Methods for simultaneous control of lignin content and composition, and cellulose content in plants

Vincent Lee C. Chiang
Laigeng Li

Follow this and additional works at: http://digitalcommons.mtu.edu/patents
Part of the Biological Engineering Commons

Recommended Citation
http://digitalcommons.mtu.edu/patents/88
METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND COMPOSITION, AND CELLULOSE CONTENT IN PLANTS

Inventors: Vincent Lee C. Chiang, Hancock, MI (US); Laigeng Li, Houghton, MI (US)

Assignee: Board of Control of Michigan Technological University, Houghton, MI (US)

Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 125 days.

Prior Publication Data

References Cited
U.S. PATENT DOCUMENTS
6,015,943 A * 1/2000 Boudet et al. 800/298
6,066,780 A 5/2000 Boudet et al.

OTHER PUBLICATIONS
Hu et al., 1998, PNAS 95:5407–5412.*
Tsai et al., 1994, Plant Cell Report 14:94.
Boudet et al., 1995, New Phyto. 129:203.
Brasileiro et al., 1992, Transgenic Res. 1:133.
Daneckar et al., 1987, Bio/Technology 5:587.

Lawton et al., 1987, Plant Mol. Biol. 9:315.
Li et al., 1999, Plant Mol. Biol. 40:555.
Li et al., 2000, J. Biol. Chem. 275:6537.
McGranahan et al., 1988, Bio/Technology 6:800.
McGranahan et al., 1990, Plant Cell Reports 8:512.
Parsons et al., 1986, Bio/Technology 4:533.
Walker et al., 1987, PNAS USA 84:6624.
Yang et al., 1990, PNAS USA 87:4144.
Yamazaki et al., 1993, Bio/Technology 11:137.

(End of List continued on next page.)

Primary examiner—David T. Fox
Assistant examiner—Stuart F. Baum

(74) Attorney, Agent, or Firm—Michael Best + Friedrich LLP

ABSTRACT

The present invention relates to a method of concurrently introducing multiple genes into plants and trees is provided. The method includes simultaneous transformation of plants with multiple genes from the phenylpropanoid pathways including 4CL, CAld5H, AldOMT, SAD and CAD genes and combinations thereof to produce various lines of transgenic plants displaying altered agronomic traits. The agronomic traits of the plants are regulated by the orientation of the specific genes and the selected gene combinations, which are incorporated into the plant genome.

30 Claims, 25 Drawing Sheets
OTHER PUBLICATIONS

Ebert et al. 1987, PNAS USA, 84:5745–5749.
Minocha et al., 1986, Proc. TAPPI Research and Development Conference, TAPPI Press, Atlanta, 89.

Sambrook et al., 2nd ed. 1982.
EMBL Acc#X62096 Bugos et al., 1991; Alignment with SEQ ID No: 6.

* cited by examiner
Phenylalanine

Cinnamate

\[ \text{C}_{4}\text{H}_{4} \]

4-Coumarate

\[ \text{C}_{3}\text{H}_{4} \]

Caffeate

\[ \text{C}_{4}\text{H}_{6} \]

Caffeoyl-CoA

\[ \text{CCoA} \text{OMT} \]

Feruloyl-CoA

\[ \text{CCO} \text{R} \]

Coniferaldehyde

\[ \text{CAD} \]

\[ \text{CAD} \]

Coniferyl alcohol

\[ \text{GUAIACYL-SYRINGYL lignin} \]

FIG. 1
FIG. 2A  SAD cDNA sequence

1  TTTTTTTTTT TTCTCTAGCC TTCTTTCTCG ACGATATTTC TCTATCTGAA
51  GCAAGCAGCA TGTCCTAAGTC ACCAGAAGAA GAACACCCCTG TGAAGGCCYT
101  CCGGTGGGCT GCTAGGGAAC AATCTGTTCA TCTTTTCTCC TCTAACTTCT
151  CCAGGAGGCA AACCTGTTGGA GAGGATGTGA GGTTCAGGCT GCTGATCTGC
201  GGGATATGCT ATTCGACATC TCAAATGACT GGGGCTTCTC
251  CATGTAACCTT TTGTTTCTCTG GGCATGAAAT TGTGAGGGAA GTGACAGAG
301  TTGGAGGCAA GGTGAAAAAG GTTAAATGAGT GAGACAAGGT GGGGCTGGGA
351  TCGTTGATGT GTGCATGTCG GCTCTGAGAG CTGGTGAGGA ATGATCTTGA
401  AAATTACTGT CAAAATATGA TCCTGACATA CAGCTTCACT TACCATGAGC
451  GAACCAGCAG TCAGCTGTTGC TACTCGAGTC ACATGTTGCG TACCGAAGC
501  TACATCATCC GATTCGGCGG TAATATGCCT CGTGAACGGG TGGCTCTCTT
551  CTTTTGTGCC GCCATTACAC TGTATAGTCC TCTGAAATAT TTTGGACTAG
601  ATGAAACCGG TAAAGCATATC GGTATCGTTG GCTTAGGTGG ACTTGGCTAC
651  GTGGCTGCTA AATTGCAAGG GCCTTTGAGA TCTAAATGAG TAGATTTT
701  TACCTCCCC TCCCAAAGAGG AGGAGGCTTT GAAAGACCTC GGTGCAGACT
751  CATTTTTGAAT TAGTCCTGAG CAAAGAGCAA TGCAGGCTGC CGCAGGAACA
801  TTAGATGCA TAATCGCTGC ACGCGATCCA ATGGGCTGTTG GCTGGCGCTC
851  GTTTTGACTG TTGAAAGTCTC ACAGGAGACG TATCTTGGTG GGGCACCAG
901  AAAGAGCTCT TGAGCTACTC GCCCTTTTCTTG TGATGCTGG AGGAGAGATA
951  GTGTGCGGGA GTGTATGTTG AGGCATGAAG GACACACAAG AGATGATTGA
1001 TTGTGAGGCA AAACAAAACA TCACAGCAGA TATCGAGATT ATTTCAACGG
1051 ACTATCTTAA TAGGCGGATA GAACGGTTGG TAAAGAACGA TGTCAGATAC
1101 CGATTCGCTCA TTGACGGTTG CAATACTCTG GCCAGCTAGC AGCCTAAGG
1151 AGAAGATCCC ATGCTCTGCA ACCCTTTATA AAATCTGATA ACATGTTTGG
1201 ATTTTATGAA TAAATGATT ATCTTTGAGA TTTTTTTTAA AAACAAGGAAG
1251 TGTTCTGCAAA AACTTTACAT CGGAAATACC CTTGCAGCTA CGAGAAACGC
1301 TTATAGATGTG TTGATAGATGT TCTTTTATTA GGGGATGACC ATGCTCTCTG
1351 GTCTTTTCTGA AGATCCATTT ATAGGTTGCCG GAATGATATG ACAAACAAAAA
1401 ATGTTTGCGG CCTCTCTTCA AAAAAA AAAAAA AAAAAA
FIG. 2B  SAD protein sequence

1 MSKSPEEEHP VKAFGWAARD QSGHLSPFNF SRRATGEEDV RFKVLYCGIC
51 HSDLHSIKND WGFSMYPLVP GHEIVGEVTE VGSKVKKVN V G DKVGVGCLV
101 GACHSCESCA NDLENYC PKM ILTYASIYHD GTITYGGYSD HMVANERYII
151 RFPDNMPLDG GAPLLCAGIT VYSPLKYFGL DEPGKHIGIV GLGGGLHVAV
201 KFACKAFGSKV TVISTSPSKK EEA KNFGAD SFLVSRDQE Q MQAAAGTLDG
251 I1DTVSAVHP LLPLFGLLLKS HGL LILDVAP EKPLELPFAS LIA GRKIVAG
301 SGGGMKETQ EMIDFAAKHN ITADI EVIST DYLNTAIERL AKNDVRYRFV
351 IDVGNTLAAT KP*
FIG. 3A  Aspen (P. tremuloides) PtCAD5H cDNA sequence

1  TAAAGTCTTG  TGGATTACAC  AAAATACAGA  CTGAAAACAT  CCATAGGCAC
51  CACACATATA  ACCATCCATG  GATTTCTCTTG  TCCAATCTTT  GAAGCTTCA
101  CCATGTCCTC  TCTTCGCTATG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
151  CTCCTCTTGG  CTCCGGCGAA  GATTGCCATA  TCCACGAGG  CCAAAAGGTT
201  TGCCACTTGTG  AGGTAGCAGT  AAAATACAGA  CTGAAAACAT  CCATAGGCAC
251  TTATCTAAGCA  ATATGGTGGG  CTCTTTCATA  GTACTTGCAT
301  GTACCTTGGAT  TCAAGACGGT  AAATGGGAG  TCTTGACTCA  CCAACCTTCC
351  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
401  AGTCTAGTCTT  GATCTCTCTCT  TCACTCTTCT  CCTCTCTCGC  CTTCGCCGAA
451  AGTTACTTAAAC  ATACGTTGGG  TCTTGAGGCT  GTACTTGCAT
501  AAAGGCCGCGA  ATACGTTGGG  TCTTGAGGCT  GTACTTGCAT
551  AAGAGTGGCC  ATACGTTGGG  TCTTGAGGCT  GTACTTGCAT
601  TTTTACGTTG  ACCATGAACA  TCACTTACAG  AGCAGCTTTC  ATGAAGGACA
651  ATGATGAGGCT  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
701  ATGATGAGGCT  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
751  GATTTCTTCTG  TCAAGACGGT  AAATGGGAG  TCTTGACTCA  CCAACCTTCC
801  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
851  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
901  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
951  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1001  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1051  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1101  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1151  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1201  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1251  GTGCGCAAGTG  TGGGATTGCA  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1301  GACAAAGGAA  ATACGTTGGG  TCTTGAGGCT  GTACTTGCAT
1351  GCACAAAGGAA  ATACGTTGGG  TCTTGAGGCT  GTACTTGCAT
1401  CGGCTGCTGG  CTGGATTGGG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1451  CGGCTGCTGG  CTGGATTGGG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1501  CGGCTGCTGG  CTGGATTGGG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1551  CGGCTGCTGG  CTGGATTGGG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1601  TGCGCGCGCGA  ATACGTTGGG  TCTTGAGGCT  GTACTTGCAT
1651  CGGCTGCTGG  CTGGATTGGG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1701  CGGCTGCTGG  CTGGATTGGG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG
1751  CGGCTGCTGG  CTGGATTGGG  GATTTCTTCT  TCCAATCTTT  TGCAGCTTGG

Figure 3A shows the cDNA sequence of the PtCAD5H gene in Aspen (P. tremuloides). The sequence is presented in a tabular format, with each nucleotide position indicated and the corresponding nucleotide base shown. The sequence is organized in rows and columns, with each row representing a segment of the cDNA sequence. The figure emphasizes the genetic information encoded within this sequence, which is crucial for understanding the molecular biology and genetic engineering of this plant species.
FIG. 3B  Aspen (*P. tremuloides*) PtCAld5H protein sequence

```
1  MDSLVQSLQA SPMSLFLIVI SSLFFFGLLS RLRRRLPYPP GPKGLPLVGS
51  MHHMDQITHR GLAKLAKQYG GLFHMROMGYL HMVTSSPEI ARQVLQVQDN
101 IFSNRPANIA ISYLTYDRAI MAFAYHGPFW RQMRKLCVMK LFSRKRASEW
151 ESVRDEVDSM LKTVANIGK PVNLGELIFT LTMNITYRAA FGAKNEGQDE
201 FIKILQEFSDK LFGAFNMSDF IPWLGWDIPQ GLSARLVKAR KALDRFIDS
251 IDDHIQKRKQ NKFSEDAETD MVDDMLAFYG EEARKVDESID DLQKAISLT
301 DNIKAIMDVE MFGGTETVAS AIEWVMAMELM KSPEDQKRVDQ QELAEVUGLE
351 RRVEESDIDK LTLKFKECLKE TLRRHHPIPL LLHETSEADAE VAGYIPKQT
401 RVMINAYAIG RDKNSWEDPD AFKPSRFLKP GVDFDKGNNHF EFPFPGSSGR
451 SCPGMLGLY TDLAVAHLL HCFTWELPDG MKFSEDMDTMD MFLTAPRAT
501 RLVAVPSKRVLCP*
```
FIG. 4A  Aspen (P. tremuloides) PtAldOMT cDNA sequence

GenBank accession number: X62096

1 tcacctctctt tccttacacc ttcttcacc cttttttccc ttgtgaatt
51 caaatctcag caagatgggt tcaacaggtt aaactcagat gactcccaatc
101 cagttatcag atgaagaggc aacacctcttt gccatgcaac tagccagtgcc
151 ttcagtttct caaatgatcct tcaaaaacgc cattgaaacct gacccctcttg
201 aaatcagtggc taagaaggtgc cctgtccttct tctgtcctac atctgtgatga
251 gctttctcacc tccctacaaa aacaacctgat ggcctctgca tgttagacgc
301 tatctcggcc ctcctgcgta ctactcccat tcttcctcgct tctctgaaag
351 atctctcctgta ttgggaagtt gagagactgt atggcctcgc tctctgtttgt
401 aaatcttttgaa ccaagaaccga ggacggtgctc tctgctcagcc ctctctgtct
451 catgaacccag gcaaggtgcc tcatggaaag ctggttatat ttgaaagatg
501 caattctttga tggaggaatt ccactttaaca aggcctatgg gtagactgca
551 tttgaatatct atggcagcggc tccagagattc aacaagatcc tcaacacaggg
601 aatgtcttgac cacctctacca ttaccatgaa gaagatctctt gagacctaca
651 aagcgctttgga aggcctcacc tcctttgggtg atgttgtggtg ttgggaactgga
701 gccgctgatta aacacacgct ctttaaatac ctttcacata agggcattaa
751 cttctgatgct cccacgctca ttgaggtatcg cccatctctat cccccgttgg
801 agcatgtttgg tggcgcacatg tttgttagtg tggccccaaagc agatgccccg
851 ttctatgagat ggtatgcca tgattggagcc gacgcaccct gcctaaattt
901 cttgaagatt tgtctatgacg cggtgccggga aacaaggcaag gtgattacttg
951 ttagtggcatt ctctccctgtgc ctctcgctac caaagcc tgtgcacccac
1001 gttgcgtgcag tttgtagtcat cattgtgagg cacaaccccg gttggaagaag
1051 gagggcggag aagaaatttg agggtttagc taagggagct gctttcacaag
1101 gtttttggaag tatttgctgt gctattcaca ccaatgatat gtagctgccc
1151 aagaaagccct aagcccccag tccagctctc aaggtacttg gggttttgca
1201 gangacgttgg cttgctgcttc tgccttttggatt gtttttttatt gttttttttt
1251 atacagaggag tagctatctc ttatggaaca tgtaagagta agatgtgctt
1301 ttagttgagctt cattctccct cattactcct cccatcttcat cccatctttt
1351 atatcgatgtg aaaaagatct ctattggtcat tttgccctca acaagttgg
1401 ctctctgtaacc gggaaagaaa caatctcag tggtagttat cttcagatg
1451 tatgagtatt ttcttaagcct tttaagtatt gttcagaaaaa aaaaaaaaaa
1501 aaaa
FIG. 4B  Aspen (*P. tremuloides*) PtAldOMT protein sequence

GenBank accession number: X62096

```
1  MGSTGETQMT PTQVSDEEAH LFAMQLASAS VLPMLKTAI ELDLLEIMAK
51  AGPGAPFLSTS EIAUSHPTKN PDAPVMDLRI LRLLASYSIL TCSLKDLPDG
101  KVERLYGLAP VCKPLTKNED GVSVSPLCLM NQQKVLMEW YYLKDAILDG
151  GIPFNKAYGM TAFEYHTGDP RFNKVFNKM GM SDHSTITMKK ILETYKGFEG
201  LTSLVDVGGS TGAVVNTIVS KYPSIKGINF DLPVIEDAP SYPGVHEVGG
251  DMFVSVPKAD AVFMKWICHD WSDAHCLKFL KNCYDLPEN GKVILVECIL
301  PVAPDSLAT KGVVHDVIM LAHNPGBKER TEKFEGLAK GAGFQGFEM
351  CCAFNTVIE FRKKA
```
FIG. 5A  4CL polynucleotide DNA sequence

cctcggcga aacctgaaaa cagagagcac ctaaaactca ccatctctcc cctgtgatct 60
ccctcgcga actccgaaaa cagagagcac ctaaaactca ccatctctcc cctgtgatct 112
ttagccgca atggacgcca ca atg aat cca caa gaa ttc atc ttt cgc tca 160
gtt ctt cag aac tgt tct aca cat tca tca aaa cct tcg ata aat 208
ggc gac aat gga gat tgc tac acc tat gct gat gtt gag ctc aca gca 256
agt cag att gtt cct gct gta gac aag att ggt att cca aca ggt gac 304

gtt atc atg ctc ttc caa aat gtt cta cct gaa ttc tgg ctc gct ttc 352
ttg ctc cta cca agt tca cct gaa ttc tgg ctc gct ttc 400
tcc acc ccc cct gcc gag cta gca aaa cct gcc aag gcc tct cca aag 448

cct ctt ata cca cag gct tgt tac tac gag aag gtt aca gat ttt gcc 496
cga gaa aat gat gtt aat gtc ctc tgt gtc gct gcc ctc gcc gag gtt 544
gtt cta ctt ccc cca ggt ctt aca cag gca gac gaa aat gaa gtt ctc 592
cag gtc gac att agt cct gat gat gtc gta gca ttc cct tct cca tca 640
ggg act aca ggg tgt cca aaa ggg gtc atg taa acg caa aca ggg cta 688
att acc aat gtt gtt cct gct gta gat gga gac aat cca aac tgt tat 736

agt aag tac aag gta ctt ata gcc cca gtt gtt cca cct gtt ctc aag att 928
gta att gct aag tca cct gat ctt gcc gag cat gcc ctc tct tct tgt 976
agg atg ata aca ccc cca tcc gga ggg gct cca ttc cgg gac gaa ctc gac 1024

act gtc aga gct aag ttg cct cag gct aga ctt ggt cag gga tat gga 1072
agt acc gag gca gga cct gtt cta gca atg tgc tgt gca ttt gcc aag 1120
gaa cca ttc gcc gac ata aca cca ggt gca tgt gga act gta gtc aag aat 1168

gca gag atg aag att gtt gac cca aca aca ggg gtc tct cta ccc agg 1216
aac cag cct ggt gac atc tgc atc cgg ggt gat cag atc atg aca gga 1264
tat ctt aat gac ccc gcc gac gcc acc tca aga atc aca aga aca tga cag 1312

ttg ctc ccc aca gcc gct atc gcc tac att gat gat gat gag ctt 1360
ttc atc gtt gac aga tgt aag gaa tgt atc aag taa aca ggg tgt tgt 1408

gtt gct ccc ccc cct gcc gct aag aag tca ccc cag gat ttc gtt aat ctt 1456
tcc gat gct gct gta gta gga tgt aat cgg gat ggg gga aac aca aag 1504
cct gtt gca ttt gta tgt aca gaa tca gaa cag ccc ggc gac gaa gat 1552
gaa att aag cag tat att tca aag cag ctt gta atc ttc tac aca aca tga 1600

aaa cgg gtt ttc ttc att gaa gcc att ccc aag gcc cca tca ggc aag 1648
gtt gct ctt tgt gta gta gga tca ggc ccc ggc cca cca gga gaa 1704
ttc gct gcc ccc cca gga gaa gtt gcc ccc tca gca gtc ctc gtc gct 1760
tggttaactgaa cattgcttctc gtttactctc tcttttaaattcttctggagatgcttctctgtct 1817
tgaaccaaga atgctttgaa aagacaagtga ccttactctt gttcttgaggactcttctctc 1883
tgacttact ttaatgctggtgttgtaa aacactagtgtgtaa gatgatattctatatgta 1915
FIG. 5B 4CL Aspen (P. tremuloides) amino acid sequence

