

Article

# Comparative Transcriptome Analysis of Male and Female Conelets and Development of Microsatellite Markers in *Pinus bungeana*, an Endemic Conifer in China

Dong Duan <sup>†</sup>, Yun Jia <sup>†</sup>, Jie Yang and Zhong-Hu Li <sup>\*</sup>

Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an 710069, China; Dongduan@nwu.edu.cn (D.D.); jy878683@163.com (Y.J.); jieyang@nwu.edu.cn (J.Y.)

<sup>\*</sup> Correspondence: lizhonghu@nwu.edu.cn; Tel./Fax: +86-29-8830-2411

<sup>†</sup> These authors contributed equally to this work.

Received: 28 September 2017; Accepted: 12 December 2017; Published: 19 December 2017

**Abstract:** The sex determination in gymnosperms is still poorly characterized due to the lack of genomic/transcriptome resources and useful molecular genetic markers. To enhance our understanding of the molecular mechanisms of the determination of sexual recognition of reproductive structures in conifers, the transcriptome of male and female conelets were characterized in a Chinese endemic conifer species, *Pinus bungeana* Zucc. ex Endl. The 39.62 Gb high-throughput sequencing reads were obtained from two kinds of sexual conelets. After de novo assembly of the obtained reads, 85,305 unigenes were identified, 53,944 (63.23%) of which were annotated with public databases. A total of 12,073 differentially expressed genes were detected between the two types of sexes in *P. bungeana*, and 5766 (47.76%) of them were up-regulated in females. The Kyoto Encyclopedia of Genes and Genomes (KEGG) enriched analysis suggested that some of the genes were significantly associated with the sex determination process of *P. bungeana*, such as those involved in tryptophan metabolism, zeatin biosynthesis, and cysteine and methionine metabolism, and the phenylpropanoid biosynthesis pathways. Meanwhile, some important plant hormone pathways (e.g., the gibberellin (GA) pathway, carotenoid biosynthesis, and brassinosteroid biosynthesis (BR) pathway) that affected sexual determination were also induced in *P. bungeana*. In addition, 8791 expressed sequence tag-simple sequence repeats (EST-SSRs) from 7859 unigenes were detected in *P. bungeana*. The most abundant repeat types were dinucleotides (1926), followed by trinucleotides (1711). The dominant classes of the sequence repeat were A/T (4942) in mononucleotides and AT/AT (1283) in dinucleotides. Among these EST-SSRs, 84 pairs of primers were randomly selected for the characterization of potential molecular genetic markers. Finally, 19 polymorphic EST-SSR primers were characterized. We found low to moderate levels of genetic diversity ( $N_A = 1.754$ ;  $H_O = 0.206$ ;  $H_E = 0.205$ ) across natural populations of *P. bungeana*. The cluster analysis revealed two distinct genetic groups for the six populations that were sampled in this endemic species, which might be caused by the fragmentation of habitats and long-term geographic isolation among different populations. Taken together, this work provides important insights into the molecular mechanisms of sexual identity in the reproductive organs of *P. bungeana*. The molecular genetic resources that were identified in this study will also facilitate further studies in functional genomics and population genetics in the *Pinus* species.

**Keywords:** *Pinus bungeana*; sexual recognition; population structure; SSR marker; transcriptome

## 1. Introduction

In plants, totipotent meristematic cells usually experience a long development period, and then undergo the reproductive stage to form flowers, which are complex sexual organs [1]. The sexual dimorphism of plants (including hermaphrodite, monoecious, and dioecious, etc.) is related to morphological and physiological characteristics that differentiate male and female plant reproductive organs [2,3]. Generally, differential gene expression is considered as an important factor for sexually dichotomous phenotypes. The development and maintenance of sex-specific phenotypes are under a series of metabolic pathways and regulatory genetic networks where various connected sex differences in expression genes, transcription factors (TFs), and other regulators are associated [4,5]. Based on the recently detected genomic/transcriptome information resources, the morphological differences between the sexes are considered to be largely affected by the sex differences in the gene regulatory and expression pattern [6,7]. Some studies have suggested that the determinants of sex differentiation in plants (i.e., *Salix suchowensis* and cucumber) are significantly involved in the expression of sex chromosomes and sex determination genes [8,9]. Some other studies have found that plant hormones signal transduction (e.g., *ACS*, *ASR1*, *IAA2*, and *AUX* gene networks) also affected the gender differentiation and plant development process [10]. Recently, the transcriptome analysis for the complete flowers of cucumber showed that the genes participating in sexual differentiation were significantly related to the ethylene synthesis, carotenoid, and auxin biosynthesis pathways [8]. In addition, the downstream metabolic pathways and genetic networks that are essential for sex differentiation in plants may be controlled by upstream sex-determining genes [4]. However, up to now, many studies of the determinants of sexual identity have mainly focused on the model angiosperms species [7,11], and the molecular genetic mechanisms of sex recognition of gymnosperms are largely unclear.

In general, the gymnosperms have long generation times, large effective population sizes, and complex genomes, and these characteristics have hindered the accurate investigation and characterization of the sex differentiation genes at a genomics level [6,12]. In recent years, with the advance of next generation sequencing technology, transcriptome sequencing has proven to be an efficient and rapid method for determining the expression of different sexes. For example, expression pattern analysis has suggested that the *TERMINAL FLOWER 1 (TFL1)*-like genes played an important role in the development process of reproductive organs and sex determinations of gymnosperms [13]. In addition, comparative transcriptome analysis of the two *Pinus tabulaeformis* sexes indicated the occurrence of a sex-biased expression pattern in gymnosperms [7,11]. However, some other metabolism pathways that are affecting sex differentiation in gymnosperms still remain unknown.

*Pinus bungeana* Zucc. ex Endl., belonging to Sect. *Parrya* Mayr, is an economically and ecologically important soft conifer species, with a key role in local forest ecosystems. This species is mainly distributed in the mountain areas in Shaanxi, Shanxi, and Henan Provinces with an altitude of 500–1800 m, and is an endemic conifer in China. It is one of the main conifer species that have adapted to calcareous loess and mild saline soil in coniferous species. In addition, *P. bungeana* has strong resistance to sulfur dioxide and soot pollution in nature. In recent years, due to the over-cutting and fragmentations of natural habitats, the wild resources of *P. bungeana* populations have increasingly declined. Meanwhile, *P. bungeana* is outcrossing, anemophilous, and has different development positions of male and female conelets. These characteristics make this tree species an excellent model for studying differential gene expression and sexual recognition between the two sexes in conifer species.

Most previous studies on *P. bungeana* have mainly focused on its phylogenetic position [14–16], physiological ecology, and phylogeographic structure [17]. Meanwhile, some researchers have also investigated the genomic/transcriptome information and population genetics of *P. bungeana* [15,17]. On the other hand, the development and application of molecular genetic information is fundamental for the conservation of the wild species. In recent years, transcriptome sequencing provides a rapid and powerful tool for the development of molecular genetic tools, such as co-dominant simple sequence

repeats (SSR) markers [8]. These transcriptome-based markers have been widely used to research phylogenetic evolution and species conservation of organisms [18,19].

In this study, we characterized the transcriptomes of male and female conelets of *P. bungeana* using Illumina high-throughput sequencing technology. This study was designed to enrich the molecular genetic resources for *P. bungeana*, to identify new candidate genes that are involved in sex determination and differentiation, and to detail the different expression patterns between the sexes. Furthermore, differentially expressed genes (DEGs) that are involved in metabolisms were further determined and analyzed. In addition, a set of novel expressed sequence tag-simple sequence repeats (EST-SSR) markers were developed from the transcriptome data. To effectively manage the wild resources of *P. bungeana*, the population genetic diversity and structure were also evaluated for the six sampled populations across most ranges of its natural distributions, using these new developed EST-SSR markers.

## 2. Materials and Methods

### 2.1. Plant Materials

*Pinus bungeana* is a monoecious conifer species, which produces male and female cones during March–May. Three male and three female conelets (female conelet 1, female conelet 2, female conelet 3, male conelet 1, male conelet 2, and male conelet 3) were separately collected from three normal trees at the campus of Northwest University, China in April 2013. We selected expanded (but un-pollinated conelets), that were frozen in liquid nitrogen for RNA isolation. In addition, we sampled fresh needles from six natural populations (Table 1, Figure S1) across most of the distributional ranges of *P. bungeana* to investigate the polymorphism of microsatellite primers and population structure.

**Table 1.** Information on the geographical distribution of wild populations of *Pinus bungeana*.

Code	Location	Longitude (E)	Latitude (N)	Altitude (m)	N
W	Wuzi Mountain, Shaanxi	107.83	32.93	700	9
A	Ankang, Shaanxi	108.94	32.97	577	10
Y	Huozhou, Shanxi	111.87	36.60	900	8
Q	Qinyang, Henan	112.79	35.22	972	9
C	Taiyuan, Shanxi	112.32	37.72	1100	8
L	Mianyang, Sichuan	105.09	32.22	778	20

Note: N = number of individuals.

