



Article Genome Wide Association Study Identifies Candidate Genes Related to the Earlywood Tracheid Properties in *Picea crassifolia* Kom.

Chengcheng Zhou ^{1,2,3}, Yingtian Guo ^{1,2,3}, Yali Chen ^{1,2,3}, Hongbin Zhang ⁴, Yousry A. El-Kassaby ⁵ and Wei Li ^{1,2,3,*}

- ¹ National Engineering Laboratory of Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China; zcc99610@bjfu.edu.cn (C.Z.); gytxiaozi@bjfu.edu.cn (Y.G.); yalichen123321@gmail.com (Y.C.)
- ² Key Laboratory of Genetics and Breeding in Forest Trees and Ornamental Plants of Ministry of Education, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China;
- ³ The Tree and Ornamental Plant Breeding and Biotechnology Laboratory of National Forestry and Grassland Administration, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China
- ⁴ Gansu Province Academy of Qilian Water Resource Conservation Forests Research Institute, Zhangye 734031, China; hongbinzhang123123@gmail.com
- ⁵ Department of Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada; y.el-kassaby@ubc.ca
- Correspondence: bjfuliwei@bjfu.edu.cn

Abstract: Picea crassifolia Kom. is one of the timber and ecological conifers in China and its wood tracheid traits directly affect wood formation and adaptability under harsh environment. Molecular studies on P. crassifolia remain inadequate because relatively few genes have been associated with these traits. To identify markers and candidate genes that can potentially be used for genetic improvement of wood tracheid traits, we examined 106 clones of P. crassifolia, and investigated phenotypic data for 14 wood tracheid traits before specific-locus amplified fragment sequencing (SLAF-seq) was employed to perform a genome wide association study (GWAS). Subsequently, the results were used to screen single nucleotide polymorphism (SNP) loci and candidate genes that exhibited a significant correlation with the studied traits. We developed 4,058,883 SLAF-tags and 12,275,765 SNP loci, and our analyses identified a total of 96 SNP loci that showed significant correlations with three earlywood tracheid traits using a mixed linear model (MLM). Next, candidate genes were screened in the 100 kb zone (50 kb upstream, 50 kb downstream) of each of the SNP loci, whereby 67 candidate genes were obtained in earlywood tracheid traits, including 34 genes of known function and 33 genes of unknown function. We provide the most significant SNP for each trait-locus combination and candidate genes occurring within the GWAS hits. These resources provide a foundation for the development of markers that could be used in wood traits improvement and candidate genes for the development of earlywood tracheid in P. crassifolia.

Keywords: Picea crassifolia; earlywood tracheid; SLAF-seq; SNP; GWAS

1. Introduction

Picea crassifolia Kom. (Qinghai spruce) is an evergreen conifer native to the Qilian Mountain areas of northwestern China that are known to be among the world's arid and semi-arid mountains. *P. crassifolia* is the most dominant tree species in this mountainous area and acts as a natural green reservoir for regulating and conserving water resources [1–3]. Drought is the primary factor for limiting tree growth in arid and semi-arid regions, greatly influencing xylem anatomical traits [4]. *P. crassifolia* is well adapted to these environmental conditions and produces high quality wood; however, the mechanism of wood formation



Citation: Zhou, C.; Guo, Y.; Chen, Y.; Zhang, H.; El-Kassaby, Y.A.; Li, W. Genome Wide Association Study Identifies Candidate Genes Related to the Earlywood Tracheid Properties in *Picea crassifolia* Kom. *Forests* **2022**, *13*, 332. https://doi.org/10.3390/ f13020332

Academic Editor: Jianquan Liu

Received: 17 January 2022 Accepted: 14 February 2022 Published: 18 February 2022

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



Copyright: © 2022 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/). under these conditions is still unclear. Studies conducted on P. crassifolia sampled from different altitudes of the Tibetan plateau, northwestern China, found that the development of tracheid radial diameter is closely related to temperature and precipitation, and trees could change their internal characteristics to adapt to changing climate [1,5]. Gymnosperms xylem development, especially tracheid radial growth, have a significant influence on wood formation and adaptability under different environments [6]. As tracheid morphology and cell wall structure influence wood and fibers flexibility, interactions among fibers, as well as the mechanical, physical, and optical properties of the end-products, are important [7]. In Pinus tabuliformis Carrière and P. crassifolia, the variation in tracheid diameter affects the duration of cell enlargement, and in turn influences their ability to adapt to drought [4]. However, no attempts have been made to explore the genetic factors underlying wood tracheid development variation. Thus, wood traits improvement and adaptability enhancement under harsh environments have become selection targets of tree breeding programs [8]. Therefore, uncovering the genetic basis of *P. crassifolia* wood tracheid traits is relevant for exploring the molecular mechanism of wood development in response to environmental changes.

The morphological characteristics of wood tracheid in conifers species, such as length, width, length-width ratio, wall thickness, lumen diameter, and wall-lumen diameter ratio, are an important basis for wood fiber utilization and key indicators in wood formation and development [9]. These are complex quantitative traits that are substantially affected by environmental factors, even though they are mainly controlled by genetic factors [10]. Although an earlier study identified quantitative trait loci (QTLs) for wood density variation in loblolly pine using restriction fragment length polymorphism (RFLP) marker genotypes [11], but previous methods were partially translated or not implemented in practical tree breeding due to QTLs are often family specific, generally explaining a small amount of phenotypic variation, and the need for very large QTL mapping families to recover desirable combinations of QTL alleles for more than five or six loci [12–14]. Fortunately, it has achieved success in some forest breeding that using genome wide association analysis (GWAS) to generally detect (as could be expected) markers closely related to the target traits with small effects at population level (across many families) in *Populus trichocarpa* Torr. & A.Gray ex Hook [15], Cryptomeria japonica (Thunb. ex L.f.) D.Don [16], Eucalyptus [17], white spruce [18], and oil palm [19], suggesting that GWAS is an appropriate research method to understand the genetics of complex traits in woody species [20,21].

Previous GWAS have mostly been based on SNP array technology, which can detect known SNP loci, however, new loci could not be accommodated [22]. In light of this apparent limitation, a high-throughput sequencing-based technology known as specific locus-amplified fragment sequencing (SLAF-seq) to offer ample SNP loci for such GWAS analyses is expected to overcome this limitation [23]. GWAS based on SLAF-seq has a series of advantages, including generating high-density SNP loci numbering in millions and detecting novel SNP loci in unknown mutation-harboring loci compared with SNP arrays [24]. Therefore, SNP loci generated from GWAS based on SLAF-seq provides an effective way to understand the genetic mechanism of *P. crassifolia* wood tracheid traits. More specifically, we developed SNP markers of these clones at the whole genome level using SLAF-seq to genotype P. crassifolia population comprising of 106 clones and utilized GWAS to determine those loci underpinning wood tracheid traits. The objectives of this study are to: (1) determine the genetic structure of this population, (2) identify SNP loci associated with wood tracheid traits, and (3) explore the candidate genes associated with wood tracheid traits. These results are expected to provide useful information for P. crassifolia breeding improvement.

