A sequence mutation in the cinnamyl alcohol dehydrogenase gene associated with altered lignification in loblolly pine

Geoffrey P. Gill1, Garth R. Brown1 and David B. Neale1,2,*

1Department of Environmental Horticulture, University of California, Davis, CA 95616, USA
2Institute of Forest Genetics, Pacific South-west Research Station, US Department of Agriculture Forest Service, Davis, CA 95616, USA

Summary
Evidence for the molecular basis of a null allele of cinnamyl alcohol dehydrogenase (CAD) has been discovered in the loblolly pine (Pinus taeda L.) clone 7-56. The mutation is a two-base pair adenosine insertion located in exon 5 that causes a frame-shift which is predicted to result in premature termination of the protein. For routine detection of the mutation, a diagnostic assay was developed utilizing Template-directed Dye-terminator Incorporation and Fluorescence Polarization detection (FP-TDI). Loblolly pine is the most important commercial tree species in the USA, being harvested for pulp and solid wood products. Chemical pulping could be increased in efficiency by selecting for trees having a two-base pair adenosine insertion, by use of the rapid diagnostic assay developed in this study.

Introduction
Loblolly pine (Pinus taeda L.) is the most intensively grown tree species in the USA for pulp and solid wood products, with plantations exceeding 12 million ha. The extraction of lignin from wood during the production of pulp and paper requires the use of costly chemicals that are toxic to the environment. Significant progress towards increasing pulping efficiency has been achieved in poplar trees through genetic manipulation of the genes involved in lignin biosynthesis (Baucher et al., 1996; Hu et al., 1999; Pilate et al., 2002). One of the key enzymes to be successfully targeted, cinnamyl alcohol dehydrogenase (CAD), catalyses the final step in the synthesis of monolignols by converting cinnamaldehydes to cinnamyl alcohols. Field-grown transgenic poplar trees with reduced CAD were easier to delignify, using smaller amounts of chemicals and yielding more high quality pulp without an adverse effect on growth (Pilate et al., 2002).

A null CAD allele (cad-n1) has been discovered in the loblolly pine clone 7-56 (MacKay et al., 1997) during the screening of a germplasm collection with different isozymes. Clone 7-56 was shown to be heterozygous for the cad-n1 allele, based on megagametophytes segregating 1 : 1 for CAD activity. Homozygous seedlings (cad-n1/cad-n1) obtained by selfing, contain between 0 and 1% of wild-type CAD activity (MacKay et al., 1997) and display a brown-red wood phenotype. The expression level of cad transcript in shoot, megagametophyte and xylem tissues was ~20-fold less in cad-n1 homozygous plants compared to wild-type (MacKay et al., 1997).

Deficiency of CAD in cad-n1 homozygotes only slightly reduces lignin content but drastically alters lignin composition (Lapierre et al., 2000; MacKay et al., 1997; MacKay et al., 2001; Ralph et al., 1997). The major compositional change in lignin was attributed to the incorporation of dihydroconiferyl alcohol (DHCA), a minor component of most lignins, but elevated to 10-fold greater levels in cad-n1 homozygous trees. Coniferaldehyde, the substrate of CAD, and vanillin are also present at increased levels, while the coniferyl alcohol component of normal lignin decreased.

The mutation has a variable effect on pulping efficiency, depending on the age of the trees and whether the mutation is present in a homozygous or heterozygous state. In totally CAD-deficient trees (cad-n1/cad-n1), delignification was significantly easier but the pulp yields were relatively low (~33%) compared to normal trees (48%) (Dimmel et al., 2001). In 4–6-year-old partially CAD-deficient trees...
(heterozygous) delignification increased in efficiency by ~20% and yields were similar to the wild-type (Dimmel et al., 2002). In contrast to these younger trees, a small sample of 14-year-old partially CAD-deficient trees displayed no major differences in ease of delignification and pulp yield (Dimmel et al., 2002).

In addition to lignin composition changes, the cad-n1 allele appears to be associated with increased stem-growth traits in heterozygous trees (Wu et al., 1999). This growth promotion correlates to an increase in the debarked volume of 4-year-old trees (14%) (Wu et al., 1999) that is also observed in 14-year-old trees (Dimmel et al., 2002). A possible explanation could be that trees harbouring the cad-n1 allele may invest fewer resources into the production of monolignols, allowing reallocation of resources towards growth. Promotion of growth was also observed in transgenic poplar, where the lignin biosynthetic enzyme 4-coumarate:coenzyme A ligase (4CL) was down-regulated (Hu et al., 1999).