### 2.2. RNA Isolation and Quality Assay

The TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) was used to extract the total RNA of the samples, according to the manufacturer's instructions. The quantity and quality of RNA were assessed by 1% gel electrophoresis and NanoPhotometer<sup>®</sup> spectrophotometer (IMPLEN, Westlake Village, CA, USA). The RNA integrity was accurately assessed by the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA).

### 2.3. Transcriptome Sequencing

Individual isolated RNA was used to construct cDNA libraries for transcriptome sequencing. In brief, mRNA was enriched using the NEBNext Poly(A) mRNA Magnetic Isolation Module (E7490, NEB, Ipswich, UK) from 3 µg of total RNA. Double-stranded cDNA was synthesized, and sequencing adaptors were ligated according to the manufacturer's instructions (Illumina, San Diego, CA, USA). The ligated products were then purified with AMPureXP beads (Beckman Coulter, Brea, CA, USA) and were amplified for the construction of cDNA libraries. Library insert sizes ranged from 100 to 200 bp. The completed libraries were sequenced on the Illumina HiSeq 2000 platform. The original RNA-seq data was deposited into the Sequence Read Archive (SRA) of the National

Centre of Biotechnology Information (NCBI) under the accession numbers SRR6015000, SRR6015001, and SRR5832159 to SRR58321162.

#### 2.4. Data Processing and Assembly

Clean data (clean reads) were generated by removing reads containing adapters and poly-A as well as low quality reads from the raw data (raw reads) [20]. All of the downstream analyses were based on clean data of high quality. De novo assembly was carried out with the program Trinity using clean reads [21]. The high-quality transcripts were obtained after filtering and assembling. These transcripts were utilized for the further process of sequence clustering with Corset to create unigenes [22]. Next, a follow-up analysis was carried out with these unigenes, which were assembled from scratch. The published transcriptomes of other *Pinus* species from the NCBI non-redundant (Nr) [23], nucleotide sequences (Nt), protein family (Pfam), protein sequence database (Swissprot, [24], Gene Ontology (GO) [25], euKaryotic Ortholog Groups (KOG) [26], and the Kyoto Encyclopedia of Genes and Genomes (KEGG) databases [27] were used as references to obtain functional annotation information of unigenes by performing BLASTX with a cut-off  $E$  value of the best hit of  $\leq 10^{-5}$ .

#### 2.5. Differential Expression Analysis (DEG)

Read counts of genes were calculated using RNA-Seq by Expectation-Maximization (RSEM) software [28]. The above results were translated into FPKM values (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced), which is currently the most commonly utilized method for estimating gene expression levels [29]. Differential gene expression between the two different libraries (female and male) was analyzed by DESeq [30]. The false discovery rate (FDR) was adjusted by  $q$ -values. The thresholds value of differentially expressed genes was set as  $q < 0.005$  and  $\log_2$  (fold change)  $> 1$ . The GO and KEGG functional enrichment analysis of DEGs was implemented to show the main biochemical and signal transduction pathways.

#### 2.6. Identification of Expressed Sequence Tag-Simple Sequence Repeats (EST-SSRs)

We used the program MicroSAteLLite (MISA) [31] to detect potential EST-SSRs from unigenes in *P. bungeana*. The determining criteria were as follows: at least five repeats for the di-, and four repeats for the tri-, tetra-, penta- and hexanucleotide motifs. In addition, we used the software Primer 3 [32] to design the EST-SSR primer pairs. The setting parameters were the product size of PCR ranging from 100 to 300 bp, GC content of the primer from 40–60%, the length of primer ranging from 18–25 bp, and melting temperature from 50–65 °C. PCR amplification of SSR markers was performed in a 10  $\mu$ L reaction volume, including 30–50 ng template DNA, 5  $\mu$ L 2 $\times$  Taq PCR Master Mix (Runde, Xi'an, China), 1  $\mu$ L 2.5  $\mu$ M of each primer, and 4  $\mu$ L sterile water. The PCR procedure included an initial denaturation of 94 °C for 4 min, followed by 32 cycles of 45 s at 94 °C, 40 s at an annealing temperature of 50–60 °C for each primer, and 45 s at 72 °C, ending with a final extension of 5 min at 72 °C. We detected and validated the PCR products using silver-stained nondenaturing polyacrylamide gels [33]. In addition, we determined the allele sizes of each SSR genetic marker using the program Quantity One (Bio-Rad, Hercules, CA, USA).

#### 2.7. Validation of EST-SSR Markers and Population Genetic Analysis

To assess the polymorphism of EST-SSR primers in *P. bungeana*, we randomly selected 84 pairs of primers to amplify genomic DNA. We used MICROCHECKER v2.2.3 to test the presence of null alleles for all loci [34]. The program ARLEQUIN v.3.11 [35] was implemented to detect loci violating assumptions of neutrality. The population genetic parameters, including the number of alleles ( $N_A$ ), observed heterozygosity ( $H_O$ ), and expected heterozygosity ( $H_E$ ) were calculated using the software POPGENE v.1.32 [36]. We also used the software GenAIEx 6 [37] to detect the Hardy-Weinberg equilibrium (HWE). The source of genetic variation was analyzed with AMOVA (Analysis of Molecular Variation) in ARLEQUIN v.3.11 [34]. The UPGMA (unweighted pair-group method with arithmetic

averaging) analysis that was based on Nei's (1987) [38] genetic distances among populations was performed using Mega software [39]. In addition, STRUCTURE was used to infer the population structure with an admixture model based on the Bayesian clustering approach [40]. The population genetic clusters ( $K$ ) ranged from 1–10, and 10 independent runs were performed for each  $K$  with 10,000 burn-in and 100,000 Markov chain Monte Carlo (MCMC) replicates. The most likely number of  $K$  clusters was estimated using the  $\Delta K$  statistics method with the program STRUCTURE HARVESTR [41].

### 3. Results

#### 3.1. Transcriptome Sequencing and De Novo Assembly of *Pinus bungeana*

A total of 39.62 Gb high quality sequencing data (clean reads) were generated from six samples (female conelet 1, male conelet 1, female conelet 2, male conelet 2, female conelet 3, and male conelet 3) in *P. bungeana*. Among the sequencing datasets of female and male types, more than 96.14% and 96.59% of bases had a  $q$ -value > 20. The mean GC contents were 44.63% and 44.33% for female and male conelets, respectively, which suggests that the results of sequencing were relatively good (Table 2).

**Table 2.** Results of quality statistics of transcriptome database of *Pinus bungeana*.

Sample	Raw Reads	Clean Reads	Clean Bases (bp)	Q20 (%)	Q30 (%)	GC (%)
Female 1	56000000	54951878	6.87G	96.14	92.25	44.26
Male 1	53506322	52712306	6.59G	96.64	93.36	44.05
Female 2	56000000	55262192	6.91G	96.74	93.48	44.28
Male 2	54024824	53089856	6.64G	96.63	93.28	44.32
Female 3	41215900	39528272	5.93G	96.69	91.84	45.35
Male3	46220380	44510796	6.68G	96.59	91.64	44.63

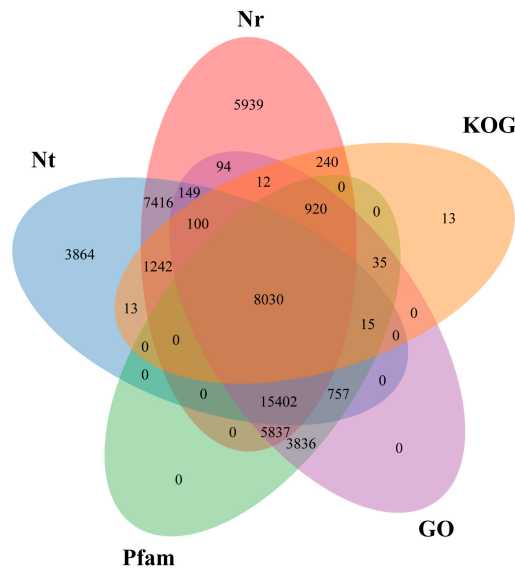
Note: Q20 = the percentage of bases with a phred value >20; Q30 = the percentage of bases with a phred value >30.

Using the Trinity software to assemble clean reads, we obtained a total 85,305 unigenes that were assembled with a mean length of 1199 bp and an N50 value of 1942 bp for *P. bungeana*, and the GC content was 44.48%. Most of the genes were relatively longer, with 34,343 (40.26%) unigenes greater than 500 bp. The size distribution of assembled unigenes was presented in Figure S2. The above results suggested that the quality of transcriptome sequencing and de novo assembly was relatively good, which were good enough to be used to carry out the subsequent bioinformatics analysis.