Met Asn Pro Glu Phe Ile Phe Arg Ser
1  5  10
Lys Leu Pro Asp Ile Tyr Ile Pro Lys Asn Leu Pro Leu His Ser Tyr
15 20 25
Val Leu Glu Asn Leu Ser Lys His Ser Ser Lys Pro Cys Leu Ile Asn
30 35
Gly Ala Asn Gly Asp Val Tyr Thr Tyr Ala Asp Val Glu Leu Thr Ala
45 50 55
Arg Arg Val Ala Ser Gly Leu Asn Lys Ile Gly Ile Gin Gin Gly Asp
60 65 70
Val Ile Met Leu Phe Leu Pro Ser Ser Pro Glu Phe Val Leu Ala Phe
75 80 85 90
Leu Gly Ala Ser His Arg Gly Ala Met Ile Thr Ala Ala Asn Pro Phe
95 100 105
Ser Thr Pro Ala Glu Leu Ala Lys His Ala Lys Ser Arg Ala Lys
110 115 120
Leu Leu Ile Thr Gin Ala Cys Tyr Tyr Glu Lys Val Lys Asp Phe Ala
125 130 135
Arg Glu Ser Asp Val Lys Val Met Cys Val Asp Ser Ala Pro Asp Gly
140 145 150
Ala Ser Leu Phe Arg Ala His Thr Gin Ala Asp Glu Asn Val Pro
155 160 165 170
Gln Val Asp Ile Ser Pro Asp Val Val Ala Leu Pro Tyr Ser Ser
175 180 185
Gly Thr Gly Leu Pro Lys Gly Val Met Leu Thr His Lys Gly Leu
190 195 200
Ile Thr Ser Val Ala Gin Glu Val Asp Gly Asp Pro Leu Thr Tyr
205 210 215
Phe His Ser Glu Asp Val Ile Leu Cys Val Leu Pro Met Phe His Ile
220 225 230
Tyr Ala Leu Asn Ser Met Met Met Cys Gly Lou Arg Val Gly Ala Ser
235 240 245 250
Ile Leu Ile Met Pro Lys Phe Glu Ile Gly Ser Leu Leu Gly Leu Ile
255 260 265
Glu Lys Tyr Lys Val Ser Ile Ala Pro Val Val Pro Pro Val Met Met
270 275 280
Ala Ile Ala Lys Ser Pro Asp Leu Asp His Asp Leu Ser Ser Leu
285 290 295
Arg Met Ile Lys Ser Gly Ile Pro Leu Gly Lys Glu Leu Asp
300 305 310
Thr Val Arg Ala Lys Phe Pro Glu Ala Arg Leu Gly Gin Gly Tyr Gly
315 320 325 330
Met Thr Glu Ala Gly Pro Val Leu Ala Met Cys Leu Ala Phe Ala Lys
335 340 345
Glu Pro Phe Asp Ile Lys Pro Gly Ala Cys Gly Thr Val Val Arg Asn
350 355 360
Ala Glu Met Lys Ile Val Asp Pro Glu Thr Gly Val Ser Leu Pro Arg
365 370 375
Asn Gin Pro Glu Gly Ile Arg Gly Asp Gin Ile Met Lys Gly
380 385 390
 Tyr Leu Asn Asp Pro Glu Ala Thr Ser Arg Thr Ile Asp Lys Gly Glu
395 400 405 410
Trp Leu His Thr Gly Asp Ile Gly Tyr Ile Asp Asp Asp Glu Leu
415 420 425
Phe Ile Val Asp Arg Leu Lys Glu Leu Ile Lys Tyr Lys Gly Phe Glu
430 435 440
Val Ala Pro Thr Glu Leu Glu Ala Leu Ile Ala His Pro Glu Ile
445 450 455
Ser Asp Ala Ala Val Val Gly Leu Lys Asp Glu Ala Gly Glu Val
460 465 470
Pro Val Ala Phe Val Val Lys Ser Glu Lys Ser Gin Ala Thr Glu Asp
475 480 485 490
Glu Ile Lys Gin Tyr Ile Ser Lys Gin Val Ile Phe Tyr Lys Arg Ile
495 500 505
Lys Arg Val Phe Phe Ile Glu Ala Ile Pro Lys Ala Pro Ser Gly Lys
510 515 520
Ile Leu Arg Lys Asn Leu Lys Gly Lys Leu Pro Gly Ile
525 530 535
FIG. 6A  Aspen (*P. tremuloides*) PtCAD protein sequence

GenBank accession number: AF217957

```
1  MGSLETERKI VGWAATDSTG HLAPYTSLR DTGPEDVLIK VISCGICHTD
51  IHQIKNDLGM SHYPMVFHGHE VVGEVVEVGS DVTKFKAGDV VGVGVIVGSC
101 KNCHPCKSEL EQYCNIKIWS YNDVYTGDGP TQGFAESMV VDGKVVVP
151 DGMSPEQAAP LLCAGLTVYS PLKHFGKLQS GVRGGILGGL GVGDGVGKIA
201 KAMGGHVTVI SSSDKKREEA MEHLGAEVL VSSDVMQK AADQLDYIID
251 TVPVVHPLEP YLSLLKLDGK LILMGVINTP LFQFVSPMVML GRLSITGSPFI
301 GSMKETEEML EFCKEKGLAS MIEVIKMDYI NTAFERLEKN DVYRFVVDV
351 AGSKLIP*
```
FIG. 6B  Aspen (P. tremuloides) PtCAD cDNA sequence

GenBank accession number: AF217957

1  AAACTCCATC CCTCTCTCTT AGCCTCGTTG TTTCAAGAAA ATGGGTAGCC
31  TTGAAACAGA GAGAAAAATT GTAGGATGGG CAGCAACAGA CTCAACTGGG
101  CATCTCGCTC CTTCACCTA TAGTCACAGA CATACGGGGC TCTTATCAAG
151  GTTTATCAAG GTTATCAGCT GTGGAATTGG CCAATTCCGAT ATCCACCAA
201  TCAGAAATGA TCTTGGCATG TCACACTATC CTATGGTCCC TGCCCATGAA
251  GTTGTTGGTG AGTTGTTGGT GGTGGATCA GATGTCGCAA GATTTCAAACG
301  TGGGATGTTT GGTGGTTTGG GAGTCATCGG TGGAAGCTGC ATCCATGCAA
351  ATCAGAGCTT GAGCAATACT GCAACAAGAA TACCTGGTCT
401  TACAATGTAG TCTACACTGA TGGCAAAACC ACCCAAGGAG TGGTTCACCC
451  ATCCATGGTT GTCGATCAAA AGTTTGTGGT GAGAATTCCT CACCAGAACA
501  AGCAGCGCCG CTGTTGTGCG CTGGATTGAC CCACTCAAAC ACTTTGGACT
551  GAAACAGAGT GGGCTAAGAG AGGAGATGCAA ATCAGAGCTT GAGCAATACT
601  TACAATGTAG TCTACACTGA TGGCAAAACC ACCCAAGGAG TGGTTCACCC
651  ATCCATGGTT GTCGATCAAA AGTTTGTGGT GAGAATTCCT CACCAGAACA
701  AGCAGCGCCG CTGTTGTGCG CTGGATTGAC CCACTCAAAC ACTTTGGACT
751  GAAACAGAGT GGGCTAAGAG AGGAGATGCAA ATCAGAGCTT GAGCAATACT
801  TACAATGTAG TCTACACTGA TGGCAAAACC ACCCAAGGAG TGGTTCACCC
851  ATCCATGGTT GTCGATCAAA AGTTTGTGGT GAGAATTCCT CACCAGAACA
901  AGCAGCGCCG CTGTTGTGCG CTGGATTGAC CCACTCAAAC ACTTTGGACT
951  GAAACAGAGT GGGCTAAGAG AGGAGATGCAA ATCAGAGCTT GAGCAATACT
1001 ATGAAACACTC TTGGTCTGTA TGAATACTCTG TGTAGCTGGG ATGGTGAAG
1051 CATGCAAAAC GCTGCTGACCC ATACCGAAAT GATGTGAGAT ATAGATTCGT
1101 TGGATTTCAG CTCAGACTATC TATCGGAGATG TGGATGCAAG TGGTTCACCC
1151 GAAACAGAGT GGGCTAAGAG AGGAGATGCAA ATCAGAGCTT GAGCAATACT
1201 TGGATTTCAG CTCAGACTATC TATCGGAGATG TGGATGCAAG TGGTTCACCC
1251 ATGAAACACTC TTGGTCTGTA TGAATACTCTG TGTAGCTGGG ATGGTGAAG
1301 CATGCAAAAC GCTGCTGACCC ATACCGAAAT GATGTGAGAT ATAGATTCGT
1351 TGGATTTCAG CTCAGACTATC TATCGGAGATG TGGATGCAAG TGGTTCACCC

ATGGGTAGCC CTCAACTGGG CAGAAGATGT ATCCACCAA
Fig. 7. pBK*pt*4CL* Pt4CL1-a construct

Fig. 8. pBK*pt*4CL* PtCAld5H-s construct
FIG. 9-1 The alignment of plant AldOMT protein sequences

<table>
<thead>
<tr>
<th>1</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>~~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS</td>
</tr>
<tr>
<td>2</td>
<td>~~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS</td>
</tr>
<tr>
<td>3</td>
<td>~~~~~~~~~~ MG STG..ETQMT PTQVSDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>4</td>
<td>~~~~~~~~~~ MG STG..ETQMT PTQVSDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>5</td>
<td>~~~~~~~~~~ MG STG..ETQIT PTHISDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>6</td>
<td>~~~~~~~~~~ MG STG..ETQMT PTQVSDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>7</td>
<td>~~~~~~~~~~ MG STG..ETQMT PTQVSDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>8</td>
<td>~~~~~~~~~~ MG STG..ETQMT PTQVSDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>9</td>
<td>~~~~~~~~~~ MG ST..ETQMT PTQVSDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>10</td>
<td>~~~~~~~~~~ MG ST..ETQMT PTQVSDEEAAN LFAMQLASAS</td>
</tr>
<tr>
<td>11</td>
<td>MESTLAFNSG SNSMNSQFSS SAEFNSPVPE TIPKSEETTF VFAHNTSSA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>51</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VLPMTLKTAA ELDLIEIMA A...GPGAFL STSEIASHLTP TNKDAPVVM</td>
</tr>
<tr>
<td>2</td>
<td>VLPMTLKTAA ELDLIEIMA A...GPGAFL STSEIASHLTP TNKDAPVVM</td>
</tr>
<tr>
<td>3</td>
<td>VLPVLKAAI ELDLIEIMA A...GPGVFL SPTDIASLQP TNKDAPVVM</td>
</tr>
<tr>
<td>4</td>
<td>VLPVLKAAI ELDLIEIMA A...GPGSFL SPSDLASLQP TKNPEAPVVM</td>
</tr>
<tr>
<td>5</td>
<td>VLPVLKSAI ELDLIEIKA A...GPGAQI SPSIEIASLQP TTNPDAPVVM</td>
</tr>
<tr>
<td>6</td>
<td>VLPVLKAAI ELDLIEIMA A...GPGAFL SPSVEAQLQP TQNPKEAPVVM</td>
</tr>
<tr>
<td>7</td>
<td>VLPVLKSAI ELDLIEIKA SPSHEAQLQP SPSVEAQLQP TQNPKEAPVVM</td>
</tr>
<tr>
<td>8</td>
<td>VLPVLKSAI ELDVLEIMA A...GPAHI STSDIASKLTP TKNPDAAVML</td>
</tr>
<tr>
<td>9</td>
<td>VLPMLKSAI ELDLIEIMA A...NGPSM STTEIASLTP TKNPKEAPVVM</td>
</tr>
<tr>
<td>10</td>
<td>VLPVLKSAV ELDLIEIMA A...GPAAI SPSELAQLQS TQNPKEAPVVM</td>
</tr>
<tr>
<td>11</td>
<td>VLPMLKSAI ELDLIEIKA A...GPGAFL STSEIAAKIT KRNPKAPVVM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>101</th>
<th>150</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DRLRLLSASILTCSLKDLDPGKVERLYGLAPVCFLFTK NEDGVSVSLPL</td>
</tr>
<tr>
<td>2</td>
<td>DRLRLLSASILTCSLKDHDPGKVERLYGLAPVCFLFTK NEDGVSVSLPL</td>
</tr>
<tr>
<td>3</td>
<td>DRLRLLSASILTCSLRTLADGKVERLYLGPPVCKFLFTK NEEGVSIAAPL</td>
</tr>
<tr>
<td>4</td>
<td>DRLRLLSASILTCSLRTLADGKVERLYLGPPVCKFLFTK NEDGVSIAAL</td>
</tr>
<tr>
<td>5</td>
<td>DRLRLLSASYILTCPVRDOQDGKVQRLYGLATVAKYLVK NEDGVSIAL</td>
</tr>
<tr>
<td>6</td>
<td>DRIFRLLSASILTCTRLNLPGKVERLYLAPVCFLFTK NEDGVSIAAAL</td>
</tr>
<tr>
<td>7</td>
<td>DRIFRLLSASILTCSNRLPSGDFVERYLGPPVCKYLT NEDGVSIAAAL</td>
</tr>
<tr>
<td>8</td>
<td>DRIFRLLSASILTCSNRLPSGDFVERYLGPPVCKYLT NADGVSVAPL</td>
</tr>
<tr>
<td>9</td>
<td>DRIFRLLTSLDPPSKRLSADGKVERLYLAPVCFLFTK NEDGVSIAAAL</td>
</tr>
<tr>
<td>10</td>
<td>DRIFRLLTSLDPPSSRLSADGKVERLYLAPVCFLFTK NADGVSVAPL</td>
</tr>
<tr>
<td>11</td>
<td>DRIFRLLTSLDPPSKRLSADGKVERLYLAPVCFLFTK NEDGVSIAAAL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>151</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CLMNQDKVML ES.WYVLKDA ILDDGIPFNNK AYGMTAFEYH GTPRFKVF</td>
</tr>
<tr>
<td>2</td>
<td>CLMNQDKVML ES.WYVLKDA ILDDGIPFNNK AYGMTAFEYH GTPRFKVF</td>
</tr>
<tr>
<td>3</td>
<td>CLMNQDKVLL ES.WYHLKDA VLDDGIPFNNK AYGMTAFEYH GTPRFKVF</td>
</tr>
<tr>
<td>4</td>
<td>CLMNQDKVLL ES.WYHLKDA VLDDGIPFNNK AYGMTAFEYH GTPRFKVF</td>
</tr>
<tr>
<td>5</td>
<td>NLMNQDKVLM ES.WYHLKDA VLDDGIPFNNK AYGMTAFEYH GTPRFKVF</td>
</tr>
<tr>
<td>6</td>
<td>NLMNQDKKLM ES.WYVLKDA VLDDGIPFNNK AYGMTAFEYH GTPRFKVF</td>
</tr>
</tbody>
</table>
### FIG. 9-2

| 7 | CLMNQDKVLM ES.WYYLKDA ILDGGIPFNLK AYGMTAEFYH GTDPRFNKVF |
| 8 | SLMNQDKVLM ES.WYHLTEA VLEGGIPFNLK AYGMTAEFYH GTDPRFNTVF |
| 9 | CLMNQDKVLM ES.WYHLKDA ILDGGIPFNLK AYGMSAFEYH GTDPRFNKVF |
| 10 | LLMNQDKVLM ES.WYHLKDA VLDGGIPFNLK AYGMTAEFYH GTDPRFNKVF |
| 11 | LLMNQDKVET QSKRYHLKDA VLDGGIPFNLK AYGMTDFEYH GTEPRFNKVF |

| 201 | |
| 1 | NKGMSDHSTI TMKKILETYK GFEGLTSLVD VGGGTGAVVN TIVSKYFSIK |
| 2 | NKGMSDHSTI TMKKILETYK GFEGLTSLVD VGGGTGAVVN TIVSKYFSIK |
| 3 | NRGMSNHSTI TMKKILETYK GFEGLTSVVD VGGGTGAVLN MIVSKYFSIK |
| 4 | NKGMSDHSTI TMKKILETYT GFEGLKSIVD VGGGTGAVIN TIVSKYFTIK |
| 5 | NRGMSDHSTI TMKKILETYK GFEGLETVVD VGGGTGAVLS MIVAKYFSIK |
| 6 | NRGMSDHSTI TMKKIFEMYT GFEALNTIVD VGGGTGAVLS MIVAKYFSIK |
| 7 | NNGMSNHSTI TMKKILETYK GFEGLSVVD VGGGTGAVLN MIVAKYFSIK |
| 8 | NNGMSNHSTI TMKKILETYK GFEGLSVVD VGGGTGAVLN MIVAKYFSIK |
| 9 | NRGMSDHSTI TMKKILETYK GFEGLSVVD VGGGTGAVLN MIVAKYFSIK |
| 10 | NRGMSDHSTI TMKKILETYK GFEGLSVVD VGGGTGAVLN MIVAKYFSIK |
| 11 | NNGMSNHSTI TMKKILEAYK GFEGLTSIVD VGGGTGAVLN MIIKSYFTIK |