2. Materials and Methods

2.1. Plant Materials

A diverse collection of 106 *P. crassifolia* clones was used for this study (Table S1). All 106 clones were 20 years old (planted in 1999 and sampled in 2019) growing in the Longqu

National Improved Variety Base, Zhangye City, Gansu Province, China ($100^{\circ}13'42''$ N, $38^{\circ}48'41''$ W). These clones were members of seven different clonal seed orchards and were classified according to their seed orchard sources: XS: Xi Shui, LC: Lian Cheng, DHS: Da Huangshan, HX: Ha Xi, DDS: Dong Dashan, LCH: Long Changhe, GC: Gu Cheng, DHK: Da Hekou, QL: Qi Lian, XYH: Xi Yinghe, and SDL: Shi Dalong. Within each orchard, clones were planted at 5×5 m within and between rows on the same soil type and managed similarly in a complete randomized block design with 18 replications.

2.2. Wood Tracheid Traits Phenotyping

A set of 14 wood tracheid traits were phenotyped and used as the research targeted traits (Table 1). From every clone, bark to pith 5 mm wood cores were extracted from the south and north directions at 1.3 m height. Tracheid length, diameter, and lumen diameter were measured by Motic 2.0 software under light microscope. A total of 15–20 tracheids were measured for each clone. Additionally, tracheid wall thickness, tracheid lumen–diameter ratio, tracheid wall–lumen ratio, and tracheid length–diameter ratio were calculated using software EXCEL 2020.

Table 1. List of the phenotypes, their abbreviations, and measurement unit.

Phenotype	Abbreviation	Unit
Earlywood tracheid length	EL	μm
Earlywood tracheid diameter	ED	μm
Earlywood tracheid lumen diameter	ELD	μm
Earlywood tracheid wall thickness	EWT	μm
Latewood tracheid length	LL	μm
Latewood tracheid diameter	LD	μm
Latewood tracheid lumen diameter	LLD	μm
Latewood tracheid wall thickness	LWT	μm
Earlywood tracheid lumen-diameter ratio	ELDR	
Earlywood tracheid wall-lumen ratio	EWLR	
Earlywood tracheid length-width ratio	ELWR	
Latewood tracheid lumen-diameter ratio	LLDR	
Latewood tracheid wall-lumen ratio	LWLR	
Latewood tracheid length-width ratio	LLWR	

Phenotypic traits indexes (mean value, standard deviation, coefficient of variation, skewness, kurtosis, and other parameters) were analyzed by SPSS 19.0 statistical software. The Origin software (version 8.2; https://www.originlab.com/https://www.originlab.com/, accessed on 20 August 2021) was used to display the frequency distribution of each trait. Correlations among traits were analyzed and visualized by psych and ggplot2 packages in R software 4.1.0.

2.3. DNA Isolation, Construction of SLAF Library, and Genome Sequencing

Total DNA was extracted from young needle tissues of each clone using a modified CTAB method [25]. Extracted genomic DNA concentration exceeded 20 ng/L, meeting the required quality for database construction. Due to the lack of *P. crassifolia* (L.) H.Karst. genome sequence, *Picea abies* genome was used as a reference genome (ftp://plantgenie.org/Data/ConGenIE/Picea_abies/v1.0/, accessed on 12 December 2020). The reference genome was 12 G in size with 37.88% GC content. The digestion of genomic DNA and the construction of SLAF-seq library according to the protocol by Biomarker Technologies Co. (Beijing, China). Clusters of libraries were loaded into an Illumina HiSeq for paired-end sequencing. The obtained clean reads were compared to the reference genome by BWA software [26], and the number of SLAF-tags and polymorphic SLAF-tags were counted. SNP markers were developed by GATK [27] and Samtools [28]. Afterwards, SNP heterozygosity was calculated

by PLINK software [29], the final SNP data with MAF < 0.05, heterozygosity > 0.02, and SNPs with more than two alleles were used to filter out.

2.4. Population Genetic Analyses and Linkage Disequilibrium

Based on the high-quality SNP markers, the software MEGA X [30] was used to construct the NJ tree among the 106 *P. crassifolia* studied clones. Principal component analysis (PCA) and calculation of a relative kinship matrix were performed using the software EIGENSTRAT [31] and GCTA [32], respectively. Additionally, population structure analysis using 12,275,765 SNPs to infer the genetic background of clonal cluster membership under a given number of populations (K) was carried out. The number of genetic clusters was predefined as K = 1-10for all clones and was calculated using Admixture software [33,34].

Linkage Disequilibrium (LD) was estimated by calculating the squared allele frequency correlation coefficient (r^2) between pairs of all SNP markers distributed throughout the genome using the vcftools software package [35]. The r^2 values were plotted against corresponding genetic distances in pairwise distance (bp). A nonlinear fitted-curve was drawn using second-degree locally weighted polynomial regression (LOESS) by applying the "loess" function in the R statistical program (http://www.r-project.org, accessed on 20 August 2021).

2.5. Association Genetics Analysis

A total of 12,275,765 SNPs from 106 clones were used in the GWAS. The efficient model was performed with both general linear model (GLM) and mixed linear model (MLM) using TASSEL [36], FaSTLMM [37], and EMMAX [38]. The population structure matrix generated from Admixture was used as the Q matrix for the GLM model, while the Q values and the K values of the kinship coefficient matrix were calculated by SPAGeDi [33] software using MLM model. *p*-values of $p \le 1.268 \times 10^{-7}$ (p = 0.01/n; n = total markers used, which is roughly a Bonferroni correction, corresponding to $-\log_{10}(p) = 7$) and $p \le 1.268 \times 10^{-8}$ (p = 0.1/n; n = total markers used, which is roughly a Bonferroni correction, corresponding to $-\log_{10}(p) = 7$) and $p \le 1.268 \times 10^{-8}$ (p = 0.1/n; n = total markers used, which is roughly a Bonferroni correction, corresponding to $-\log_{10}(p) = 7$) and $p \le 1.268 \times 10^{-8}$ (p = 0.1/n; n = total markers used, which is roughly a Bonferroni correction, corresponding to $-\log_{10}(p) = 8$) were defined as the genome wide control threshold and suggestive threshold, respectively. A standard interval of 100 kb (50 kb upstream and downstream) was explored for each candidate locus and adjusted according to the extent of local linkage disequilibrium with the candidate SNP ($R^2 < 0.8$). All candidate genes were annotated by GO, KEGG, NR, KOG, Pfam, and Swissprot databases. Manhattan plots were generated using the R package "CMplot" (https://github.com/YinLiLin/R-CMplot accessed on 20 August 2021).

