

Article

Allelic Variation in Cinnamyl Alcohol Dehydrogenase (*LoCAD*) Associated with Wood Properties of *Larix olgensis*

Yanhong Wang^{1,2}, Qinbin Jia¹, Lei Zhang¹, Zhen Zhang¹ and Hanguo Zhang^{1,*}

¹ Key Laboratory of Forest Tree Genetic Improvement and Biotechnology, Northeast Forestry University, Harbin 150040, China;

E-Mails: wangyanhong@neau.edu.cn (Y.W.); 729217823@qq.com (Q.J.); 109477986@qq.com (L.Z.); 413003320@qq.com (Z.Z.)

² College of Animal Science and Technology, Northeast Agricultural University, Harbin 150030, China

* Author to whom correspondence should be addressed; E-Mail: hanguo Zhang1@sina.com; Tel.: +86-451-8219-1536.

Academic Editors: Om P. Rajora and Eric J. Jokela

Received: 16 April 2015 / Accepted: 30 April 2015 / Published: 12 May 2015

Abstract: Cinnamyl alcohol dehydrogenase (CAD) catalyzes the key step in the lignin monomer biosynthesis pathway, but little is known about CADs in larch (*Larix olgensis*). Larch is one of the most important conifer plantation species and is used worldwide for reforestation and paper making. However, the presence of lignin is a significant barrier in the conversion of plant biomass to bioethanol. In the current study, 240 individuals from the Northeast Forest University provenance progeny trial population were evaluated, and 47 single-nucleotide polymorphisms (SNPs) were identified in the CAD gene. We used a candidate gene-based association mapping approach to identify CAD gene allelic variants that were associated with growth and wood property traits in *L. olgensis*. We found that *LoCAD* harbors high single nucleotide polymorphism (SNP) diversity ($\pi_T = 0.00622$ and $\theta_w = 0.00646$). The results of an association analysis indicated that nine SNPs and six haplotypes were significantly associated with wood property and growth traits, explaining between 1.35% and 18.4% of the phenotypic variance. There were strong associations between SNP (g.590G > T) and SNP (g.1184A > T) in *LoCAD*. These SNPs might represent two quantitative trait nucleotides that are important for the analysis of lignin content.

Keywords: lignin; *Larix olgensis*; haplotype; single-nucleotide polymorphisms (SNPs); gene

1. Introduction

Larch (*Larix olgensis*) belongs to the family Pinaceae and is a rapidly growing and economically important conifer crop that produces wood for lumber, pulp, paper, and biofuel [1,2]. In Northern China, larch is an ecologically and economically important tree, and consequently, many breeding programs have aimed to increase the quantity and quality of its wood products [1,3–5].

Lignin (a biopolymer that accumulates in wood) is one of the most abundant biopolymers in trees and its use is becoming mandatory to fulfill many industrial applications such as pulp and paper production [6–8]. However, in the pulp and paper industry, lignin is an undesirable component in the conversion of wood to pulp and paper and should be eliminated by delignification during chemical pulping [8]. Both the lignin content and composition influence the chemical pulping process, and trees with an altered lignin quantity or quality, or those with a desirable balance between lignin content, wood density and carbon storage are important in forest tree breeding programs.

Due to its economic importance, lignin biosynthesis has been well characterized during the last two decades [6]. Significant progress has been made in the understanding of the biochemistry of the enzymes involved in lignin biosynthesis [9]. The genes for each step of the lignin biosynthesis pathway have been cloned, and the impact on lignin amount and composition has been explored in various species through the use of mutants or reverse genetics [6,10–12]. However, little is known about the genetic variability within the genes coding for biosynthesis of lignin or other secondary cell wall components in larch.

Cinnamyl alcohol dehydrogenase (CAD) enzymes in the lignin biosynthesis pathway catalyze the final step in the synthesis of monolignols by converting cinnamaldehydes to cinnamyl alcohols [13–15]. Due to its important role in determining lignin content and composition, the involvement of CAD in lignin biosynthesis has been studied in several plant species, including dicot and monocot plants and coniferous species [8,16–23]. The CAD genes belong to small multigene families. Sibout *et al.* (2005) surveyed the complete CAD gene family in *Arabidopsis*, and found that *CAD-C* and *CAD-D* act as the primary genes involved in lignin biosynthesis [6]. Fiona *et al.* (2003) identified one SNP in CAD exons that affected a highly conserved amino acid within a protein sequence in *Eucalyptus globulus* and it may be possible to link this molecular variation to the variation found in lignin profiles [10]. Seven CAD homologs (*LtuCAD1* to *LtuCAD7*) were identified by mining a comprehensive *Liriodendron tulipifera* EST dataset. The *LtuCAD1* homolog appeared to restore the total lignin content and decrease the S/G lignin ratio [8]. However, little is known about the genetic variability within the genes that code for biosynthesis of lignin or other secondary cell wall components in larch. To develop markers associated with lignin content, we employed three methods: single SNPs, combinational-SNPs, and haplotype-based associations with wood properties and growth traits in a provenance trial. These methods were used to investigate whether significant associations could be found in a Chinese larch landrace.

2. Experimental Section

2.1. Association Population

Ten mainland populations of larch had been maintained as a provenance progeny trial located at Maoershan Forest, Shang Zhi City, Heilongjiang Province, China (45°21′–45°34′ N; 127°30′–127°39′ E). The provenance trial was designed using a randomized complete block with five replications, each provenance contained 30 populations. This trial was established in 1982. A total of 1500 individual trees that grew from seeds were collected from natural populations of *L. olgensis* in 1980. These populations represent 10 areas (provenances): Baidao, Baihe, Dahai, Dashi, Helong, Jixi, Lushui, Muling, Tianqiao, and Xiaobei (Figure S1), and these areas contain the entire natural distribution of *L. olgensis* (40°–48° N, 124°–135° E). The trial was thinned twice in year 1995 and 2001, and only 500 individual trees with the best vigor and stem were retained. In June 2011, needles were collected from four of five blocks; each block contained 10 provenances, and each provenance contained six individuals. A total of 240 individuals representing 240 different families were selected.

2.2. Phenotypic Data

Eight traits were measured, *i.e.*, wood density, tree height, diameter at breast height (DBH), stem volume, carbon content (CB), carbon concentration (CCR), lignin content and cellulose content. For the measurement of wood traits, wood core samples (15 cm long and 10 mm in diameter) containing bark and pith at breast height (1.3 m above ground level) were collected from each individual in 2011. Lignin was estimated using the Klason lignin method described in Kirk *et al.* (1988), and the lignin content was expressed as a percentage [24]. The cellulose content was estimated using the wet chemistry techniques described in Porth *et al.* (2011) [25]. The carbon concentration (CCR) was measured using a MultiN/C2100 analyzer (HT1300 Solids Module, Analytik Jena AG, Germany) as described in Zhu *et al.* [26], and the carbon content was calculated from the carbon concentration. The minimum, maximum, and mean values, and the coefficient of phenotypic variation (CV (%)) for *L. olgensis* traits are shown in Table S1.

2.3. DNA Extraction and CAD Gene Identification

Total genomic DNA was extracted from young needles of each of the 240 individuals using the modified CTAB method [27]. The CAD gene was isolated and sequenced by Sangon Biotech (Shanghai, China). Candidate CAD gene in *L. olgensis* was identified by amino acid homology with known proteins in *L. olgensis* and other plant systems using a BLAST search of the NCBI databases. The multiple amino acid sequences were aligned using Clustal W (<http://www.mbio.ncsu.edu/Clustal/Clustal.html>) and Dnaman (<http://www.mbio.ncsu.edu/dnaman/dnaman.html>) (Figure 2). The sequence data from this article can be found in the GenBank/EMBL databases under the following accession numbers: *LsCAD* in *Larix sibirica* (EU441954) [28], *PtaCADb* in *Pinus taeda* (Z37992), *PtCAD* in *Populus trichocarpa* (XM_002300175), *AtCAD* in *Arabidopsis thaliana* (NM_112832.3),

OsCAD in *Oryza sativa* (NM_001052667.1), and *BdCAD5* in *Brachypodium distachyon* (XM_003570926) [29].