**Motif I**

| 251 | |
| 1 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 2 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 3 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 4 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 5 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 6 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 7 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 8 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 9 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 10 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |
| 11 | GINFDLPHVI EDAPSGEVE HVGGDMFVSV PKADAVFMKW ICHDWSDAHC |

**Motif II**

| 301 | |
| 1 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 2 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 3 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 4 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 5 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 6 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 7 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 8 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 9 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |

**Motif III**

| 350 | |
| 1 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 2 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 3 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 4 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 5 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 6 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 7 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 8 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
| 9 | LKFLKNCYDA LEPNGKIVLV ECILPVAPDT SLATKGVVHV DVIMLAHNPG |
FIG. 9-3

10 LKFLNKCYEA LPANGKVIIA ECILPEAPDT SLATKNIVHV DIVMLAHNPG
11 LKFLENCYQA LPDNGKVIVA ECILPPVPDT SLATKSAVHI DIVILAYNTG

motif III

351
1 GKERTEKEFE GLAKGAGFQG FEVMCCAFNT HVIEFRKKA
2 GKERTEKEFE GLAKGAGFQG FEVMCCAFNT HVIELRKN-
3 GKERTEQEFE ALAKGAGFQG FNVACSAFNT YVIEFLKKN
4 GKERTEQEFE ALAKGAGFQG IRVCCDAFNT YVIEFLKKI
5 GKERTEQEFE DLAKGAGFQG FKVCNAFNT YIMEFLKKV
6 GKERTEKEFE TLAKGAGFQG FQVMCCAFGT YVMEFLKTA
7 GKERTEKEFE ALAIAGFGKG FKVACCAFNT YVIMEFLKTA
8 GKERTEKEFE ALAKGAGFGG FRVALCAYNT WIIIEFLKKI
9 GKERTEKEFE ALAKGAGFGG IKVVCDAFGV NLIEELLKL
10 GKERTEKEFE ALAKGAGFGG FARLVALTTL GSWNSTSN-
11 GKERTEKEFE ALAKGAGFGG FKVVCCAFNS WIMEFCKTA

Plant AldOMTs from

1) Aspen, X62096
2) Poplar, M73431
3) Almond, X83217
4) Strawberry, AF220491
5) Alfalfa, M63853
6) Eucalyptus, X74814
7) Clarkia breweri, AF006009
8) Sweetgum, AF139533
9) Arabidopsis, U70424
10) Tobacco, X74452
11) Vitis vinifera, AF239740
FIG. 10-1 The alignment of full length plant CAD protein sequences available in the GenBank database

Zn2

1
MGSLE.TEKT VTGYAARDSS GHLSPYTYNL RKKGPEDVIV KVIYCGICH
2 MGSLE.SEKT VTGYAARDSS GHLSPYTYNL RKKGPEDVIV KVIYCGICH
3 MGSLE.SEKT VTGYAARDSS GHLSPYTYNL RKKGPEDVIV KVIYCGICH
4 MGSLE.SERT VTGYAARDSS GHLSPYTYNL RKKGPEDVIV KVIYCGICH
5 MGLS.ASERK VVGWAARDAT GHLSPYTYTL RNTGPEDVYI KIHYCGVCH
6 MGLS.ASERK VVGWAARDAT GHLAPYTYTL RNTGPEDVYI KIHYCGVCH
7 MGLS.ASERK VVGWAARDAT GHLAPYTYTL RNTGPEDVYI KIHYCGVCH
8 MGSL.EAERT TVGLAAKDPES GILTPYTYTL RNTGPEDVYI KIHYCGVCH
9 MGSL.EAERT TVGLAAKDPES GILTPYTYTL RNTGPEDVYI KIHYCGVCH
10 MGSL.EK.ERT TTGWAARDPS GVLSPYTYSL RNTGPEDVYI KVLSCGICH
11 MGSL.EK.ERT TTGWAARDPS GVLSPYTYSL RNTGPEDVYI KVLSCGICH
12 MGGEV.EKT TIGWAARDPS GVLSPYTYSL RNTGPEDVYI KVLSCGICH
13 MGSLDV.EKS AIGWAARDPS GLLSPYTYTL RNTGPEDVYV KVLYCGVCH
14 MGSLDV.EKS AIGWAARDPS GLLSPYTYSL RNTGPEDVYV KVLYCGVCH
15 MGSLDV.EKS AIGWAARDPS GLLSPYTYSL RNTGPEDVYV KVLYCGVCH
16 MGSLDV.EKS AIGWAARDPS GLLSPYTYSL RNTGPEDVYV KVLYCGVCH

51
1 DLVQMRNEMG MSYPMVPGHE EVGIVTEIG SEVKKFKVGE HVGVCIVGS
2 DLVQMRNEMG MSYPMVPGHE EVGIVTEIG SEVKKFKVGE HVGVCIVGS
3 DLVQMRNEMG MSYPMVPGHE EVGIVTEIG SEVKKFKVGE HVGVCIVGS
4 DLVQMRNEMG MSYPMVPGHE EVGIVTEIG SEVKKFKVGE HVGVCIVGS
5 DHQAKNHNLG ASKYPMPVHG EVVGEVTEIG PEVAKYGVGD VVGVGVIGVGC
6 DHQAKNHNLG ASKYPMPVHG EVVGEVTEIG PEVAKYGVGD VVGVGVIGVGC
7 DHQAKNHNLG ASKYPMPVHG EVVGEVTEIG PEVAKYGVGD VVGVGVIGVGC
8 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
9 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
10 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
11 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
12 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
13 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
14 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
15 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC
16 DLHQIKNDLG MSYPMPVHG EVGGEVTEGL SNVTFRKVMG IVGVGVIGVGC

101
1 CRSCGNCNSQ MEQYCSKRIW TYNDVHNHDT PTQGGFASTM VVDQMFVVRRI
2 CRSCGNCNSQ MEQYCSKRIW TYNDVHNHDT PTQGGFASTM VVDQMFVVRRI
3 CRSCGNCNSQ MEQYCSKRIW TYNDVHNHDT PTQGGFASTM VVDQMFVVRRI
4 CRSCGNCNSQ MEQYCSKRIW TYNDVHNHDT PTQGGFASTM VVDQMFVVRRI
5 CRECSFCAN VEYCYNKIKI SYNDVYTGDR PTQGGFASTM VVDQKFVVKI
6 CRECSFCAN VEYCYNKIKI SYNDVYTGDR PTQGGFASTM VVDQKFVVKI
7 CRECSFCAN VEYCYNKIKI SYNDVYTGDR PTQGGFASTM VVDQKFVVKI
8 CKSCRACDSE IEYQNKKIKI SYNDVYTDGK ITQGGFAEST VVEQKFVVKI
9 CKSCRACDSE IEYQNKKIKI SYNDVYTDGK ITQGGFAEST VVEQKFVVKI

Zn2
FIG. 10-2

10 CRSCSPCNSD QEYCNKKIW NYNDVYTDGK PTQGGFAGEI VVGERFVVKI
11 CRSCSPCNSD QEYCNKKIW NYNDVYTDGK PTQGGFAGEI VVGERFVVKI
12 CRNCGPKRDI EIQCNKKIW NCNVDYTDGK PTQGGFAESM VVQKFVVRI
13 CRNCGPKRE IEYCNKKIW NCNVDYTDGK PTQGGFAESM VVQKFVVRI
14 CKNCHPCKSE LEQCNKKIW SYNDVYTDGK PTQGGFAESM VVQKFVVRI
15 CKNCHPCKRE IEYCNKKIW SYNDVYTDGK PTQGGFAESM VVQKFVVRI
16 CKTCRPCKAD VEYCNKKIW SYNDVYTDGK PTQGGFAGSM VVQKFVVRI

Zn2

151 PENLPLEQAA PLLCAGVTVF SPMKHFMTE .PGKKGILG LGGVGHMVK
2 PENLPLEQAA PLLCAGVTVF SPMKHFMTE .PGKKGILG LGGVGHMVK
3 PENLPLEQAA PLLCAGVTVF SPMKHFMTE .PGKKGILG LGGVGHMVK
4 PENLPLEQAA PLLCAGVTVF SPMKHFMTE .PGKKGILG LGGVGHMVK
5 PAGLAPWEQA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
6 PAGLAPWEQA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
7 PAGLAPWEQA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
8 PAGLAPWEQA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
9 PAGLAPWEQA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
10 PDGLEXEQAA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
11 PDGLEXEQAA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
12 PDGLEXEQAA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
13 PDGLEXEQAA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
14 PDGLEXEQAA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
15 PDGLEXEQAA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK
16 PDGLEXEQAA PLLCAGTVYY SPLLHFGL.T TPGLRGGILG LGGVGHMVK

NADP

201 IAKAFGLHVT VISSSDKKKK EAMEVLGADA YLVSDFTEKM MEAASESLDYI
2 IAKAFGLHVT VISSSDKKKK EAMEVLGADA YLVSDFTEKM MEAASESLDYI
3 IAKAFGLHVT VISSSDKKKK EAMEVLGADA YLVSDFTEKM MEAASESLDYI
4 IAKAFGLHVT VISSSDKKKK EAMEVLGADA YLVSDFTEKM MEAASESLDYI
5 VAKAMGHVVI VISSSDKKKK EMADLHGLA YLVDSSADAM GAPAASLDYI
6 VAKAMGHVVI VISSSDKKKK EMADLHGLA YLVDSSADAM GAPAASLDYI
7 VAKAMGHVVI VISSSDKKKK EMADLHGLA YLVDSSADAM GAPAASLDYI
8 VAKALGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
9 VAKALGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
10 VAKAMGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
11 VAKAMGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
12 VAKAMGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
13 VAKAMGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
14 VAKAMGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
15 VAKAMGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
16 VAKAMGHVVI VISSSDKKKK EALSEDGLADN YLVDSSDTGVM QEAASSLDYI
FIG. 10-3

251  300
1  MDTIPVHPL EPYLALLKTN GKLVMLGVPV EPLHVFTPLL ILGRSSIAGS
2  MDTIPVHPL EPYLALLKTN GKLVMLGVPV EPLHVFTPLL ILGRSSIAGS
3  MDTIPVHPL EPYLALLKTN GKLVMLGVPV EPLHVFTPLL ILGRSSIAGS
4  MDTIPVHPL EPYLALLKTN GKLVMLGVPV EPLHVFTPLL ILGRSSIAGS
5  IDTVPVHPL EPYLALLKLD GKLNGTHVIGE TLPSFVSPPMV MLGRKAITGS
6  IDTVPVHPL EPYLALLKLD GKLNGTHVIGE TLPSFVSPPMV MLGRKAITGS
7  IDTVPVHPL EPYLALLKLD GKHLLGTIGE TLPSFVSPPMV MLGRKAITGS
8  IDTVPVHPL EPYLALLKID GKLILGTVIN TPLQFVTPMV MLGRSKITGS
9  IDTVPVHPL EPYLALLKID GKLILGTVIN TPLQFVTPMV MLGRSKITGS
10 FDTIPVHPL EPYLALLKLD GKLIGTVIN TPLQFISPMV MLGRSKITGS
11 FDTIPVHPL EPYLALLKLD GKLILTVIN TPLQFISPMV MLGRSKITGS
12 FDTIPVHPL EPYLALLKLD GKLILTVIN TPLQFISPMV MLGRSKITGS
13 FDTIPVHPL EPYLALLKLD GKLILTVIN TPLQFISPMV MLGRSKITGS
14 FDTIPVHPL EPYLALLKLD GKLILTVIN TPLQFISPMV MLGRSKITGS
15 FDTIPVHPL EPYLALLKLD GKLILTVIN TPLQFISPMV MLGRSKITGS
16 FDTIPVHPL EPYLALLKLD GKLILTVIN TPLQFISPMV MLGRSKITGS

301  350
1  FIGSMEEETQE TLDFCAEKKV SSMIEVVDLD YINTAMERLE KNDVRYRFV
2  FIGSMEEETQE TLDFCAEKKV SSMIEVVDLD YINTAMKRLLE KNDVRYRFV
3  FIGSMEEETQE TLDFCAEKKV SSMIEVVDLD YINTAMERLE KNDVRYRFV
4  FIGSMEEETQE TLDFCAEKKV SSMIEVVDLD YINTAMERLE KNDVRYRFV
5  FIGSIDATEAVLQFCDVDKLQ TGSLIEVVKLMG YVNEALERLE KNDVRYRFV
6  FIGSIDATEAVLQFCDVDKLQ TGSLIEVVKLMG YVNEALERLE KNDVRYRFV
7  FIGSIDATEAVLQFCDVDKLQ TGSLIEVVKLMG YVNEALDRLE KNDVRYRFV
8  FVGTVEKETEMLFKEKEKGL SSMIEVTMRD YINKFAERLE KNDVRYRFV
9  FVGTVEKETEMLFKEKEKGL SSMIEVTMRD YINKFAERLE KNDVRYRFV
10 FIGSMKEETEMLEFCVEKGL TGQIEVIMPD YVNTAERLE KNDVRYRFV
11 FIGSMKEETEMLEFCVEKGL TGQIEVIMPD YVNTAERLE KNDVRYRFV
12 FIGSMKEETEMLEFCVEKGL TGQIEVIMPD YVNTAERLE KNDVRYRFV
13 FIGSMKEETEMLEFCVEKGL TGQIEVIMPD YVNTAERLE KNDVRYRFV
14 FIGSMKEETEMLEFCVEKGL ASMEVIMKD YINTAERLE KNDVRYRFV
15 FIGSMKEETEMLEFCVEKGL ASMEVIMKD YINTAERLE KNDVRYRFV
16 FIGSMKEETEMLEFCVEKGL ASMEVIMKD YINTAERLE KNDVRYRFV

351  370
1 DVGSKLDN* ~~~~~~~~~~
2 DVAASKLDN* ~~~~~~~~~~
3 DVGSSLDN* ~~~~~~~~~~
4 DVAASNLDK* ~~~~~~~~~~
5 DVAGSNVAAE AAAAAAASN*
6 DVAGSNVAAE AAAAAAASN*
7 DVAGSNVAAE AAAAAAASN*
8 DVKSKFEY* ~~~~~~~~~~
9 DVKSKFEY* ~~~~~~~~~~
10 DVVSKLD* ~~~~~~~~~~
FIG. 10-4

11 DVVGSKLD* ~~~~~~~~~~
12 DVIGSKLDQ* ~~~~~~~~~~
13 DVAGSKLDQ* ~~~~~~~~~~
14 DVAGSKLIP* ~~~~~~~~~~
15 DVAGSKLIH* ~~~~~~~~~~
16 DVAGSKLDQE T~~~~~~~~~

Full length plant CADs from

1) Radiata pine, U62394                                9) Lucerne, Z19573
2) Loblolly pine, Z37992                               10) Eucalyptus, AF038561
3) Loblolly pine, Z37991                               11) Eucalyptus, X65631
4) Norway spruce, X72675                                12) Tobacco, X62343
5) Maize, aj005702                                     13) Tobacco, X62344
6) Maize, Y13733                                       14) Aspen, AF217957
7) Sugarcane, AJ231135                                  15) Cottonwood, Z19568
8) Lucerne, AF083332                                    16) Udo, D13991
FIG. 11-1  The alignment of full length plant CAld5H protein sequences

1) Aspen; 2) Poplar, AJ010324; 3) Sweetgum, AF139532; 4) Arabidopsis, U38416.* Heme-binding signature

1
MDSLVQSLQ AS...PMFL IVSSLFFFG LLSRLRRRLP YPPGKGLPL
2
~MDSLQSLQ TL...PMFL IISSIFFLG LISLRRRSSP YPPGKGFPL
3
MDSLHEALQ P..PMTEP I...PLLPLLGV VSLRQQLP YPPGKGLPL
4
MESSISQTLS KLSDPTTLSLV IVVSLFIFIS FTRLRRPP YPPGPRGWI

50
1
51
VGSMHMDQTI THRGLAKLAK QYGGLFHMRM GYLHMVTSS PEIARQVQLQ
2
IGSMHMDQTI THRGLAKLAK QYGGLFHMRM GYLHMASGSS PEVARQVQLQ
3
IGNLHMDQTI THRGLAKLAK QYGGLFHLMK GFLHMVAST PDARQVQLQ
4
IGNLHMDQTI THRGLAKLAK QYGGLCHLRM GFLHMAYSS PEVARQVQLQ