3. Results

3.1. Sequencing Results

The SLAF-seq sequencing, resulted in a total of 1375.57 Mb of clean data from the 106 *P. crassifolia* clones. GC content ranged from 39.19 to 43.62%, with an average of 40.27%. The sequencing quality value Q30 ranged from 90.97 to 97.82%, with an average of 95.51% (Table S2). SNP detection was performed, based on 4,058,883 SLAF tags detected in the sequencing, producing a total of 12,275,765 SNP loci with a minor allele frequency >0.05 were generated, and SNP integrity of each sample ranged from 15.40 to 36.56%, and the SNP heterozygosity ranged from 5.41 to 10.99% (Table S3). Our analysis yielded a total of 1,573,899 polymorphic SLAF markers with a mean depth of 21.21 (Table S4). After quality control, a total of 12,275,765 SNPs were used for subsequent GWAS analyses.

3.2. Phenotypic Characterization of 14 Wood Tracheid Traits

Tracheid phenotypic data exhibited abundant variation among clones as the 106 clones originated from different forest regions (Figure 1). Coefficients of variation (CVs) of the 14 traits exhibited values ranged between 5.62 and 89.47%, with the largest and smallest CV values were observed for ELDR, and EWLR, respectively, and average CV values across the 14 traits was 17.54% (Table 2). The average CVs of earlywood tracheid properties were

relatively large, indicating that earlywood tracheid properties are easily affected by the environment. In general, the coefficient of variation in a half of 14 wood tracheid traits were more than 10%, indicating the present of large phenotypic variation among the studied clones.



Figure 1. Frequency distribution of 14 wood tracheid traits in 106 clones of *P. crassifolia*. The green lines indicate fitting curve using GaussAmp Non-linear fitting model.

Trait	Min	Max	Mean	SD	CV (%)	Skewness	Kurtosis
EL	2303.61	4131.40	3237.20	348.74	10.77	-0.42	-0.01
ED	44.37	66.46	54.14	5.26	9.71	-0.40	0.36
ELD	20.17	59.62	46.78	5.97	12.76	2.71	-0.42
EWT	5.28	30.29	7.36	2.40	32.64	80.35	8.41
ELDR	43.44	76.76	60.98	6.99	11.47	-0.38	-0.03
EWLR	0.41	0.91	0.86	0.05	5.62	69.75	-7.56
EDLR	0.10	1.74	0.17	0.16	89.47	98.73	9.77
LL	2434.23	4317.94	3505.45	339.43	9.68	0.28	-0.21
LD	34.20	47.83	40.45	2.57	6.34	0.39	0.07
LLD	14.82	26.64	20.83	2.45	11.76	-0.33	-0.01
LWT	15.21	23.69	19.62	1.90	9.66	-0.29	-0.26
LLDR	67.79	101.16	87.21	6.59	7.56	0.13	-0.44
LWLR	0.40	0.62	0.51	0.04	8.71	0.10	-0.04
LDLR	0.63	1.67	1.02	0.20	19.47	0.75	0.72

Table 2. Phenotype statistics of traits in P. crassifolia clones.

The absolute values of skewness and kurtosis of most tracheid traits in the studied population were less than 1, and most traits were normally distributed, indicating that these traits were quantitative traits controlled by multiple genes. Correlation analysis showed a highly significant relationship existed mainly among earlywood tracheid traits (Figure 2). In earlywood, significant correlations were observed among ED, EL, ELD, EWLR, ELWR, and EWT, ranging from -0.57 to 0.94 (p < 0.01), whereas ED was significantly positively correlated with ELD (traits correlation coefficient (r_t) = 0.94, p < 0.01), indicating that the diameter and lumen diameter of tracheid in development were increased simultaneously. Additionally, EWT was significantly negatively correlated with EWLR ($r_t = -0.86$, p < 0.01), suggesting that tracheid wall thickness decreased as the lumen diameter of tracheid in earlywood increased. Finally, similar correlation results were observed among latewood tracheid traits.



Figure 2. Correlation analysis of 14 wood tracheid traits. ** indicates that the correlation was significant at the p < 0.01 level. E and D indicate earlywood and latewood, respectively. L indicates tracheid length; D indicates tracheid diameter; LD indicates tracheid lumen diameter; WT indicates tracheid wall thickness; LWR indicates tracheid length–width ratio; WLR indicates tracheid wall-lumen ratio; and LDR indicates tracheid length–diameter ratio.

3.3. Analysis of Population Structure and Linkage Disequilibrium

The Admixture software was used to analyze population clustering and structure of the 106 clones (Figure 3a,b). Specifically, clustering was first performed assuming that the number of clusters (K) was between 1 and 10. Then, the results were cross-validated to determine that the optimal K-value was 1 (according to the valley of the error rates of cross-validation). In other words, our results implied that the collection most likely originated from the same ancestors. Principal component analysis showed that the first two principal components explained only 2.86% and 2.63% of the total variance (Figure 3c). The top 18 PCA components cumulatively contribute about 30% of the total marker variation. This means that the population structure of the association panel was weak and the population structure cannot be explained by a few principal components, which might be attributed to



the extensive exchanges of *P. crassifolia* breeding materials. Additionally, we observed family relationships along the diagonal with a scattered distribution of closely related individuals, and the remained part of the relationship matrix indicated low kinship (Figure 3d).

Figure 3. The analysis of population structure. (**a**) Population structure plot (ordinate represents cross-validation error rate (CV-value); *x*-axis represents number of clusters (K)). (**b**) Population clustering analyses by admixture software. (**c**) Principal component analyses. (**d**) Heatmap showing pairwise Kinship matrix. The *x*-axis and *y*-axis in QQ-plots indicate the kinship value.

The LD in *P. crassifolia* was estimated using all squared correlation coefficient (r^2) values and the physical distances between the same SNP pair. The nonlinear fitted-curve indicated that the LD is low in *P. crassifolia*, rapidly decaying by over 50% (from 0.50 to 0.10) (Figure 4). The average distance associated with the LD decline for $r^2 = 0.02$ varied roughly from 50 to 3000 bp. This result is expected and consistent with the trend of rapid LD decay in conifers, including Norway spruce [14] and loblolly pine [39]. Additionally, we detected high LD extending up to 100 kb. In woody plants, especially conifers, there are relatively high LD exist, for example, LD extend up to 140 kb and >145 kb in Norway spruce [14] and *Shorea platyclados* Slooten ex x Endert [40], respectively. This indicates that associations between phenotypic traits and markers in LD can be more easily and feasibly detected with GWAS than with analysis of quantitative trait loci (QTLs) [40].



Figure 4. Linkage disequilibrium (LD) decay in the 106 *P. crassifolia* clones. Pairwise LD (r^2) values plotted against the physical distance (bp) between all pairs of SNPs. The trend line of the nonlinear regressions against physical distance is given by the red line.

3.4. Association Analysis of 14 Wood Tracheid Traits

The GLM and MLM models identified more significant SNPs for the former than the later with some SNPs overlapped between the two models. However, the accuracy of MLM was better than that of GLM. A total of 96 SNP loci ($p < 1.11 \times 10^{-8}$) randomly distributed on 887,836 loci (Tables S5 and S6) of *Picea abies* genome [41] were identified on the MLM as significantly associated with three wood tracheid traits (ELDR, EWLR, and EWT). All SNP loci contributed to more than 10% to phenotypic variation. Among them, 9 were simultaneously detected for ELDR, EWLR, and EWT, and 11 SNP were simultaneously detected for ELDR, EWLR, supporting the observed phenotypic correlation and pleiotropic effect between phenotypes.