For the phylogenetic analysis, an alignment of the protein sequences was performed using ClustalW. An unrooted dendrogram was then generated using the neighbor-joining method in MEGA version 3.1 [30] (Figure S2). The sequence data from this article can be found in the GenBank/EMBL databases under the following accession numbers: *Arabidopsis thaliana AtCAD1* (NP_195643); *AtCAD-A* (NP_195510); *AtCAD-B1* (CAA48027); *AtCAD-B2* (NP_195512); *AtCAD-C* (NP_188576); *AtCAD-D* (NP_195149); *AtCAD-E* (NP_179765); *AtCAD-F* (NP_179780); *AtCAD-G* (NP_177412); *EgCAD2* in *Eucalyptus gunnii* (X65631.1); and *OsCAD2* in *Oryza sativa* (DAA02237).

2.4. SNP Discovery and Genotyping

The CAD gene was sequenced and analyzed in 61 *L. olgensis* individuals to identify SNPs. Dnaman and BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>) were used for sequence alignment, and manual editing was used to confirm sequence quality and to remove primer sequences. Subsequently, 47 SNPs were detected in the entire gene, and 29 of 47 SNPs were considered common (frequency > 0.10). The 29 common SNPs were successfully genotyped across 240 individuals using contigs software (<http://www.mbio.ncsu.edu/contig/contig.html>). Homozygous genotypes showed a unimodal pattern, whereas heterozygous genotypes showed overlapping double peaks. The genotype was named for each mutant using the mutation and base name, such as a T to C mutation, e.g., TT, CC and TC.

2.5. Nucleotide Diversity and Linkage Disequilibrium

Nucleotide diversity: The 240 genomic clones were aligned using DNA sequence polymorphism (DNASP) software (version 5.10) (<http://www.ub.edu/dnasp>), and this software was also used to calculate the summary statistics for the SNP polymorphisms. The number of segregating sites (S), nucleotide diversity π [31] and θ (per site) [32] were estimated for all 240 individuals.

Linkage disequilibrium: LD was measured as the squared correlation of allele frequencies R^2 , which was affected by both recombination and differences in allele frequencies between sites [33]. The R^2 value between each pair of 29 common SNPs (minor allele frequencies > 0.10) in *LoCAD* was calculated using Haploview software (<http://www.broad.mit.edu/mpg/haploview.html>).

2.6. Association Analysis

$$Y = \mu + s + P + A + e \quad (1)$$

$$Y = \mu + h + P + A + e \quad (2)$$

$$Y = \mu + c + P + A + e \quad (3)$$

Association analysis was conducted using the GLM program in SAS 9.1.3. The model used for the association analysis was fitted using the following variables: single-SNP genotypes (s), haplotypes (h), combination genotypes (c), provenance (P) and tree age (A). The models, Equations (1)–(3), were used for single SNPs, haplotypes and combination genotype associations with lignin-related traits,

respectively, where Y is a vector representing the observed dependent variable (trait), μ is the population mean, and e is the random error.

To estimate the phenotypic and genetic additive correlations, we considered a bivariate analysis using the following model:

$$\sigma_G^2 = \sum_{i=1}^{n_g} q_i G_i^2 - \left(\sum_{i=1}^{n_g} q_i G_i \right)^2 \quad h^2 = \frac{\sigma_G^2}{\sigma_P^2} \quad (4)$$

(where q_i is the frequency of the i th genotype in the candidate gene, G_i is the effect of the i th genotype, and n_g is the total number of genotypes).

Haplotype analysis: Candidate gene-based LD mapping approaches have been used to dissect complex growth and wood property traits in forest trees [34]. Generally, the power of a single-marker association test is limited because the LD information contained in flanking markers is frequently ignored. Intuitively, haplotypes (a block of linked ordered markers) may be more powerful than individual markers [35,36]. In this study, haplotypes were constructed based on a sliding “window” that consisted of three markers; the “window” was allowed to slide one marker at a time (Figure S3, Table S2) [33]. The haplotypes of each “window” were constructed using the HAPLOTYPE procedure in SAS 9.1.3 software (Chapel Hill, NC, USA). Using the *LoCAD* gene and 29 common SNPs, 27 “windows” were constructed. The number of haplotypes in each “window” was eight in theory, but certain haplotypes did not exist in the studied population. The associations of the haplotypes in each “window” with wood properties and growth traits were analyzed.

Single-SNP models: The associations between the wood property traits and 29 SNPs were tested using the GLM program of SAS 9.1.3 (Chapel Hill, NC, USA).

Combinational-SNP models: The combination genotypes of the two polymorphisms were constructed. The term “combination genotype” is defined as follows: if the genotypes of a given individual at two loci are GG and AA, the combination genotype of the individual at the two loci is GGAA [33].

3. Results

3.1. Identification and Phylogenetic Analysis of *LoCAD*

The CAD gene of *L. olgensis* is 1675 bp long with a 1061 bp exon and 614 bp intron (Figure 1). Multiple amino acid sequence alignment was performed using the *L. olgensis* CAD and other functionally characterized CAD proteins in *Pinus taeda* (*PtaCAD*), *Populus trichocarpa* (*PtCAD*), *Arabidopsis thaliana* (*AtCAD*), *Oryza sativa* (*OsCAD*), and *Brachypodium distachyon* (*BdCAD*). The alignment results indicated a high degree of similarity in conserved domains and binding residues characteristic of alcohol dehydrogenases (Figure 2). The CAD gene in *L. olgensis* contains the Zn-1 binding domain motif GHE(X)2G(X)5G(X)2V and the conserved Zn-1 catalytic residues C (in 47 bp), H (in 69 bp), and C (in 163 bp). The sequence of the Zn-1 binding motif is most highly conserved in *LoCAD*, with 99.3% homology to the motif in the other aforementioned grass species CAD proteins. A glycine-rich repeat GXG(X)2G, involved in NADP(H) cosubstrate binding, is conserved among all CAD proteins. The consensus sequence GD(X)9,10C(X)2C(X)2C(X)7C for binding the Zn-2 metal

ion is preserved in *LoCAD*. In addition, 12 amino acids have been identified as substrate-binding residues in the bona fide CADs of various plant species [20,21,37–40]. Among the CAD in *L. olgensis*, *LoCAD* contains nine of these 12 conserved residues. Interestingly, *LoCAD* contains both active substrate binding residues, *i.e.*, W119 and F298, which determine specificity for aromatic alcohols, and the conserved S212 residue that determines NADP(H) binding at that position, as observed in *OsCAD* in rice [41]. Pair-wise sequence alignments with the *LoCAD* protein revealed a high percentage identity to *Pinus taeda* (96.3%). Based on amino acid sequences, it appears that *LoCAD* contains the conserved functional and structural features of a medium chain dehydrogenase/reductase specific to enzymes involved in lignin biosynthesis in secondary cell walls.