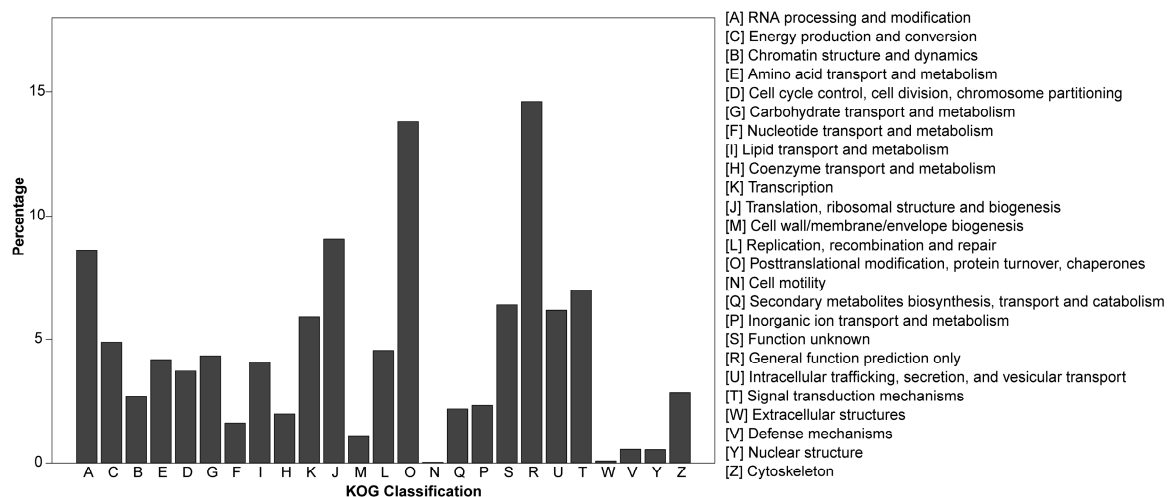
#### 3.2. Gene Annotation of *Pinus bungeana*

The assembled unigenes of *P. bungeana*, of which 53,944 (63.23%) were aligned and annotated with Nr, Nt, Pfam, Swissprot, KOG, GO, and KEGG databases, where 45,381 (53.19%), 36,988 (43.35%), and 34,319 (40.23%) unigenes had significant matches in Nr, Nt, and a manually annotated and Swissprot, respectively. Among all of them, 8030 unigenes could be common mapped by Nr, Nt, Pfam, GO, and KOG (Figure 1). The top-hit species in the annotated distribution was *Picea sitchensis* (18,903), followed by *Amborella trichopoda* (5203), and *Nelumbo nucifera* (3280) (Figure S3). The remaining 31,361 potential unigenes showed no homology to known sequences that are deposited in these databases.

Based on the alignments, a total number of 10,620 (12.44%) annotated unigenes were identified from KOG. The database represents an attempt at phylogenetic classification of proteins encoded in complete genomes. Among the 25 KOG categories, the cluster in the assembly of male and female libraries for "General function prediction only" represented the largest group, followed by the "Posttranslational modification, protein turnover, chaperones", "Translation, ribosomal structure and biogenesis" and "RNA processing and modification" clusters (Figure 2).



**Figure 1.** Venn diagram of annotation against the National Center for Biotechnology Information (NCBI) non-redundant (Nr), nucleotide sequences (Nt), Protein family (Pfam), Gene Ontology (GO) and euKaryotic Ortholog Groups (KOG) databases for the *Pinus bungeana* unigenes.



**Figure 2.** KOG functional classification of *P. bungeana* unigenes.

Furthermore, analysis of the GO categories showed that most of the unique sequences (35,187, 41.24%) were mapped to biological processes, followed by the molecular functions, and the least was the cellular component. The cellular process (19,470, 55.33%), binding (20,269, 57.60%), and cell (10,112, 28.74%) were the largest highly represented categories in these three main ontologies (Figure S4). All of the assembled unigenes were further annotated based on KEGG pathways. There were 15,721 (18.82%) unigenes that were divided and mapped into 19 functional pathways, with the 'Metabolism' cluster representing the largest group (Figure S5).

### 3.3. Analysis of Differently Expressed Genes in Male and Female Conelets of *Pinus bungeana*

We tested the global quality of the RNA-seq dataset by checking the reproducibility between each pair of samples (female and male conelets) [42], and found that reproducibility among the technical replicates of the gene expression was generally high (Figure S6).

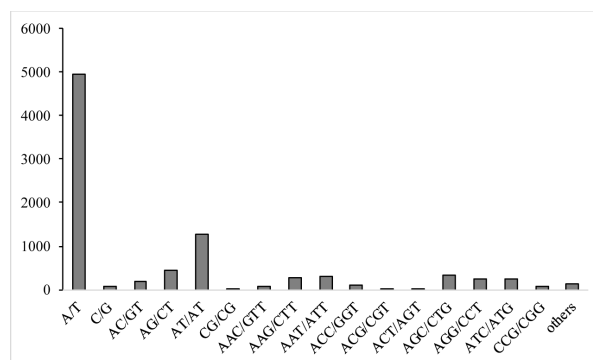
Through the screening of differentially expressed genes (standard = 2X and FDR < 0.05), 12,073 unigenes were identified as differentially expressed between male and female conelets, which

was comprised of 5766 unigenes that were up-regulated and 6307 unigenes that were down-regulated (Figure S7). Differential expression analysis indicated that there were more significant differences between the inter-gender than intra-gender variations by the three biological replicates from the female and male conelets, respectively (Figure S8). Moreover, we performed an enrichment analysis of GO and KEGG terms for these genes. For the GO term enrichment of DEGs, the six most highly represented terms were ‘metabolic process’, ‘catalytic activity’, ‘single-organism process’, ‘single-organism metabolic process’, ‘oxidation-reduction process’, and ‘oxidoreductase activity’ (Figure S9).

To further investigate the biological pathways that are active in sexual dimorphism, KEGG enrichment was performed based on the DEGs. The results indicated that 981 DEGs were significantly enriched in the top 20 KEGG pathways ( $q$ -value < 0.05), such as the 113 unigenes distributed in plant hormone signal transduction, the 132 unigenes allocated in starch and sucrose metabolism, and the 110 unigenes assigned to photosynthesis metabolism (Figure S10). DEGs in plant hormone signal transduction encoding the *Arabidopsis* response regulators (*ARR*), small auxin-up RNA (*SAUR*), and ethylene response factor (*ERF*) participated in the regulation of several hormone homeostasis and reproductive processes. These DEGs were involved in tryptophan metabolism (Ko00380), zeatin biosynthesis (Ko00908), cysteine, methionine (Ko00270), gibberellin (GA) pathway (Ko00904), carotenoid biosynthesis (Ko00906) and brassinolide (BR) pathway (Ko00905) (Figure S11, Table S2). Furthermore, we found that the female-biased cinnamoyl-CoA reductase (*CRR*) (EC: 1.2.1.44) was enriched in the phenylpropanoid biosynthesis pathways (Ko00940) (Figure S12, Table S2). The general phenylpropanoid components have been well shown to affect pollen development and male sterility [43]. Overall, these annotations provided a substantial resource for investigating specific processes, functions, and pathways during sexual determination.

#### 3.4. Polymorphism of EST-SSR Markers and Population Genetic Structure

With the purpose of developing novel molecular markers, the 85,305 unigenes of *P. bungeana* were used to mine for potential SSR markers. In total, 8791 SSRs were detected in 7859 unigenes. The most abundant repeat types were mononucleotides (5019), followed by dinucleotides (1926) and trinucleotides (1711). The dominant classes of the sequence repeat were A/T (4942) in the mononucleotides and AT/AT (1283) in the dinucleotides (Figure 3).



**Figure 3.** The numbers and motifs of simple sequence repeats (SSR) of *P. bungeana* unigenes.

To verify the reliability of these SSR primers, we randomly selected 84 pairs of primers to amplify the genomic DNA of *P. bungeana*. Sixty-four individuals of six natural populations of *P. bungeana* were sampled to determine the polymorphism of these microsatellite markers. Among these, 84 primer pairs, 40 resulted in successful PCR amplification and showed the predicted PCR products, 19 of which showed polymorphisms among 64 individuals (Table 1 and Table S1, Figure S1). We found no evidence for the existence of null alleles. Two EST-SSR loci (1314 and 22,642) departed significantly from the simulated  $F_{ST}$  distribution, indicating that they could be under disruptive selection or linked to a locus

under selection. These two loci were therefore removed in subsequent analysis. The mean number of alleles ( $N_A$ ), observed heterozygosity ( $H_O$ ), and expected heterozygosity ( $H_E$ ) ranged from 1.333–3, 0.033–0.492, and 0.067–0.455 for each loci, respectively (Table 3). For each population, the mean value of the number of alleles ( $N_A$ ), observed heterozygosity ( $H_O$ ), and expected heterozygosity ( $H_E$ ) ranged from 1.579–1.842, 0.112–0.256, and 0.168–0.240, respectively (Tables 3 and 4). The probabilities of deviation from the HWE proved that most of the EST-SSR markers significantly violated the HWE (Table 3). Population genetic differentiation was also significant across all loci ( $F_{ST} = 0.252^{***}$ ) (Table S3). An Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrogram that was based on Nei's genetic distance showed that all of the populations were divided into two different clusters, cluster I for populations W, A, Q, and Y, and cluster II, for populations C and L (Figures S13 and S14).

