.526	1.574	0.570	0.699	0.705	0.180
	20%	0.174	10.810	4.591	1.558	0.558	0.686	0.689	0.179
	30%	0.187	11.476	4.490	1.550	0.568	0.682	0.684	0.159
	40%	0.368	12.000	4.478	1.544	0.565	0.679	0.680	0.161
	50%	0.376	12.429	4.461	1.532	0.566	0.674	0.675	0.151
R	10%	0.131	8.643 *	4.118	1.441	0.551	0.651	0.657	0.143
	20%	0.174	10.157	4.250	1.479	0.557	0.658	0.661	0.144
	30%	0.242	11.043	4.237	1.480	0.554	0.655	0.657	0.144
	40%	0.318	11.490	4.255	1.487	0.552	0.656	0.658	0.149
	50%	0.373	12.057	4.255	1.490	0.552	0.657	0.658	0.151

3.2. Construction of Core Collection

We comprehensively considered the homology of the core collections and the preservation of alleles, A–NE (E–NE) and CV were given different weights, and the core collections were constructed under a 30% sampling intensity (Table 2). The results showed that A–NE&CV73 and E–NE&CV55 could minimize A–NE and maximize E–NE under the

condition of complete preservation of alleles. Finally, two core collections were built using these two methods.

Table 2. Comparison of A–NE&CV and E–NE&CV under different weights. Bold font represents the optimal weight in the current method.

Method (Weight)	A–NE	CV	Method (Weight)	E–NE	CV
A–NE	0.3650	0.8392	E–NE	0.6712	0.8426
A–NE&CV 19	0.3676	1.0000	E–NE&CV 19	0.6545	1.0000
A–NE&CV 28	0.3676	1.0000	E–NE&CV 28	0.6540	1.0000
A–NE&CV 37	0.3676	1.0000	E–NE&CV 37	0.6541	1.0000
A–NE&CV 46	0.3676	1.0000	E–NE&CV 46	0.6542	1.0000
A–NE&CV 55	0.3678	1.0000	**E–NE&CV 55**	**0.6542**	**1.0000**
A–NE&CV 64	0.3677	1.0000	E–NE&CV 64	0.6557	0.9967
A–NE&CV 73	**0.3676**	**1.0000**	E–NE&CV 73	0.6580	0.9902
A–NE&CV 82	0.3670	0.9902	E–NE&CV 82	0.6627	0.9707
A–NE&CV 91	0.3660	0.9577	E–NE&CV 91	0.6682	0.9186

3.3. Evaluation of Core Collections

The genetic diversity parameters (Na, Ne, I, Ho, He, uHe, F) of the two core collections created based on A–NE&CV73 and E–NE&CV55 showed no significant difference compared with the whole seed orchard germplasms, including 100% of alleles of the original population. In addition, no first-degree relationship was detected in the core collection (Table 3). Principal coordinate analysis (PCoA) was used to analyze the two core collections and the germplasms of the entire Chinese fir seed orchard to determine the representativeness of the core collection. We detected that the core collections A–NE&CV73 and E–NE&CV55 were evenly distributed in the principal coordinates of the entire Chinese fir seed orchard resources (Figure 2).

Table 3. Comparison of genetic diversity of core collections, reserved collections, and whole collection. P, homogeneity (P = individuals with first-degree relationships within populations/the total number of populations). N, number of entries; Na, observed number of alleles; Ne, effective number of alleles; I, Shannon diversity index; Ho, observed heterozygosity; He, expected heterozygosity; uHe, unbiased expected heterozygosity; F, fixation index. A–NE&CV73-R (E–NE&CV55-R) represents the remaining entries except for the entries of A–NE&CV73 (E–NE&CV55).