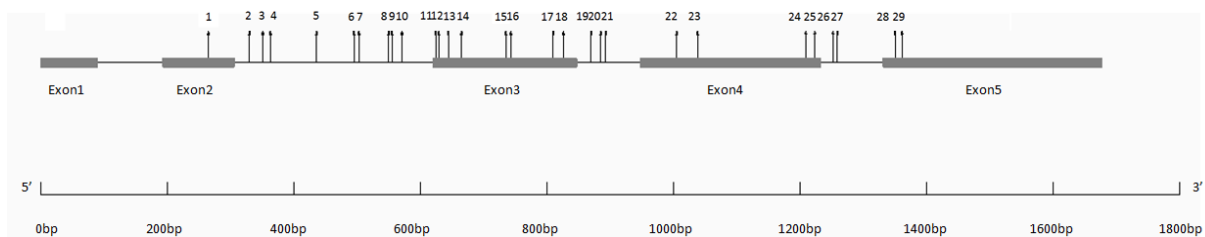


Figure 1. Genomic organization of the *LoCAD* (Cinnamyl alcohol dehydrogenase). Exons are shown as boxes, and introns are shown as lines. The positions of common SNPs (single-nucleotide polymorphisms) are shown as vertical lines.

		10	20	30	40	50	60	70	80
<i>LoCAD</i>	--MGSLES-E	KTVTGYAARD	SSGHLSPYTY	TLRNGKPEDV	IVRVYICGI	HSDLVQMRNE	MGM SHYPMVP	GHEVVGIVTE	
<i>PtaCAD</i>	--MGSLES-E	KTVTGYAARD	SSGHLSPYTY	NLRKKGPEVD	IVKVIYCGI	HSDLVQMRNE	MGM SHYPMVP	GHEVVGIVTE	
<i>PtCAD</i>	MAKSPVEVHP	HKAFGWAARD	SSGVLSPFHF	SRRDNGVEDV	TIKILYCGV	HSDLHAAKNE	WGF SRYFLVP	GHEIVGIVTK	
<i>AtCAD</i>	--MGSVEAGE	KKALGWAARD	PSGVLSPYSY	TLRSTGADDV	YIKVICCGI	HTDIHOAKND	LGM SNYPMVP	GHEVVGIVLE	
<i>OsCAD</i>	--MGSIAA-E	KTVTGWAAARD	ASGHLTPYNY	TLRKTGPEDV	VVKVLYCGI	HTDIHOAKNH	LGA SKYPMVP	GHEVVGIVVE	
<i>BdCAD</i>	--MGSIASEE	RTVTGWAAARD	ADGHLSPYTY	TLRKTGPEDV	LVKVLYCGI	HTDIHOAKNH	LGA SKYPMVP	GHEVVGIVVE	
		90	100	110	120	130	140	150	160
<i>LoCAD</i>	VGSEV	KKFKV	GENVGVG	CLV	GSCRS	CGN	QSMEQ	YCSKR	IW
<i>PtaCAD</i>	IGSEV	KKFKV	GEHVGVG	CLV	GSCRS	CGN	QSMEQ	YCSKR	IW
<i>PtCAD</i>	IGSNV	KKFKV	DDQVGVG	VLV	NSCRS	CEYCD	ODLEN	YCPKM	IFT
<i>AtCAD</i>	VGSDV	SKFTV	GDVVG	GVV	GCCGS	CKPCS	SELEQ	YCNKR	IWS
<i>OsCAD</i>	VGSEV	TKYSA	GDVVG	VIV	GCCRC	CHPCK	ANVEQ	YCNKR	IWS
<i>BdCAD</i>	VGSEV	SKYSA	GDVVG	VIV	GCCRC	CRPCK	ASVEQ	YCNKR	IWS
		170	180	190	200	210	220	230	240
<i>LoCAD</i>	AAPLLCAGVT	VFSPMKHFGM	TEPGKRCGIL	GLGGV	GHM	GV	KIAKAFGLHV	TVIS	SDRKK
<i>PtaCAD</i>	AAPLLCAGVT	VFSPMKHFAM	TEPGKRCGIL	GLGGV	GHM	GV	KIAKAFGLHV	TVIS	SDRKK
<i>PtCAD</i>	AAPLLCAGIT	VYSPMKYYGM	TEPGKHLGIV	GLGGL	GHVAV	KIGKAFGLKV	TVIS	SSRKE	SEALDRLGAD
<i>AtCAD</i>	AAPLLCAGVT	VYSPLSHFGL	MASGLRGGIL	GLGGV	GHM	GV	KIAKAMGHHV	TVIS	SDRKK
<i>OsCAD</i>	AAPLLCAGLT	VYSPLKHFGL	MSPGLRGGVL	GLGGV	GHM	GV	KVAKSMGHHV	TVIS	SARKR
<i>BdCAD</i>	AAPLLCAGVT	VYSPLKHFGL	MSPGLRGGIL	GLGGV	GHM	GV	KVAKSMGHHV	TVIS	SGKKR
		250	260	270	280	290	300	310	320
<i>LoCAD</i>	MMEAAESLDY	ILDITPVSHF	LEPYLALLKT	NGKLVMLGVV	PEPLHFVTPL	LILGRRS	IAG	SFIGSMEETQ	ETLDFCAEKK
<i>PtaCAD</i>	MMEAAESLDY	IMDITPVVHP	LEPYLALLKT	NGKLVMLGVV	PEPLHFVTPL	LILGRRS	IAG	SFIGSMEETQ	ETLDFCAEKK
<i>PtCAD</i>	MKAAPGMDY	IIDTVSAVHA	LAPLLSLLKT	NGKLVTLGLP	EKPLELPFIP	LVLGRKLVGG	SDIGGVKETQ	EMLDFCARHN	
<i>AtCAD</i>	MQRADLSLDY	IIDTVVVFHP	LDPYLACLKL	DGKLIILMGVI	NTPLOFVTPL	VILGRKVISG	SFIGSKIKETE	EVLVFCCKEK	
<i>OsCAD</i>	MAAAGDSLDY	IIDTVVVFHP	LEPYLALLKL	DGKLIILMGVI	NQPLSFI	SPM	VMLGRKAITG	SFIGSMAETE	EVLNFCVCKG
<i>BdCAD</i>	MAAAADSLDY	VIDTVVVKHP	LEPYLSLLKM	DGKLVLMGVI	GEPLSFV	SPM	VMLGRKSITG	SFIGSMEETE	EVLKFCVCKG
		330	340	350	360	370			
<i>LoCAD</i>	VSSMIEVVGL	DYINTAMKRL	EKNDVRYRFV	VDVAASK---	-----				
<i>PtaCAD</i>	VSSMIEVVGL	DYINTAMKRL	EKNDVRYRFV	VDVAASKLDN	-----				
<i>PtCAD</i>	ITSDEVVIRM	DQINTAMDRL	AKSDVRYRFV	IDVANSLSQS	QL-----				
<i>AtCAD</i>	LTSTIEVTKI	DELNIAFERL	RKNDVRYRFV	VDVAGSNLVE	EAATTTN				
<i>OsCAD</i>	LTSQIEVVVM	DYVNQALERL	ERNVDVRYRFV	VDVAGSNIDD	ADAPPA-				
<i>BdCAD</i>	LTSQIEVVVM	DYLNQALERL	ERNVDVRYRFV	VDVAGSKIED	AA-----				

Figure 2. Sequence alignment of *LoCAD* with CAD of other species. The amino acid sequence comparison of *LoCAD* proteins with other functionally characterized CAD proteins in *Pinus taeda* (*PtaCAD*), *Populus trichocarpa* (*PtCAD*), *Arabidopsis thaliana* (*AtCAD*), *Oryza sativa* (*OsCAD*), and *Brachypodium distachyon* (*BdCAD*). The Zn-1 binding domain motif is labeled in pink, and the Zn-1 catalytic residues are labeled in yellow. The NADP(H) cosubstrate-binding domain motif is labeled in red, and the Zn-2 structural motif is labeled in green. Substrate-binding residues are labeled in gray, and the Ser212-specific NADP(H) binding residue is labeled in purple.