100
1
QDNIFSNRPA NIAISYLTMD RADMAFAHYG FEWRQMRKLC VMKFGSARKRA
2
QDNMFNSRPA NIAISYLTMD RADMAFAHYG FEWRQMRKLC VMKFGSARKRA
3
QDNIFSNRPA TIAISYLTMD RADMAFAHYG FEWRQMRKLC VMKFGSARKRA
4
QDSVFSNRPATIAISYLTMD RADMAFAHYG FEWRQMRKVC VMKFGSARKRA

150
1
ESWESVRDEV DSKKMTVEAN IGKPVNLGEL IFTLFMNITY RAAFGA.KNE
2
ESWESVRDEV DSKYKTVESN IGKPVVNLGEL IFTLFMNITY RAAFGA.KNE
3
ESWESVRDEV DSAVRVVAGN IGSTNNIHEL VFAALTKNITY RAAZGATSHC
4
ESWASVRDEV DKMVRSVSCN VGKPINVEGQ IFALTERNITY RAAZGASEK

200
1
GQDEFIKILQ EFSKLFQAFN MSDFIPWLGW IDPQLSRLA VKARKALKRF
2
GQDEFIKILQ EFSKLFQAFN MSDFIPWLGW IDPQLSRLA VKARKALKRF
3
GQDEFVAILQ EFSQLFQAFN IADFIPWKLW V.PQGINVRL NKARGALDFG
4
GQDEFIRILQ EFSKLFQAFN VAFDIPYGFW IDPQGINKRL VKARNDLDGF

250
1
IDSIIDDIHQ KRKQNKFSED ...AETDMVD DMLAFYGEA RVDVSDSDLQ
2
IDSIIDDIHQ KRKQNNYSEE ...AETDMVD DMLTFYSEET .KVQNSDDLQ
3
IDSIIDDIHQ KGSKN...SEE ...VDMDV DLMFYGEA .KVESESDLQ
4
IDSIIDDIHQ KKQNFQAVDD GDVVDTDMVD DLLAFYGEA KLVSETSDLQ

300
1
KAIKTLDNIA KAIIMDVMFG GTETVSAIE WVMAELMKSP EDQKRVQVEL
2
NAIKLTDNIA KAIIMDVMFG GTETVSAIE WAMAEPLKSP EDIKRVQVEL
3
NSIKLTDNIA KAIIMDVMFG GTETVSAIE WAMELMKSP EDIKRVQVEL
4
NSIKLTDNIA KAIIMDVMFG GTETVSAIE WATELLRSP EDIKRVQVEL
FIG. 11-2

351
1. AEVVGLERRV EESIDKLTFL KCALKETLRL MHPPIPILLH ETSEDFAVAG
2. ADVVGLERRV EESDFKLTFL KFCTKETLRL LHPPIPILLH ETSEDFAVAG
3. AVVVGLDRRV EEKDFEKLTY LKCVKKEVLR LHPPIPILLH ETAEDAEVG
4. AEVVGLDRRV EESIEKLTLY LKCTKETLRL MHPPIPILLH ETAEDTSIDG

400
1. YFIPKQTRVM INAYAIGRDK NSWEDPDAFK PSRFKPGVP DFKGNHFEFI
2. YYVPPKTRVM INAYAIGRDK NSWEDPDSFK PSRFLEPGVP DFKGNHFEFI
3. YYIPAKSRVM INACAIGRDK NSWADPDFTFR PSRFKGDGVP DFKGNNFEFI
4. FFIPKKSVM INAFAIHRDP TSWTDPDTFR PSRFLEPGVP DFKGSNFIEF

450
1. PGSGRRSCP GMQLGLYLYLD LAVAHLHCFTWELPGMKP SELDMNTDFM
2. PGSGRRSCP GMQLGLYLYLD LAVAHLHCFTWELPGMKP SELDMNTDFM
3. PGSGRRSCP GMQLGLYALE TTVAHLHCFTWELPGMKP SELEMNDVFG
4. PGSGRRSCP GMQLGLYALD LAVAHLHCFTWKLPGMKP SELDMNDVFG

500
1. LTAPRATRLV AVPSRVLCP L*
2. LTAPRATRLV AVPRKRVVCPL~~
3. LTAPRAIRLT AVPSRRLCPLY*
4. LTAPKATRLF AVPTTRLILA~~
FIG. 12-1  PLANT 4CL AMINO ACID SEQUENCE ALIGNMENTS

(1) 1:-------------------------------MPQ-EFFFRSLPDIYIPKNLFLHSYVLENLKSXHPCLI 41
(2) 1:-------------------------------MDA-IMNSQEFEFRSLPDIYIPKNLFLHSYVLENLKSXHPCLI 46
(3) 1:-------------------------------MGDCVAKKELFRSLPDIYIPKNLFLHSYVLENLKSXHPCLI 46
(4) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(5) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(6) 1:-------------------------------MEKDTKSDIDFRSLPDIYIPKNLFLHSYVLENLKSXHPCLI 48
(7) 1:-------------------------------MGDCVAPKEDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(8) 1:-------------------------------MGDCVAPKEDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(9) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(10) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(11) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(12) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(13) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(14) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(15) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48
(16) 1:-------------------------------MPMUTETKQSGDLIFRFSDLPIYIPKNLFLHSYVLENLKSXHPCLI 48

(1) 42:NG-ANGDYTVYADVETDIARRVAGSGLNIGIQGQVDVIMLFLPSSPEFLVAFGLASHGAM 99
(2) 47:NG-ANGDYTVYADVETDIARRVAGSGLNIGIQGQVDVIMLFLPSSPEFLVAFGLASHGAM 104
(3) 47:NG-ANGDYTVYADVETDIARRVAGSGLNIGIQGQVDVIMLFLPSSPEFLVAFGLASHGAM 104
(4) 49:DG-ANDRIYYAELTVSRKVA-VGGLKIGQIQKDTIMLLPNCPEFVFAVIGASAGIL 99
(5) 46:NG-TG0DIHTYAKKRLTARRVAGSGLNLGKIEKGDVMILLLNPNTSEFVAFGLASHGAM 104
(6) 49:DG-ANDRIYYAELTVSRKVA-VGGLKIGQIQKDTIMLLPNCPEFVFAVIGASAGIL 104
(7) 49:DG-ANDRIYYAELTVSRKVA-VGGLKIGQIQKDTIMLLPNCPEFVFAVIGASAGIL 104
(8) 58:DGATGVLYLTVADVRLSRLRAALRAPPGLRRGVMVMSLRRNSPEFLVSGFAARVLAGA 117
(9) 61:VG-PASKTFTYADTDHLLSKIAAGSLNLIGKLIGVMILLQNSAFVDFSLIAISMIGA 118
(10) 59:NGPTGVYTVYDSDVHISQIAANFHK-VGNQNDVVMILPNCPEFVSLAASFRGAT 116
(11) 45:DG-ATDRTYCFESEELERIVRVA-AGLAKLGLQQGQVVMLLPNCIEFVAFVGMASVRAI 102
(12) 46:DG-ATDRTYCFESEELERIVRVA-AGLAKLGLQQGQVVMLLPNCIEFVAFVGMASVRAI 103
(13) 1:-------------------------------A-------------------------------A- 3
(14) 1:-------------------------------A-------------------------------A- 3
(15) 1:-------------------------------A-------------------------------A- 3
(16) 1:-------------------------------A-------------------------------A- 3

(1) 100:ITAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 155
(2) 105:VTAANPFSTPAELAKHAKFPRKLLKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 160
(3) 105:VTAANPFSTPAELAKHAKFPRKLLKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 160
(4) 107:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(5) 107:VTAANPFSTPAELAKHAKFPRKLLKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(6) 107:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(7) 107:VTAANPFSTPAELAKHAKFPRKLLKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(8) 118:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(9) 119:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(10) 119:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(11) 119:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(12) 120:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(13) 120:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(14) 120:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(15) 120:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162
(16) 120:VTAANPFSTPAELAKHAKASRAKLLITQACYYEKVK--DFAERSDVKMCDMS-AED-PA-1A 162

(1) 156:SLFRAHTQADENEVPQV-------------DISPDVVALFYSSGTTGLPKGVMTHKLGLTTSVA 207
(2) 161:LSHSLRTQADENEVPQV-------------DFSPDVVALFYSSGTTGLPKGVMTHKLGLTTSVA 212
(3) 161:LSHSLRTQADENEVPQV-------------DFSPDVVALFYSSGTTGLPKGVMTHKLGLTTSVA 212
(4) 163:VHSFELIQSDHEDIPD-------------KQIPDVVALFYSSGTTGLPKGVMTHKLGLTTSVA 214
(5) 165:LSHSLRTQADENEVPQV-------------DISPDVVALFYSSGTTGLPKGVMTHKLGLTTSVA 212
FIG. 12-3

(10) 353: GYGMTEAGPVLAMSLGFKEFPFVKSGACGTVNRNAEMKIIVDPDTGDSLSRNPGEICIR

(11) 331: GYGMTEAGPVLAMLAFKNNPFPVKSGCSGTGVRTYNAQIKLDTGTESEGELHPFVAGIEICIR

(12) 332: GYGMTEAGPVLAMNFANPNFPVKSGCSGTGVRNAQIKLDTGTESEGELHPFVAGIEICIR

(13) 213: GYGMTEAGPVL----------------------------------------------- 223

(14) 221: GYGMTEAGPVL----------------------------------------------- 230

(15) 289: GYGMTEAGPVLAMLAFKKEFVKSGS----------------------------- 317

(16) 291: GYGMTEAGPVLAMLAFKKEFVKSGS------------------------------- 318

******* *

(1) 388: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPTE

(2) 393: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(3) 393: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(4) 395: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(5) 393: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(6) 395: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(7) 391: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(8) 412: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(9) 410: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(10) 413: GQQIMKGLYNPEATSRTIDKEGWLHTGDIYIDDDDELFIHDLKELIKYKGFQVAPAE

(11) 391: GPEIMKGIYNDPEATSTIDEGEWLHTGDEYIDDDDELFIHDLKELIKYKGFQVAPAE

(12) 392: GPEIMKGIYNDPEATSTIDEGEWLHTGDEYIDDDDELFIHDLKELIKYKGFQVAPAE

(13) 224: -----------------------------------------------

(14) 231: -----------------------------------------------

(15) 318: -----------------------------------------------

(16) 319: -----------------------------------------------

(1) 448: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(2) 453: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(3) 453: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(4) 451: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(5) 453: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(6) 451: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(7) 452: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(8) 472: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(9) 470: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(10) 473: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(11) 451: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(12) 452: LEALLIAHPESDAAVVLKDFDEAGFVPAFVKESEKATEDEIKYSKQVYFQKIR

(13) 224: -----------------------------------------------

(14) 231: -----------------------------------------------

(15) 318: -----------------------------------------------

(16) 319: -----------------------------------------------

(1) 508: RVFFIEAPKAPSGKILRKNLKEEI-PG----

(2) 513: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(3) 513: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(4) 515: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(5) 513: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(6) 515: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(7) 512: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(8) 529: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(9) 530: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(10) 533: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(11) 509: RVFFIEAPKAPSGKILRKNLRETL-PGI----

(12) 510: RVFFIEAPKAPSGKILRKNLRETL-PGI----

* * * *
FIG. 12-4

(13) 224:----------------------------------------
(14) 231:----------------------------------------
(15) 318:----------------------------------------
(16) 319:----------------------------------------

1: aspen AF041049 9: soybean x69955
2: Hybrid populus AF283552 10: Ara AF106084
3: Parsley X13324 11: Pinus teada U12012
4: potato M62755 12: Pinus teada U12013
5: Rubus idaeus AF239687 13: Larix AF144513
6: solanum AF150686 14: PseudolarixAF144528
7: Tobacco D43773 15: Pseudotsuga AF144511
8: rice x52623 16: Tsuga AF144526
METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND COMPOSITION, AND CELLULOSE CONTENT IN PLANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/230,086, filed on Sep. 5, 2000, and is incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with United States government support awarded by the Energy Biosciences Program, United States Department of Energy, and the United States Department of Agriculture research grant numbers USDA 99-35103-7986, USDA 01-03749, and DOE DE-FG02-01ER15179. The United States government has certain rights in this invention.

BACKGROUND OF THE INVENTION

The invention provides a method of introducing two or more genes, involved in lignin biosynthesis, into plant cells. The method of the invention employs either an Agrobacterium-mediated or other appropriate plant gene delivery system by which multiple genes together with a single selectable marker gene are simultaneously transferred and inserted into the genome of plants with high frequencies.

The ability to introduce foreign genes into plants is a prerequisite for engineering agronomic traits in plants. Many systems have been developed for introducing a foreign gene into plant cells, which involve mainly either Agrobacterium- or microprojectile bombardment-mediated transformation (Christou, 1996). The principle of all these systems involves the insertion of a target gene into the host plant genome together with a selectable marker gene encoding either antibiotic or herbicide resistance to aid in the selection of transgenic cells from non-transgenic cells. These systems generally are only effective for introducing a single target gene into the host plant.

To alter agronomic traits, which generally are polygenic in nature, multiple genes involved in complex biosynthetic pathways must be introduced and expressed in plant cells. In this context, the traditional single-gene transfer systems are essentially useless for the following two reasons: 1) it is impractical to introduce multiple genes by repetitive insertion of single genes into transgenic plants due to the time and effort required for recovery of the transgenic tissues; in particular, a repetitive single-gene approach is highly impractical for plant species such as trees which, depending upon the species, require two to three years for transgenic tissue selection and regeneration into a tree; and 2) the presence of a selectable marker gene in a transgenic line precludes the use of the same marker gene in subsequent transformations of plant material from that line. Moreover, the number of available marker genes is limited, and many plant species are recalcitrant to regeneration unless appropriate antibiotic or herbicide selection is used.

Chen et al. (1998) recently reported the genetic transformation of rice with multiple genes by cocoinfected bacteria of several gene constructs into embryogenic suspension tissues. In contrast, Agrobacterium-mediated gene transfer and whole plant regeneration through organogenesis is a simple process and a less species-dependent than bombardment-mediated transformation and regeneration via embryogenesis. However, the introduction of more than one gene in a plasmid vector via Agrobacterium may be technically troublesome and limited by the number or the size of the target genes (Chen et al., 1998). For example, Tricoli et al. (1995) reported the transfer of three target genes to squash via Agrobacterium-mediated gene transfer. A binary plasmid vector containing the three target genes was incorporated into an Agrobacterium strain, which was subsequently used to infect the leaf tissue of squash. As only one line was recovered from numerous infected squash tissues that contained all of the target genes, the use of a single binary vector with a number of genes appears to be a highly inefficient method to produce transgenic plants with multiple gene transfers. Therefore, it was commonly accepted that transfer of multiple genes via Agrobacterium-mediated transformation was impractical (Ebinuma et al., 1997), until success of multiple gene transfer via Agrobacterium was first reported in co-pending, commonly owned PCT application, PCT/US/0027704, filed Oct. 6, 2000, entitled “Method of Introducing a Plurality of Genes into Plants” by Chiang et al, incorporated herein by reference. However, homologous tissue-specific preparation of transgenic trees to specifically alter lignin content, increase S/G (syringyl:guaiacyl) lignin ratio and increase cellulose quantity, as compared to an untransformed plant was unsuccessful.

Yet, the altering of lignin content and composition in plants has been a goal of genetically engineered traits in plants. Lignin, a complex phenolic polymer, is a major part of the supportive structure of most woody plants including angiosperm and gymnosperm trees, which, in turn, are the principal sources of fiber for making paper and cellulosic products. Lignin generally constitutes about 25% of the dry weight of the wood, making it the second most abundant organic compound on earth after cellulose. Lignin provides rigidity to wood for which it is well suited due, in part, to its resistance to biochemical degradation.

Despite its importance to plant growth and structure, lignin is nonetheless problematic to post-harvest, cellulosic-based wood/crop processing for fiber, chemical, and energy production because it must be removed or degraded from cellulose at great expense. Certain structural constituents of lignin, such as the guaiacyl (G) moiety, promote monomer cross-linkages that increase lignin resistance to degradation (Sarkanen, 1971; Chang and Sarkanen, 1973; Chang and Funakosa, 1990). In angiosperms, lignin is composed of a mixture of guaiacyl (G) and syringyl (S) monolignols, and can be degraded at considerably less energy and chemical cost than gymnosperm lignin, which consists almost entirely of guaiacyl moieties (Freudenberg, 1965). It has been estimated that, if syringyl lignin could be genetically incorporated into gymnosperm guaiacyl lignin or into angiosperms to increase the syringyl lignin content, the annual saving in processing of such genetically engineered plants as opposed to their wild types would be in the range of $6 to $10 billion in the U.S. alone. Consequently, there has been a long-standing incentive to understand the biosynthesis of syringyl monolignol to genetically engineer plants to contain more syringyl lignin, thus, facilitating wood/crop processing (Trotter, 1990; Bugos et al., 1991; Boudet et al., 1995; Hu et al., 1999).

Depending on the use for the plant, genetic engineering of certain traits has been attempted. For some plants, as indi-
cated above, there has been a long-standing incentive to genetically modify lignin and cellulose to decrease lignin and increase cellulose contents. For example, it has been demonstrated that the digestibility of forage crops by ruminants is inversely proportional to the quantity of S lignin content in plants (Buxton and Roussel, 1988, Crop. Sci., 28, 553-558; Jung and Vogel, 1986, J. Anim., Sci., 62, 1703-1712). Therefore, decreased lignin and high cellulose plants are desirable in forage crops to increase their digestibility by ruminants, thereby providing the animal with more nutrients per unit of forage.