Population	N	P	Na	Ne	I	Ho	He	uHe	F
Whole collection	607	0.563	13.619	4.282	1.497	0.553	0.657	0.657	0.149
A–NE&CV73	182	0	13.619	4.395	1.534	0.562	0.666	0.668	0.147
E–NE&CV55	182	0	13.619	4.601	1.583	0.527	0.683	0.685	0.224
A–NE&CV73-R	425	0.348	11.048	4.212	1.471	0.549	0.652	0.653	0.149
E–NE&CV55-R	425	0.541	10.952	4.122	1.446	0.565	0.644	0.645	0.111

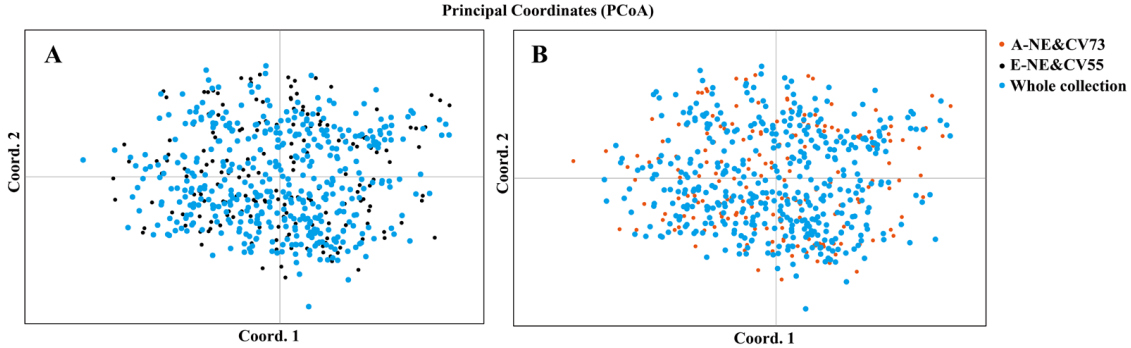


Figure 2. Principal coordinate analysis (PCoA) of constructed core collections and whole collection. (**A**) E–NE&CV55; (**B**) A–NE&CV73.

By comparing the distribution of accessions in the two core collections from nine seed orchards, ZJ1 (48) and HN2 (8) contributed the most and fewest accessions, respectively, to A–NE&CV73; ZJ1 (43) and HN3 (9) contributed the most and the fewest entries, respectively, to E–NE&CV55 (Figure 3A). In A–NE&CV73, the proportion of accessions contributed by FJ1 (30.89%) to the original population was slightly higher than that of FJ2 (23.91%) and FJ3 (25.37%). The same situation was also shown for ZJ (ZJ1, 40.00%; ZJ2, 26.51%; ZJ3, 35.48%). The proportion of accessions of HN3 (36.67%) was higher than that of HN1 (23.88%) and HN2 (20.51%) (Figure 3B). The coefficient of variation of the proportion of A–NE&CV73 accessions in their respective populations was 0.164. In E–NE&CV55, the proportion of accessions of FJ1, FJ2, and FJ3 in their respective populations were similar (FJ1, 28.46%; FJ2, 28.31%; FJ3, 31.34%). The proportion of accessions of HN1 was lower than that of HN2 and HN3, and the proportion of accessions of ZJ1 and ZJ3 in their population was higher than that of ZJ2 (Figure 3B). The coefficient of variation of accessions of E–NE&CV55 in their population was 0.232. Overall, the first-generation seed orchard contributed 30.48% of its entries to build A–NE&CV73, which was slightly higher than for the second-generation (28.14%) and the third-generation (29.52%) seed orchards. In the E–NE&CV55, the proportion of seed orchards contributing accessions increased with the generation (first, 27.90%; second, 30.17%; third, 32.27%).