To analyze the phylogenetic relationship between *LoCAD* and the CADs from *Arabidopsis*, rice and *Eucalyptus*, we performed a phylogenetic analysis using 12 protein sequences from *A. thaliana*, *O. sativa* and *E. gunnii* based on 1000-replicate bootstrap values (Figure S2). Phylogenetic analysis suggested that the cloned *LoCAD* belongs to the class II CAD subfamily. The *AtCAD-C*, *AtCAD-D*, *EgCAD2* and *OsCAD2* proteins belong to the same class (Figure S2).

3.2. SNP Diversity and Genotyping

To identify polymorphisms for association mapping, we re-sequenced the *LoCAD* region in 61 unrelated individuals from the provenance trial. The statistical analysis of the nucleotide polymorphisms (excluding indels) over different regions of *LoCAD* are summarized in Table 1. Across the samples, 47 SNPs were detected in the entire gene at a frequency of approximately one SNP every 37 bp (Table 1). Altogether, 29 of 47 SNPs (61.7%) were considered common (frequency > 0.10). A total of 15 of these SNPs were found in exons, and 14 SNPs were found in introns (Table 1, Figure 1). Of the SNPs located in exons, 12 led to non-synonymous changes, and the other three SNPs were synonymous mutations (Table 1).

Table 1. Nucleotide polymorphism at the *LoCAD* locus.

Region	Length (bp)	Number of Polymorphic Sites	Frequency (bp ⁻¹)	Nucleotide Diversity	
				π_T	θ_w
Exon 1	90	0	–	–	–
Intron 1	103	0	–	–	–
Exon 2	113	1	113	0.00141	0.00219
Intron 2	312	9	35	0.01018	0.00738
Exon 3	227	8	28	0.00420	0.00498
Intron 3	100	3	33	0.01012	0.00994
Exon 4	285	4	71	0.00373	0.00276
Intron 4	99	2	50	0.00524	0.00406
Exon 5	346	2	173	0.01515	0.01836
Synonymous	141.14	3	40	0.00269	0.00271
Nonsynonymous	572.86	12	47	0.00303	0.00304
Total exon	1061	15	71	0.00377	0.00474
Total intron	614	14	44	0.00731	0.00667
Total	1675	29	58	0.00622	0.00646

3.3. LD Analysis

Using genotypic data from 29 common SNPs located in *LoCAD*, the R^2 values were pooled to assess the overall behavior of LD using Haploview 4.2 software (<http://www.broad.mit.edu/mpg/haploview.html>). The average value of R^2 was 0.0489, with a range from 0.0 (equilibrium) to 1.0 (disequilibrium), which indicates that the LD levels were quite low for loci located far away from each other. Several high LD distinct haplotype blocks ($R^2 > 0.8$; $p < 0.001$) across the sequenced regions only existed between relatively close loci. These findings indicate that for the common SNPs in the CAD gene in larch, the linkage disequilibrium decayed rapidly (Figure 3).

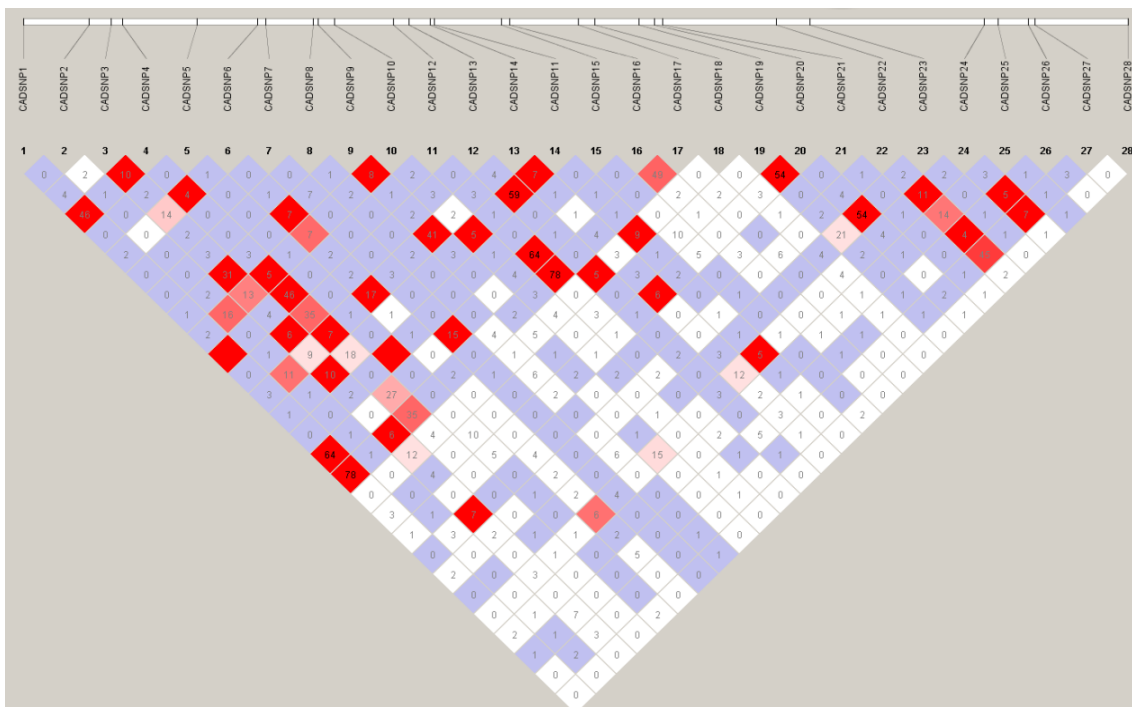


Figure 3. Pairwise linkage disequilibrium (R^2) between SNPs of *LoCAD*. The color from red to white represent LD levels from high to low.

3.4. Association Analysis

Haplotype-based associations: We used all 240 individuals for these analyses. Haplotype-based association tests were performed to identify significant associations of haplotypes with wood properties and growth traits in this low LD tree species (Table 2). The results of the haplotype analysis indicated that the polymorphisms of *LoCAD* were associated with lignin p -values < 0.05 (Table 2). The “windows” 12–14 and 25–27 were significantly associated with lignin p -values < 0.05 (Table 2). The *LoCAD* “window” 12–14 contained SNP12–16, and “window” 25–27 contained SNP25–29. This finding suggests that the association with lignin content might be located between SNP12–16 and SNP25–29 or might occur in linkage disequilibrium with this region. The value of “window” 14 was more significant, indicating that SNP14 (g.549A > C) might be significantly associated with lignin content. The values of “windows” 25 and 27 were lower, indicating that SNP25 (g.1148A > G) and SNP27 (g.1184A > T) might be significantly associated with lignin content (Table 2).

Table 2. Association analysis between different haplotypes in each “window” and wood traits in the CAD gene (p -value).

Windows	Lignin	Cellulose	Height	DBH	Density	Volume	CB	CCR
1(5)	0.1362	0.5297	0.0437 *	0.8234	0.1904	0.8755	0.2961	0.4595
2(5)	0.1374	0.7240	0.0480 *	0.8310	0.2954	0.8970	0.2965	0.2854
3(4)	0.1219	0.4307	0.0256 *	0.9428	0.1015	0.9627	0.0842	0.2266
4(4)	0.1332	0.5183	0.0185 *	0.9457	0.1033	0.9580	0.0573	0.1805
5(4)	0.0932	0.5082	0.0481 *	0.9395	0.0360 *	0.9465	0.0541	0.4621
6(4)	0.1908	0.7644	0.0121 *	0.9316	0.0611	0.8589	0.0957	0.3051

Table 2. Cont.