In other plants, genetically increasing the S/G ratio of the lignin has been sought. As noted above, lignin in angiosperms is composed of guaiacyl (G) and syringyl (S) monomeric units, whereas gymnosperm lignin consists entirely of G units. The structural characteristics of G units in gymnosperm lignin promote monomer cross-linkages that increase lignin resistance to chemical extraction during wood pulping. However, the S units present in angiosperm lignin prevent such chemical resistant cross-links. Therefore, without exception, chemical extraction of G lignin in pulping of gymnosperms is more difficult and requires more chemicals, longer reaction times and higher energy levels than the extraction of G-S lignin during pulping of angiosperms (Sarkanen, K. V., 1971, in Lignins: Occurrence, Formation, Structure and Reaction, Sarkanen, K. V. & Ludwig, C. H., eds., Wiley-Interscience, New York; Chang, H. M. and Sarkanen, K. V., 1973, TAPPI, 56:132-136). As a rule, the reaction rate of extracting lignin during wood pulping is directly proportional to the quantity of the S unit in lignin (Chang, H. M. and Sarkanen, K. V., 1973, TAPPI, 56:132-136). Hence, altering lignin into more reactive G-S type in gymnosperms and into high S/G ratio in angiosperms would represent a pivotal opportunity to enhance current pulping and bleaching efficiency and to provide better, more economical, and more environmentally sound utilization of wood.

Recent results have indicated that high S/G ratio may also add further mechanical advantages to plants, balancing the likely loss of sturdiness of plants with severe lignin reduction (Li et al., 2001, Plant Cell, 13:1567-1585). Moreover, a high S/G lignin ratio would also improve the digestibility of forage crops by ruminants (Buxton and Roussel, 1988, Crop. Sci., 28, 553-558; Jung and Vogel, 1986, J. Anim., Sci., 62, 1703-1712).

In some applications, both a high lignin content and high S/G ratio have been sought (i.e., combining these two traits in plants). For example, it has been demonstrated that when lignin is extracted out of wood during chemical pulping, lignin in the pulping liquor is normally used as a fuel source to provide energy to the pulping and bleaching operations. This lignin-associated energy source, which is not necessary for pulp mills using purchased fuel for energy, is essential to some pulp mills which depend upon internal sources, such as extracted lignin, to be self-sufficient in energy. Therefore, for this purpose, it may be desirable to increase lignin content in pulpwod species, and at the same time to increase the S/G ratio in these species to facilitate the extraction of more lignin to be used as fuel.

Additionally, for grain production and other non-related purposes, increased lignin content and/or S/G lignin ratio are desirable to provide extra sturdiness in plants to prevent the loss of socially and economically important food crops due to defoliation and due to damage to the aerial parts of the plant.

The plant monolignol biosynthetic pathway is set forth in FIG. 1 and will be explained in more detail hereinbelow. The key lignin control sites in the monolignol biosynthetic pathway are mediated by genes encoding the enzymes 4-coumarate-CoA ligase (4CL) (Lee et al., 1997), coniferyl alcohol 5-hydroxylase (CAD) (Osakabe et al., 1999) and S-adenosyl-L-methionine (SAM)-dependent 5-hydroxyconiferaldehyde O-methyltransferase (AldOMT) (Li et al., 2000), respectively, for the formation of sinapaldehyde (see, FIG. 1). Further, coniferyl alcohol dehydrogenase (CAD) (MacKay et al., 1997) catalyzes the reaction including the substrate coniferaldehyde to coniferyl alcohol. It has recently been discovered that sinapyl alcohol dehydrogenase (SAD) enzymatically converts sinapaldehyde into sinapyl alcohol, the syringyl monolignol, for the biosynthesis of syringyl lignin in plants (see, FIG. 1). See, concurrently filed, commonly owned U.S. non-provisional application entitled “Genetic Engineering of Syringyl-Enriched Lignin in Plants,” incorporated herein by reference. It should be noted that the gene encoding the enzyme sinapyl alcohol dehydrogenase (SAD) represents the last gene that is indispensable for genetic engineering of syringyl lignin in plants.

A summary of the conserved regions contained within the coding sequence of each of the above listed proteins is described below. Because SAD is a recently discovered enzyme in Aspen, sequence alignments with other representative species were unable to be performed.

The protein sequence alignments of plant AlDOMTs are shown in FIG. 9. All AlDOMTs have three conserved sequence motifs (I, II, and III) which are the binding sites of S-adenosyl-L-methionine (SAM), the co-substrate or methyl donor for the OMT reaction (Ibrahim, 1997, Trends Plant Sci., 2:249-250; Li et al., 1997, Proc. Natl. Acad. Sci. USA, 94:5461-5466; Joshi and Chiang, 1998, Plant Mol. Biol., 37:663-674). These signatures motifs and the high sequence homology of these proteins to PtAldOMT attest to their function as an AlDOMT specific for converting 5-hydroxyconiferaldehyde into sinapaldehyde (Li et al., 2000, J. Biol. Chem., 275:6537-6545), the content of which is incorporated by reference, herein, in its entirety. This AlDOMT, like CAD5H, also operates at the aldehyde level of the plant monolignol biosynthetic pathway.

The protein sequence alignments of plant CADs are shown in FIG. 10. It was recently proven that CADs are actually guaiacyl monolignol pathway specific (Li et al., 2001, Plant Cell, 13:1567-1585). Based on high sequence homology, the alignment program picked up CADs from angiosperms as well as gymnosperms (radiata pine, loblolly pine and spruce) which have only G-lignin. All CADs have the Zn1 binding motif and structural Zn2 consensus region, as well as a NAD(P) binding site (Jornvall et al., 1987, Eur. J. Biochem., 167:195-201; MacKay et al., 1995, Mol. Gen. Genet., 247:537-545). All these sequence characteristics and high sequence homology to PtCAD1 attest to their function as a G-monolignol specific CAD (Li et al., 2001, Plant Cell, 13:1567-1585).

The protein sequence alignments of plant CalD5Hs are shown in FIG. 11. Although, there are different types of 5-hydroxylases, i.e., F5H, CalD5H is the sole enzyme catalyzing specifically the conversion of coniferaldehyde into 5-hydroxyconiferaldehyde. All full-length CalD5Hs have the proline-rich region located from amino acid 40 to 45 which is believed to be involved in the process of correct folding of microsomal P450s and is also important in heme incorporation into P450s (Yamazaki et al., 1993, J. Biochem., 114:652-657). Also they all have the heme-binding domain (PF00440XXXG) that is conserved in all P450 proteins (Nelson et al. 1996, Pharmacogenetics, 6:1-41). These signature sequences and the high sequence homology of these
proteins to PtCAld5H their function as a 5-hydroxylase that is specific for converting coniferaldehyde into 5-hydroxyconiferaldehyde (Osakabe et al., 1999, Proc. Natl. Acad. Sci. USA, 96:8955–8960).

The protein sequence alignment of plant 4CLs is shown in FIG. 12. In general, 4CL catalyzes the activation of the hydroxycinnamic acids to their corresponding hydroxycinnamoyl-CoA esters. 4CL has the highest activity with p-coumaric acid. 4CL cDNA sequences have been reported from a number of representative angiosperms and gymnosperms, revealing two highly conserved regions, a putative AMP-binding region (SSGTTLPGKGV), and a catalytic motif (GEICIRG). The amino acid sequences of 4CL from plants contain a total of five conserved Cys residues.

Despite recognition of these key enzymes in lignin biosynthesis, there continues to be a need to develop an improved method to simultaneously control the lignin quantity, lignin compositions, and cellulose contents in plants by introducing multiple genes into plant cells.

**BRIEF SUMMARY OF THE INVENTION**

The invention provides a method of introducing two or more genes involved in lignin biosynthesis present in one or more independent vectors into plant cells. The method of the invention suitably employs an Agrobacterium-mediated or another gene delivery system by which multiple genes together with a single selectable marker gene are simultaneously transferred and inserted into the genome of plants with high frequencies.

If an Agrobacterium-mediated gene delivery system is used, each gene of interest is present in a binary vector that has been introduced into Agrobacterium to yield an isolated Agrobacterium strain comprising the binary vector. Moreover, more than one gene of interest may be present in each binary vector. Plant materials comprising plant cells, e.g., plant seed, plant parts or plant tissue including explant materials such as leaf discs, from a target plant species are suitably inoculated with at least two, preferably at least three, and more preferably at least four or more, of the isolated Agrobacterium strains, each containing a different gene of interest. A mixture of the strains is suitably contacted with plant cells. At least one of the binary vectors in the isolated Agrobacterium strains contains a marker gene, and any marker gene encoding a trait for selecting transformed cells from non-transformed cells may be used. Transformed plant cells are regenerated to yield a transgenic plant, the genome of which is augmented with DNA from at least two, preferably at least three, and more preferably at least four, and even more preferably at least five of the binary vectors.

The method of the invention is thus applicable to all plant species that are susceptible to the transfer of genetic information by Agrobacterium or other gene delivery system. Suitable plant species useful in the method of the invention include agriculture and forage crops, as well as monocots. In particular, plant species useful in the method of the invention include trees, e.g., angiosperms and gymnosperms, and more suitably a forest tree, but are not limited to the tree.

The method of the invention is suitably employed to enhance a desired agronomic trait by altering the expression of two or more genes. Such traits include alterations in lignin biosynthesis (e.g., reduction, augmentation and/or structural changes), cellulose biosynthesis (e.g., augmentation, reduction, and/or quality including high degree of polymerization and crystallinity), growth, wood quality (e.g., high density, low juvenile wood, high mature wood, low reaction wood, desirable fiber angle), stress resistance (e.g., cold-, heat-, and salt-tolerance, pathogen-, insect- and other disease-resistance, herbicide-resistance), sterility, high grain yield (for forage and food crops), and increased nutrient level.

Thus, the present invention advantageously provides gymnosperm and angiosperm plants with decreased lignin content, increased syringyl/guaiacyl (S/G) lignin ratio and increased cellulose content in which a single trait or multiple traits are changed.

In another aspect, the invention provides gymnosperm plants with syringyl enriched lignin and/or increased lignin content and/or increased syringyl/guaiacyl (S/G) lignin ratio.

Similarly, the present invention also provides angiosperm plants with increased lignin content.

Other advantages and a fuller appreciation of specific attributes and variations of the invention will be gained upon an examination of the following detailed description of exemplary embodiments and the like in conjunction with the appended claims.

**BRIEF DESCRIPTION OF THE DRAWING(S)**

FIG. 1 is a schematic representation of plant monolignol pathways for production of coniferyl alcohol and sinapyl alcohol;

FIG. 2 is the SAD polynucleotide DNA sequence (SEQ ID NO: 1) and the SAD amino acid sequence (SEQ ID NO: 2) respectively FIGS. 2A and 2B;

FIG. 3 is the CAld5H polynucleotide DNA sequence (SEQ ID NO: 3) and the CAld5H amino acid sequence (SEQ ID NO: 4) respectively FIGS. 3A and 3B;

FIG. 4 is the AldOMT polynucleotide DNA sequence (SEQ ID NO: 5) and the AldOMT amino acid sequence (SEQ ID NO: 6) respectively FIGS. 4A and 4B;

FIG. 5 is the 4CL polynucleotide DNA sequence (SEQ ID NO: 7) and the 4CL amino acid sequence (SEQ ID NO: 8) respectively FIGS. 5A and 5B;

FIG. 6 is the CAD polynucleotide DNA sequence (SEQ ID NO: 9) and the CAD amino acid sequence (SEQ ID NO: 10) respectively FIGS. 6A and 6B;

FIG. 7 is a map of the DNA construct, pBKPlacPta4CL1-a, positioned in a plant transformation binary vector.

FIG. 8 is a map of the DNA construct, pBKPlacPta5ICa5Hs, positioned in a plant transformation binary vector.

FIG. 9 is the protein sequence alignment of AldOMTs for representative species of plants.

FIG. 10 is the protein sequence alignment of CADs for representative species of plants.

FIG. 11 is the protein sequence alignment of CAld5Hs for representative species of plants.

FIG. 12 is the polynucleotide DNA sequence of 4CLs for representative species of plants.

It is expressly understood that the figures of the drawing are for the purposes of illustration and description only and are not intended as a definition of the limits of the invention.

**DETAILED DESCRIPTION OF THE INVENTION**

The present invention provides a method and DNA constructs useful for the transformation of plant tissue for the
alteration of lignin monomer composition, increased syringyl/guaiacyl (S/G) lignin ratio and increased cellulose content and transgenic plants resulting from such transformations. The present invention is of particular value to the paper and pulp industries because lignin containing higher syringyl monomer content is more susceptible to chemical delignification. Woody plants transformed with the DNA constructs provided herein offer a significant advantage in the delignification process over conventional paper feedstocks. Similarly, modification of the lignin composition in grasses by the insertion and expression of a heterologous SAD gene offers a unique method for increasing the digestibility of grasses and is of significant potential economic benefit to the farm and agricultural industries.

The terms used in this specification generally have their ordinary meanings in the art, within the context of the invention and in the specific context where each term is used. Certain terms are discussed below, or elsewhere in the specification, to provide additional guidance to the person of skill in the art in describing the compositions and methods of the invention and how to make and use them. It will be appreciated that the same thing can be said in more than one way. Consequently, alternative language and synonyms may be used for any one or more of the terms discussed herein, nor is any special significance to be placed upon whether or not a term is elaborated or discussed herein. Synonyms for certain terms are provided. A recital of one or more synonyms does not exclude the use of other synonyms. The use of examples anywhere in this specification, including examples of any terms discussed herein, is illustrative only, and in no way limits the scope and meaning of the invention or of any exemplified term. Likewise, the invention is not limited to the preferred embodiments.

As used herein, “gene” refers to a nucleic acid fragment that expresses a specific protein including the regulatory sequences preceding (5' noncoding) and following (3' noncoding) the coding region or coding sequence (Sec., below). “Native” gene refers to the gene as found in nature with its own regulatory sequences.

“Endogenous gene” refers to the native gene normally found in its natural location in the genome.

“Transgene” refers to a gene that is introduced by gene transfer into the host organism.

“Coding sequence” or “Coding Region” refers to that portion of the gene that contains the information for encoding a polypeptide. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, for example, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA, and even synthetic DNA sequences.

“Promoter” or “Promoter Sequence” refers to a DNA sequence, in a given gene, which sequence controls the expression of the coding sequence by providing the recognition site for RNA polymerase and other factors required for proper transcription. Most genes have regions of DNA sequence that are promoter sequences which regulate gene expression. Promoter regions are typically found in the 5' flanking DNA sequence upstream from the coding sequence in both prokaryotic and eukaryotic cells. A promoter sequence provides for regulation of transcription of the downstream gene sequence and typically includes from about 50 to about 2000 nucleotide base pairs. Promoter sequences also contain regulatory sequences such as enhancer sequences that can influence the level of gene expression. Some isolated promoter sequences can provide for gene expression of heterologous DNAs, that is DNA different from the natural homologous DNA. Promoter sequences are also known to be strong or weak or inducible. A strong promoter provides for a high level of gene expression, whereas a weak promoter provides for a very low level of gene expression. An inducible promoter is a promoter that provides for turning on and off of gene expression in response to an exogenously added agent or to an environmental or developmental stimulus. An isolated promoter sequence that is a strong promoter for heterologous DNAs is advantageous because it provides for a sufficient level of gene expression to allow for easy detection and selection of transformed cells, and provides for a high level of gene expression when desired. A promoter may also contain DNA sequences that are involved in the binding of protein factors which control the effectiveness of transcription initiation in response to physiological or developmental conditions.

“Regulatory sequence(s)” refers to nucleotide sequences located upstream (5'), within, and/or downstream (3') of a coding sequence, which control the transcription and/or expression of the coding sequences in conjunction with the protein biosynthetic apparatus of the cell. Regulatory sequences include promoters, translation leader sequences, transcription termination sequences and polyadenylation sequences.

“Encoding” and “coding” refer to the process by which a gene, through the mechanisms of transcription and translation, provides the information to a cell from which a series of amino acids can be assembled into a specific amino acid sequences to produce an active enzyme. It is understood that the process of encoding a specific amino acid sequence includes DNA sequences that may involve base changes that do not cause a change in the encoded amino acid, or which cause base changes which may alter one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. It is therefore understood that the invention encompasses more than the specific exemplary sequences. Modifications to the sequences, such as deletions, insertions or substitutions in the sequence which produce silent changes that do not substantially affect the functional properties of the resulting protein molecule are also contemplated. For example, alterations in the gene sequence which affect the degeneracy of the genetic code, or which result in the production of a chemically equivalent amino acid at a given site, are contemplated. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a biologically equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. In some cases, it may in fact be desirable to make mutants of the sequence to study the effect of retention of biological activity of the protein. Each of these proposed modifications is well within the routine skill in the art, as is the determination of retention of biological activity in the encoded products. Moreover, the skilled artisan recognizes that sequences encompassed by this invention are also defined by their ability to hybridize, under stringent condition, with the sequences exemplified herein.
“Expression” is meant to refer to the production of a protein product encoded by a gene. “Overexpression” refers to the production of a gene product in transgenic organisms that exceed levels of production in normal or non-transformed organisms.

“Functional portion” or “functional fragment” or “functional equivalents” of an enzyme is that portion, fragment or equivalent section which contains the active site for binding one or more reactants or is capable of improving or regulating the rate of reaction. The active site may be made up of separate portions present on one or more polypeptide chains and will generally exhibit high substrate specificity.

“Enzyme encoded by a nucleotide sequence” includes enzymes encoded by a nucleotide sequence which includes partial isolated DNA sequences.

“Transformation” refers to the transfer of a foreign gene into the genome of a host organism and its genetically stable inheritance.