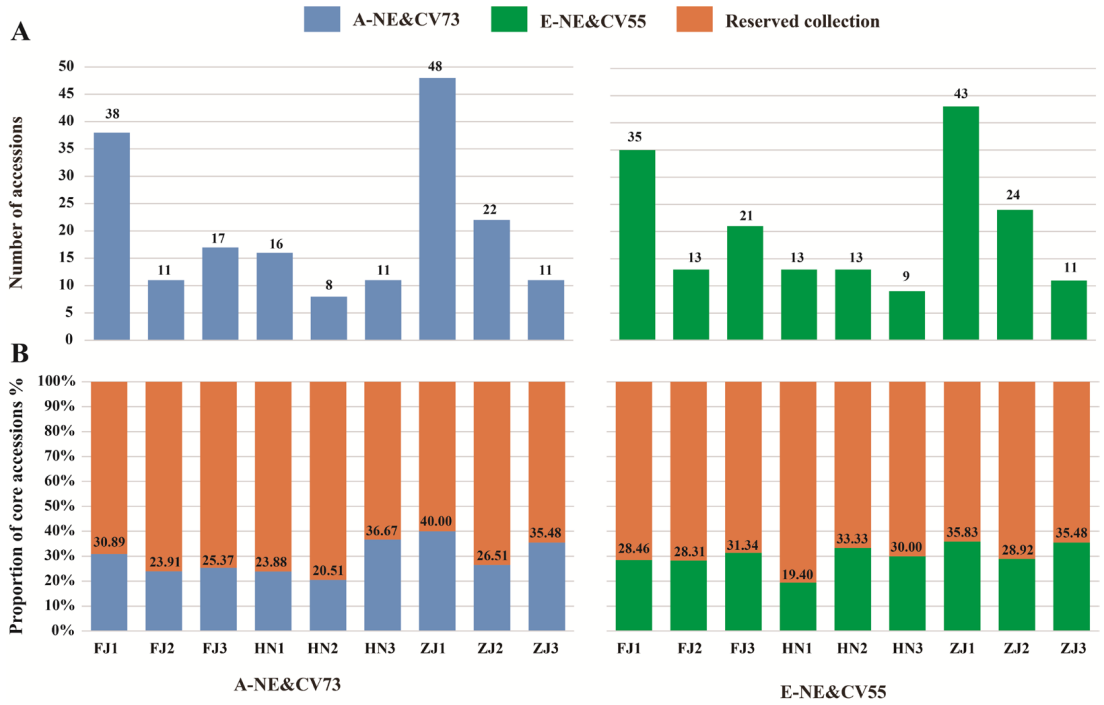


Figure 3. Distribution of entries of core collection in nine seed orchards. (**A**) Number of entries in the two core collections. (**B**) The proportion of entries in two core collections in their respective groups.

4. Discussion

The core collection has been proven to be a useful tool for the rapid screening of germplasm collections to obtain ideal traits [12]. For perennial trees, it is difficult to collect and preserve resources, and reliable phenotypic trait data need to be evaluated for a long period of time [36]. Molecular marker technology is less affected by the external environment and the growth and development stages of plants; therefore, it is more suitable to create core collections and evaluate genetic diversity [37,38]. The strategies for constructing the core collection from molecular marker data can be summarized into two categories [2]. The first category is based on the M strategy, which maximizes the number of alleles observed at each marker locus, reduces redundancy, and captures most of the genetic diversity [17,39,40]. This strategy is widely used in the core collection of plant

development [15,41–43]. The second category is based on the strategy of genetic distance, which stratifies or clusters the test materials, and then uses different allocation methods to select entries from each group [19,30,44]. The strategy based on distance can reduce the correlation between entries in the core collection. However, the strategy of building a core collection depends on the purpose of the core collection [2]. The category based on genetic distance is more suitable for plant breeding scholars [2], and includes items with low homology are elite materials for hybrid seed production and the creation of new germplasms. However, based on the M strategy of allele abundance, the retention of rare and local alleles attracts interested taxonomists and geneticists [2,39].