This paper describes the discovery of a sequence mutation likely to be responsible for the loss of function associated with the cad-n1 allele identified during SNP discovery within the cad gene of loblolly pine. We have subsequently developed a genotyping assay for the mutant allele utilizing Template-directed Dye-terminator Incorporation and Fluorescence Polarization detection (FP-TDI) (Hsu et al., 2001; Kwok, 2002). Until now, the mutation could only be accurately diagnosed using CAD isozyme analysis on haploid megagametophytes obtained from seed. Isozyme analysis cannot be applied to diploid tissue, such as needles, because the null mutation is recessive and heterozygotes cannot be distinguished from wild-type homozygotes. The diagnostic tool developed here will allow breeders to accurately determine the presence and copy number of the cad-n1 allele in their germplasm at any age and from any tissue type. Early and rapid screening for this economically important mutation is therefore now possible.

Results and discussion

Discovery of a sequence mutation in the cad-n1 allele
SNP discovery within the cad gene was performed on haploid megagametophyte DNA from clone 7-56 and 31 other unrelated individuals. A two-base pair adenosine insertion was identified as being unique to clone 7-56, known to be deficient in CAD activity. The insertion was located in the second codon of exon five and creates a frame-shift that generates a premature stop codon (Figure 1). Seventeen haploid megagametophytes from the heterozygous 7-56 clone were assayed by isozyme gel electrophoresis and DNA sequence analysis to confirm that the discovered sequence mutation was associated with CAD-deficiency. In every case, the two-base pair adenosine insertion corresponded with an absence of CAD activity (data not shown).

The transcript and hence active protein of the cad-n1 allele is barely detectable compared to wild-type (MacKay et al., 1997). Sequence data obtained over the entire coding region for the wild-type and cad-n1 allele (data not shown) did not reveal any further sequence differences that could account for the decreased levels of cad-n1 transcript. The lack of cad-n1 transcript could be attributed to the double adenosine mutation discovered introducing a premature termination codon that subjects cad-n1 mRNA to nonsense-mediated decay (for review, see Hentz and Kulozik, 1999). Although the mutation presented in this paper is most likely responsible for the null cad-n1 allele, the possibility of a further sequence mutation residing in the promoter region which drastically reduces expression could not be excluded at this point.

Identification of a second cad locus
During amplification of the sequencing template, two additional products were amplified in 4 out of 32 individuals...
Sequence mutation associated with altered lignification in loblolly pine

The secondary products were not amplified in clone 7-56. To remove any potential interference from the amplification of additional cad loci occurring in the genotyping assays this was investigated further. A second cad locus (cad-ps1) has been identified by RFLP analysis which maps to linkage group 2 in loblolly pine (Figure 2). Progeny of the mapping population were analysed for the same co-segregation pattern using the primers CADF4 and CADR4 (Table 1) to determine whether the amplified secondary products were from the same locus as the cad-ps1 RFLP locus. The 210 bp and 370 bp secondary amplification products co-segregated with the allele from the cad-ps1 RFLP locus and mapped to linkage group 2 in loblolly pine (Figure 2). The 370 bp fragment was inferred to be a heteroduplex formed with the cad-ps1 and the primary cad locus. The product size of cad-ps1 differed by ~100 bp, which corresponded to a fragment lacking intron four. As expected, a primer (CADF8) designed within intron four failed to amplify the cad-ps1 locus, enabling the primary CAD locus to be specifically amplified for the genotyping assays.

Our data supports previous evidence showing that there is only one functional cad locus (MacKay et al., 1995) in loblolly pine, but also provides evidence for one other cad locus that is most likely a pseudogene and present in about ~15% of individuals. The cad-ps1 locus lacks at least introns three, four and five (data not shown) and thus belongs to the ‘processed’ class of pseudogenes that arise from reverse transcribed mRNA that is reintegrated into the genome (Vanin, 1985).