The regional Manhattan plots and quantile–quantile plots (QQ plots) of GLM and MLM for three earlywood tracheid traits (ELDR, EWLR, and EWT) and one latewood tracheid trait (LWLR) are presented in Figure 5. These plots include candidate genes within 100 kb of the significant SNP marker (50 kb upstream, 50 kb downstream), yielding a total of 67 candidate genes (Table 3), including 33 with unknown functions and 34 with predicted function annotation. Among these SNPs, some were highly associated with LWLR for the GLM, but not with the MLM. Three earlywood tracheid traits (ELDR, EWLR, and EWT) were associated with seven candidate genes (MA_10430313g0010, MA_10429843g0020, MA_10434936g0010, MA_11137g0010, MA_119933g0010, MA_17692g0010, and MA_19953g0020). The synonymous SNP MA_119933 is located on the gene MA_119933g0010 homologous with wall-associated receptor kinase, this ability to bind and respond to several types of pectins correlates with a demonstrated role for WAKs in both the pathogen response and cell expansion during plant development [42]. Therefore, the gene homologous with MA_119933g0010 in *P. crassifolia* may be involved in the cell wall development in earlywood tracheid.



Figure 5. The Manhattan plots and QQ plots of four highly associated wood tracheid traits based on the GLM and MLM. The red points indicate the significantly associated SNP loci $(-\log_{10}(p) \text{ values } > 6)$ and the green points indicate the moderately associated SNP loci $(6 > -\log_{10}(p) \text{ values } > 4)$. Loci 1 indicates 323,277 loci (MA_1~MA_199993); Loci 2 indicates 78,521 loci (MA_2~MA_29998); Loci 3 indicates 65,557 loci (MA_3~MA_399967); Loci 4 indicates 75,052 loci (MA_4~MA_49999); Loci 5 indicates 52,270 loci (MA_5000~MA_5999942); Loci 6 indicates 55,725 loci (MA_60~MA_69999); Loci 7 indicates 61,650 loci (MA_7000058~MA_79999); Loci 8 indicates 77,603 loci (MA_8~MA_899996); and Loci 9 indicates 98,009 loci (MA_90~MA_9999978) in Norway spruce genome. The QQ-plots inset—right with observed $\text{Log}_{10}(p)$ values on the *y*-axis and expected $\text{Log}_{10}(p)$ values on the *x*-axis.

Two earlywood tracheid traits, EWLR and EWT, were associated with two candidate genes (MA_10303051g0010 and MA_10883g0010). The gene MA_10883g0010 was homologous with EXORDIUM-like 2 (EXL2), which may play a role in a brassinosteroid-dependent regulation of growth and development, and the extracellular EXORDIUM protein mediates cell expansion in Arabidopsis leaves [43,44]. Therefore, the homologous gene of MA_10883g0010 in *P. crassifolia* may be involved in earlywood tracheid cell expansion. Additionally, a number of genes related to metabolism and transportation were associated highly with EWLR, including carboxylase, reductases, phosphatase, and transporters. This suggests that various physiological and biochemical reactions and material transportation in tracheid probably affect the earlywood tracheid wall–lumen ratio.

Trait	Gene ID	Loci	Gene-start	Gene-end	SNP location	Allele	Annotation
ELDR, EWLR, EWT	MA_10430313g0010	MA_10428216	15911	16486	8522	T/A	Serine/threonine-protein kinase
	MA_10429843g0020	MA_10429843	27084	33339	14836	G/T	Aminodeoxychorismate synthase
	MA_10434936g0010	MA_10434936	3785	12707	16639	T/C	Poly (ADP-ribose) glycohydrolase 1
	MA_11137g0010	MA_11137	25232	25570	17526	G/A	Histone H1
	MA_119933g0010	MA_119933	7320	8147	15072	C/A	Wall-associated receptor kinase
	MA_17692g0010	MA_17692	82061	82593	87850	G/A	Serine carboxypeptidase
	MA_19953g0020	MA_19953	81799	91867	38054	C/T	Pathogenesis-related protein 5
FWIR FWT	MA_10303051g0010	MA_10303051	23536	23793	2008	G/A	Multicopper oxidase
EWER, EWI	MA_10883g0010	MA_10883	53514	53744	17064	G/A	EXORDIUM-like 2
	MA_101170g0020	MA_101170	24857	25888	32445	C/T	Carboxylesterase 17
	MA_10427040g0010	MA_10427040	22428	23383	2504	A/G	Heat shock 22 kDa protein
	MA_10427391g0010	MA_10427391	21043	42330	58343	G/A	MAG2
	MA_10431643g0010	MA_10431643	1	11637	12312	A/G	Proteasome subunit alpha type-5
	MA_10432762g0010	MA_10432762	17888	18253	1144	G/A	Telomerase reverse transcriptase
	MA_10433801g0010	MA_10433801	551	15049	64877	C/T	E3 ubiquitin-protein ligase
	MA_10437109g0030	MA_10437109	94725	96103	91706	G/C	ABC transporter D family member 1
	MA_10437270g0010	MA_10437270	42477	60307	1265	G/T	IPG1
	MA_114300g0010	MA_114300	1314	3462	4171	A/G	STRUBBELIG-RECEPTOR FAMILY 8
	MA_174203g0010	MA_174203	12346	13811	5106	A/G	Sodium/metabolite cotransporter BASS1
	MA_17795g0010	MA_17795	42316	42738	57812	T/A	Histone H2B.1
FWLR	MA_182831g0010	MA_182831	20010	21544	31133	C/T	Hyoscyamine 6-dioxygenase
EWER	MA_20154g0020	MA_20154	67781	78419	57944	C/T	Phosphoenolpyruvate carboxylase
	MA_31g0010	MA_31	21393	40460	18526	A/G	Ubiquitin carboxyl-terminal hydrolase 12
	MA_34680g0010	MA_34680	1	23548	847	G/A	Peroxiredoxin-2F
	MA_39532g0010	MA_39532	4950	6096	29400	A/G	Copper transport protein CCH
	MA_417207g0010	MA_417207	3307	5128	5643	G/T	Purple acid phosphatase 3
	MA_482451g0010	MA_482451	1	1106	5136	C/T	Ribulose-1,5 bisphosphate carboxylase
	MA_5201g0010	MA_5201	4440	9434	12271	G/A	Proline transporter 2
	MA_588g0010	MA_588	18671	26200	22351	T/A	Abscisic acid hydroxylase 1
	MA_74441g0010	MA_74441	1775	11928	1579	A/T	Aldo-keto reductase family 4 member C9
	MA_80965g0010	MA_80965	17882	22398	39315	C/G	Flavin-containing monooxygenase 1
	MA_951514g0010	MA_951514	262	1867	3347	A/G	2-alkenal reductase
	MA_958g0010	MA_958	40918	71141	41171	G/T	MODIFIER OF SNC1 1
	MA_99302g0010	MA_99302	18687	19646	28542	C/T	Thioredoxin-like protein CDSP32

Table 3. The candidate genes of significantly associated regions for three earlywood tracheid traits on the MLM.