Ideally, a "good" core collection should represent the whole collection in terms of systematic classification and geographical origin, without redundant materials. Its scale should also be easy to manage and allocate [12,32]. The representativeness of type 1 and type 2 core collections based on A–NE and E–NE have been confirmed compared with those that were arbitrarily selected [32]. At present, core collections of strawberry (*F.* × *ananassa*) [12], hazelnut (*Corylus avellana* L.) [2], and other species [45,46] have been successfully constructed using this standard. Previous studies have included the establishment of the core collection of Chinese fir mainly based on geographic origin [21,47]. However, there may be complex genetic relationships in Chinese fir seed orchard germplasms [35] with the promotion of seed orchard generations, introduction, and cultivation between provinces and regions. Therefore, a core collection of Chinese fir seed orchards should limit the homology between accessions. In this study, we compared and analyzed nine methods for constructing core collections. E–NE, A–NE, and MLS performed well in homology, but some alleles were lost. Power Core and O_CV were superior to other methods in preserving alleles. They could preserve all alleles of the whole collection at a 20% sampling intensity, but they performed poorly in homology. Based on the performance of various methods of core construction under different sampling intensities, the minimum homology and maximum number of alleles could be maintained under a 30% sampling intensity. Therefore, we believe that 30% is the optimal sampling intensity for constructing a core collection of Chinese fir seed orchard germplasms. This sampling intensity is also consistent with popular views (20%–30%) [4,41,48]. We optimized both E–NE (A–NE) and O_CV according to the respective advantages of E–NE, A–NE, and O_CV methods. In general, a core collection optimized for multiple standards usually performs worse in a single metric [13,49]. However, compared with E–NE (A–NE), we obtained a slightly higher allele preservation rate (100%) and lower homology (0) achieved by simultaneously optimizing E–NE (A–NE) and O_CV at a 30% sampling intensity.

We observed that giving them different weights (A–NE, 0.7; CV, 0.3 and E–NE, 0.5; CV, 0.5) could effectively improve the preservation rate of alleles and control the correlation between accessions in the core collection, resulting in it being lower than that of the first-degree relationship. Therefore, optimization of the E–NE (A–NE) and O_CV strategies was selected as the best strategy to build the final core collections of Chinese fir seed orchards. The two core collections we constructed effectively represented the whole collection of germplasms (Figure 1). In addition, core collections provide the opportunity to study both the variability in the whole collection and differential genotype × environment interactions by testing a minimal number of accessions [13]. Their genotype data results in these two core collections greatly improving the resources of Chinese fir research and breeding. Although the core collections created represent the genetic diversity of the entire Chinese fir seed orchard resources to a large extent, the entries of the core collections and the method of constructing the core collections should be kept dynamic. In addition, they should be revised regularly to include the addition of new entries, information regarding new characterization methods (such as functional markers), and new methods aimed at improving their representativeness [2].

The entries of A–NE&CV73 and E–NE&CV55 were largely from the first-generation seed orchard, and the least came from the third-generation seed orchard, which was related to the initial population size of the seed orchards. The proportion of the entries of

core collections in the original population reflected the redundancy of these populations. In the two core collections we built, the largest proportion was ZJ1 (A–NE&CV73, 40%; E–NE&CV55, 35.84%), which may be related to the large sample size from ZJ1. Furthermore, the lowest proportions were HN1 (A–NE&CV73, 20.51%) and HN2 (E–NE&CV55, 19.40%), which may be related to the selection of parental clones in the HN seed orchard. Most (95.5%) of the parental clones of HN1 came from Huitong and Jingzhou, Hunan, China. Only a small number of parental clones were introduced from other regions. Additional supplementary selection measures already existed for HN2 and HN3. HN2 was supplemented by superior trees in other regions of Hunan province, and HN3 was expanded by super clones.