Windows	Lignin	Cellulose	Height	DBH	Density	Volume	CB	CCR
7(4)	0.1077	0.6644	0.0377 *	0.8474	0.0546	0.7996	0.1449	0.3078
8(4)	0.1074	0.6561	0.0384 *	0.6958	0.1778	0.7087	0.0268 *	0.2043
9(4)	0.1358	0.5000	0.1033	0.3932	0.1642	0.6393	0.0243 *	0.3486
10(4)	0.1359	0.4249	0.0266 *	0.5502	0.1112	0.7128	0.0252 *	0.2026
11(4)	0.1075	0.5509	0.0210 *	0.6249	0.1101	0.7647	0.0267 *	0.2015
12(4)	0.0236 *	0.3801	0.0120 *	0.9362	0.1398	0.9437	0.2360	0.3917
13(4)	0.0322 *	0.3151	0.0600	0.9130	0.0981	0.9511	0.0666	0.3859
14(5)	0.009 **	0.1297	0.0508	0.8631	0.0140 *	0.8735	0.1053	0.4675
15(5)	0.1062	0.5070	0.0310 *	0.8980	0.0695	0.9631	0.0782	0.2674
16(6)	0.1331	0.7762	0.0346 *	0.9634	0.2690	0.9860	0.2884	0.2946
17(7)	0.1112	0.7294	0.0285 *	0.8164	0.1287	0.7927	0.2488	0.3193
18(5)	0.1447	0.6631	0.0904	0.7772	0.1651	0.7777	0.1084	0.6097
19(4)	0.1016	0.7278	0.0862	0.9731	0.1398	0.9749	0.1615	0.5112
20(4)	0.1106	0.3637	0.0828	0.3963	0.1726	0.5819	0.1678	0.5536
21(4)	0.1160	0.3853	0.0840	0.3870	0.2175	0.5734	0.1414	0.2992
22(4)	0.1154	0.2899	0.0721	0.3841	0.1054	0.5740	0.0228 *	0.3710
23(4)	0.1112	0.4683	0.0834	0.9241	0.1014	0.9148	0.0111 *	0.3512
24(4)	0.1193	0.4125	0.0462 *	0.5816	0.1267	0.5440	0.0169 *	0.5200
25(4)	0.0486 *	0.2998	0.0398 *	0.5227	0.0832	0.4852	0.0269 *	0.3868
26(6)	0.0494 *	0.3519	0.0788	0.3555	0.1523	0.2635	0.0484 *	0.1284
27(5)	0.0144 *	0.1307	0.0673	0.4304	0.0666	0.3238	0.0852	0.1256

CB, carbon storage; CCR, carbon content rate; * Significant at the $p < 0.05$ level; ** Significant at the $p < 0.01$ level; The number in parentheses in the “windows” column is the number of haplotypes in each “window”.

Single-SNP based associations: The associations between 29 common SNPs and wood properties and growth traits were analyzed. In total, nine associations were found in *LoCAD*, representing 31% of all tests at a threshold of $p < 0.05$. Two SNPs, SNP (g.590G > T) and SNP (g.1184A > T), were significantly associated at a threshold of $p < 0.01$ (Table 3). The result was found for the haplotype-based associations. Of these nine unique SNPs, three were non-synonymous, one was synonymous, and five were noncoding SNPs (Table 3). All of these SNPs were associated with multiple traits, including lignin, height, density and CB, but they were not associated with cellulose, DBH, volume or CCR.

Table 3. Significant associations of single-nucleotide polymorphisms in the CAD gene with wood quality traits (*p*-value).

Mutation	Position	Frequency	Lignin	Cellulose	Height	DBH	Density	Volume	CB	CCR
g.361C > T	Intron 2	0.41(T)	0.0462 *	0.5060	0.0299 *	0.7597	0.5967	0.2466	0.0257 *	0.3391
g.428C > T	Intron 2	0.21(T)	0.0193 *	0.7256	0.0341 *	0.8435	0.0353 *	0.2583	0.2247	0.3809
g.480A > G	Intron 2	0.18(A)	0.0484 *	0.5938	0.0350 *	0.4847	0.5695	0.1710	0.2165	0.4751
g.549A > C ^{ns}	Exon 3	0.21(C)	0.0142 *	0.6063	0.0349 *	0.3165	0.5172	0.1638	0.0429 *	0.2023
g.554C > T ^s	Exon 3	0.11(T)	0.0112 *	0.6569	0.0077 *	0.7681	0.4352	0.3279	0.0991	0.2857
g.590G > T ^{ns}	Exon 3	0.34(T)	0.0043 **	0.6359	0.0290 *	0.5402	0.5988	0.1353	0.1669	0.3398
g.1148A > G ^{ns}	Exon 4	0.12(G)	0.0496 *	0.7501	0.0272 *	0.6813	0.1691	0.5196	0.0233 *	0.4252
g.1177A > G	Intron 4	0.13(G)	0.0473 *	0.7426	0.0185 *	0.6347	0.1710	0.5186	0.1917	0.3304
g.1184A > T	Intron 4	0.26(A)	0.0024 **	0.2962	0.0225 *	0.5077	0.0407 *	0.3126	0.2012	0.3576

CB, carbon storage; CCR, carbon content rate; ns, non-synonymous polymorphism; s, synonymous polymorphism. * Significant at the *p* < 0.05 level; ** Significant at the *p* < 0.01 level.

The non-synonymous marker SNP (g.590G > T) in exon 3, which has the minor allele (T), results in an amino acid change from Cys to Phe, which is significantly associated with lignin content and other traits (Table 3, Figure 4). Another non-synonymous marker, SNP (g.549A > C) in exon3, has the minor allele (C) that results in an amino acid change from Lys to Gln (Table 3). The other non-synonymous marker, SNP (g1148A > G) in exon 4, has the minor allele (G) that results in an amino acid change from Arg to Gly.

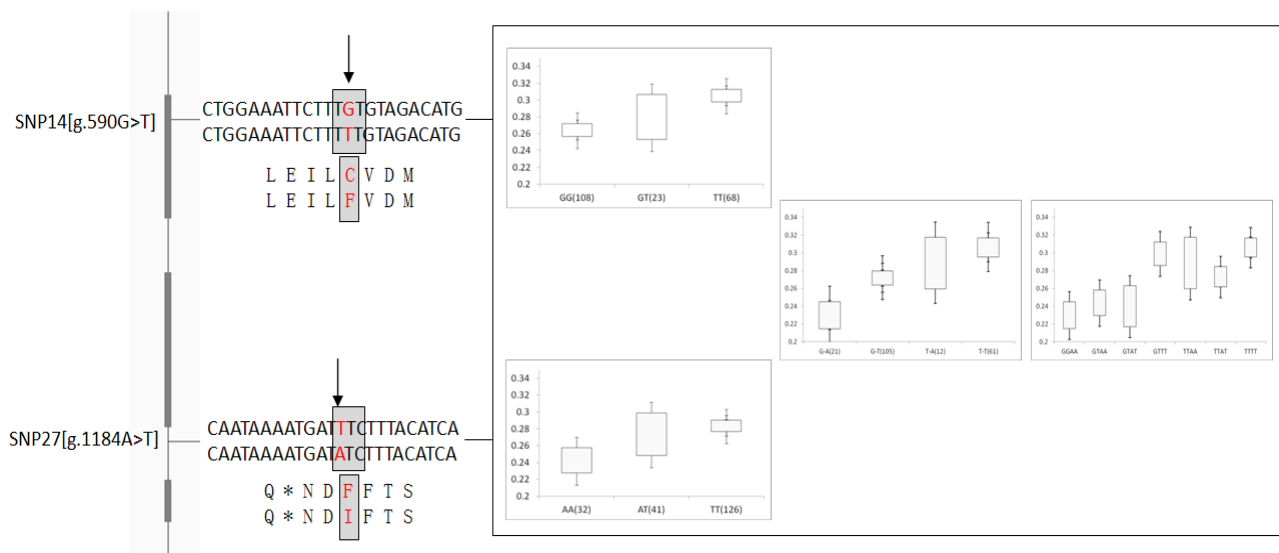


Figure 4. Association analysis of the genotypic effects of the significant single-nucleotide polymorphisms (SNPs) in the CAD gene on the lignin trait.