**FP-TDI assay for the cad-n1 allele**

The designs of the forward and reverse FP-TDI assays are shown in Figure 3. Trial testing of the assay was performed on 167 plants obtained from nine different crosses involving clone 7-56 or progeny from 7-56. Results from a subset of 96 plants using the forward and reverse FP-TDI assay are shown in Figure 4. Controls were included that consisted of all three possible genotype classes and blanks that contained no DNA. Samples that did not fall clearly into a genotype cluster (1–2%) were not scored. When both the forward and reverse

<table>
<thead>
<tr>
<th>Function</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery (PCR and sequencing)</td>
<td>CADF2-CTCTGTATATGCAAGGGCTTACA</td>
<td>CADR2-CGAAGTGCAACGCTCTGG</td>
</tr>
<tr>
<td>Pseudogene analysis (PCR)</td>
<td>CADF4-TCGGTTAGAGGCGCTGG</td>
<td>CADR4-AGGTAGACCTGAATGCTGG</td>
</tr>
<tr>
<td>FP-TDI (PCR)</td>
<td>CADF8-TGAAAAGATGATGCGGCCA</td>
<td>CADR2-CGAAGTGCAACGCTCTGG</td>
</tr>
<tr>
<td>FP-TDI assay</td>
<td>CAD1528F-ATCCGTGTTGAGCACGAAAA</td>
<td>CAD1528R-GTAATCTAGCCTCTCTGCTTT</td>
</tr>
</tbody>
</table>

Figure 2  Identification of a second cad locus in loblolly pine. Co-segregation of the cad-ps1 RFLP locus (lower panel) with secondary products amplified using the CADF4 and CADR4 primers (top panel) in 25 individuals from a QTL mapping population. The larger PCR product (top panel) is an inferred heteroduplex formed between the cad and cad-ps1 locus. The cad locus maps to linkage group 9 and the cad-ps1 locus to linkage group 2. Lanes marked M1 and M2 correspond to the 100 bp (Promega) and 1 kb (Gibco BRL) DNA size ladders, respectively.
reaction results were combined, all the plants were accurately assigned to a genotype class and no contradictory genotypes were observed. The absence of homozygous cad-n1 clones was expected based on the parental genotypes used to construct the nine crosses tested.

Analysing an indel mutation by single-base extension has the potential to give a false result if a substitution occurs in the position examined (Figure 3). For example, if the first nucleotide of codon 241 (G) was substituted to an adenosine (forward assay) or the first base of codon 240 (G) was substituted to an adenosine (reverse assay) a false positive result for the cad-n1 allele would occur. Both of these positions require non-synonymous amino acid changes to occur, alanine to threonine in the forward and glutamine to lysine in the reverse. These non-synonymous changes were not observed in any of the clones present on the SNP discovery panel nor in a selection of 242 first-generation clones. If both the forward and reverse assays are performed, the probability of an error occurring due to nucleotide substitutions would be extremely low.

Since the FP-TDI assay is based on single-base extension, it should be amenable to other platforms such as the SureScore SNP Genotyping Kit (Invitrogen, Carlsbad, CA) and SNaPSHOT (Applied Biosystems, Foster City, CA).

**Frequency of the cad-n1 allele**

The frequency of the cad-n1 allele was estimated by analysing the 242 first generation clones that were distributed across the present-day range of loblolly pine (from Texas to Florida and extending north to Delaware). The mutation was not detected in any of the clones analysed using the forward FP-TDI assay, confirming the rarity of this mutation. The frequency of cad-n1 might be higher in some populations, such as in the region where 7-56 was discovered (Williamsburg, NC), however, much more extensive sampling would be required.

The frequency of cad-n1 in loblolly pine breeding populations and plantations will likely increase due to the inclusion of 7-56 as an elite parent in numerous co-operative and private breeding programmes. The diagnostic tool presented here will allow breeders to rapidly screen for the presence of
the cad-n1 allele in their germplasm. Screening of additional loblolly pine populations could be performed to identify new select trees harbouring the cad-n1 allele.

**Methods**

**Plant material, CAD isozyme assays and DNA extractions**

Four plant material sources were used for the identification and testing of the presence of the cad-n1 allele or the cad-ps1 pseudogene: (i) a panel of 32 loblolly pine megagametophytes (Weyerhaeuser Company Federal Way, WA), including one megagametophyte from clone 7-56, was used for SNP discovery within the cad gene, (ii) 167 clones (CellFor Inc., Vancouver, BC, Canada) resulting from nine crosses, using clone 7-56 or 7-56 offspring as parents, was used for testing the FP-TDI assay, (iii) a selection of 242 first-generation clones (North Carolina State University Cooperative Tree Improvement Program and Weyerhaeuser Company Federal Way, WA) from the natural range of loblolly pine was used to estimate the frequency of the cad-n1 allele, and (iv) 96 progeny from the QTL pedigree (IFGVEQ) used in a study to identify QTLs for wood specific gravity (Neale et al., 2002) were used to investigate the cad-ps1 locus.