**Table 3.** The parameters of genetic diversity of the 17 expressed sequence tag-simple sequence repeats (EST-SSR) primers of *P. bungeana*.

Primers	N	$N_A$	$N_E$	$H_E$	$H_O$	P
5358	64	1.5	1.216	0.134	0.188	0.518
7309	64	1.833	1.528	0.294	0.208	0.021 *
24177	64	1.667	1.236	0.173	0.164	0.235
67970	64	1.333	1.148	0.097	0.126	0.634
10335	64	1.5	1.108	0.087	0.033	0.000 ***
73317	64	2	1.568	0.350	0.489	0.009 **
72763	64	1.5	1.32	0.189	0.208	0.033 *
66538	64	2	1.169	0.112	0.045	0.000 ***
60339	64	1.5	1.078	0.067	0.072	0.835
34533	64	1.667	1.494	0.262	0.413	0.049 *
33255	64	1.333	1.314	0.161	0.158	0.000 ***
10373	64	1.5	1.205	0.130	0.135	0.025 *
11371	64	1.667	1.191	0.117	0.056	0.000 ***
19808	64	1.833	1.583	0.339	0.492	0.008 **
7028	64	1.667	1.096	0.083	0.089	0.000 ***
6545	64	3	2.404	0.455	0.345	0.072
7029	64	1.833	1.559	0.321	0.240	0.988
Mean		1.754	1.383	0.205	0.206	

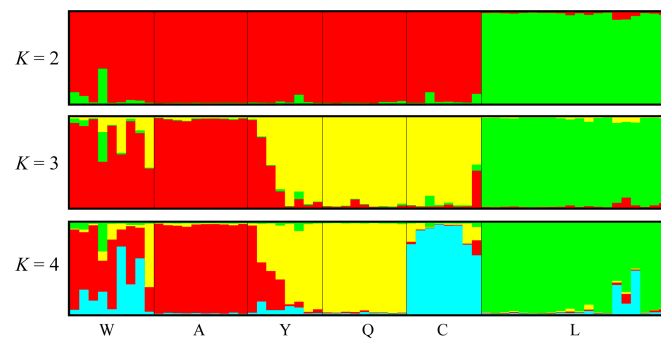
Note: N = number of individuals;  $N_A$  = number of alleles;  $N_E$  = number of effective alleles;  $H_E$  = expected heterozygosity;  $H_O$  = observed heterozygosity; P = Tests for Hardy-Weinberg Equilibrium (\*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$ ).

**Table 4.** Genetic diversity parameters of six natural populations of *P. bungeana*.

Population	N	$N_A$	$N_E$	$H_E$	$H_O$
W	9	1.789	1.450	0.231	0.231
A	10	1.632	1.310	0.189	0.256
Y	8	1.842	1.510	0.240	0.255
Q	9	1.842	1.436	0.221	0.187
C	8	1.579	1.292	0.168	0.197
L	20	1.842	1.298	0.181	0.112
Mean		1.754	1.383	0.205	0.206

For the population genetic structure of *P. bungeana*, the most likely population cluster K was two with the STRUCTURE analysis (Figure S13). The six populations of *P. bungeana* were assigned into two distinct groups: one group included the populations W, A, Y, Q, and C, while the other included only population L (Figure 4). At  $K = 4$ , samples from population C were further subdivided into an independent group. This is consistent with the UPGMA dendrogram.





**Figure 4.** Bayesian clustering analysis of population structure of *P. bungeana*.

## 4. Discussion

### 4.1. Transcriptome Characterization

Transcriptome sequencing is an effective and rapid method to identify genomic resources for non-model plants [7], especially for the conifer species, which possess complex and large genomes [6,13]. In this study, a total of 39.62 Gb clean reads were obtained from the transcriptome sequencing of the conifer species *P. bungeana*. The total number of assembled unigenes (85,305) was more than that of the other conifer species *P. tabuliformis* (46,584 unigenes) [11], also the N50 length of the unigenes, 1942 bp, was longer than that of *P. tabuliformis* (N50 = 744 bp) [11], which suggested that the assembly procedure for *P. bungeana* had good quality in this study. Intriguingly, few unigenes (1124, 2.48%) were annotated for an important *Pinus* species, *P. taeda*. This difference indicated the potential to discover novel genes that are specific to *P. bungeana*. Additionally, *Pinus taeda* belongs to Subgen. *Pinus*, whereas *P. bungeana* belongs to the Subgen. *Strobus* (Sweet) Rehd. The two *Pinus* species were divided into different groups, as well as different wild geographic distributions of two conifer species, which might lead to interspecific differences between them.

### 4.2. The Pathway Analysis of Differentially Expressed Genes

#### 4.2.1. The Plant Hormones Pathway Analysis

Plant hormones are endogenous regulators with multiple signal functions that affect nearly all aspects of plant growth and development [44–46]. Some studies have indicated that various phytohormones were likely to participate in the regulation of sex-determining genes and developmental pathways in unisexual flowers [47]. For example, some researchers have found that plant hormones and gibberellins (GAs) could promote early flowering in conifers and enhance the regularization of seed production [48–50]. In addition, the auxin-responsive protein (IAA), ethylene, and kinetin could also elicit a feminization effect on the sex of hemp [51]. In our study, the signaling pathways of several hormones, including auxin (Tryptophan metabolism, Ko00380), ethylene (cysteine and methionine metabolism, Ko00270), and cytokinin (zeatin biosynthesis, Ko00908), steroid hormones (brassinolide pathway, Ko00905) were enriched by pathway-based analysis (Figure S10). The characterization and future analysis of critical genes responsible for plant hormone production and signaling would greatly facilitate studies on the complex genetic network of sexual differentiation in *P. bungeana*.

#### 4.2.2. Tryptophan Metabolism

Tryptophan (TRP)- dependent metabolism (Ko00380) is considered to be one of the main pathways of auxin (IAA) biosynthesis [52]. The primary auxin response genes consisted of members of three gene families, the auxin influx carrier (*AUX/IAA*), small auxin up RNA (*SAUR*), and gretchen hagen3 (*GH3*) [53], which may participate in sex differentiation in *P. bungeana* (Figure S11, Table S2).

The auxin-signaling pathway, as mediated by *AUX1* and auxin response factor (*ARF*), was up-regulated in female conelets of *P. bungeana*. The auxin influx carrier (Cluster-2735.16335 (3.2832)) (belonging to the *AUX1 LAX* family) has been demonstrated to encode a high-affinity auxin influx carrier and plays a major role in many aspects of plant growth and development [54]. *ARF* (Cluster-2735.24659 (2.3325)), such as *ARF6* and *ARF8* were required to promote inflorescence stem elongation and late stages of petal, stamen, and gynoecium development in *Arabidopsis thaliana* [53]. In addition, it also has a conserved role in controlling the growth and development of vegetative and flower organs [55]. Therefore, the function of this auxin influx carrier (Cluster-2735.16335 (3.2832)) and auxin response factor (Cluster-2735.24659 (2.3325)) in sex differentiation is worthy of further study and exploration.

#### 4.2.3. Cysteine and Methionine Metabolism

Ethylene has an extensive regulation role in the plant growth and development process, especially in the aspect of sexual differentiation [56,57]. In this pathway (Ko00270), one unigene assembly (Cluster-2735.8266 (-1.8084)) was annotated to serine/threonine-protein kinase (*CTR1*), which showed an even closer similarity to the *A.s thaliana CTR1* gene (Figure S11, Table S2). The ethylene receptor (*ETR2*) acts upstream of *CTR1*, coding for a Raf-related protein kinase, which is ubiquitously expressed and has a higher level of expression in some tissues, including inflorescence, floral meristems, petals, and ovules [58].

#### 4.2.4. Zeatin Biosynthesis

Cytokinins (CTK) are another key plant hormone to affect plant gender expression, which comes from zeatin biosynthesis (Ko00908) (Figure S11). The cytokinin signal is perceived by three membrane-located receptors named *Arabidopsis* histidine kinase 2 (*AHK2*), *AHK3*, and *AHK4/CRE1* [59]. These receptors in the sporophyte are indispensable for anther dehiscence, pollen maturation, the induction of pollen germination by the stigma, and female gametophyte formation and maturation [60]. In our study, the histidine kinase 2-like (Cluster-2735.33324 (3.6499)), histidine kinase 3 (Cluster-2735.50860 (5.7972)), histidine kinase 4 isoform X2 (Cluster-2735.21997 (3.1149)), and the histidine kinase 4 isoform X1 (Cluster-2735.22398 (6.2557)) were identified and up-regulated in females. Furthermore, the histidine-containing phosphotransfer protein (*AHP*) (Cluster-2735.25394 (7.0138)) was also up-regulated in females, which was the mediator in a multistep phosphorelay pathway for cytokinin signaling [61]. It also negatively regulates the thickening of the secondary cell wall of the anther endothecium [61]. These results suggested that CTK was expected to play important roles of inducing female in the sexual differentiation of *P. bungeana*. A total of 15 unigenes encoding *ARRs* were enriched (Table S2). There was a functional overlap among the *ARRs*, which can act as positive regulators of cytokinin signal transduction [62]. These genes might be useful in identifying the system that induced sexual differentiation and possibly respond to the levels of cytokinin established *P. bungeana*.