The noncoding marker SNP (g.1184A > T) located on intron 4 was significantly associated with lignin content, tree height and density (Table 3). This SNP explained 18.4% of phenotypic variance in terms of lignin content, 5.15% in terms of tree height and 2.29% in terms of density. The other noncoding markers, SNP (g.361A > T), SNP (g.428A > T) and SNP (g.480A > T) from intron 2 and SNP (g.1177A > G) from intron 4 were significantly associated with lignin, tree height, density, and

CB, with small single-SNP effects ranging from 1.07%–7.03%. A similar phenomenon has been identified whereby silent SNPs should not be considered potential false positives a priori because they may affect transcript levels and codon usage.

Combinational-SNP models: To determine whether *LoCAD* expression was altered in the different genotypic classes, two significantly associated SNPs were tested using multiple analysis. The results indicated that g.590G > T and g.1184A > T in the CAD gene had significantly lower values in the genotypes GG and AA compared with the genotypes TT (g.590G > T) and TT (g.1184A > T) (Figure 4). The haplotype G-A and combination genotype GGAA had the same effects as the genotypes AA (g.590G > T) and GG (g.1184A > T) (Figure 4, Tables S3–S5). The percentage of phenotypic variance that the single SNPs, haplotypes and combination genotypes could explain was 16.90%–18.40%, 25.34% and 26.47%, respectively. The phenotypic variance explanation (%) in *Larix* was associated mainly with lignin-related traits, then carbon storage traits, and finally wood volume traits.

4. Discussion

4.1. Nucleotide Diversity in *L. olgensis*

LoCAD, a highly targeted candidate gene belonging to key pathway of lignin biosynthesis, was investigated using a more comprehensive approach than those used in previous studies of conifers [42]. The collected sequencing data based on 11 amplicons from a sample of 240 trees provided useful information. In general, the *LoCAD* locus has high nucleotide diversity (π_T), where $\pi_T = 0.00622$ and $\theta_w = 0.00646$ (Table 1). In the provenance trial, the estimates of nucleotide diversity for the different gene regions ranged from 0.00141 (Exon 2) to 0.01515 (Exon 5), and θ_w varied between 0.00219 (Exon 2) and 0.01836 (Exon 5) (Table 1). These value are high compared with those obtained thus far in other forest tree species for different sets of genes and for samples of individuals representative of species ranges [29]. In terms of SNP densities, the average nucleotide diversity level in the coding regions (π was 0.00377) was substantially lower than in the non-coding regions (π was 0.00731) (Table 1). This result suggests the action of purifying selection on exonic regions is directly linked to the primary structure of the protein. Compared with previous reports on other genes detected in forest tree species, the nucleotide diversity (π was 0.00622) of *L. olgensis* was higher than that of CAD4 (π was 0.0012) in *Populus nigra* and nine different genes (π was 0.0018) in *Populus trichocarpa* [21,43], somewhat lower than that of *PtoCesA7* (π was 0.0091) [44] and *PtGA20Ox* (π was 0.0099) [45], and similar to that of *PtACT1* (π was 0.0079) [35] and *PtCOBL4* (π was 0.0080) [46]. This suggested that a similar pattern of genetic variance exists in *LoCAD*, similar to the candidate genes in *Populus*.

4.2. Detection of Phenotype-Genotype Associations in *L. olgensis*

In this study, a comparison of single-marker and haplotype-based association demonstrated that the power of single-marker-based association analyses is either similar to or greater than that of haplotype-based tests (Table 2). These results suggest that single-marker-based tests are preferable to haplotype-based tests for avoiding uncertainty in haplotype determination associated with diploid SNP

datasets. Therefore, combining these two methods may provide greater potential for detecting functional allelic variance underlying quantitative traits in association populations.

CAD is a key enzyme in the lignin biosynthesis pathway that catalyzes the final step in the synthesis of monolignols by converting cinnamaldehydes to cinnamyl alcohols [13–15]. Downregulation of CAD notably affects lignin structure and composition [13,18]. The present study identified nine SNPs significantly associated with lignin content, and two SNPs (g.590G > T, g.1184A > T) were also significantly associated with lignin content based on the haplotype analysis. SNP (g.590G > T) was significantly associated with lignin content and tree height. This non-synonymous SNP is located on exon 3 of *LoCAD* and represents a missense mutation. This mutation causes an amino acid change from Cys to Phe. The differences in the lignin content among the three SNP (g.590G > T) genotypes were significant (0.2405% for GG, 0.2800% for GT and 0.3052% for TT). Moreover, most of the common haplotypes significantly associated with wood traits surround SNP (g.590G > T). This SNP explained 16.9% of the phenotypic variance in terms of the lignin content and 1.94% of the phenotypic variance in terms of tree height. A polymorphism may cause variation in protein expression, stability, efficiency, *etc.* In the same way, these results together strongly suggest that SNP (g.590G > T) may cause variation in the CAD capacity. However, SNP (g.1184A > T) located on intron 4 in the non-coding regions of *LoCAD* was significantly consistent with the lignin content, tree height and density. This SNP explained 18.4% of phenotypic variance in terms of the lignin content, 5.15% in terms of tree height and 2.29% in terms of density. Larch with the AA genotype (g.1184A > T) had a significantly lower lignin content, greater height, larger DBH, greater volume and CCR compared with the trees with the TT genotype. In forest trees, rapid growth, a low lignin content and a high cellulose content are needed in the pulping industry [8]; therefore, people are usually interested in the simultaneous selection of both wood quantity and growth traits. The associations between the lignin content and the two SNPs (g.590G > T and g.1184A > T) provided evidence that these SNPs can serve as important markers in breeding for improved lignin structure and composition of *L. olgensis*.

In our study, two SNPs (g.428C > T and g.1184A > T) were significantly associated with wood density, and SNP (g.428C > T) was also significantly associated with density based on the haplotype analysis. SNP (g.428C > T) was located on intron 2 in the non-coding regions of *LoCAD*. This SNP explained 2.26% of the phenotypic variance in terms of the density trait. Previous studies have also provided evidence of the effect of CAD on wood density in other conifer species. Yu *et al.* (2006) found that the mutant null allele in *CAD-nl* genes was associated with wood density in heterozygous loblolly pine trees [20]. Gonzalez-Martinez *et al.* (2007) found four genes (*cad*, *sams-2*, *lp3-1* and *a-tubulin*) associated with different wood property traits in *Pinus taeda*, and allelic variation in CAD SNPM28 was associated with earlywood [13,47]. Dillon *et al.* (2010) found that CAD and COMT2 were associated with similar traits (early and latewood density) in a *P. radiata* discovery population [42].

Of these nine unique SNPs, one SNP was synonymous and five SNPs were located in non-coding regions. Although the synonymous mutation did not result in an amino acid change, silent SNPs can cause changes via other mechanisms because they may affect gene expression and exon splicing [44,48]. Previous studies have shown that significant SNPs in non-coding regions are associated with phenotypic traits. For example, Gonzalez-Martinez *et al.* (2006) used association

genetics and found a strong association between SNPM10 located on intron1 and the earlywood microfibril angle in *Pinus taeda* [13]. In addition, Gonzalez-Martinez *et al.* (2008) found a strong association between silent SNP Q1 and the carbon isotope discrimination (CID) trait in *Pinus taeda* [49]. Tian *et al.* (2014) found SNP 7 (located in the 5' UTR of *PtUGDH*) was significantly associated with the holocellulose content in *Populus tomentosa* [50]. Similarly, a synonymous SNP on exon 5 of *EniCOBL4A* was associated with cellulose content [51] and a synonymous SNP (SNP60) in *PAL1* of *Pinus radiata* was significantly associated with wood density [42].