4. Discussion

In typical GWAS, phenotype and genotype data are collected for a large sample of assembled individuals [45]. However, a representative sample size combined with high-throughput sequencing and appropriate algorithms is sufficient to generate a relatively rich set of SNP and association loci in forest trees [46,47]. While it is possible that a detected genetic marker resides within a causative gene for the phenotype of interest, this is often not the case. Instead, GWAS rely on linkage disequilibrium (LD) between markers under testing and functional polymorphisms of causative genes [48,49]. The large number of SNPs in GWAS of conifers species is a reflection of their large and complex genomes and dramatically increased marker density would enable said markers to better track LD with causal variants in these large, genetically diverse genomes [50]. A large number of SNP loci can be identified in relatively small number of samples in conifers. For example, in lodgepole pine, more than 95,000 SNPs were obtained in 98 serotinous and nonserotinous samples from three populations [51]. De la Torre et al. [52] identified 799 significant associations of cold-related traits by GWAS in 217 samples in Douglas-fir. GWAS

in 194 maritime pines from different families provided the map position of 1671 SNPs corresponding to 1192 different loci [39]. In our study, the studied populations were mostly composed of individuals representing a species with narrow distribution range [53], we intentionally selected 106 individuals representing different natural forest populations that could represent the core germplasm of Qilian Mountain as determined by the little genetic relationship among them. The results of GWAS showed that there are abundant phenotypic variation and associated SNP loci. Thus, these obtained results demonstrate that sample size is not the most important factor in GWAS and the genetic relationship among samples and LD effects should be the focus of sample selection. Therefore, samples representing the core germplasm resources can be used as the materials for association analysis of the quantitative traits under the limited materials (i.e., small sample size).

Genomes of conifers are huge and complex and only few species have their genomes sequenced including Norway spruce [41], white spruce [54], and loblolly pine [55]. Traditional methods in developing molecular markers are usually cumbersome, time-consuming, and in most cases are unable to meet the experiment requirements. SLAF-seq-based GWAS is a fast and cost-effective approach to develop a large number of SNP markers in the absence of reference genomes. It is an effective method to determine molecular markers that influence essential traits in the absence of whole-genome markers [56,57]. As a consequence, SLAF-seq-based GWAS has been used in several conifers, for example, Bai et al. [58] used the SLAF-seq technique to analyze *Pinus massoniana* Lamb. germplasm resources in Guangdong Province, and identified 471,660 SNP markers in 599,164 SLAF polymorphic markers. Yang et al. [59] also used this approach and detected 524,662 high-quality SLAFs and identified 249,619 SNPs in hybrid clones of Taxodium species. In this study, a total of 4,058,883 SLAF-tags were detected and 12,275,765 SNP markers were developed and employed in GWAS to identify SNP loci associated with important traits and determined their candidate genes.

GLM and MLM are the most used algorithmic models in GWAS. The advantage of GLM is that it is more comprehensive and can identify more SNPs associations with the traits; however, its accuracy is lower than that of MLM [60–63]. In our study, the associated analysis results showed that the significantly associated SNPs in MLM ($p < 1.11 \times 10^{-7}$) were actually less than that of the GLM. Meanwhile, in the analysis of associated candidate gene, latewood tracheid trait LWLR was associated with few candidate genes in GLM model, but not in MLM model (Figure 5). This is because MLM model is more stringent than GLM model, and MLM model considers the influence of population structure (Q value) and genetic relationship (K value) and removes possible false associations [64]. Therefore, MLM can improve the accuracy of the analysis but can also miss some important SNP loci due to the strict screening conditions. Thus, multiple algorithmic models should be used to conduct GWAS data analysis to overcome this limitation [65,66].

Additionally, a total of 14 wood tracheid traits of *P. crassifolia* were used in GWAS. Due to the large amount of genome data and high levels of genome wide heterozygosity, the Norway spruce genome we used as reference genome did not assign identified association to the chromosome level. As the majority of the SNP loci were associated with earlywood tracheid traits, this is an indication of extensive variations in earlywood formation and maybe the presence of a more complex regulatory network. In temperate zones, earlywood forms in the spring, and the activity of cell cambium and physiological metabolism are vigorous due to suitable temperature and sufficient water [67,68]. As a result, the development of earlywood tracheid in *P. crassifolia* is expected to harbor more variation and thus SNP loci were easily associated with earlywood tracheid traits. Hence, QQ plots of the four highly associated earlywood tracheid traits (ELDR, EWLR, EWT, and LWLR) were generated to validate the accuracy of the population correction. The results of the QQ plots showed that, overall, the observed values did not match the expected values except for a few outliers at the beginnings. In other words, the highly associated SNP loci deviate from

expectation, it is considered that the deviation of the SNP observed value is caused by the genetic effect caused by the SNP mutation (i.e., true association) [69].

Next, candidate genes were screened in the 100 kb (50 kb upstream, 50 kb downstream) zone of each of the observed highly significant SNP loci. The MLM generated 67 potential candidate genes associated with three earlywood tracheid traits, ELDR, EWLR, and EWT (Figure 5). The multiple annotation databases results showed 34 candidate genes while the function of additional 33 candidate genes were unknown. The candidate gene MA_119933g0010 was associated with three earlywood tracheid traits (ELDR, EWLR, and EWT), and it is homologous with wall-associated receptor kinase. WAKs plays an important role in cell development, and our results indicated that the candidate SNP loci of gene MA_119933g0010 has a similar function in earlywood tracheid. Likewise, the gene MA_373300g0010 identified by the GWAS of the tracheid traits in Norway spruce is likely similar to WAKs [7]. In Arabidopsis, WAKs proteins are involved in cell wall expansion of leaf and regulation of leaf senescence [70–72]. Therefore, WAKs are probably key proteins and regulators in tracheid development of the spruce. The candidate gene MA_10883g0010 was associated with EWLR and EWT, and is homologous with EXORDIUM-like 2 that is active in mediating regulation of growth and development [43,44]. EXOORDIUM and WAKs proteins in P. crassifolia may have a significant effect in the development of earlywood tracheid. In addition, several candidate genes related to various enzymes and transporters were highly associated with EWLR, including carboxylesterase, phosphoenolpyruvate carboxylase, 2-alkenal reductase, ABC transporter D family member, sodium/metabolite cotransporter, and Proline transporter. The association of these enzymes and transporters also showed that metabolism and material transport are active in earlywood tracheid, and that the growth and development regulation of earlywood is important for *P. crassifolia*. We expect that these associated SNP loci with 67 candidate genes will provide essential genetic basis for earlywood tracheid traits improvement in P. crassifolia.