Seeds from loblolly pine clone 7-56 were germinated and the haploid megagametophytes were removed for CAD isozyme analysis or DNA extraction. CAD isozyme assays were performed as described by MacKay et al. (1995). All DNA extractions were performed using the Plant DNAeasy kit (Qiagen, Valencia, CA) in either a single tube or 96-well format.

**PCR**

All primers for PCR and their usage are described in Table 1 and their relative position within the cad gene shown in Figure 1. Primer CAD-F8 was designed within intron four so as not to amplify the cad-ps1 locus. Primers CAD-F4 and CAD-R4 were designed so that the cad and cad-ps1 products could be size distinguished on agarose gels more easily than products amplified with the CAD-F2 and CAD-R2 primers. All PCR reactions were performed on ~20 ng template in a total volume of 25 µL. Each reaction comprised 0.8 µM of each primer; 0.65 units of HotStarTaq DNA polymerase (Qiagen, Valencia, CA); 1× PCR buffer containing 1.5 mM Mg; and 100 µM each of dATP, dCTP, dGTP and dTTP (Applied Biosystems, Foster City, CA). Amplification was performed on a PTC100 thermocycler (MJ Research, Waltham, MA) with the following parameters: initial denaturation step of 95 °C for 15 min (for activation of HotStarTaq) followed by 37 amplification cycles of 30 s at 95 °C, 30 s at 60 °C and 2 min at 72 °C.

**DNA sequencing and analysis**

Degradation of excess primers and dNTPs was performed on 5 µL of PCR product treated with 1 U of exonuclease I (USB, Cleveland, OH) and 1 U of shrimp alkaline phosphatase (USB, Cleveland, OH) that was incubated at 37 °C for 1 h followed by a heat inactivation step of 85° for 15 min. The primers for PCR were also used for sequencing (Table 1). Cycle sequencing was performed using an ABI Prism big dye terminator mix (Applied Biosystems, Foster City, CA) using standard conditions as supplied by the manufacturer. Reactions were run on an ABI 377 Automated DNA sequencer using standard ABI protocols. Sequencher (GeneCodes, Ann Arbor, MI) was used to assemble the sequences into a contig where polymorphic differences could easily be visualized. The cad cDNA and translated protein sequence used for alignment in this study had the GenBank accession numbers Z37992 and CAA86073, respectively. The intron and exon structure of the cad gene was inferred from a Pinus radiata genomic sequence (AF060491).

**Analysis of the cad-ps1 locus**

Mapping and co-segregation analysis of the cad-ps1 locus was performed on 96 progeny from the QTL pedigree (IFGVEQ) used in a study to identify QTLs for wood specific gravity. (Neale et al., 2002). RFLP Southern blots were also obtained from this study. Products amplified from the cad-ps1 locus and the cad locus, using primers CADF4 and CAD R4 (Table 1), were resolved on a 1.5% agarose gel.

**Detection of the cad-n1 allele using FP-TDI**

The FP-TDI assay relies on a single base extension of different fluorescent labelled dideoxynucleotide terminators. FP is observed when a fluorescent molecule excited by plane-polarized light emits polarized light into a fixed plane and the molecule remains stationary between excitation and emission (Perrin, 1926). Since a large molecule has slower rotation than a smaller one, a dye-labelled terminator attached to a primer will show a higher level of polarization than an unincorporated terminator, if excited by plane polarized light.

The template for the assays was amplified using the primers CADF8 and CADR2 (Table 1), as described in the section on PCR. The assay design for the forward and reverse...
reactions is shown in Figure 3, and the primer sequences listed in Table 1. FP-TDI reactions were performed using an Acycloprime-FP SNP detection kit (Perkin Elmer Life Sciences, Boston, MA) as described by the manufacturer, except that the thermocycling conditions were altered to 25 cycles consisting of 95 °C for 15 s and 54 °C for 30 s. Fluorescence polarization was measured on a Wallac Victor2 plate reader (Perkin Elmer Life Sciences, Boston, MA) with the manufacturer's recommended filter sets and G-Factor calibration.

Acknowledgements
We are grateful to Dave Ellis, Margarita Gilbert (CellFor Inc., Vancouver, BC, Canada) and Nicholas Wheeler (Weyerhaeuser Company Federal Way, WA, USA), for providing loblolly pine tissue and DNA, and to John MacKay for a critical reading of the manuscript. Funding for this project was provided by the National Science Foundation (grant no. 9975806).

References