From the perspective of seed orchard generations, the largest proportion of the entries of A–NE&CV73 in the original population was from the first-generation seed orchards. This was to be expected, because the A–NE standard represented the genetic diversity of the whole collection [31]. Furthermore, the first-generation Chinese fir seed orchard preserved the highest genetic diversity, and the genetic diversity of the Chinese fir seed orchard decreased with an increase in generations [50]. In E–NE&CV55, the proportion of the items contributed by each seed orchard increased with an increase in generations, which may be related to the different standards for building the core collection. The collection based on the E–NE standards reflects the differences between the items in the core collection [32], indicating that the proportion of different individuals in the high-generation seed orchard of Chinese fir was high, and the redundancy was low. Although there were 78 identical entries (43%) in the two core collections, the function of the collections could meet the requirements of different breeding plans.

5. Conclusions

We compared and analyzed nine methods of constructing a core collection of Chinese fir seed orchard germplasms using SSR makers and explored the best sampling intensity of the core collections. The methods optimizing E–NE (A–NE) and CV were determined to be the best strategy for building the final core collection of Chinese fir seed orchards. Different weights were given to the E–NE (A–NE) and CV (A–NE, 0.7; CV, 0.3 and E–NE, 0.5; CV, 0.5), which could effectively improve the preservation rate of alleles and control the genetic relationship in the collection. Finally, we created two core collections from Chinese fir seed orchard resources, including all alleles of the whole collection, and no individuals in the core collection had first-degree relationships. This provides elite candidate materials for the advanced breeding of Chinese fir.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/f14020305/s1, Table S1: Passport information for the accessions evaluated and which accessions belong to the E–NE&CV55 or A–NE&CV73 core collection; Table S2: Sequence information of SSR primers.

Author Contributions: H.W.: Sampling survey, data analysis, visualization, and writing of original draft. A.D.: Writing review, project management, and funding acquisition. X.W.: Investigation and data collection. Z.C.: Investigation and writing—editing. X.Z.: Data collection and software. G.H.: Data collection and writing—editing. J.Z.: Writing—review, verification, and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported financially by Research on Breeding of High-Yield, High-Quality and High-Efficiency New Varieties of Chinese Fir, the National Key Research and Development Project of China of the 14th Five-Year Plan (Grant Number 2022YFD2200201).