#### 4.2.5. Other Important Plant Hormones

Carotenoid is a precursor of abscisic acid (*ABA*), which plays an extremely important role, as well as gibberellin, on regulating plant growth and development [63]. Research has shown that *ABA* has certain effects on affecting the plant sex expression [64,65]. As two types of receptors of *ABA*, *PYRABACTIN RESISTANCE1* (*PYR1*)/*PYR1-LIKE* (*PYL*) (Cluster-2735.52019 (-11.2) and Cluster-2735.52803 (-8.8097)) were up-regulated in males. In addition, Type 2C protein phosphatases (*PP2Cs*) (Cluster-2735.44601 (-7.1102) and Cluster-2735.38390 (-1.6378)) were also increased in males, which are vitally involved in *ABA* signaling (Figure S9, Table S2) [66]. *ABA* binds to *PYR1*, which, in turn, binds to and inhibits *PP2Cs* [66]. In summary, *ABA* perception by *PYR/PYLs* plays a major role in the regulation of seed germination and establishment, the basal *ABA* signaling that is required for vegetative and reproductive growth, stomatal aperture, and transcriptional response to the hormone [67].

Gibberellin (GA) derived from the diterpenoid biosynthesis pathway (Figure S11) is expected to influence plant sexual development. For example, in the monoecious species of *Buchloe dactyloides*, GA showed a dual effect with the induction of males and inhibition of females [68]. GA could increase the male-induced trait as the concentrations increased in the seedling of *Spinacia oleracea* [69]. Gibberellin is perceived by its nuclear receptors GA INSENSITIVE DWARF1s (GID1s), which then trigger the degradation of downstream repressors DELLAs [70]. In our study, the gibberellin receptor GID1 (Cluster-2735.41544 (4.7043)) was up-regulated in female conelets. The functional study of GID1 mutant combinations confirmed that GID1A plays a major role during fruit-set and growth, whereas GID1B and GID1C have specific roles in seed development and pod elongation, respectively [69]. GID1A was expressed throughout the whole pistil, while GID1B was expressed in ovules, and GID1C was expressed in valves. In our study, we observed that the gibberellin receptor GID1C (Cluster-2735.47107 (−3.2955)) were up-regulated in male conelets (Table S2). GA perception by GID1 causes slender rice1 (SLR1) protein degradation involving the F-box protein GID2; this triggers GA-associated responses, such as shoot elongation and seed germination [71]. There were four genes that were annotated as F-box proteins GID2 (Cluster-2735.64327 (1.9978), Cluster-2735.21997 (3.1149), Cluster-2735.60058 (Inf), and Cluster-4222.0 (Inf)) were up-regulated in female conelets.

The brassinosteroids (BR) are the most important discovery after GA as plant growth regulators [72]. BR signaling establishes an unexpected genetic pathway in the floral-regulating network [72]. Two genes encoding protein brassinosteroid insensitive 1 (BRI1) (Cluster-2735.63413 (2.7201) and Cluster-2735.44795 (1.7374)) of *P. bungeana* were involved in the brassinolide synthesis pathway and both showed high levels of expression in female conelets (Figure S11, Table S2). BRI1 was found to have the predominant function as a flowering-time enhancer, and also exhibited the elevated expression of potent floral repressor FLOWERING LOCUS C (FLC) [73].

Jasmonate (JA) was synthesized by the free  $\alpha$ -linolenic acid through the lipoxygenase (LOX) pathway [74], which is necessary for the normal development of floral organs. The biosynthesis and signal transduction process of jasmonate directly influences the developmental status of the flower organ. These studies have shown that jasmonate is mainly involved in the regulation of plant development of stamens in male flower organ [75–80]. Transcription factor MYC2 (Transcription factor MYC2-like, Cluster-2735.41360 (2.4305)) that is involved in the pathway of jasmonate synthesis was identified [80], which was up-regulated in female conelets in *P. bungeana* (Table S2). MYC2 is involved in JA-regulated plant development, lateral and adventitious root formation, flowering time, and shade avoidance syndrome [81]. Another of its notable functions is to regulate the crosstalk between the signaling pathways of JA and those of other phytohormones, such as ABA, GAs, and IAA [81].

This study preliminarily explored the co-expression of various hormones through multiple metabolic pathways and mechanisms in *P. bungeana*. In conclusion, each type of plant hormone did not independently regulate plant sexual differentiation, but a variety of hormones formed a network control mode to influence each other (Figure S11). Nearly all kinds of plant hormones could affect gender differentiation to a certain extent. However, it is important to note that the hormone role of plant sex expression is also not absolute. For example, plant hormones are hardly involved in *Silene* (*Silene latifolia*) flower development [82]. In addition, the regulation of plant hormones in female or male organs is related to the plant species, but the same hormones could have completely opposite effects in different plants [83,84].

#### 4.2.6. Photosynthesis Metabolism Pathway

We also found numerous genes in the phenylpropanoid biosynthesis pathway (Ko00940) that were differentially expressed between the two sexes (Figure S12, Table S2). Significant male expression occurred in genes encoding caffeoyl shikimate esterase (CSE, EC: 3.1.1), and ferulate-5-hydroxylase (F5H, EC: 1.2.1.44). Several genes in the phenylpropanoid pathway also exhibited significantly higher transcript abundance in females, including shikimate *O*-hydroxycinnamoyltransferase (HCT, EC: 2.3.1.133) and cinnamoyl-CoA reductase (CCR, EC: 1.2.1.44) (Figure S10, Table S2). Two genes

that were annotated as F5H (Cluster-2735.14069 (-3.2707) and Cluster-2735.56444 (-Inf)) were significantly up-regulated in male conelets. Over-expressing F5H in a line of *Arabidopsis* lacking caffeic acid *O*-methyltransferase induced a new type of lignin enriched 5-hydroxy-guaiacyl units, which had a profound impact on plant growth and development and cell-wall properties, and resulted in male sterility due to the complete disruption of the formation of the pollen wall [85]. The differential expression of numerous genes (cinnamoyl-CoA reductase, CCR) and enzymes (shikimate *O*-hydroxycinnamoyltransferase, HCT; caffeoyl shikimate esterase, CSE) that are involved in lignin biosynthesis and metabolism in the phenylpropanoid pathways could be used to further reveal the link between the biosynthetic pathways of lignin and the pollen wall-forming sporopollenin [43,86].

#### 4.3. Polymorphism of EST-SSR Markers and Population Structure of *Pinus bungeana*

In this study, 84 primer pairs of *P. bungeana* were randomly selected from 8791 primer pairs that were successfully designed for EST-SSR loci. Within these primer pairs, 40 (47.6%) resulted in successful PCR amplification and showed clear bands with the predicted PCR products, and 19 (22.6%) primers showed polymorphism among the sampled 64 individuals. As the *Pinus* species are generally diploid plants [8], the failed amplification of EST-SSR markers might be primarily due to the highly repetitive sequences and unsuitable primer sequences in conifer species (e.g., Reference [6]). Furthermore, the average number of alleles ( $N_A$ ) of 1.723 across the six populations of *P. bungeana* was lower than that reported for other gymnosperms (Table 4). These results suggested that the polymorphism level of the 19 EST-SSR markers that was developed in this study was low to moderate when compared with other conifer species [87,88].

Generally, the patterns of genetic variation and population structure are the basis of species evolution [89,90]. The evolutionary potential and adaptive ability of a natural species are largely dependent on the levels and distributions of its genetic diversity [91,92]. In this study, we detected a low level of genetic variability (an average value of  $H_O = 0.206$ ,  $H_E = 0.205$ ) at the species level for the endemic conifer. The reason for the low diversity may be due to the limited sample size of *P. bungeana* in this study. In addition, the dramatic changes of population size, genetic drift, and long generation time of species might have caused the low genetic diversity [93]. In a previous study, Yang et al. [17] concluded that the natural populations of *P. bungeana* experienced dramatic range fluctuations based on three nuclear genes. In the evolutionary history of *Pinus*, genetic drift and population size changes that were caused by climatic oscillations and geological events may have caused the low genetic variability. Meanwhile, the long generation time of the *Pinus* species and evolution process with low mutation rate may have led to the low level of diversity of *P. bungeana* [90]. On the other hand, in recent years, habitat fragmentation and human over-cutting may also have affected the genetic variation of natural populations of *P. bungeana* [17].