In our study, the finding of significant associations of *LoCAD* with both wood quality (e.g., lignin content) and growth (e.g., tree height, density, and carbon storage) traits is encouraging (Table 2). Compared with previous reports in forest tree species, the phenotypic variance explained by a single SNP association ranged from 1.35%–18.4% in *L. olgensis*. Most of the associations explained a small proportion of the phenotypic variance. These findings are consistent with other association studies, e.g., the phenotypic variance explained by three significant SNPs (1, 2, 7) in the 5' untranslated regions of *PtUGDH*, in *Populus tomentosa* ranged from 5.37%–11.97%. These SNP effects were in accordance with polygenic quantitative models of wood traits [50].

5. Conclusions

The results of the current study indicate that the *CAD* gene is a major gene that determines lignin content. Two SNPs (g.590G > T and g.1184A > T) that were identified in the *CAD* gene might be two quantitative trait nucleotides for lignin content.

Acknowledgments

This work was supported by the planning committee of “The Twelfth Five-Year-Plan” of the National Science and Technology Programs of China (No.2013AA102704).

Author Contributions

Yanhong Wang and Hanguo Zhang conceived and designed the study. Yanhong Wang, Qinbin Jia, Zhen zhang and Hanguo zhang collected needles from Maoershan Forest. Yanhong Wang performed the experiments and wrote the paper. Yanhong Wang, Lei Zhang and Hanguo Zhang reviewed and edited the manuscript. All authors read and approved the manuscript.

Conflicts of Interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. Zhu, J.; Yang, K.; Yan, Q.; Liu, Z.; Yu, L.; Wang, H. Feasibility of implementing thinning in even-aged *Larix olgensis* plantations to develop uneven-aged larch–broadleaved mixed forests. *J. For. Res.* **2010**, *15*, 71–80.

2. Guan, C.; Zhang, H.; Zhang, L.; Li, X.; Deng, J.; Jiang, T. Construction of Genetic Linkage Maps of Larch. *Biotechnol. Biotechnol. Equip.* **2014**, *25*, 2197–2202.
3. Liu, S. Nitrogen cycling and dynamic analysis of man made larch forest ecosystem. *Plant Soil* **1995**, *168–169*, 391–397.
4. Pâques, L.E.; García-Casas, M.D.C.; Charpentier, J. Distribution of heartwood extractives in hybrid larches and in their related European and Japanese larch parents: Relationship with wood colour parameters. *Eur. J. For. Res.* **2013**, *132*, 61–69.
5. Zhao, J.; Wang, B.; Wang, X.; Zhang, Y.; Dong, M.; Zhang, J. iTRAQ-based comparative proteomic analysis of embryogenic and non-embryogenic tissues of Prince Rupprecht's larch (*Larix principis-rupprechtii* Mayr). *Plant Cell Tissue Organ Cult.* **2015**, *120*, 655–669.
6. Sibout, R. CINNAMYL ALCOHOL DEHYDROGENASE-C and -D Are the Primary Genes Involved in Lignin Biosynthesis in the Floral Stem of Arabidopsis. *Plant Cell Online* **2005**, *17*, 2059–2076.
7. Zapata-Valenzuela, J.; Isik, F.; Maltecca, C.; Wegrzyn, J.; Neale, D.; McKeand, S.; Whetten, R. SNP markers trace familial linkages in a cloned population of *Pinus taeda*—Prospects for genomic selection. *Tree Genet. Genomes* **2012**, *8*, 1307–1318.
8. Xu, Y.; Thammannagowda, S.; Thomas, T.P.; Azadi, P.; Schlarbaum, S.E.; Liang, H. LtuCAD1 Is a Cinnamyl Alcohol Dehydrogenase Ortholog Involved in Lignin Biosynthesis in *Liriodendron tulipifera* L., a Basal Angiosperm Timber Species. *Plant Mol. Biol. Rep.* **2013**, *31*, 1089–1099.
9. Matsui, N.; Chen, F.; Yasuda, S.; Fukushima, K. Conversion of guaiacyl to syringyl moieties on the cinnamyl alcohol pathway during the biosynthesis of lignin in angiosperms. *Planta* **2000**, *210*, 831–835.
10. Poke, F.S.; Vaillancourt, R.E.; Elliott, R.C.; Reid, J.B. Sequence variation in two lignin biosynthesis genes, cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase 2 (CAD2). *Mol Breed.* **2003**, *12*, 107–118.
11. Guillet-Claude, C.; Birolleau-Touchard, C.; Manicacci, D.; Fourmann, M.; Barraud, S.; Carret, V.; Martinant, J.P.; Barri Re, Y. Genetic diversity associated with variation in silage corn digestibility for three O-methyltransferase genes involved in lignin biosynthesis. *Theor. Appl. Genet.* **2004**, *110*, 126–135.
12. Hancock, J.E.; Bradley, K.L.; Giardina, C.P.; Pregitzer, K.S. The influence of soil type and altered lignin biosynthesis on the growth and above and belowground biomass allocation of *Populus tremuloides*. *Plant Soil* **2008**, *308*, 239–253.
13. Gonzalez-Martinez, S.C.; Wheeler, N.C.; Ersoz, E.; Nelson, C.D.; Neale, D.B. Association Genetics in *Pinus taeda* L. I. Wood Property Traits. *Genetics* **2007**, *175*, 399–409.
14. Boudet, A. A new view of lignification. *Trends Plant Sci.* **1998**, *3*, 67–71.
15. Bonawitz, N.D.; Chapple, C. The genetics of lignin biosynthesis: Connecting genotype to phenotype. *Annu. Rev. Genet.* **2010**, *44*, 337–363.
16. Baucher, M.; Chabbert, B.; Pilate, G.; van Doorselaere, J.; Tollier, M.T.; Petit-Conil, M.; Cornu, D.; Monties, B.; van Montagu, M.; Inze, D.; *et al.* Red Xylem and Higher Lignin Extractability by Down-Regulating a Cinnamyl Alcohol Dehydrogenase in Poplar. *Plant Physiol.* **1996**, *112*, 1479–1490.