Data Availability Statement: Data are available for research upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. 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]
2. Boccacci, P.; Aramini, M.; Ordidge, M.; van Hintum, T.J.L.; Marinoni, D.T.; Valentini, N.; Sarraquigne, J.-P.; Solar, A.; Rovira, M.; Bacchetta, L.; et al. Comparison of selection methods for the establishment of a core collection using SSR markers for hazelnut (*Corylus avellana* L.) accessions from European germplasm repositories. *Tree Genet. Genomes* **2021**, *17*, 48. [CrossRef]
3. Grenier, C.; Deu, M.; Kresovich, S.; Bramel, P.; Hamon, P. Assessment of genetic giversity in three subsets constituted from the ICRISAT Sorghum collection using random vs non-random sampling procedures. *Theor. Appl. Genet.* **2000**, *101*, 197–202. [CrossRef]
4. Hoshikawa, K.; Lin, Y.-P.; Schafleitner, R.; Shirasawa, K.; Isobe, S.; Nguyen, D.C.; Ohsawa, R.; Yoshioka, Y. Genetic diversity analysis and core collection construction for *Amaranthus tricolor* germplasm based on genome-wide single-nucleotide polymorphisms. *Sci. Hortic.* **2022**, *307*, 111428. [CrossRef]
5. Frankel, O.H. Genetic perspectives of germplasm conservation. In *Genetic Manipulation: Impact on Man and Society*; Arber, W.K., Illmensee, K., Peacock, W.J., Starlinger, P., Ehrlich, S.D., Dagert, M., Romac, S., Michel, B., Levy, S.B., Goebel, W., et al., Eds.; Cambridge University Press: Cambridge, UK, 1984; pp. 161–170.
6. Brown, A.H.D. Core collections—A practical approach to genetic-resources management. *Genome* **1989**, *31*, 818–824. [CrossRef]
7. Hu, J.; Zhu, J.; Xu, H.M. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theor. Appl. Genet.* **2000**, *101*, 264–268. [CrossRef]
8. Xiao, Y.; Liu, W.; Dai, Y.; Fu, C.; Bian, Y. Using SSR markers to evaluate the genetic diversity of Lentinula edodes' natural germplasm in China. *World J. Microbiol. Biotechnol.* **2010**, *26*, 527–536. [CrossRef]
9. Zhang, R.; Hu, D.; Gu, J.; Hu, X.; Zuo, X.; Wang, H. Development of SSR markers for typing cultivars in the mushroom *Auricularia Auricula-Judae*. *Mycol. Prog.* **2012**, *11*, 587–592. [CrossRef]
10. Lassois, L.; Denancé, C.; Ravon, E.; Guyader, A.; Guisnel, R.; Hibrand-Saint-Oyant, L.; Poncet, C.; Lasserre-Zuber, P.; Feugey, L.; Durel, C.-E. Genetic diversity, population structure, parentage analysis, and construction of core collections in the French apple germplasm based on SSR markers. *Plant Mol. Biol. Report.* **2016**, *34*, 827–844. [CrossRef]
11. Zane, L.; Bargelloni, L.; Patarnello, T. Strategies for microsatellite isolation: A review. *Mol. Ecol.* **2002**, *11*, 1–16. [CrossRef]
12. Zurn, J.D.; Hummer, K.E.; Bassil, N.V. Exploring the diversity and genetic structure of the U.S. National Cultivated Strawberry Collection. *Hortic. Res.* **2022**, *9*, uhac125. [CrossRef]