5. Conclusions

This work presents the first genome wide dissection of wood tracheid traits in *P. crassifolia*. A total of 67 highly significant SNP loci were associated with three earlywood tracheid traits, ELDR, EWLR, and EWT. These SNP loci have identified a set of candidate genes that could be exploited to improve earlywood tracheid traits for regulating development of earlywood tracheid for ecological adaptability of *P. crassifolia* under arid and semi-arid conditions. Previous research on the environmental effects on the development of tracheid in conifers mainly focused on morphology, and our study provided a major opportunity for understanding the underpinning of wood tracheid traits through using 12,275,765 SNPs for GWAS and extending the work to functional mapping approach. It is also worth mentioning that no associations were detected in latewood tracheid traits. The magnitude of variation earlywood tracheid traits was also higher than in latewood tracheid traits. These results showed that SNP mutations are more likely to affect growth and development of earlywood tracheid. Therefore, studies of wood development or drought adaptability of P. crassifolia, are more likely to benefit if focus is directed to the regulation and improvement of earlywood tracheid traits. Multiple SNP loci associated with the candidate genes related to cell and cell wall development are expected to provide a genetic basis for exploring and verifying the key mechanisms regulating earlywood tracheid development in P. crassifolia.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/f13020332/s1, Table S1: The information of 106 clones in *P. crassifolia;* Table S2: The sequencing result of *P. crassifolia;* Table S3: The SNP statistics of *P. crassifolia;* Table S4: The SLAF-tag statistics of *P. crassifolia;* Table S5: The information of SNP loci in *P. crassifolia;* Table S6: The locus information of reference genome; Table S7: The SRA metadata of SLAF-seq in *P. crassifolia.*

Author Contributions: Conceptualization, W.L., Y.A.E.-K. and C.Z.; methodology, C.Z., Y.G. and Y.C.; software, C.Z.; formal analysis, C.Z. and Y.G.; investigation, Y.C.; resources, H.Z.; writing—original draft preparation, C.Z.; writing—review and editing, W.L. and Y.A.E.-K.; visualization, C.Z.; and supervision, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: The project was funded by the National Natural Science Foundation of China (31770713, 31860221).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw sequencing data and the SLAF-sequencing data (SRA accession: SUB10523923), are available in the NCBI Sequence Read Archive (SRA) database under BioProject accession number PRJNA771805, the detailed information of SRA metadata is shown in Table S7. Other datasets supporting the conclusions of this article are included within the article and its additional files.

Acknowledgments: We are grateful for the generous grant from the National Engineering Laboratory of Tree breeding of Beijing Forestry University and Gansu Province Academy of Qilian Water Resource Conservation Forests Research Institute that made this work possible.

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

References

- 1. Xu, J.M.; Lu, J.X.; Evans, R.; Downes, G.M. Climatic signal in cellulose microfibril angle and tracheid radial diameter of *Picea crassifolia* at different altitudes of the Tibetan plateau, northwest China. *Wood Sci. Technol.* **2015**, *49*, 1307–1318. [CrossRef]
- 2. Tian, Q.Y.; He, Z.B.; Xiao, S.C.; Peng, X.M.; Ding, A.J.; Lin, P.F. Response of stem radial growth of Qinghai spruce (*Picea crassifolia*) to environmental factors in the Qilian Mountains of China. *Dendrochronologia* **2017**, *44*, 76–83. [CrossRef]
- 3. Zhu, T.Q.; Hu, J.W.; Qi, S.X.; Ouyang, F.Q.; Kong, L.S.; Wang, J.H. Transcriptome and morpho-physiological analyses reveal factors regulating cone bud differentiation in Qinghai spruce (*Picea crassifolia* Kom.). *Trees* **2021**, *35*, 1151–1166. [CrossRef]
- 4. Gao, J.N.; Yang, B.; Peng, X.M.; Rossi, S. Tracheid development under a drought event producing intra-annual density fluctuations in the semi-arid China. *Agric. For. Meteorol.* **2021**, *308–309*, 108572. [CrossRef]
- 5. Song, W.Q.; Mu, C.C.; Zhang, Y.D.; Zhang, X.; Li, Z.S.; Zhao, H.Y. Moisture-driven changes in the sensitivity of the radial growth of *Picea crassifolia* to temperature, northeastern Tibetan plateau. *Dendrochronologia* **2020**, *64*, 125761. [CrossRef]
- 6. Yamashita, S.; Yoshida, M.; Takayama, S.; Okuyama, T. Stem-righting mechanism in gymnosperm trees deduced from limitations in compression wood development. *Ann. Bot.* **2007**, *99*, 487–493. [CrossRef]
- 7. Baison, J.; Zhou, L.; Forsberg, N.; Mörling, T.; Grahn, T.; Olsson, L.; Karlsson, B.; Wu, H.X.; Mellerowicz, E.J.; Lundqvist, S.O.; et al. Genetic control of tracheid properties in Norway spruce wood. *Sci. Rep.* **2020**, *10*, 18089. [CrossRef]
- 8. Xu, J.M.; Lu, J.X.; Bao, F.C.; Evans, R.; Downes, G.M. Climate response of cell characteristics in tree rings of *Picea crassifolia*. *Holzforschung* **2013**, *67*, 217–225. [CrossRef]
- 9. Arend, M.; Fromm, J. Seasonal change in the drought response of wood cell development in Poplar. *Tree Physiol.* **2007**, 27, 985–992. [CrossRef]
- Wang, Q.T.; Zhao, C.Y.; Gao, C.C.; Xie, H.H.; Qiao, Y.; Gao, Y.F.; Yuan, L.M.; Wang, W.B.; Ge, L.J.; Zhang, G.D. Effects of environmental variables on seedling-sapling distribution of Qinghai spruce (*Picea crassifolia*) along altitudinal gradients. *Forest Ecol. Manag.* 2017, 384, 54–64. [CrossRef]
- Groover, A.; Devey, M.; Fiddler, T.; Lee, J.; Megraw, R.; Mitchel-Olds, T.; Sherman, B.; Vujcic, S.; Williams, C.; Neale, D. Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* 1994, 138, 1293–1300. [CrossRef]
- 12. Gailing, O. QTL analysis of leaf morphological characters in a *Quercus robur* full-sib family (*Q. robur* x *Q. robur* ssp. slavonica). *Plant Biol.* [CrossRef]
- 13. Heslot, N.; Akdemir, D.; Sorrells, M.E.; Jannink, J.L. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor. Appl. Genet.* **2014**, 127, 463–480. [CrossRef]
- Baison, J.; Vidalis, A.; Zhou, L.; Chen, Z.Q.; Li, Z.; Sillanpää, M.J.; Bernhardsson, C.; Scofield, D.; Forsberg, N.; Grahn, T.; et al. Genome-wide association study identified novel candidate loci affecting wood formation in Norway spruce. *Plant J.* 2019, 100, 83–100. [CrossRef]
- Porth, I.; Klapšte, J.; Skyba, O.; Hannemann, J.; McKown, A.D.; Guy, R.D.; DiFazio, S.P.; Muchero, W.; Ranjan, P.; Tuskan, G.A.; et al. Genome-wide association mapping for wood characteristics in populus identifies an array of candidate single nucleotide polymorphisms. *New Phytol.* 2013, 200, 710–726. [CrossRef]