In addition, the management and protection of wild natural populations are very important to the endemic conifer species. Priority should be placed on conserving populations with high levels of genetic diversity; for example, population Y of *P. bungeana* ( $H_O = 0.255$ ,  $H_E = 0.240$ , with an average value of  $H_O = 0.206$ ,  $H_E = 0.205$ ), which had the highest diversity should be of the highest priority (Table 4, Figure 4). In addition, the cluster analysis revealed two distinct genetic groups for the sampled six populations in this endemic species, which may have been caused by the fragmentation of habitats and long term geographic isolation among the different populations. According to the STRUCTURE results, samples from population C were further subdivided into an independent group at  $K = 4$ , which was consistent with the UPGMA dendrogram. It seems that although a distinct genetic structure was identified in *P. bungeana*, we also need to further estimate genetic diversity and population structure to obtain accurate conclusions based on more samples from different natural populations.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2073-4425/8/12/393/s1](http://www.mdpi.com/2073-4425/8/12/393/s1). Figure S1: The geographical distributions of natural populations of *Pinus bungeana*. Figure S2: Length distributions of *Pinus bungeana* transcripts and unigenes. Figure S3: The number and distribution of homologous unigenes in *Pinus bungeana*. Figure S4: Gene Ontology (GO) classification of *Pinus bungeana* unigenes. Figure S5: KEGG

functional classification of *Pinus bungeana* unigenes. Figure S6: Spearman correlation matrix of experimental replications of transcriptome sequencing datasets of *Pinus bungeana*. Figure S7: Analysis of differential unigenes expression. In total, 12,073 unigenes were identified as differentially expressed between male and female conelets, which comprised 5766 unigenes that were up-regulated and 6307 unigenes that were down-regulated. Figure S8: Venn diagram of differential expression unigenes within three biological replications from male and female conelets, respectively. Figure S9: Analysis of differential unigenes expression. Figure S10: KEGG pathways enriched for differentially expressed genes of male and female conelets in *Pinus bungeana*. Figure S11: Plant hormone signal transduction in *Pinus bungeana* was adapted from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. Figure S12: Cysteine and Methionine metabolism pathway was adapted from the KEGG. Figure S13: Results of the Bayesian assignment analysis of *Pinus bungeana* using the program STRUCTURE. Figure S14: Dendrogram for natural six populations of *Pinus bungeana* based on 17 SSR loci. Table S1: Information of SSR Primers of *Pinus bungeana* used in this study. Table S2: Differently expressed genes in plant hormone signal transduction and photosynthesis metabolism pathways. Table S3: Results of the analysis of molecular variance (AMOVA) performed on six populations in *Pinus bungeana* using 17 microsatellite markers.

**Acknowledgments:** We are very grateful to Aureliano Bombarely and three anonymous reviewers for their constructive comments. This work was co-supported by the National Natural Science Foundation of China (41101058, 31470400) and the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT, No. IRT\_15R55).

**Author Contributions:** Z.-H.L. designed and conceived the work. D.D., Y.J. and J.Y. performed the experiments. Z.-H.L., D.D., Y.J. and J.Y. contributed materials and analysis tools. D.D. and Z.-H.L. wrote the manuscript. D.D., Y.J. and Z.-H.L. revised the manuscript. All authors finally approved the manuscript.

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

## References

- Walbot, V.; Evans, M.M.S. Unique features of the plant life cycle and their consequences. *Nat. Rev. Genet.* **2003**, *4*, 369–379. [[CrossRef](#)] [[PubMed](#)]
- Wellmer, F.; Graciet, E.; Riechmann, J.L. Specification of floral organs in *Arabidopsis*. *J. Exp. Bot.* **2014**, *65*, 1–9. [[CrossRef](#)] [[PubMed](#)]
- Harkess, A.; Mercati, F.; Shan, H.Y.; Sunseri, F.; Falavigna, A.; Leebens-Mack, J. Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytol.* **2015**, *207*, 883–892. [[CrossRef](#)] [[PubMed](#)]
- Chawla, A.; Stobdan, T.; Srivastava, R.B.; Jaiswal, V.; Chauhan, R.S.; Kant, A. Sex-biased temporal gene expression in male and female floral buds of seabuckthorn (*Hippophae rhamnoides*). *PLoS ONE* **2015**, *10*, e0124890. [[CrossRef](#)] [[PubMed](#)]
- Rocheta, M.; Sobral, R.; Magalhães, J.; Amorim, M.I.; Ribeiro, T.; Pinheiro, M.; Egas, C.; Morais-Cecílio, L.; Costa, M.M.R. Comparative transcriptomic analysis of male and female flowers of monoecious *Quercus suber*. *Front. Plant Sci.* **2014**, *5*, 599. [[CrossRef](#)] [[PubMed](#)]
- Nystedt, B.; Street, N.R.; Wetterbom, A.; Zuccolo, A.; Lin, Y.C.; Scofield, D.G.; Vezzi, F.; Delhomme, N.; Giacomello, S.; Alexeyenko, A.; et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* **2013**, *497*, 579–584. [[CrossRef](#)] [[PubMed](#)]
- Niu, S.; Yuan, H.; Sun, X.; Porth, I.; Li, Y.; El-Kassaby, Y.A.; Li, W. A transcriptomics investigation into pine reproductive organ development. *New Phytol.* **2016**, *209*, 1278–1289. [[CrossRef](#)] [[PubMed](#)]
- Guo, S.G.; Zheng, Y.; Joung, J.G.; Liu, S.; Zhang, Z.; Crasta, O.R.; Sobral, B.W.; Xu, Y.; Huang, S.; Fei, Z. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genom.* **2010**, *11*, 384. [[CrossRef](#)] [[PubMed](#)]
- Liu, J.; Yin, T.; Ye, N.; Chen, Y.; Yin, T.; Liu, M.; Hassani, D. Transcriptome analysis of the differentially expressed genes in the male and female shrub willows (*Salix suchowensis*). *PLoS ONE* **2013**, *8*, e60181. [[CrossRef](#)] [[PubMed](#)]
- Wu, T.; Qin, Z.; Zhou, X.; Feng, Z.; Du, Y. Transcriptome profile analysis of floral sex determination in cucumber. *J. Plant Physiol.* **2010**, *167*, 905–913. [[CrossRef](#)] [[PubMed](#)]
- Chen, J.; Uebbing, S.; Gyllenstrand, N.; Lagercrantz, U.; Lascoux, M.; Källman, T. Sequencing of the needle transcriptome from Norway spruce (*Picea abies* Karst L.) reveals lower substitution rates, but similar selective constraints in gymnosperms and angiosperms. *BMC Genom.* **2012**, *13*, 589. [[CrossRef](#)] [[PubMed](#)]
- Syring, J.; Farrell, K.; Businský, R.; Cronn, R.; Liston, A. Widespread genealogical nonmonophyly in species of *Pinus* subgenus. *Strobil. Syst. Biol.* **2007**, *56*, 163–181. [[CrossRef](#)] [[PubMed](#)]