13. Díez, C.M.; Imperato, A.; Rallo, L.; Barranco, D.; Trujillo, I. Worldwide core collection of olive cultivars based on simple sequence repeat and morphological markers. *Crop Sci.* **2012**, *52*, 211–221. [CrossRef]
14. Bernard, A.; Barreneche, T.; Donkpegan, A.; Lheureux, F.; Dirlewanger, E. Comparison of structure analyses and core collections for the management of walnut genetic resources. *Tree Genet. Genomes* **2020**, *16*, 76. [CrossRef]
15. Yang, H.; Zhang, R.; Wang, B.; Xu, Z.; Zhou, Z. Construction of core collection of *Schima superba* based on SSR molecular markers. *Sci. Silvae Sin.* **2017**, *53*, 37–46.
16. Liu, K.; Muse, S.V. PowerMarker: An integrated analysis environment for genetic marker analysis. *Bioinformatics* **2005**, *21*, 2128–2129. [CrossRef]
17. Kim, K.-W.; Chung, H.-K.; Cho, G.-T.; Ma, K.-H.; Chandrabalan, D.; Gwag, J.-G.; Kim, T.-S.; Cho, E.-G.; Park, Y. PowerCore: A program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* **2007**, *23*, 2155–2162. [CrossRef]
18. De Beukelaer, H.; Smykal, P.; Davenport, G.F.; Fack, V. Core Hunter II: Fast core subset selection based on multiple genetic diversity measures using Mixed Replica search. *BMC Bioinform.* **2012**, *13*, 312. [CrossRef]
19. De Beukelaer, H.; Davenport, G.F.; Fack, V. Core Hunter 3: Flexible core subset selection. *BMC Bioinform.* **2018**, *19*, 203. [CrossRef]
20. Hong, J.; Cheng, B. Memoir of Chinese fir provenance test. *For. Res.* **1994**, *7*, 1–25. (In Chinese)
21. Duan, H.; Cao, S.; Zheng, H.; Hu, D.; Lin, J.; Cui, B.; Lin, H.; Hu, R.; Wu, B.; Sun, Y.; et al. Genetic characterization of Chinese fir from six provinces in Southern China and construction of a core collection. *Sci. Rep.* **2017**, *7*, 13814. [CrossRef]
22. Wu, H.; Lei, J.; Li, X.; Wang, H.; Duan, A.G.; Zhang, J. Aggregation distributions across stand age in provenances of *Cunninghamia lanceolata* (Lamb.) Hook. *For. Ecol. Manag.* **2021**, *494*, 119317. [CrossRef]
23. Shi, J. Present situation of genetic improvement of *Cunninghamia lanceolata* in Fujian Province and its technical countermeasures. *J. Fujian For. Sci. Technol.* **1994**, *21*, 28–31. (In Chinese)
24. Duan, A.; Sun, J.; Deng, Z.; Zhang, J. Studies on genetic variation and evaluation of hybrid offspring of the parents in the third generation seed orchards of Chinese fir. *Acta Agric. Univ. Jiangxiensis* **2022**, *44*, 261–270. (In Chinese) [CrossRef]
25. Duan, A.; Zhang, X.; Zhang, J.; Zhong, J. Growth and genetic evaluation of 21- year -old Chinese fir clonal plantation. *For. Res.* **2014**, *27*, 672–676. (In Chinese)
26. Wen, B. Discussion on selection methods and standards of superior trees of Chinese Fir. *Guangxi For. Sci.* **1990**, *3*, 24–28. (In Chinese)
27. Li, L.; Wu, B.; Hu, D.; Liang, S.; Shen, W.; Huang, Y.; Ruan, Z. Preliminary report on regionlization test of super Chinese fir clone. *Guangdong For. Sci. Technol.* **2004**, *20*, 5–8. (In Chinese)