“% identity” refers to the percentage of the nucleotides/amino acids of one polynucleotide/polypeptide that are identical to the nucleotides/amino acids of another sequence of polynucleotide/polypeptide as identified by a program such as GAP from Genetics Computer Group Wisconsin (GCG) package (version 9.0) (Madison, Wis.). GAP uses the algorithm of Needleman and Wunsch (J. Mol. Biol. 48: 443–453, 1970) to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. When parameters required to run the above algorithm are not specified, the default values offered by the program are contemplated.

“Substantial homology” or “substantial similarity” refers to a 70% or more similarity or 70% homology wherein “% similarity” or “% homology” between two polypeptide sequences is a function of the number of similar positions shared by two sequences on the basis of the scoring matrix used divided by the number of positions compared and then multiplied by 100. This comparison is made when two sequences are aligned (by introducing gaps if needed) to determine maximum homology. The PowerBlast program, implemented by the National Center for Biotechnology Information, can be used to compute optimal, gapped alignments. GAP program from Genetics Computer Group Wisconsin package (version 9.0) (Madison, Wis.) can also be used.

“Lignin monomer composition” refers to the relative ratios of guaiacyl monomer and syringyl monomer found in lignified plant tissue.

“Plant” includes whole plants and portions of plants, including plant organs (e.g., roots, stems, leaves, etc.).

“Angiosperm” refers to plants that produce seeds encased in an ovary. A specific example of an angiosperm is *Liquidambar styraciflua* (L.)[sweetgum].

“Gymnosperm” refers to plants that produce naked seeds, i.e., seeds that are not encased in an ovary. A specific example of a gymnosperm is *Pinus taeda* (L.)[loblolly pine].

As used herein, the terms “isolated and/or purified” with reference to a nucleic acid molecule or polypeptide refer to in vitro isolation of a nucleic acid or polypeptide molecule from its natural cellular environment, and from association with other components of the cell, such as nucleic acid or polypeptide, so that it can be sequenced, replicated and/or expressed.

An “isolated” strain of *Agrobacterium* refers to cells derived from a clone of *Agrobacterium* that is transformed in vitro with an isolated binary vector.

A “vector” is a recombinant nucleic acid construct, such as plasmid, phage genome, virus genome, cosmid, or artificial chromosome to which a polynucleotide in accordance with the invention may be attached. In a specific embodiment, the vector may bring about the replication of the attached segment, e.g., in the case of a cloning vector.

“Sinapyl alcohol dehydrogenase” or “SAD”, coniferyl alcohol dehydrogenase or “CAD”, coniferaldehyde 5-hydroxylase or “CaldSH”, 5-hydroxyconiferaldehyde 0-methyltransferase or “AldOMT”, and 4-coumarate-CoA ligase or “4CL” refer to enzymes in the plant phenylpropanoid biosynthetic pathway. In the illustrated embodiments of the present invention, the DNA sequences encoding these enzymes were identified from quaking aspen *Populus tremuloides*. It is understood that each sequence can be used as a probe to clone its equivalent from any plant species by techniques (EST, PCR, RT-PCR, antibodies, etc.) well known in the art.

The Phenyl Propanoid Biosynthetic Pathway

Reference is made to FIG. 1 which shows different steps in the biosynthetic pathways from 4-coumarate (1) to guaiacyl (coniferyl alcohol (6)) and syringyl (sinapyl alcohol (9)) monolignols for the formation of guaiacyl-syringyl lignin together with the enzymes responsible for catalyzing each step. The enzymes indicated for each of the reaction steps are: 4-coumaric acid 3-hydroxylase (C3H) which converts 4-coumarate (1) to caffeate (2); 4-coumarate-CoA ligase (4CL) converts caffeate (2) to caffeoyl CoA (3) which in turn is converted to feruloyl CoA (4) by caffeoyl-CoA O-methyltransferase (CCoAOMT); cinnamoyl-CoA reductase (CCR) converts feruloyl CoA (4) to coniferaldehyde (5); coniferyl alcohol dehydrogenase (CAD) converts coniferaldehyde (5) to the guaiacyl monolignol coniferyl alcohol (6); at coniferaldehyde (5), the pathway splits wherein coniferaldehyde (5) can also be converted to 5-hydroxyconiferaldehyde (7) by coniferaldehyde 5-hydroxylase (CaldSH); 5-hydroxyconiferaldehyde 0-methyltransferase (AldOMT) converts 5-hydroxyconiferaldehyde (7) to sinapaldehyde (8) which, in turn, is converted to the syringyl monolignol, sinapyl alcohol (9) by sinapyl alcohol dehydrogenase (SAD).

DNA Constructs

According to the present invention, a DNA construct is provided which is a plant DNA having a promoter sequence, a coding region and a terminator sequence. The coding region encodes a combination of enzymes essential to lignin biosynthesis, specifically, SAD, CAD, C5H, and 4CL, and 4CL protein sequences, substantially similar sequences, or functional fragments thereof. The coding region is suitably a minimum size of 50 bases. The gene promoter is positioned at the 5'-end of a transgene (e.g., 4CL alone or together with CAD, C5H, and AldOMT, and combinations thereof, or 4CL and CAD alone, or together with C5H, SAD, CAD, and AldOMT, and combinations thereof, as described hereinafter) for controlling the transgene expression, and a gene termination sequence that is located at the 3'-end of the transgene for signaling the end of the transcription of the transgene.

The DNA construct in accordance with the present invention can be incorporated into the genome of a plant by transformation to alter lignin biosynthesis, increase syringyl/guaiacyl (S/G) lignin ratio and increase cellulose content. The DNA construct may include clones of C5H, SAD, AldOMT, CAD, and 4CL, and variants thereof such as are permitted by the degeneracy of the genetic code and the functional equivalents thereof.

The DNA constructs of the present invention may be inserted into plants to regulate production the following
enzymes: CAD5H, SAD, AldOMT, CAD, and 4CL. Depending on the nature of the construct, the production of the protein may be increased or decreased, either throughout or at specific stages in the life of the plant, relative to a similar control plant that does not incorporate the construct into its genome. For example, the orientation of the DNA coding sequence, promoter, and termination sequence can serve to either suppress lignin formation or amplify lignin formation. For the down-regulation of lignin synthesis, the DNA is in the antisense orientation. For the amplification of lignin biosynthesis, the DNA is in the sense orientation, thus to provide one or more additional copies of the DNA in the plant genome. In this case, the DNA is suitably a full-length cDNA copy. It is also possible to target expression of the gene to specific cell types of the plants, such as the epidermis, the xylem, the roots, etc. Constructs in accordance with the present invention may be used to transform cells of both monocotyledons and dicotyledons plants in various ways known in the art. In many cases, such plant cells may be cultured to regenerate whole plants which subsequently reproduce to give successive generations of genetically modified plants. Examples of plants that are suitably genetically modified in accordance with the present invention, include but are not limited to, trees such as aspen, poplar, pine and eucalyptus.

Promoters and Termination Sequences

Various gene promoter sequences are well known in the art and can be used in the DNA constructs of present invention. The promoter in the constructs in accordance with the present invention suitably provides for expression of the linked DNA segment. The promoter may also be inducible so that gene expression can be turned on or off by an exogenously added agent. It may also be preferable to combine the desired DNA segment with a promoter that provides tissue-specific expression or developmentally regulated gene expression in plants.

The promoter may be selected from promoters known to operate in plants, e.g., CaMV35S, GPAL2, GPAL3 and endogenous plant promoter controlling expression of the enzyme of interest. Use of a constitutive promoter such as the CaMV35S promoter (Odell et al. 1985), or CaMV 19S (Lawton et al., 1987) can be used to drive the expression of the transgenes in all tissue types in a target plant. Other promoters are nos (Ebert et al., 1987), Adh (Walker et al., 1987), sucrose synthase (Yang et al., 1990), A-tubulin, ubiquitin, actin (Wang et al., 1992), cab (Sullivan et al., 1989), PEPCase (Hudsphel et al., 1989) or those associate with the R gene complex (Chandler and et al., 1989). On the other hand, use of a tissue-specific promoter permits functions to be controlled more selectively. The use of a tissue-specific promoter has the advantage that the desired protein is only produced in the tissue in which its action is required. Suitable, tissue-specific promoters, such as those would confine the expression of the transgenese in developing xylem where lignification occurs, may be used in the inventive DNA constructs.

A DNA segment can be combined with the promoter by standard methods as described in Sambrook et al., 2nd ed. (1982). Briefly, a plasmid containing a promoter such as the CaMV 35S promoter can be constructed as described in Jefferson (1987) or obtained from Clontech Lab, Palo Alto, Calif. (e.g., pH121 or pH221). Typically, these plasmids are constructed to provide for multiple cloning sites having specificity for different restriction enzymes downstream from the promoter. The DNA segment can be subcloned downstream from the promoter using restriction enzymes to ensure that the DNA is inserted in proper orientation with respect to the promoter so that the DNA can be expressed.

The gene termination sequence is located 3' to the DNA sequence to be transcribed. Various gene termination sequences known in the art may be used in the present inventive constructs. These include nopaline synthase (NOS) gene termination sequence (see, e.g., references cited in co-pending, commonly-owned PCT application, PCT/US/0027704, filed Oct. 6, 2000, entitled “Method of Introducing a Plurality of Genes into Plants,” incorporated herein by reference.)

Marker Genes

A marker gene may also be incorporated into the inventive DNA constructs to aid the selection of plant tissues with positive integration of the transgene. “Marker genes” are genes that impart a distinct phenotype to cells expressing the marker gene, and thus, allow such transformed cells to be distinguished from cells that do not have the marker. Many examples of suitable marker genes are known to the art and can be employed in the practice of the invention, such as neomycin phosphotransferase II (NPT II) gene that confers resistance to kanamycin or hygromycin antibiotics which would kill the non-transformed plant tissues containing no NPT II gene (Bevan et al., 1983). Numerous other exemplary marker genes used in the method, in accordance with the present invention are listed in Table 1 of co-pending, commonly owned of PCT/US/0027704, filed Oct. 6, 2000, entitled “Method of Introducing a Plurality of Genes into Plants,” incorporated herein by reference.

Therefore, it will be understood that the following discussion is exemplary rather than exhaustive. In light of the techniques disclosed herein and the general recombinant techniques which are known in the art, the present invention renders possible the introduction of any gene, including marker genes, into a recipient cell to generate a transformed plant.

Optional Sequences in the Expression Cassette

The expression cassette containing DNA sequences in accordance with the present invention can also optionally contain other DNA sequences. Transcription enhancers or duplications of enhancers can be used to increase expression from a particular promoter. One may wish to obtain novel tissue-specific promoter sequences for use in accordance with the present invention. To achieve this, one may first isolate cDNA clones from the tissue concerned and identify those clones which are expressed specifically in that tissue, for example, using Northern blotting. Ideally, one would like to identify a gene that is not present in a high copy number, but which gene product is relatively abundant in specific tissues. The promoter and control elements of corresponding genomic clones may then be localized using the techniques of molecular biology known to those of skill in the art.

Expression of some genes in transgenic plants will occur only under specified conditions. It is known that a large number of genes exist that respond to the environment. In some embodiments of the present invention expression of a DNA segment in a transgenic plant will occur only in a certain time period during the development of the plant. Developmental timing is frequently correlated with tissue specific gene expression.

As the DNA sequence inserted between the transcription initiation site and the start of the coding sequence, i.e., the untranslated leader sequence, can influence gene expression, one can also employ a particular leader sequence. Preferred leader sequence include those which comprise sequences selected to direct optimum expression of the attached gene, i.e., to include a preferred consensus leader sequence which can increase or maintain mRNA stability and prevent inappropriate initiation of translation (Joshi, 1987). Such
sequences are known to those of skill in the art. Sequences that are derived from genes that are highly expressed in plants will be most preferred.

Additionally, expression cassettes can be constructed and employed to target the gene product of the DNA segment to an intracellular compartment within plant cells or to direct a protein to the extracellular environment. This can generally be achieved by joining a DNA sequence encoding a transit or signal peptide sequence to the coding sequence of the DNA segment. Also, the DNA segment can be directed to a particular organelle, such as the chloroplast rather than to the cytoplasm.

Alternatively, the DNA fragment coding for the transit peptide may be chemically synthesized either wholly or in part from the known sequences of transit peptides such as those listed above. The description of the optional sequences in the expression cassette, is commonly owned, co-pending PCT/US 0027704, filed Oct. 6, 2000, entitled “Method of Introducing a Plurality of Genes into Plants,” incorporated herein by reference.

Transformation

Transformation of cells from plants, e.g., trees, and the subsequent production of transgenic plants using e.g., Agrobacterium-mediated transformation procedures known in the art, and further described herein, is one example of a method for introducing a foreign gene into plants. Although, the method of the invention can be performed by other modes of transformation, Agrobacterium-mediated transformation procedures are cited as examples, herein. For example, transgenic plants may be produced by the following steps: (i) culturing Agrobacterium in low-pH induction medium at low temperature and preconditioning, i.e., coculturing bacteria with wounded tobacco leaf extract in order to induce a high level of expression of the Agrobacterium vir genes whose products are involved in the T-DNA transfer; (ii) coculturing desired plant tissue explants, including zygotic and/or somatic embryo tissues derived from cultured explants, with the incited Agrobacterium; (iii) selecting transformed callus tissue on a medium containing antibiotics; and (iv) converting the embryos into platelets.

Any non-tumorigenic A. tumefaciens strain harboring a disarmed Ti plasmid may be used in the method in accordance with the invention. Any Agrobacterium system may be used. For example, Ti plasmid-binary vector system or a cointegrative vector system with one Ti plasmid may be used. Also, any marker gene or polynucleotide conferring the ability to select transformed cells, callus, embryos or plants and any other gene, such as for example a gene conferring resistance to a disease, or one improving lignin content or structure or cellulose content, may also be used. A person of ordinary skill in the art can determine which markers and genes are used depending on particular needs.

To increase the infectivity of the bacteria, Agrobacterium is cultured in low-pH induction medium, i.e., any bacterium culture media with a pH value adjusted to from 4.5 to 6.0, most preferably about 5.2, and at low temperature such as for example about 19–30°C, preferably about 21–26°C. The conditions of low-pH and low temperature are among the well-defined critical factors for inducing virulence activity in Agrobacterium (e.g., Altombe et al., 1989; Fullner et al., 1996; Fullner and Nester, 1996).

The bacteria is preconditioned by coculturing with wounded tobacco leaf extract (prepared according to methods known generally in the art) to induce a high level of expression of the Agrobacterium vir genes. Prior to inoculation of plant somatic embryos, Agrobacterium cells can be treated with a tobacco extract prepared from wounded leaf tissues of tobacco plants grown in vitro. To achieve optimal stimulation of the expression of Agrobacterium vir genes by wound-induced metabolites and other cellular factors, tobacco leaves can be wounded and pre-cultured overnight. Culturing of bacteria in low pH medium and at low temperature can be used to further enhance the bacteria vir gene expression and infectivity. Preconditioning with tobacco extract and the vir genes involved in the T-DNA transfer process are generally known in the art.

Agrobacterium treated as described above is then cocultured with a plant tissue explant, such as for example, zygotic and/or somatic embryo tissue. Non-zygotic (i.e., somatic) or zygotic tissues can be used. Any plant tissue may be used as a source of explants. For example, cotyledons from seeds, young leaf tissue, root tissues, parts of stems including nodal explants, and tissues from primary somatic embryos such as the root axis may be used. Generally, young tissues are a preferred source of explants.

The above-described transformation and regeneration protocol is readily adaptable to other plant species. Other published transformation and regeneration protocols for plant species include Danekar et al., 1987; McGranahan et al., 1988; McGranahan et al., 1990; Chen, Ph.D. Thesis, 1991; Sullivan et al., 1993; Huang et al., 1991; Wilde et al., 1992; Minocha et al., 1986; Parsons et al., 1986; Fillatti et al., 1987; Pythoud et al., 1987; De Block, 1990; Brasileiro et al., 1991; Brasileiro et al., 1992; Howe et al., 1991; Klopfenstein et al., 1991; Leple et al., 1992; and Nilsson et al., 1992.

Characterization

To confirm the presence of the DNA segment(s) or “transgene(s)” in the regenerated plants, a variety of assays may be performed. Such assays include, for example, “molecular biological” assays well known to those of skill in the art, such as Southern and Northern blotting and PCR; “biochemical” assays, such as detecting the presence of a protein product, e.g., by immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as leaf or root assays; and also, by analyzing the phenotype of the whole regenerated plant.

1. DNA Integration, RNA Expression and Inheritance

Genomic DNA may be isolated from callus cell lines or any plant parts to determine the presence of the DNA segment through the use of techniques well known to those skilled in the art. Note that intact sequences will not always be present, presumably due to rearrangement or deletion of sequences in the cell.

The presence of DNA elements introduced through the methods of this invention may be determined by polymerase chain reaction (PCR). Using this technique, discreet fragments of DNA are amplified and detected by gel electrophoresis. This type of analysis permits one to determine whether a DNA segment is present in a stable transformant, but does not prove integration of the introduced DNA segment into the host cell genome. In addition, it is not possible using PCR techniques to determine whether transformants have exogenous genes introduced into different sites in the genome, i.e., whether transformants are of independent origin. It is contemplated that using PCR techniques it would be possible to clone fragments of the host genomic DNA adjacent to an introduced DNA segment.

Positive proof of DNA integration into the host genome and the independent identities of transformants may be determined using the technique of Southern hybridization. Using this technique, specific DNA sequences that were introduced into the host genome and flanking host DNA sequences can be identified. Hence the Southern hybridization...
tion pattern of a given transformant serves as an identifying characteristic of that transformant. In addition, it is possible through Southern hybridization to demonstrate the presence of introduced DNA segments in high molecular weight DNA, i.e., confirm that the introduced DNA segment has been integrated into the host cell genome. The technique of Southern hybridization provides information that is obtained using PCR, e.g., the presence of a DNA segment, but also demonstrates integration into the genome and characterizes each individual transformant.