13. Liu, Y.Y.; Yang, K.Z.; Wei, X.X.; Wang, X.Q. Revisiting the phosphatidylethanolamine-binding protein (PEBP) gene family reveals cryptic *FLOWERING LOCUS T* gene homologs in gymnosperms and sheds new light on functional evolution. *New Phytol.* **2016**, *212*, 730–744. [[CrossRef](#)] [[PubMed](#)]
14. Wang, X.R.; Tsumura, Y.; Yoshimaru, H.; Nagasaka, K.; Szmidt, A.E. Phylogenetic relationship of Eurasian pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. *Am. J. Bot.* **1999**, *86*, 1742–1753. [[CrossRef](#)] [[PubMed](#)]
15. Wang, B.S.; Wang, X.R. Mitochondrial DNA capture and divergence in *Pinus* provide new insights into the evolution of the genus. *Mol. Phylogenet. Evol.* **2014**, *80*, 20–30. [[CrossRef](#)] [[PubMed](#)]
16. Hao, Z.Z.; Liu, Y.Y.; Nazaire, M.; Wei, X.X.; Wang, X.Q. Molecular phylogenetics and evolutionary history of sect. *Quinquefoliae* (*Pinus*): Implications for Northern Hemisphere biogeography. *Mol. Phylogenet. Evol.* **2015**, *87*, 65–79. [[CrossRef](#)] [[PubMed](#)]
17. Yang, Y.X.; Wang, M.L.; Liu, Z.L.; Zhu, J.; Yan, M.Y.; Li, Z.H. Nucleotide polymorphism and phylogeographic history of an endangered conifer species *Pinus bungeana*. *Biochem. Syst. Ecol.* **2016**, *64*, 89–96. [[CrossRef](#)]
18. Zhang, L.; Yan, H.F.; Wu, W.; Yu, H.; Ge, X.J. Comparative transcriptome analysis and marker development of two closely related Primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genom.* **2013**, *14*, 329. [[CrossRef](#)] [[PubMed](#)]
19. Wu, J.; Cai, C.F.; Cheng, F.Y.; Cui, H.L.; Zhou, H. Characterisation and development of EST-SSR markers in tree peony using transcriptome sequences. *Mol. Breed.* **2014**, *34*, 1853–1866. [[CrossRef](#)]
20. Cox, M.P.; Peterson, D.A.; Biggs, P.J. Solexa QA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform.* **2010**, *11*, 485. [[CrossRef](#)] [[PubMed](#)]
21. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
22. Davidson, N.M.; Oshlack, A. Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* **2014**, *15*, 410. [[PubMed](#)]
23. Kitts, P.A.; Church, D.M.; Thibaud-Nissen, F.; Choi, J.; Hem, V.; Sapojnikov, V.; Smith, R.G.; Tatusova, T.; Xiang, C.; Zherikov, A.; et al. Assembly: A resource for assembled genomes at NCBI. *Nucleic Acids Res.* **2016**, *44*, D73–D80. [[CrossRef](#)] [[PubMed](#)]
24. Schneider, M.; Lane, L.; Boutet, E.; Lieberherr, D.; Tognolli, M.; Bougueleret, L.; Bairoch, A. The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. *J. Proteom.* **2009**, *72*, 567–573. [[CrossRef](#)] [[PubMed](#)]
25. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [[CrossRef](#)] [[PubMed](#)]
26. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36*, D480–D484. [[CrossRef](#)] [[PubMed](#)]
27. Mao, X.; Cai, T.; Olyarchuk, J.G.; Wei, J. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **2005**, *21*, 3787–3793. [[CrossRef](#)] [[PubMed](#)]
28. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
29. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M.J.; Salzberg, S.L.; Word, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [[CrossRef](#)] [[PubMed](#)]
30. Anders, S.; Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **2010**, *11*, R106. [[CrossRef](#)] [[PubMed](#)]
31. Thiel, T.; Michalek, W.; Varshney, R.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [[CrossRef](#)] [[PubMed](#)]
32. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—New capabilities and interfaces. *Nucleic Acids Res.* **2012**, *40*, e115. [[CrossRef](#)] [[PubMed](#)]
33. Chen, X.B.; Xie, Y.H.; Sun, X.M. Development and characterization of polymorphic genic-SSR markers in *Larix kaempferi*. *Molecules* **2015**, *20*, 6060–6067. [[CrossRef](#)] [[PubMed](#)]

34. Van Oosterhout, C.; Hutchinson, W.F.; Wills, D.P.; Shipley, P. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **2004**, *4*, 535–538. [[CrossRef](#)]
35. Excoffier, L.; Laval, G.; Schneider, S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **2007**, *1*, 47–50. [[CrossRef](#)] [[PubMed](#)]
36. Yeh, F.C.; Yang, R.C.; Boyle, T. *POPGENE. Microsoft Windows-Based Freeware for Population Genetic Analysis Release 1.31*; University of Alberta: Edmonton, AB, Canada, 1999.
37. Peakall, R.; Smouse, P.E. GenA1Ex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295. [[CrossRef](#)]
38. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987.
39. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis using Likelihood, Distance, and Parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)] [[PubMed](#)]
40. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959. [[PubMed](#)]
41. Earl, D.; vonHoldt, B. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **2012**, *4*, 359–361. [[CrossRef](#)]
42. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-cabrero, D.; Cervera, A.; Mcpherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.G.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 13. [[CrossRef](#)] [[PubMed](#)]
43. Van der Meer, I.M.; Stam, M.E.; van Tunen, A.J.; Mol, J.N.; Stuitje, A.R. Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. *Plant Cell* **1992**, *4*, 253–262. [[CrossRef](#)] [[PubMed](#)]
44. Gray, W.M. Hormonal regulation of plant growth and development. *PLoS Biol.* **2004**, *2*, e311. [[CrossRef](#)] [[PubMed](#)]
45. Papadopoulou, E.; Little, H.A.; Hammar, S.A.; Grumet, R. Effect of modified endogenous ethylene production on sex expression, bisexual flower development and fruit production in melon (*Cucumis melo* L.). *Sex. Plant Reprod.* **2005**, *18*, 131–142. [[CrossRef](#)]
46. Little, H.A.; Papadopoulou, E.; Hammar, S.A.; Grumet, R. The influence of ethylene perception on sex expression in melon (*Cucumis melo* L.) as assessed by expression of the mutant ethylene receptor, At-etr1-1, under the control of constitutive and floral targeted promoters. *Sex. Plant Reprod.* **2007**, *20*, 123–136. [[CrossRef](#)]
47. Diggle, P.K.; Di Stilio, V.S.; Gschwend, A.R.; Golenberg, E.M.; Moore, R.C.; Russell, J.R.W.; Sinclair, J.P. Multiple developmental processes underlie sex differentiation in angiosperms. *Trends Genet.* **2011**, *27*, 368–376. [[CrossRef](#)] [[PubMed](#)]
48. Pharis, R.P. Manipulation of flowering in conifers through the use of plant hormones. In *Modern Methods in Forest Genetics*; Springer: Berlin, Germany, 1976; pp. 265–282.
49. Pharis, R.P.; Kuo, C.G. Physiology of gibberellins in conifers. *Can. J. Forest Res.* **1997**, *7*, 299–325. [[CrossRef](#)]
50. Pharis, R.P.; Ross, S.D.; Wample, R.L.; Owens, J.N. Promotion of flowering in conifers of the *Pinaceae* by certain of the gibberellins. In *ISHS Acta Horticulturae 56: Symposium on Juvenility in Woody Perennials*; ISHS: Leuven, Belgium, 1976.
51. Galoch, E. The hormonal control of sex differentiation in dioecious plants of hemp (*Cannabis sativa*). The influence of plant growth regulators on sex expression in male and female plants. *Acta Soc. Bot. Pol.* **1978**, *47*, 153. [[CrossRef](#)]
52. Mano, Y.; Nemoto, K. The pathway of auxin biosynthesis in plants. *J. Exp. Bot.* **2012**, *63*, 2853–2872. [[CrossRef](#)] [[PubMed](#)]
53. Hagen, G.; Guilfoyle, T. Auxin-responsive gene expression: Genes, promoters and regulatory factors. *Plant Mol. Biol.* **2002**, *49*, 373–385. [[CrossRef](#)] [[PubMed](#)]
54. Péret, B.; Swarup, R. AUX/LAX genes encode a family of auxin influx transporters that perform distinct functions during Arabidopsis development. *Plant Cell* **2012**, *24*, 2874–2885. [[CrossRef](#)] [[PubMed](#)]
55. Liu, N.; Wu, S.; Van, H.J.; Wang, Y.; Ding, B.; Fei, Z.; Clarke, T.H.; Reed, J.W.; van der Knaap, E. Down-regulation of auxin response factors 6 and 8 by microRNA 167 leads to floral development defects and female sterility in tomato. *J. Exp. Bot.* **2014**, *65*, 2507–2520. [[CrossRef](#)] [[PubMed](#)]