17. Lapierre, C.; Pollet, B.; MacKay, J.J.; Sederoff, R.R. Lignin structure in a mutant pine deficient in cinnamyl alcohol dehydrogenase. *J. Agric. Food Chem.* **2000**, *48*, 2326–2331.
18. Pilate, G.; Guiney, E.; Holt, K.; Petit-Conil, M.; Lapierre, C.; Leple, J.C.; Pollet, B.; Mila, I.; Webster, E.A.; Marstorp, H.G.; *et al.* Field and pulping performances of transgenic trees with altered lignification. *Nat. Biotechnol.* **2002**, *20*, 607–612.
19. Sibout, R. Expression Pattern of Two Paralogs Encoding Cinnamyl Alcohol Dehydrogenases in Arabidopsis. Isolation and Characterization of the Corresponding Mutants. *Plant Physiol.* **2003**, *132*, 848–860.
20. Yu, Q.; Li, B.; Nelson, C.D.; McKeand, S.E.; Batista, V.B.; Mullin, T.J. Association of the *cad-n1* allele with increased stem growth and wood density in full-sib families of loblolly pine. *Tree Genet. Genomes* **2006**, *2*, 98–108.
21. Marroni, F.; Pinosio, S.; Zaina, G.; Fogolari, F.; Felice, N.; Cattonaro, F.; Morgante, M. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. *Tree Genet. Genomes* **2011**, *7*, 1011–1023.
22. Rahantamalala, A.; Rech, P.; Martinez, Y.; Chaubet-Gigot, N.; Grima-Pettenati, J.; Pacquit, V. Coordinated transcriptional regulation of two key genes in the lignin branch pathway—CAD and CCR—Is mediated through MYB-binding sites. *BMC Plant Biol.* **2010**, *10*, 130.
23. Rogers, L.A.; Dubos, C.; Surman, C.; Willment, J.; Cullis, I.F.; Mansfield, S.D.; Campbell, M.M. Comparison of lignin deposition in three ectopic lignification mutants. *New Phytol.* **2005**, *168*, 123–140.
24. Kirk, T.K.; Obst, J.R. Lignin Determination. In *Methods in Enzymology*; Willis, A., Wood, S.T.K., Eds.; Academic Press: Waltham, MA, USA, 1988; Volume 161, pp. 87–101.
25. Porth, I.; Klapste, J.; Skyba, O.; Lai, B.S.; Gerald, A.; Muchero, W.; Tuskan, G.A.; Douglas, C.J.; El-Kassaby, Y.A.; Mansfield, S.D. *Populus trichocarpa* cell wall chemistry and ultrastructure trait variation, genetic control and genetic correlations. *New Phytol.* **2013**, *197*, 777–790.
26. Zhu, H.Y.; Weng, Y.H.; Zhang, H.G.; Meng, F.R.; Major, J.E. Comparing fast- and slow-growing provenances of *Picea koraiensis* in biomass, carbon parameters and their relationships with growth. *For. Ecol. Manag.* **2013**, *307*, 178–185.
27. Saghai-Marouf, M.A.; Soliman, K.M.; Jorgensen, R.A.; Allard, R.W. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 8014–8018.
28. Araki, N.H.T.; Khatab, I.A.; Hemamali, K.K.G.U.; Inomata, N.; Wang, X.; Szmidt, A.E. Phylogeography of *Larix sukaczewii* Dyl. and *Larix sibirica* L. inferred from nucleotide variation of nuclear genes. *Tree Genet. Genomes* **2008**, *4*, 611–623.
29. Mandrou, E.; Hein, P.R.G.; Villar, E.; Vigneron, P.; Plomion, C.; Gion, J. A candidate gene for lignin composition in Eucalyptus: Cinnamoyl-CoA reductase (CCR). *Tree Genet. Genomes* **2012**, *8*, 353–364.
30. Kumar, S.; Tamura, K.; Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* **2004**, *5*, 150–163.
31. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595.

32. Watterson, G.A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **1975**, *7*, 256–276.
33. Zhang, H.; Liu, S.H.; Zhang, Q.; Zhang, Y.D.; Wang, S.Z.; Wang, Q.G.; Wang, Y.X.; Tang, Z.Q.; Li, H. Fine-mapping of quantitative trait loci for body weight and bone traits and positional cloning of the RB1 gene in chicken. *J. Anim. Breed. Genet.* **2011**, *128*, 366–375.
34. Thumma, B.R.; Southerton, S.G.; Bell, J.C.; Owen, J.V.; Henery, M.L.; Moran, G.F. Quantitative trait locus (QTL) analysis of wood quality traits in *Eucalyptus nitens*. *Tree Genet. Genomes* **2010**, *6*, 305–317.
35. Zhang, D.; Du, Q.; Xu, B.; Zhang, Z.; Li, B. The actin multigene family in *Populus*: Organization, expression and phylogenetic analysis. *Mol. Genet. Genomics* **2010**, *284*, 105–119.
36. Zhang, D.; Xu, B.; Yang, X.; Zhang, Z.; Li, B. The sucrose synthase gene family in *Populus*: Structure, expression, and evolution. *Tree Genet. Genomes* **2011**, *7*, 443–456.
37. Selman-Housein, G.; López, M.A.A.; Hernández, D.; Civardi, L.; Miranda, F.; Rigau, J.; Puigdomènech, P. Molecular cloning of cDNAs coding for three sugarcane enzymes involved in lignification. *Plant Sci.* **1999**, *143*, 163–171.
38. Lynch, D.; Lidgett, A.; McInnes, R.; Huxley, H.; Jones, E.; Mahoney, N.; Spangenberg, G. Isolation and characterisation of three cinnamyl alcohol dehydrogenase homologue cDNAs from perennial ryegrass (*Lolium perenne* L.). *J. Plant Physiol.* **2002**, *159*, 653–660.
39. Hawkins, S.; Boudet, A.; Grima-Pettenati, J. Characterisation of caffeic acid O-methyltransferase and cinnamyl alcohol dehydrogenase gene expression patterns by *in situ* hybridisation in *Eucalyptus gunnii* Hook. plantlets. *Plant Sci.* **2003**, *164*, 165–173.
40. Eudes, A.; Pollet, B.; Sibout, R.; Do, C.; Séguin, A.; Lapierre, C.; Jouanin, L. Evidence for a role of AtCAD 1 in lignification of elongating stems of *Arabidopsis thaliana*. *Planta* **2006**, *225*, 23–39.
41. Hirano, K.; Aya, K.; Kondo, M.; Okuno, A.; Morinaka, Y.; Matsuoka, M. OsCAD2 is the major CAD gene responsible for monolignol biosynthesis in rice culm. *Plant Cell Rep.* **2012**, *31*, 91–101.
42. Dillon, S.K.; Nolan, M.; Li, W.; Bell, C.; Wu, H.X.; Southerton, S.G. Allelic Variation in Cell Wall Candidate Genes Affecting Solid Wood Properties in Natural Populations and Land Races of *Pinus radiata*. *Genetics* **2010**, *185*, 1477–1487.
43. Gilchrist, E.J.; Haughn, G.W. TILLING without a plough: A new method with applications for reverse genetics. *Curr. Opin. Plant Biol.* **2005**, *8*, 211–215.
44. Tian, J.; Chang, M.; Du, Q.; Xu, B.; Zhang, D. Single-nucleotide polymorphisms in PtoCesA7 and their association with growth and wood properties in *Populus tomentosa*. *Mol. Genet. Genomics* **2014**, *289*, 439–455.
45. Tian, J.; Du, Q.; Chang, M.; Zhang, D. Allelic variation in PtGA20Ox associates with growth and wood properties in *Populus* spp. *PLoS ONE* **2012**, *7*, e53116.
46. Zhang, D.; Yang, X.; Zhang, Z.; Li, B. Expression and nucleotide diversity of the poplar COBL gene. *Tree Genet. Genomes* **2010**, *6*, 331–344.
47. Thumma, B.R. Polymorphisms in Cinnamoyl CoA Reductase (CCR) Are Associated With Variation in Microfibril Angle in *Eucalyptus* spp. *Genetics* **2005**, *171*, 1257–1265.
48. Xu, N.; Loflin, P.; Chen, C.Y.; Shyu, A.B. A broader role for AU-rich element-mediated mRNA turnover revealed by a new transcriptional pulse strategy. *Nucleic Acids Res.* **1998**, *26*, 558–565.

49. Gonzalez-Martinez, S.C.; Huber, D.; Ersoz, E.; Davis, J.M.; Neale, D.B. Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity* **2008**, *101*, 19–26.
50. Tian, J.; Du, Q.; Li, B.; Zhang, D. Single-nucleotide polymorphisms in the 5' UTR of UDP-glucose dehydrogenase (PtUGDH) associate with wood properties in *Populus tomentosa*. *Tree Genet. Genomes* **2014**, *10*, 339–354.
51. Thumma, B.R.; Matheson, B.A.; Zhang, D.; Meeske, C.; Meder, R.; Downes, G.M.; Southerton, S.G. Identification of a Cis-Acting Regulatory Polymorphism in a Eucalypt COBRA-Like Gene Affecting Cellulose Content. *Genetics* **2009**, *183*, 1153–1164.

© 2015 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 license (<http://creativecommons.org/licenses/by/4.0/>).