It is contemplated that by using the techniques of dot or slot blot hybridization which are modifications of Southern hybridization techniques, one could obtain the same information that is derived from PCR, e.g., the presence of a DNA segment.

Both PCR and Southern hybridization techniques can be used to demonstrate transmission of a DNA segment to progeny. In most instances the characteristic Southern hybridization pattern for a given transformant will segregate in progeny as one or more Mendelian genes (Spencer et al., 1992; Laursen et al., 1994) indicating stable inheritance of the gene.

Whereas DNA analysis techniques may be conducted using DNA isolated from any part of a plant, RNA may only be expressed in particular cells or tissue types, and hence, it will be necessary to prepare RNA for analysis from these tissues. PCR techniques may also be used for detection and quantitation of RNA produced from introduced DNA segments. In this application of PCR, it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances, PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique will demonstrate the presence of an RNA species and give information about the integrity of that RNA. The presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and demonstrate only the presence or absence of an RNA species.

2. Gene Expression

While Southern blotting and PCR may be used to detect the DNA segment in question, they do not provide information as to whether the DNA segment is being expressed. Expression may be evaluated by specifically identifying the protein products of the introduced DNA segments or evaluating the phenotypic changes brought about by their expression.

Assays for the production and identification of specific proteins may make use of physical-chemical, structural, functional, or other properties of the proteins. Unique physical-chemical or structural properties allow the proteins to be separated and identified by electrophoretic procedures, such as native or denaturing gel electrophoresis or isoelectric focussing, or by chromatographic techniques such as ion exchange or gel exclusion chromatography. The unique structures of individual proteins also offer opportunities for use of specific antibodies to detect their presence in formats such as an ELISA assay. Combinations of approaches may be employed with even greater specificity such as western blotting in which antibodies are used to locate individual gene products that have been separated by electrophoretic techniques. Additional techniques may be employed to absolutely confirm the identity of the product of interest such as evaluation by amino acid sequencing following purification. Although these are among the most commonly employed, other procedures may be additionally used.

Assay procedures may also be used to identify the expression of proteins by their functionality, especially the ability of enzymes to catalyze specific chemical reactions involving specific substrates and products. These reactions may be followed by providing and quantifying the loss of substrates or the generation of products of the reactions by physical or chemical procedures. Examples are as varied as the enzyme to be analyzed and may include assays for PAT enzymatic activity by following production of radiolabelled acetylated phosphinothricin from phosphinothricin.

Very frequently the expression of a gene product is determined by evaluating the phenotypic results of its expression. These assays also may take many forms including but not limited to analyzing changes in the chemical composition, morphology, or physiological properties of the plant. Chemical composition may be altered by expression of DNA segments encoding enzymes or storage proteins which change amino acid composition and may be detected by amino acid analysis, or by enzymes which change starch quantity which may be analyzed by near infrared reflectance spectrometry. Morphological changes may include greater stature or thicker stalks. Most often changes in response of plants or plant parts to imposed treatments are evaluated under carefully controlled conditions termed bioassays.

The invention will be further described by the following non-limiting examples.

**EXAMPLE 1**

**Preparation of Transgenic Aspen**

**Construction of Binary Vectors**

pBKPtACT1 a- Asp 4CL1 xylem specific promoter (PPT4CL1 1.1 kb, GenBank AF041051) was prepared and linked to aspen 4CL1 cDNA (Pt4CL1, GenBank AF041049) which was orientated in the antisense direction. Then the cassette containing aspen 4CL1 promoter and antisense aspen 4CL1 cDNA was positioned in a plant transformation binary vector, as shown in FIG. 1 (pBKlPtACT1 Pt44CL1-a construct).

pBKPTACT1 PtCAlD5H-s From pBKPTACT1 Pt44CL1-a construct, the antisense Pt4CL1 was replaced with PtCAlD5H cDNA in a sense orientation, yielding a pBKPTACT1 PtCAlD5H-s transformation binary construct, as shown in FIG. 8.

Also, Example 1 of PCT application PCT/US/0027704, filed Oct. 6, 2000, entitled “Method of Introducing a Plurality of Genes into Plants,” incorporated herein by reference, describes a number of other gene constructs for preparing transgenic plants. The plants are transformed with a gene from the phenylpropanoid pathway (i.e., 4CL, AOMT, CoAOMT, and CAld5H) using an operably linked to either a homologous or a heterologous and either a constitutive or tissue-specific promoter.

**Incorporation of Binary Vector into Agrobacterium**

According to the protocol described in Tsai et al. (1994, Plant Cell Reports, 14:94–97) Agrobacterium C58/pMP90 strain was grown in LB with selection of gentamicin at 28° C. overnight. Cells were harvested by centrifugation at 10,000 rpm for 10 minutes at 4° C. The cell pellet was washed with 0.5 volume of ice-cold 20 mM CaCl2, and centrifuged again. The cells were then resuspended in 0.1 volume of ice-cold 20 mM CaCl2 in a sample tube. About 1 µg of binary vector DNA was added to 200 µL of the cell suspension and mixed by pipetting. The sample tube was chilled in liquid N2 for 5 minutes and thawed at 37° C in a water bath for 5 minutes. One mL of LB medium was added and the mixture was incubated at 28° C. for 3 hours with gentle shaking. Twenty µL of the cells were spread onto a LB
plate containing 25 μg/mL gentamicin and 50 μg/mL kanamycin and incubated at 28°C for 2 days. PCR
(amplification conditions, cycling parameters and primers are described below) was used to verify the presence of DNA from the vector in the transformed colonies.

Simultaneous Transformation of Aspen with Multiple Genes Via Engineered Agrobacterium Strains

For simultaneous transformation of multiple genes, pBKpTT4c, Pt4CL-a and pBKpTT4c, PtCal5H Agrobacterium clones were cultured in LB medium at 28°C overnight separately. The Agrobacterium strains were subcultured individually by a 1:100 fold dilution into 50 mL of LB (pH 5.4) containing 50 μg/mL kanamycin, 25 μg/mL gentamicin and 20 μM acetosyringone (in DMSO), and grown overnight at 28°C with shaking. An equal volume of the same density of individually cultured Agrobacterium strains was then mixed. Leaves excised from sterile tobacco plants were cut into pieces with a size of about 5 mm² and the leaf discs were then immered in the Agrobacterium mixture for 5 minutes.

After removing excess Agrobacterium cells, the treated leaf discs were placed on callus induction medium (WPM: Woody Plant Medium, BA: 6-benzyladenine, 2,4-D: 2,4-dichlorophenoxyacetic acid; Tsai et al. 1994, Plant Cell Reports, 14:94-97) and cultured for 2 weeks. The pre-cultured leaf discs were rinsed with sterile water several times to remove the Agrobacterium cells and washed in 1 mg/mL clafor an and 1 mg/mL ticarcillin with shaking for 3 hours to kill Agrobacterium. After briefly blot-drying, the pre-cultured and washed leaf discs were cultured on callus induction medium containing 50 μg/mL kanamycin and 300 μg/mL claforan for selection of transformed cells. After 2 to 3 subcultures (10 days/subculture), the calli grown on the leaf discs were excised and transferred onto shoot induction medium (WPM+TDZ: N-phenyl-N’-1,2,3-thiadiazol-5-yl-urca) containing 50 μg/mL kanamycin and 300 μg/mL claforan for regenerant shoots. After shoots were grown to about 0.5 cm high, they excised and planted to rooting media (WPM with kanamycin and claforan). Whole plants about 7 cm high were transplanted into soil and maintained in a greenhouse for subsequent molecular characterization.

Genomic DNA Isolation

Genomic DNA was isolated according to Hu et al. (1998). About 100 mg of young leaves were collected from each plant growing in the greenhouse and ground in liquid N₂ to fine powder for DNA isolation using QIAGEN plant DNA isolation kit (Valencia, Calif.). Specifically, the powdered tissue was added to extract buffer containing 2% hexade- cyltrimethylammonium bromide (CTAB), 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 1.4 M NaCl and 30 mM β-mercaptoethanol at 5 mL/g tissue. The extraction mixture was incubated in a tube at 60°C for 1 hour with occasional shaking. One volume of chloroform-isooamyl alcohol (24:1) was added and mixed gently. The two phases were separated by centrifugation at 10,000g for 10 minutes. The aqueous phase was transferred to a new tube and extracted with chloroform in the presence of 1% CTAB and 0.7 M NaCl. The DNA was precipitated by addition of 2/3 volume of 5 M ammonia acetate and 10 mL of 95% ethanol at -20°C for 20 minutes. After centrifugation, the pellet was washed with 70% ethanol. After a brief drying, genomic DNA was dissolved in TE buffer.

PCR Verification of Foreign Gene Insertion in Host Plant Genome

PCR was used to verify the integration of the gene constructs in the genome of transgenic plants. Two specific primers were synthesized for each construct and used to PCR-amplify the corresponding construct in genome of transgenic Aspen. For the PBKpTT4c, Pt4CL-a construct, two specific primers were synthesized that amplify a 4CL cDNA fragment. Pt4CL1 promoter sense primer (5’CAGGAATGCTCTGACTCTG3’) (SEQ ID NO:11) and Pt4CL1 sense primer (5’ATGAATCCACAGAATCCA’3’) (SEQ ID NO:12), at the translation start region. Primers for PCR verification of pBKpTT4c, PtCal5H-s construct are Pt4CL1 promoter sense primer (5’CAGGAATGCTCTGACTCTG3’) (SEQ ID NO:13) and PtCal5H antisense primer (5’TITAGAGGACAGAGCAGACG3’) (SEQ ID NO:14) at translation stop region.

The PCR reaction mixture contained 100 ng genomic DNA of transformed aspen, and 0.2 μM of each primer, 100 μM of each deoxyribonucleotide triphosphate, 1×PCR buffer and 2.5 Units of Taq DNA polymerase (Promega Madison, Wis.) in a total volume of 50 μL. The cycling parameters were as follows: 94°C for 1 minute, 56°C for 1 minute (for 4CL and Cal5H or can vary between cDNA templates used) according to different gene checked) and 72°C for 2 minutes, for 40 cycles, with 5 minutes at 72°C extension. The PCR products were eleetrophoresized on a 1% agarose gel.

EXAMPLE 2
Preparation of Other Transgenic Plants

It is important to recognize that there is a substantial percentage of sequence homology among the plant genes involved in the lignin biosynthetic pathway, discussed herein. This substantial sequence homology allows the method in accordance with the invention disclosed herein to be applicable to all plants that possess the requisite genes involved in the lignin biosynthetic pathway. To demonstrate the substantial sequence homology among plant genes, the percentage sequence homology is set forth in tabular form, for example, Cal5H genes (Table 1), AldOMT genes (Table 2), CAD genes (Table 3), and 4CL genes (See FIG. 12). Therefore, it is possible to alter lignin monomer composition, increase S/G lignin ratio, and increase cellulose content in all plants by using the method in accordance with the invention, described herein.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Protein sequence homology (%) of plant Conifer Aldehyde 5-hydroxylase (Cal5H) from 1) Aspen; 2) Poplar, AB010324: 3) Sweetgum, AF179557: 4) Ashbialog (Ferulic Acid 5-hydroxylase, F5H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>2</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>81</td>
</tr>
</tbody>
</table>
incorporated into plant cells by Agrobacterium mediated transformation, as set forth by the method in accordance with the invention and after PCR confirmation of transgene integration, 14 positive transgenic trees were randomly selected, representing three different transgenic groups, i.e., Groups I, II and III. Group I (plant #21, 22, 23, 25, and 37) consists of those with the integration of only antisense Pt4CLl cDNA, whereas Group III plants (#71, 72, 74, and 141) contained both antisense Pt4CLl and sense PtCald5H transgenes. These transgenic trees were then further analyzed for their lignin and cellulose contents and lignin S/G ratio (Table 4). It is clear that, when compared with the control, untransformed aspen, transgenic plants (#21, 22, 23, 25, and 37) engineered for the suppression of 4CL gene with antisense Pt4CLl transgene had drastic reductions in their lignin content, with significant increases in their cellulose content. Transgenic plants (#32, 84, 93, 94, and 108) engineered for the overexpression of

To further demonstrate the versatility of this invention in transferring a variety of foreign genes and the applicability of this invention to plants other than the herbaceous species, different binary vectors were constructed and transferred into aspen (Populus tremuloides) tree. Two binary vectors, each containing a cDNA sequence and a neomycin phosphotransferase (NPT II) cDNA encoding kanamycin resistance, were constructed. Each vector was then individually mobilized into Agrobacterium strain CS8 to create two isolated (engineered) Agrobacterium strains. It should be noted that about 50 transgenic tobacco plants were generated by the same technique harboring 4 different sets of foreign genes, as described in the PCT application PCTUS0027704 filed Oct. 6, 2000, entitled “Method of Introducing a Plurality of Genes into Plants,” incorporated herein by reference.

Table 4 summarizes the numerical results from simultaneous suppression of 4CL gene and overexpression of CAld5H gene, transgenic plants (#71, 72, 74, and 141) all exhibited low lignin content, high S/G ratio and elevated cellulose quantity. In summary, these results show that multiple genes carried by individual Agrobacterium strains can be integrated simultaneously into the plant genome.

Moreover, it was demonstrated as shown herein below, that transgenic plants with a nearly 30% increase in cellulose content and over 50% lignin quantity reduction, accompanied with a significant augmentation of the S/G ratio, can be easily produced. It is conceivable that more genes can also be efficiently transferred at one time. Only one suitable marker gene is required for this system, although a number of marker genes can also be employed.

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein sequence homology (%) of plant AldOMTs from 1) Aspen, X62096; 2) Poplar, M73431; 3) Almond, X63217; 4) Strawberry, AF220491; 5) Alfalfa, M63853; 6) Eucalyptus, X74614; 7) Cuckee breweri, AF060609; 8) Sweetgum, AF165583; 9) Ambrosia, U70424; 10) Tobacco, X74582; 11) Viva vinifera, AF397440</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
</tbody>
</table>
TABLE 4

Simultaneous manipulating xylem-specific expression of 4CL and CAld5H in transgenic aspen.

<table>
<thead>
<tr>
<th>Plant #</th>
<th>Control</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>25</th>
<th>32</th>
<th>84</th>
<th>93</th>
<th>108</th>
<th>71</th>
<th>72</th>
<th>74</th>
<th>141</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene integrated</td>
<td>4CL-s</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Lignin content (%)</td>
<td>22.4</td>
<td>16.0</td>
<td>15.3</td>
<td>14.4</td>
<td>13.1</td>
<td>14.9</td>
<td>22.4</td>
<td>21.6</td>
<td>21.1</td>
<td>20.7</td>
<td>19.7</td>
<td>13.2</td>
<td>13.7</td>
</tr>
<tr>
<td>Lignin S/G ratio</td>
<td>2.2</td>
<td>2.1</td>
<td>2.0</td>
<td>2.2</td>
<td>2.3</td>
<td>2.1</td>
<td>4.8</td>
<td>4.0</td>
<td>5.5</td>
<td>4.9</td>
<td>3.0</td>
<td>3.3</td>
<td>3.6</td>
</tr>
<tr>
<td>Cellulose content (%)</td>
<td>41.4</td>
<td>43.1</td>
<td>ND</td>
<td>ND</td>
<td>47.3</td>
<td>ND</td>
<td>40.0</td>
<td>ND</td>
<td>44.7</td>
<td>ND</td>
<td>ND</td>
<td>49.2</td>
<td>ND</td>
</tr>
</tbody>
</table>

ND: not determined

EXAMPLE 3

Production of Commercially Desirable Agronomic Traits in Transformed Plants

The following genetic transformations illustrate the production of commercially desirable agronomic traits in plants.

Gymnosperms

A. To produce syringyl-enriched lignin in gymnosperm plants, gymnosperm plants are genetically transformed with SAD, CAld5H, and AldOMT genes in the sense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These five genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

B. To produce decreased lignin content, increased syringyl/guaiacyl (S/G) lignin ratio and increased cellulose quantity in gymnosperm plants, gymnosperm plants are genetically transformed with SAD, CAld5H and AldOMT genes in the sense orientation and 4CL gene in either sense or antisense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These four genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

C. To produce decreased lignin content, increased syringyl/guaiacyl (S/G) lignin ratio and increased cellulose quantity in gymnosperm plants, gymnosperm plants are genetically transformed with SAD, CAld5H and AldOMT genes in the sense orientation and 4CL gene in either sense or antisense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These five genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

D. To produce increased lignin content in gymnosperm plants, gymnosperm plants are genetically transformed with 4CL gene in the sense orientation driven by any appropriate promoter and via any appropriate genetic transformation system.

E. To produce increased lignin content and increased syringyl/guaiacyl (S/G) lignin ratio in gymnosperm plants, gymnosperm plants are genetically transformed with SAD, CAld5H, AldOMT, and 4CL genes in the sense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These four genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

F. To produce increased lignin content, increased syringyl/guaiacyl (S/G) lignin ratio in gymnosperm plants, gymnosperm plants are genetically transformed with SAD, CAld5H, AldOMT, and 4CL genes in the sense orientation and CAD gene in the antisense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These four genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

Angiosperms

A. To produce increased S/G lignin ratio in angiosperm plants, angiosperm plants are genetically transformed with either CAld5H, AldOMT, or SAD genes in sense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These three genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

B. To produce decreased lignin content, increased S/G lignin ratio and increased cellulose quantity in angiosperm plants, angiosperm plants are genetically transformed with either SAD, CAld5H, or AldOMT genes in the sense orientation and 4CL gene in the sense or antisense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These four genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

C. To produce decreased lignin content, increased S/G lignin ratio and increased cellulose quantity in angiosperm plants, angiosperm plants are genetically transformed with either SAD, CAld5H, or AldOMT genes in the sense orientation and 4CL gene in the sense or antisense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These four genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.
in the sense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These four genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

F: To produce increased lignin content and increased S/G ratio in angiosperm plants, angiosperm plants are genetically transformed with 4CL in the sense orientation and either SAD, CALD5H, or ALDOMT genes also in the sense orientation and CAD in the antisense orientation driven by any appropriate promoter and via any appropriate genetic transformation system. These four genes can be transferred into the host plant either simultaneously (in one or individual constructs) or sequentially (in individual constructs) in any order.