56. Boualem, A.; Fergany, M.; Fernandez, R.; Troadec, C.; Martin, A.; Morin, H.; Sari, M.A.; Collin, F.; Flowers, J.M.; Pitrat, M.; et al. A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. *Science* **2008**, *321*, 836–838. [[CrossRef](#)] [[PubMed](#)]
57. Iqbal, N.; Khan, N.A.; Ferrante, A.; Trivellini, A.; Francini, A.; Khan, M.I.R. Ethylene role in plant growth, development and senescence: Interaction with other phytohormones. *Front. Plant Sci.* **2017**, *8*, 475. [[CrossRef](#)] [[PubMed](#)]
58. Sakai, H.; Hua, J.; Chen, Q.G.; Chang, C.; Medrano, L.J.; Bleecker, A.B.; Meyerowitz, E.M. Etr2 is an etr1-like gene involved in ethylene signaling in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5812–5817. [[CrossRef](#)] [[PubMed](#)]
59. Bartrina, I.; Jensen, H.; Novak, O.; Strnad, M.; Werner, T.; Schmölling, T. Gain-of-function mutants of the cytokinin receptors AHK2 and AHK3 regulate plant organ size, flowering time and plant longevity. *Plant Physiol.* **2017**, *173*, 1783–1797. [[CrossRef](#)] [[PubMed](#)]
60. Kinoshita-Tsujimura, K.; Kakimoto, T. Cytokinin receptors in sporophytes are essential for male and female functions in *Arabidopsis thaliana*. *Plant Signal. Behav.* **2011**, *6*, 66–71. [[CrossRef](#)] [[PubMed](#)]
61. Jung, K.W.; Oh, S.I.; Kim, Y.Y.; Yoo, K.S.; Cui, M.H.; Shin, J.S. *Arabidopsis* histidine-containing phosphotransfer factor 4 (AHP4) negatively regulates secondary wall thickening of the anther endothecium during flowering. *Mol. Cells* **2008**, *25*, 294–300. [[PubMed](#)]
62. Irish, E.E.; Nelson, T. Sex determination in monoecious and dioecious plants. *Plant Cell* **1989**, *1*, 737–744. [[CrossRef](#)] [[PubMed](#)]
63. Richards, D.E.; King, K.E.; Ait-Ali, T.; Harberd, N.P. How gibberellin regulates plant growth and development: A molecular genetic analysis of gibberellin signaling. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **2001**, *52*, 67–88. [[CrossRef](#)] [[PubMed](#)]
64. Friedlander, M.; Atsmon, D.; Galun, E. Sexual differentiation in cucumber: The effects of abscisic acid and other growth regulators on various sex genotypes. *Plant Cell Physiol.* **1977**, *18*, 261–269.
65. Rudich, J.; Halevy, A.H. Involvement of abscisic acid in the regulation of sex expression in the cucumber. *Plant Cell Physiol.* **1974**, *15*, 635–642. [[CrossRef](#)]
66. Park, S.Y.; Fung, P.; Nishimura, N.; Jensen, D.R.; Fujii, H.; Zhao, Y.; Lumba, S.; Santiago, J.; Rodrigues, A.; Chow, T.F.; et al. Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **2009**, *324*, 1068–1071. [[CrossRef](#)] [[PubMed](#)]
67. Gonzalez-Guzman, M.; Pizzio, G.A.; Antoni, R.; Vera-Sirera, F.; Merilo, E.; Bassel, G.W.; Fernández, M.A.; Holdsworth, M.J.; Perez-Amador, M.A.; Kollist, H.; et al. *Arabidopsis* PYR/PYL/RCAR receptors play a major role in quantitative regulation of stomatal aperture and transcriptional response to abscisic acid. *Plant Cell* **2012**, *24*, 2483–2496. [[CrossRef](#)] [[PubMed](#)]
68. Yin, T.; Quinn, J.A. Tests of a mechanistic model of one hormone regulating both sexes in *Cucumis sativus* (Cucurbitaceae). *Am. J. Bot.* **1995**, *82*, 1537–1546. [[CrossRef](#)]
69. Ei-Gizawy, A.M.; El-Oksh, I.; Sharaf, A.; El-Habar, M. Effect of gibberellic acid and alar on flowering and seed yield of spinach. *Egypt. J. Hortic.* **1992**, *19*, 191–200.
70. Gallego-Giraldo, C.; Hu, J.; Urbez, C.; Gomez, M.D.; Sun, T.; Perez-Amador, M.A. Role of the gibberellin receptors *GID1* during fruit-set in *Arabidopsis*. *Plant J.* **2014**, *79*, 1020–1032. [[CrossRef](#)] [[PubMed](#)]
71. Ueguchi-Tanaka, M.; Hirano, K.; Hasegawa, Y.; Kitano, H.; Matsuoka, M. Release of the repressive activity of rice DELLA protein SLR1 by gibberellin does not require SLR1 degradation in the *gid2* mutant. *Plant Cell* **2008**, *20*, 2437–2446. [[CrossRef](#)] [[PubMed](#)]
72. Benschop, J.J.; Millenaar, F.F.; Smeets, M.E.; van Zanten, M.; Voeselek, L.A.; Peeters, A.J. Abscisic acid antagonizes ethylene-induced hyponastic growth in *Arabidopsis*. *Plant Physiol.* **2007**, *143*, 1013–1023. [[CrossRef](#)] [[PubMed](#)]
73. Domagalska, M.A.; Schomburg, F.M.; Amasino, R.M.; Vierstra, R.D.; Nagy, F.; Davis, S.J. Attenuation of brassinosteroid signaling enhances *FLC* expression and delays flowering. *Development* **2007**, *134*, 2841–2850. [[CrossRef](#)] [[PubMed](#)]
74. Vick, B.A.; Zimmerman, D.C. The biosynthesis of jasmonic acid: A physiological role for plant lipoxygenase. *Biochem. Biophys. Res. Commun.* **1983**, *111*, 470–477. [[CrossRef](#)]
75. Stintzi, A.; Browse, J. The *Arabidopsis* male-sterile mutant, *opr3*, lacks the 12-oxophytodienoic acid reductase required for jasmonate synthesis. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10625–10630. [[CrossRef](#)] [[PubMed](#)]



76. Ishiguro, S.; Kawai-Oda, A.; Ueda, J.; Nishida, I.; Okada, K. The *DEFECTIVE IN ANther DEHISCENCE1* gene encodes a novel phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis, which synchronizes pollen maturation, anther dehiscence, and flower opening in *Arabidopsis*. *Plant Cell* **2001**, *13*, 2191–2209. [[CrossRef](#)] [[PubMed](#)]
77. Caldelari, D.; Wang, G.; Farmer, E.E.; Dong, X. *Arabidopsis lox3 lox4* double mutants are male sterile and defective in global proliferative arrest. *Plant Mol. Biol.* **2011**, *75*, 25–33. [[CrossRef](#)] [[PubMed](#)]
78. Park, W.; Li, J.J.; Song, R.; Messing, J.; Chen, X. CARPEL FACTORY, a dicer homolog, and HEN1, a novel protein, act in microRNA in metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **2002**, *12*, 1484–1495. [[CrossRef](#)]
79. Chung, H.S.; Howe, G.A. A critical role for the TIFY motif in repression of jasmonate signaling by a stabilized splice variant of the JASMONATE ZIM-domain protein JAZ10 in *Arabidopsis*. *Plant Cell* **2009**, *21*, 131–145. [[CrossRef](#)] [[PubMed](#)]
80. Song, S.; Qi, T.; Huang, H.; Ren, Q.; Wu, D.; Chang, C.; Peng, W.; Liu, Y.; Peng, J.; Xie, D. The Jasmonate-ZIM domain proteins interact with the R2R3-MYB transcription factors MYB21 and MYB24 to affect Jasmonate-regulated stamen development in *Arabidopsis*. *Plant Cell* **2011**, *23*, 1000–1013. [[CrossRef](#)] [[PubMed](#)]
81. Kazan, K.; Manners, J.M. MYC2: The master in action. *Mol. Plant* **2013**, *6*, 686–703. [[CrossRef](#)] [[PubMed](#)]
82. Dellaporta, S.L.; Calderon-Urrea, A. Sex determination in flowering plants. *Plant Cell.* **1993**, *5*, 1241–1251. [[CrossRef](#)] [[PubMed](#)]
83. Harvey, C.F.; Gill, G.P.; Fraser, L.G.; McNeilage, M.A. Sex determination in *Actinidia*. 1. Sex-linked markers and progeny sex ratio in diploid *A. chinensis*. *Sex. Plant Reprod.* **1997**, *10*, 149–154. [[CrossRef](#)]
84. Yang, Q.; Ren, L.; Du, G. Effects of ethephon, GA<sub>3</sub> and nutrient elements on sex expression of Chinese chestnut. *Sci. Horti.* **1985**, *26*, 209–215.
85. Weng, J.K.; Mo, H.; Chapple, C. Over-expression of F5H in COMT-deficient *Arabidopsis* leads to enrichment of an unusual lignin and disruption of pollen wall formation. *Plant J.* **2010**, *64*, 898–911. [[CrossRef](#)] [[PubMed](#)]
86. Matsuda, N.; Tsuchiya, T.; Kishitani, S.; Tanaka, Y.; Toriyama, K. Partial male sterility in transgenic tobacco carrying antisense and sense PAL cDNA under the control of a tapetum-specific promoter. *Plant Cell Physiol.* **1996**, *37*, 215–222. [[CrossRef](#)]
87. Liu, L.; Zhang, S.; Lian, C. De novo transcriptome sequencing analysis of cDNA library and large-scale unigene assembly in Japanese red pine (*Pinus densiflora*). *Int. J. Mol. Sci.* **2015**, *16*, 29047–29059. [[CrossRef](#)] [[PubMed](#)]
88. Sun, R.X.; Lin, F.R.; Huang, P.; Zheng, Y.Q. Moderate genetic diversity and genetic differentiation in the relict tree *Liquidambar formosana* Hance revealed by genic simple sequence repeat markers. *Front. Plant Sci.* **2016**, *7*, 1411. [[CrossRef](#)] [[PubMed](#)]
89. Mayr, E. *Systematics and the Origin of Species*; Columbia University Press: New York, NY, USA, 1942.
90. Hewitt, G.M. The genetic legacy of the Quaternary ice ages. *Nature* **2000**, *405*, 907–913. [[CrossRef](#)] [[PubMed](#)]
91. Gao, J.; Wang, B.; Mao, J.F.; Ingvarsson, P.; Zeng, Q.Y.; Wang, X.R. Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan plateau. *Mol. Ecol.* **2012**, *21*, 4811–4827. [[CrossRef](#)] [[PubMed](#)]
92. Wang, B.; Zhao, W.; Wang, X.R.; Mao, J.F.; Gao, J. Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Mol. Ecol.* **2011**, *20*, 3769–3811. [[CrossRef](#)] [[PubMed](#)]
93. Wright, S.I.; Gout, B.S. Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* **2005**, *22*, 506–519. [[CrossRef](#)] [[PubMed](#)]