28. Wen, Y.; Han, W.; Zhou, H.; Xu, G. SSR mining and development of EST-SSR markers for *Cunninghamia lanceolata* based on transcriptome sequences. *Sci. Silvae Sin.* **2015**, *51*, 41–49. (In Chinese) [CrossRef]
29. Li, Y.; Wang, Z.; Sui, J.; Zeng, Y.; Duan, A.; Zhang, J. Isolation and characterization of microsatellite loci for *Cunninghamia lanceolata* (Lamb.) Hook. *Genet. Mol. Res.* **2015**, *14*, 453–456. [CrossRef]
30. Perrier, X.; Jacquemoud-Collet, J.P. DARwin Software. 2006. Available online: http://darwin.cirad.fr/ (accessed on 21 October 2022).
31. Odong, T.L.; van Heerwaarden, J.; Jansen, J.; van Hintum, T.J.L.; van Eeuwijk, F.A. Statistical techniques for defining reference sets of accessions and microsatellite markers. *Crop Sci.* **2011**, *51*, 2401–2411. [CrossRef]
32. Odong, T.L.; Jansen, J.; van Eeuwijk, F.A.; van Hintum, T.J.L. Quality of core collections for effective utilisation of genetic resources review, discussion and interpretation. *Theor. Appl. Genet.* **2012**, *126*, 289–305. [CrossRef]
33. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef]
34. Wang, J. COANCESTRY: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol. Ecol. Resour.* **2011**, *11*, 141–145. [CrossRef]
35. Wu, H.; Zhao, S.; Wang, X.; Duan, A.; Zhang, J. Mating pattern and pollen dispersal in an advanced generation seed orchard of *Cunninghamia lanceolata* (Lamb.) Hook. *Front. Plant Sci.* **2022**, *13*, 1042290. [CrossRef]
36. El Bakkali, A.; Haouane, H.; Moukhli, A.; Costes, E.; Van Damme, P.; Khadari, B. Construction of core collections suitable for association mapping to optimize use of Mediterranean olive (*Olea europaea* L.) genetic resources. *PLoS ONE* **2013**, *8*, e61265. [CrossRef]
37. van Treuren, R.; Tchoudinova, I.; van Soest, L.J.M.; van Hintum, T.J.L. Marker-assisted acquisition and core collection formation: A case study in barley using AFLPs and pedigree data. *Genet. Resour. Crop Evol.* **2006**, *53*, 43–52. [CrossRef]
38. Zhang, Y.; Zhang, Q.L.; Yang, Y.; Luo, Z.R. Development of Japanese persimmon core collection by genetic distance sampling based on SSR narkers. *Biotechnol. Biotechnol. Equip.* **2009**, *23*, 1474–1478. [CrossRef]
39. Marita, J.; Rodriguez, J.; Nienhuis, J. Development of an algorithm identifying maximally diverse core collections. *Genet. Resour. Crop Evol.* **2000**, *47*, 515–526. [CrossRef]
40. Gouesnard, B.; Bataillon, T.; Decoux, G.; Rozale, C.; Schoen, D.; David, J. MSTRAT: An algorithm for building germ plasm core collections by maximizing allelic or phenotypic richness. *J. Hered.* **2001**, *92*, 93–104. [CrossRef]
41. Hu, J.; Wang, P.; Su, Y.; Wang, R.; Li, Q.; Sun, K. Microsatellite diversity, population structure, and core collection formation in melon germplasm. *Plant Mol. Biol. Report.* **2014**, *33*, 439–447. [CrossRef]
42. Liang, W.; Dondini, L.; De Franceschi, P.; Paris, R.; Sansavini, S.; Tartarini, S. Genetic diversity, population structure and construction of a core collection of apple cultivars from Italian germplasm. *Plant Mol. Biol. Report.* **2014**, *33*, 458–473. [CrossRef]
43. Öztürk, S.C.; Balık, H.İ.; Balık, S.K.; Kızılcı, G.; Duyar, Ö.; Doğanlar, S.; Frary, A. Molecular genetic diversity of the Turkish national hazelnut collection and selection of a core set. *Tree Genet. Genomes* **2017**, *13*, 113. [CrossRef]
44. van Hintum, T.J.L.; Brown, A.H.D.; Spillane, C.; Hodgkin, T. *Core Collections of Plant Genetic Resources*; IPGRI Technical Bulletin No. 3; International Plant Genetic Resources Institute: Rome, Italy, 2000.
45. Kaur, V.; Aravind, J.; Manju; Jacob, S.R.; Kumari, J.; Panwar, B.S.; Pal, N.; Rana, J.C.; Pandey, A.; Kumar, A. Phenotypic characterization, genetic diversity assessment in 6,778 accessions of barley *(Hordeum vulgare* L. ssp. *vulgare) germplasm conserved in national genebank of India and development of a core set. Front. Plant Sci.* **2022**, *13*, 771920. [CrossRef]
46. Risliawati, A.; Suwarno, W.; Lestari, P.; Trikoesoemaningtyas, T.; Sobir, S. A strategy to identify representative maize core collections based on kernel properties. *Genet. Resour. Crop Evol.* **2022**, 1–12. [CrossRef]
47. Zheng, H.; Hu, D.; Wei, R.; Yan, S.; Wang, R. Chinese fir breeding in the high-throughput sequencing era: Insights from SNPs. *Forests* **2019**, *10*, 681. [CrossRef]
48. Gong, F.; Geng, Y.; Zhang, P.; Zhang, F.; Fan, X.; Liu, Y. Genetic diversity and structure of a core collection of Huangqi (*Astragalus* ssp.) developed using genomic simple sequence repeat markers. *Genet. Resour. Crop Evol.* **2022**. [CrossRef]
49. Thachuk, C.; Crossa, J.; Franco, J.; Dreisigacker, S.; Warburton, M.; Davenport, G.F. Core Hunter: An algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinform.* **2009**, *10*, 243. [CrossRef] [PubMed]
50. Li, X.; Wang, L.; Wen, Y.; Lin, J.; Wu, X.; Yuan, M.; Zhang, Y.; Wang, M. Genetic diversity of of Chinese fir (*Cunninghamia lanceolata*) breeding population among different generations. *Sci. Silvae Sin.* **2020**, *56*, 53–61. (In Chinese)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.