- 16. Uchiyama, K.; Iwata, H.; Moriguchi, Y.; Ujino-Ihara, T.; Ueno, S.; Taguchi, Y.; Tsubomura, M.; Mishima, K.; Iki, T.; Watanabe, A.; et al. Demonstration of genome-wide association studies for identifying markers for wood property and male strobili traits in Cryptomeria japonica. *PLoS ONE* **2013**, *8*, e79866. [CrossRef]
- 17. Resende, R.T.; Resende, M.D.; Silva, F.F.; Azevedo, C.F.; Takahashi, E.K.; Silva-Junior, O.B.; Grattapaglia, D. Regional heritability mapping and genome-wide association identify loci for complex growth, wood and disease resistance traits in eucalyptus. *New Phytol.* 2017, 213, 1287–1300. [CrossRef]
- 18. Lamara, M.; Raherison, E.; Lenz, P.; Beaulieu, J.; Bousquet, J.; MacKay, J. Genetic architecture of wood properties based on association analysis and co-expression networks in white spruce. *New Phytol.* **2016**, *210*, 240–255. [CrossRef]
- 19. Ithnin, M.; Xu, Y.; Marjuni, M.; Mohamed, S.N.; Din, A.; Low, L.; Tan, Y.C.; Yap, S.J.; Ooi, L.C.L.; Nookiah, R.; et al. Multiple locus genome-wide association studies for important economic traits of oil palm. *Tree Genet. Genomes* **2017**, *13*, 103. [CrossRef]
- Fahrenkrog, A.M.; Neves, L.G.; Resende, M.F., Jr.; Vazquez, A.I.; de Los, C.G.; Dervinis, C.; Sykes, R.; Davis, M.; Davenport, R.; Barbazuk, W.B.; et al. Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phytol.* 2016, 213, 799–811. [CrossRef]
- Josephs, E.B.; Stinchcombe, J.R.; Wright, S.I. What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytol.* 2017, 214, 21–33. [CrossRef] [PubMed]
- 22. Zhang, T.; Hu, Y.; Wu, X.; Ma, R.; Jiang, Q.; Wang, Y. Identifying liver cancer-related enhancer SNPs by integrating GWAS and histone modification ChIP-seq data. *Biomed. Res. Int.* **2016**, 2395341. [CrossRef] [PubMed]
- Sun, X.; Liu, D.; Zhang, X.; Li, W.; Liu, H.; Hong, W.; Jiang, C.; Guan, N.; Ma, C.; Zeng, H.; et al. SLAF-seq: An efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* 2013, *8*, e58700. [CrossRef] [PubMed]
- 24. Geng, X.; Jiang, C.; Yang, J.; Wang, L.; Wu, X.; Wei, W. Rapid identification of candidate genes for seed weight using the SLAF-seq method in *brassica napus*. *PLoS ONE* **2016**, *11*, e0147580. [CrossRef]
- 25. Zhang, J.F.; Stewart, J.M. Economical and Rapid Method for Extracting Cotton Genomic DNA. J. Cotton Sci. 2000, 4, 193–201.
- Li, H.; Durbin, R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 2009, 25, 1754–1760. [CrossRef]
- McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010, 20, 1297–1303. [CrossRef]
- Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/Map format and SAMtools. *Bioinformatics* 2009, 25, 2078–2079. [CrossRef]
- Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007, *81*, 559–575. [CrossRef]
- Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 2018, 35, 1547–1549. [CrossRef]
- Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 2006, *38*, 904–909. [CrossRef] [PubMed]
- 32. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 2010, 88, 76–82. [CrossRef] [PubMed]
- Hardy, O.J.; Vekemans, X. SPAGeDi: A versatile computer program to analyze spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes.* 2002, 2, 618–620. [CrossRef]
- Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009, 19, 1655–1664. [CrossRef]
- 35. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; De Pristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef]
- Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 2007, 23, 2633–2635. [CrossRef]
- 37. Lippert, C.; Listgarten, J.; Liu, Y.; Kadie, C.M.; Davidson, R.I.; Heckerman, D. FaST linear mixed models for genome-wide association studies. *Nat. Methods* **2011**, *8*, 833–835. [CrossRef]
- Kang, H.M.; Sul, J.H.; Service, S.K.; Zaitlen, N.A.; Kong, S.Y.; Freimer, N.B.; Sabatti, C.; Eskin, E. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 2010, 42, 348–354. [CrossRef]
- 39. Neale, D.B.; Savolainen, O. Association genetics of complex traits in conifers. Trends Plant Sci. 2004, 9, 325–330. [CrossRef]
- Naoki, T.; Mohammad, N.; Widiyatno, S.I.; Kentaro, U.; Rempei, S.; Kevin, K.S.N.; Soon, L.L.; Yoshihiko, T. Potential of genomewide association studies and genomic selection to improve productivity and quality of commercial timber species in tropical rainforest, a Case Study of *Shorea platyclados. Forests* 2020, 11, 239. [CrossRef]
- Nystedt, B.; Street, N.R.; Wetterbom, A.; Zuccolo, A.; Lin, Y.C.; Scofield, D.G.; Vezzi, F.; Delhomme, N.; et al.; Giacomello, S.; Alexeyenko, A.; et al. The Norway spruce genome sequence and conifer genome evolution. *Nature* 2013, 497, 579–584. [CrossRef] [PubMed]

