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

Yanhong Wang 1,2, Qinbin Jia 1, Lei Zhang 1, Zhen Zhang 1 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: hanguozhang1@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 (πT = 0.00622 and θW = 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.
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), and other plant systems.
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 $\pi$ [31] and $\theta$ (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), \( \mu \) 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:

\[
\sigma_g^2 = \sum_{i=1}^{n} q_i G_i^2 - (\sum_{i=1}^{n} q_i G_i)^2 \quad h^2 = \frac{\sigma_g^2}{\sigma_p^2} \tag{4}
\]

(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.

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.

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

#### 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 Haploviev 4.2 software (http://www.broad.mit.edu/mpg/haploviev.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).
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).

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

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

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

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).

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

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 exon 3, 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 ($\pi$), 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 $\theta_W$ 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 ($\pi$ was 0.00377) was substantially lower than in the non-coding regions ($\pi$ 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 ($\pi$ was 0.00622) of L. olgensis was higher than that of CAD4 ($\pi$ was 0.0012) in Populus nigra and nine different genes ($\pi$ was 0.0018)in Populus trichocarpa [21,43], somewhat lower than that of PtoCesA7 ($\pi$ was 0.0091) [44] and PtGA20Ox ($\pi$ was 0.0099) [45], and similar to that of PtACT1 ($\pi$ was 0.0079) [35] and PtCOBL4 ($\pi$ 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-*n1* 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
Forests 2015, 6

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


© 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/).