- 42. Kohorn, B.; Kohorn, S. The cell wall-associated kinases, waks, as pectin receptors. *Front. Plant Sci.* **2012**, *3*, 88. [CrossRef] [PubMed]
- 43. Majda, M.; Robert, S. The role of auxin in cell wall expansion. Int. J. Mol. Sci. 2018, 19, 951. [CrossRef] [PubMed]
- 44. Schröder, F.; Lisso, J.; Lange, P.; Müssig, C. The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. *BMC Plant Biol.* 2009, *9*, 20. [CrossRef]
- 45. Tibbs, C.L.; Zhang, Z.; Yu, J. Status and prospects of genome-wide association studies in plants. *Plant Genome* **2021**, *14*, e20077. [CrossRef]
- Allwright, M.R.; Payne, A.; Emiliani, G.; Milner, S.; Viger, M.; Rouse, F.; Keurentjes, J.J.B.; Bérard, A.; Wildhagen, H.; Faivre-Rampant, P.; et al. Biomass traits and candidate genes for bioenergy revealed through association genetics in coppiced european *Populus nigra* (L.). *Biotech. Biofuels* 2016, *9*, 195. [CrossRef]
- Cappa, E.P.; El-Kassaby, Y.A.; Garcia, M.N.; Acuña, C.; Borralho, N.M.; Grattapaglia, D.; Marcucci-Poltri, S.N. Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: A case study in *Eucalyptus globulus*. *PLoS ONE* 2013, *8*, e81267. [CrossRef]
- 48. Lipka, A.E.; Kandianis, C.B.; Hudson, M.E.; Yu, J.; Drnevich, J.; Bradbury, P.J.; Gore, M.A. From association to prediction: Statistical methods for the dissection and selection of complex traits in plants. *Curr. Opin. Plant Biol.* **2015**, 24, 110–118. [CrossRef]
- 49. Visscher, P.M.; Wray, N.R.; Zhang, Q.; Sklar, P.; McCarthy, M.I.; Brown, M.A.; Yang, J. 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **2017**, *101*, 5–22. [CrossRef]
- Thistlethwaite, F.R.; Gamal El-Dien, O.; Ratcliffe, B.; Klápště, J.; Porth, I.; Chen, C.; Stoehr, M.U.; Ingvarsson, P.K.; El-Kassaby, Y.A. Linkage disequilibrium vs. pedigree: Genomic selection prediction accuracy in conifer species. *PLoS ONE* 2020, 15, e0232201. [CrossRef]
- 51. Parchman, T.L.; Gompert, Z.; Mudge, J.; Schilkey, F.D.; Benkman, C.W.; Buerkle, C.A. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* 2012, *21*, 2991–3005. [CrossRef] [PubMed]
- De La Torre, A.R.; Wilhite, B.; Puiu, D.; St Clair, J.B.; Crepeau, M.W.; Salzberg, S.L.; Langley, C.H.; Allen, B.; Neale, D.B. Dissecting the polygenic basis of cold adaptation using genome-wide association of traits and environmental data in Douglas-fir. *Genes* 2021, 12, 110. [CrossRef]
- Plomion, C.; Chancerel, E.; Endelman, J.; Lamy, J.B.; Mandrou, E.; Lesur, I.; Ehrenmann, F.; Isik, F.; Bink, M.C.; van Heerwaarden, J.; et al. Genome-wide distribution of genetic diversity and linkage disequilibrium in a mass-selected population of maritime pine. *BMC Genomics* 2014, 15, 171. [CrossRef] [PubMed]
- 54. Ta, F.; Liu, X.D.; Huang, D.L.; Wang, L.; Liu, R.H.; Zhao, W.J.; Jing, W.M. Quantitative dynamics of *Picea crassifolia* population in Dayekou basin of Qianlian mountains. *Acta Ecol. Sin.* **2021**, *41*, 6871–6882. (In Chinese)
- 55. Birol, I.; Raymond, A.; Jackman, S.D.; Pleasance, S.; Coope, R.; Taylor, G.A.; Yuen, M.M.; Keeling, C.I.; Brand, D.; Vandervalk, B.P.; et al. Assembling the 20-Gb white spruce (*Picea glauca*) genome from whole-genome Shotgun sequencing data. *Bioinformatics* 2013, 29, 1492–1497. [CrossRef] [PubMed]
- 56. Zimin, A.; Stevens, K.A.; Crepeau, M.W.; Holtz-Morris, A.; Koriabine, M.; Marçais, G.; Puiu, D.; Roberts, M.; Wegrzyn, J.L.; de Jong, P.J.; et al. Sequencing and assembly of the 22-Gb loblolly pine genome. *Genetics* **2014**, *196*, 875–890. [CrossRef]
- Gu, X.; Feng, C.; Ma, L.; Song, C.; Wang, Y.; Da, Y.; Li, H.; Chen, K.; Ye, S.; Ge, C.; et al. Genome-wide Association Study of Body Weight in Chicken F2 Resource Population. *PLoS ONE* 2011, 6, e21872. [CrossRef]
- 58. Liu, R.; Sun, Y.; Zhao, G.; Wang, F.; Wu, D.; Zheng, M.; Chen, J.; Zhang, L.; Hu, Y.; Wen, J. Genome-wide association study identifies loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. *PLoS ONE* **2013**, *8*, e61172. [CrossRef]
- 59. Bai, Q.; Cai, Y.; He, B.; Liu, W.; Pan, Q.; Zhang, Q. Core set construction and association analysis of *Pinus massoniana* from Guangdong province in southern China using SLAF-seq. *Sci. Rep.* **2019**, *9*, 1–13. [CrossRef]
- Yang, Y.; Xuan, L.; Yu, C.; Wang, Z.; Xu, J.; Fan, W.; Guo, J.; Yin, Y. High-density genetic map construction and quantitative trait loci identification for growth traits in (*Taxodium distichum* var. distichum × *T. mucronatum*) × *T. mucronatum*. *BMC Plant Biol*. **2018**, *18*, 263. [CrossRef]
- 61. Huang, X.; Wei, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [CrossRef] [PubMed]
- Yang, X.; Yan, J.; Shah, T.; Warburton, M.L.; Li, Q.; Li, L.; Gao, Y.; Chai, Y.; Fu, Z.; Zhou, Y.; et al. Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theor. Appl. Genet.* 2010, 121, 417–431. [CrossRef] [PubMed]
- 63. Zhang, Z.; Ersoz, E.; Lai, C.Q.; Todhunter, R.J.; Tiwari, H.K.; Gore, M.A.; Bradbury, P.J.; Yu, J.; Arnett, D.K.; Ordovas, J.M.; et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360. [CrossRef] [PubMed]
- 64. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* **2016**, *12*, e1005767. [CrossRef] [PubMed]
- Dhanapal, A.P.; Crisosto, C.H. Association genetics of chilling injury susceptibility in peach (Prunus persica (L.) Batsch) across multiple years. 3 Biotech 2013, 3, 481–490. [CrossRef] [PubMed]
- Hecht, B.C.; Campbell, N.R.; Holecek, D.E.; Narum, S.R. Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Mol. Ecol.* 2013, 22, 3061–3076. [CrossRef] [PubMed]

- 67. Zhang, H.; Fan, X.C.; Zhang, Y.; Jiang, J.F.; Liu, C.H. Identification of favorable SNP alleles and candidate genes for seedlessness in *Vitis vinifera* L. using genome-wide association mapping. *Euphytica* **2017**, *213*, 136. [CrossRef]
- Kudo, K.; Nabeshima, E.; Begum, S.; Yamagishi, Y.; Nakaba, S.; Oribe, Y.; Yasue, K.; Funada, R. Formation of new networks of earlywood vessels in seedlings of the deciduous ring-porous hardwood *Quercus Serrata* in springtime. *Trees Struct. Funct.* 2018, 32, 725–734. [CrossRef]
- 69. Martin, H.; Hanuš, V. Comparison of earlywood vessel variables in the wood of *Quercus robur* L. and *Quercus petraea* (Mattuschka) Liebl. growing at the same site. *Dendrochronologia* **2014**, *32*, 284–289. [CrossRef]
- 70. Tsai, D.M.; Yang, C.H. A quantile–quantile plot based pattern matching for defect detection. *Pattern Recognit. Lett.* 2005, 26, 1948–1962. [CrossRef]
- 71. Li, L.; Li, K.; Ali, A.; Guo, Y. AtWAKL10, a cell wall associated receptor-like kinase, negatively regulates leaf senescence in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* **2021**, *22*, 4885. [CrossRef] [PubMed]
- 72. Daniel, J.C. Plant cell walls: Wall-associated kinases and cell expansion. Curr. Biol. 2001, 11, R558–R559. [CrossRef]