All publications, patents and patent applications cited herein are incorporated herein by reference. While in the foregoing specification, this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details herein may be varied considerably without departing from the basic principles of the invention. Accordingly, it is intended that the present invention be solely limited by the broadest interpretation that can be accorded the appended claims.

REFERENCES

Tsai et al., 1994, Plant Cell Report 14:94.
Boudet et al., 1995, New Phytol. 129:203.
Brasileiro et al., 1992, Transgenic Res. 1:133.
Chandler et al., 1989.
Danekar et al., 1987, Biotechnolgy 5:587.
Ebert et al. 1987.
Freudenberg, 1965.

Lawton et al., 1987, Plant Mol. Biol. 9:31F.
Li et al., 1999, Plant Mol. Biol. 40:555.
Li et al., 2000, J. Biol. Chem. 275:6537.
MacKay et al., 1997.
McGranahan et al., 1988, Bio/Technology 6:800.
McGranahan et al., 1990, Plant Cell Reports 8:512.
Minocha et al., 1986, Proc. TAPPI Research and Development Conference, TAPPI Press, Atlanta, 89.
Parsons et al., 1986, Bio/Technology 4:533.
Sambrook et al., 1982.
Tsai et al., 1998, Plant Cell Reports 14:94.
Tricoli et al., 1995.
Walker et al., 1987, PNAS USA 84:6624.
Yang et al., 1990, PNAS USA 87:4144.
### SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 14

<210> SEQ ID NO 1
<211> LENGTH: 1446
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremulous
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: SAD

<400> SEQUENCE: 1

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tttttttttt</td>
<td>ttctctgcct ctctctctgc agatatttc</td>
</tr>
<tr>
<td>ttgcacgac taccacacct gcaagcgcct gcggtgggttg gcgggtgggc</td>
<td>120</td>
</tr>
<tr>
<td>caatctagtc tctctctctc ttctctcttc ccagagagcc cactggtgct gagatgtgc</td>
<td>180</td>
</tr>
<tr>
<td>ggttcacagt acgtacgcct cggagtatgc ttcctctgt ctacagatgc aagaattgc</td>
<td>240</td>
</tr>
<tr>
<td>ggggctcccct cgtgacccct ttggtgcttg ccagtcgacg ctcctattgc</td>
<td>300</td>
</tr>
<tr>
<td>tgggtagcag atctggtgcag tggggcgcac ccgcaagtgc gggggag</td>
<td>360</td>
</tr>
<tr>
<td>tgggtgaacc gggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>420</td>
</tr>
<tr>
<td>tggggcagct cggacgccgc atctggtgcag tggaaaattgc</td>
<td>480</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>540</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>600</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>660</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>720</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>780</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>840</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>900</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>960</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1020</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1080</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1140</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1200</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1260</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1320</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1380</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1440</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1500</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1560</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1620</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1680</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1740</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1800</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1860</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1920</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>1980</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2040</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2100</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2160</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2220</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2280</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2340</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2400</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2460</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2520</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2580</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2640</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2700</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2760</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2820</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2880</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>2940</td>
</tr>
<tr>
<td>tggggggtgcttg gctggtgcttg gactggtgcttg</td>
<td>3000</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 2
<211> LENGTH: 362
<212> TYPE: PRT
<213> ORGANISM: aspen populus tremulous

<400> SEQUENCE: 2

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Ser Lys Ser Pro Glu Glu Glu His Pro Val Lys Ala Phe Gly Trp</td>
<td>1 5 10 15</td>
</tr>
<tr>
<td>Ala Ala Arg Asp Gln Ser Gly His Leu Ser Pro Phe Aen Phe Ser Arg</td>
<td>20 25 30</td>
</tr>
</tbody>
</table>
Arg Ala Thr Gly Glu Glu Asp Val Arg Phe Lys Val Leu Tyr Cys Gly
35  40  45
Ile Cys His Ser Asp Leu His Ser Ile Lys Asn Asp Trp Gly Phe Ser
50  55  60
Met Tyr Pro Leu Val Pro Gly His Glu Ile Val Gly Glu Val Thr Glu
65  70  75  80
Val Gly Ser Lys Val Lys Lys Val Aen Val Gly Asp Lys Val Gly Val
85  90  95
Gly Cys Leu Val Gly Ala Cys His Ser Cys Glu Ser Cys Ala Aen Asp
100 105 110
Leu Glu Asn Tyr Cys Pro Lys Met Ile Leu Thr Tyr Ala Ser Ile Tyr
115 120 125
His Asp Gly Thr Ile Thr Tyr Gly Tyr Ser Asp His Met Val Ala
130 135 140
Aen Glu Arg Tyr Ile Ile Arg Phe Pro Aen Met Pro Leu Aep Gly
145 150 155 160
Gly Ala Pro Leu Leu Cys Ala Gly Ile Thr Val Tyr Ser Pro Leu Lys
165 170 175
Tyr Phe Gly Leu Asp Glu Pro Gly Lys His Ile Gly Ile Val Gly Leu
180 185 190
Gly Gly Leu Gly His Val Aal Ala Val Lys Phe Ala Lys Ala Phe Gly Ser
195 200 205
Lys Val Thr Val Ile Ser Thr Ser Pro Ser Lys Gly Glu Ala Leu
210 215 220
Lys Ala Phe Gly Ala Asp Ser Phe Leu Val Ser Arg Asp Glu Glu Glu
225 230 235 240
Met Gln Ala Ala Ala Gly Thr Leu Asp Gly Ile Ile Asp Thr Val Ser
245 250 255
Ala Val His Pro Leu Leu Pro Leu Phe Gly Leu Leu Lys Ser His Gly
260 265 270
Lys Leu Ile Leu Val Gly Ala Pro Glu Lys Pro Leu Gly Leu Pro Ala
275 280 285
Phe Ser Leu Ile Ala Gly Arg Lys Ile Val Ala Gly Ser Gly Ile Gly
290 295 300
Gly Met Lys Glu Thr Glu Met Ile Asp Phe Ala Ala Lys His Asn
305 310 315 320
Ile Thr Ala Asp Ile Glu Val Ile Ser Thr Asp Tyr Leu Aen Thr Ala
325 330 335
Ile Glu Arg Leu Ala Lys Aen Val Arg Tyr Phe Val Ile Asp
340 345 350
Val Gly Asn Thr Leu Ala Ala Thr Lys Pro
355 360
<210> SEQ ID NO 3
<211> LENGTH: 1764
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremuloides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: CAld5H
<400> SEQUENCE: 3
taaagtcttg tgtatt tacttt actaaactgttt cctagagccc tcaaataaca
60
acccatcagt gattctttt ttcaaccttc acctagtttccc acccttacgttcttttctt tttctttgattcatacagc
120
cgtagttgtctt ctactt tctctgtt ctctotocgcc ototogcaga gattgocata
180
-continued

tccaccaggg cotaaggtt tgccacttgt aggttagcag caactgatgg acacaataac 240
tcaacgggt ttagcaacca tagtaaga atatgtgagg ctctttcata tgcgcctgag 300
gtaacatcat atgtgctc ttgcaagcgc gcaatgtcc tggaggtcag 360
ggcacagtt tacccaca gaacagccaa ctagcoca aytactttaa cctagtattg 420
tgcagatag gctttgccca acatcaggtc ttttggcga cagatgcgta agctotgctg 480
cattgcttt ttagcgcga aaaggttga tctctggaag tctggtgagg ataggggtaa 540
tccagggcct gagcatcagag acaaccctat agagcaagct ggtcagatgg 600
ttttcagtyg acaattgaca taacattcag agacgtttc gggcttaaaa ataagagca 660
ggatgccct atcaaggttt tggccagtt tttcgagcat tcaagacgtg 720
tgatctcatt cctggctcgg gctttgctga ccccaagcgg cctggctgcta atcgctgca 780
gggctgag catgtagttt aagttgctt aaccagatgg aatgtagttt atctgtagctt 840
aaacagatg aagcatcttg aagcatcttg aaaccagatg gctgctagca tctgtagctt 900
tatgctggaa gagaacctg aagctagatt atcgctagatt ttacaaaaag ctcgacgctt 960
tatcagattg acactcagag gcaatctttg gagctttagg ttgcttgagaa gggccaggtt 1020
ggcctgggga ataggtgtct cctagcgcga tctgctctag aacgacagct atcggagctt 1080
agttggcgat ggtgctgttt gctctgctga accgttagtt gatgtagggt aagtttagat 1140
tggccacact gcgctgtgat gctgtagttt gggcttgatt gtatgtagtt ttcgctgctt 1200
ccctctctc tctagattaa cttcgctgaa tctctgtcgt gcgctgtatt ttcgctgctt 1260
gcacaagtg ggcctgacta aagctgtatt ttagggga gaaccagatt catgggaga 1320
tcctgctgctg ttaagccttt gaaaccaggg gtctgctgatt ttaaggggaa 1380
tccttcgctc ttaacatttt ctggctgttt gctgctgatt gctatgccttg tggctgctt 1440
gtttttctc cttgactgga cgtgactgga cttgactgga ggtttaggtt atgtgctgta 1500
tctgctctcc gacagctgtt gaactgttgc atgactgagc ctcgctgctt atgactgagc 1560
gacatcgcga cttgatcttg tctgctgctg agctgctgct gttctacagt atgactgagc 1620
aaagttggcg atgagttgga gcctgactt gctgactgga cttgactgga ggtttaggtt 1680
agcgctctg gaaaccaggg caaattgtag ttttttttt tttttttttt tttttttttt atatataaa 1740
ttcgctgact gaaaaaacat atatcttttt gtttttttttt tttttttttt atatataaa 1764

<210> SEQ ID NO 4
<211> LENGTH: 514
<212> TYPE: PRT
<213> ORGANISM: aspen populus tremuloides

<400> SEQUENCE: 4

<table>
<thead>
<tr>
<th>Met</th>
<th>Asp</th>
<th>Ser</th>
<th>Leu</th>
<th>Val</th>
<th>Gln</th>
<th>Ser</th>
<th>Leu</th>
<th>Gln</th>
<th>Ala</th>
<th>Ser</th>
<th>Pro</th>
<th>Met</th>
<th>Ser</th>
<th>Leu</th>
<th>Phe</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ile</td>
<td>Val</td>
<td>Ile</td>
<td>Ser</td>
<td>Leu</td>
<td>Phe</td>
<td>Phe</td>
<td>Gly</td>
<td>Leu</td>
<td>Leu</td>
<td>Ser</td>
<td>Arg</td>
<td>Leu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Arg</td>
<td>Leu</td>
<td>Pro</td>
<td>Tyr</td>
<td>Pro</td>
<td>Pro</td>
<td>Gly</td>
<td>Pro</td>
<td>Lys</td>
<td>Gly</td>
<td>Leu</td>
<td>Pro</td>
<td>Leu</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Ser</td>
<td>Met</td>
<td>His</td>
<td>Met</td>
<td>Asp</td>
<td>Gln</td>
<td>Ile</td>
<td>Thr</td>
<td>His</td>
<td>Arg</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
<td>Gln</td>
<td>Tyr</td>
<td>Gly</td>
<td>Gly</td>
<td>Leu</td>
<td>Phe</td>
<td>His</td>
<td>Met</td>
<td>Arg</td>
<td>Met</td>
<td>Gly</td>
<td>Tyr</td>
<td>Leu</td>
</tr>
<tr>
<td>65</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Met</td>
<td>Val</td>
<td>Thr</td>
<td>Val</td>
<td>Ser</td>
<td>Pro</td>
<td>Glu</td>
<td>Ile</td>
<td>Ala</td>
<td>Arg</td>
<td>Gln</td>
<td>Val</td>
<td>Leu</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Val Gln Asp Asn Ile Phe Ser Asn Arg Pro Ala Asn Ile Ala Ile Ser 100 105 110  
Tyr Leu Thr Tyr Asp Arg Ala Asp Met Ala Phe Ala His Tyr Gly Pro 115 120 125  
Phe Trp Arg Gln Met Arg Lys Leu Cys Val Met Lys Leu Phe Ser Arg 130 135 140  
Lys Arg Ala Glu Ser Trp Glu Ser Val Arg Asp Glu Val Asp Ser Met 145 150 155 160  
Leu Lys Thr Val Glu Ala Asn Ile Gly Lys Pro Val Asn Leu Gly Glu 165 170 175  
Leu Ile Phe Thr Leu Thr Met Asn Ile Thr Tyr Arg Ala Ala Phe Gly 180 185 190  
Ala Lys Asn Glu Gly Gln Asp Gln Phe Ile Lys Ile Leu Gln Glu Phe 195 200 205  
Ser Leu Leu Phe Gly Ala Phe Asn Met Ser Asp Phe Ile Pro Trp Leu 210 215 220  
Gly Trp Ile Asp Pro Gln Gly Leu Ser Ala Arg Leu Val Lys Ala Arg 225 230 235 240  
Lys Ala Leu Asp Arg Phe Ile Asp Ser Ile Ile Asp Asp His Ile Gln 245 250 255  
Lys Arg Lys Gln Asn Lys Phe Ser Glu Asp Ala Glu Thr Asp Met Val 260 265 270  
Asp Asp Met Leu Ala Phe Tyr Gly Glu Gln Ala Arg Lys Val Asp Glu 275 280 285  
Ser Asp Leu Gln Lys Ala Ile Ser Leu Thr Lys Asp Asn Ile Lys 290 295 300  
Ala Ile Ile Met Asp Val Met Phe Gly Gln Thr Glu Thr Val Ala Ser 305 310 315 320  
Ala Ile Glu Trp Val Met Ala Glu Leu Met Lys Ser Pro Glu Asp Gln 325 330 335  
Lys Arg Val Gln Gln Glu Leu Ala Glu Val Val Gly Leu Glu Arg Arg 340 345 350  
Val Glu Glu Ser Asp Asp Lys Leu Thr Phe Leu Lys Cys Ala Leu 355 360 365  
Lys Glu Thr Leu Arg Met His Pro Ile Pro Leu Leu Leu His Glu 370 375 380  
Thr Ser Glu Asp Ala Glu Val Ala Gly Tyr Phe Ile Pro Lys Gln Thr 385 390 395 400  
Arg Val Met Ile Asn Ala Tyr Ala Ile Gly Arg Asp Lys Asn Ser Trp 405 410 415  
Glu Asp Pro Asp Ala Phe Lys Pro Ser Arg Phe Leu Lys Pro Gly Val 420 425 430  
Pro Asp Phe Lys Gly Asn His Phe Glu Phe Ile Pro Phe Gly Ser Gly 435 440 445  
Arg Arg Ser Cys Pro Gly Met Gln Leu Gly Leu Tyr Thr Leu Asp Leu 450 455 460  
Ala Val Ala His Leu Leu His Cys Phe Thr Trp Glu Leu Pro Aep Gly 465 470 475 480  
Met Lys Pro Ser Glu Leu Asp Met Thr Asp Met Phe Gly Leu Thr Ala 485 490 495  
Pro Arg Ala Thr Arg Leu Val Ala Val Pro Ser Lys Arg Val Leu Cys 500 505 510  
Pro Leu
<210> SEQ ID NO 5
<211> LENGTH: 1503
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremuloides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: AldOMT; GenBank accession number: X62096
<400> SEQUENCE: 5
tcactttcct ttccttaccc ttctttttcc ttgtttagacct cacttctgat 60
cacagtggct tacacaggtg asactcgat gactccacct caggttacag atgaagagcc 120
aacatcccc gcaagctgcc tctagtttta ccagttgatt aggacacag 180
cattgaacct gacatctgct aatacttgggt caatgcggct tctttgcaca 240
atttcggat gttcttcacc ttccttaccc asacactcgat gctccttctgca tcttagaccc 300
tatgctgcag ctctgggctg gtaactccat cctttaccaag atcttttctga 360
tgggaagctg gagagagctgt aggccctgctg cctttgcagaag atctttggga 420
gacagtggct tacacaggtg atctgtgttc atggagagac ggtgccttccct 480
cctgcttatt tggaaagatg caattttgca gttttttgttctgacccgctgtc 540
gacagtggct ttccttaccc gctccttctgca tccacttgcag aatcttcccag 600
aatgttcgac cactcttacca ttcacatgaa gaatattttg gactccacta caggttcttg 660
agatgctcag agatgctcag ctctttctccct caaccatgctg ggtgccttccct 720
catcttttct caccttcttct caccatgctg ggtgccttccct caccatgctg 780
cccacttctt caccatgctg ggtgccttccct caccatgctg ggtgccttccct 840
agatgctcag agatgctcag ctctttctccct caaccatgctg ggtgccttccct 900
cattttgctg gttttttgttctgacccgctgtc tccacttgcag aatcttcccag 960
ttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1020
ttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1080
ttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1140
ttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1200
ttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1260
ttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1320
nttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1380
nttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1440
nttccttaccc gctccttctgca tccacttgcag aatcttcccag tccacttgcag 1500
aaa 1503

<210> SEQ ID NO 6
<211> LENGTH: 365
<212> TYPE: PRT
<213> ORGANISM: aspen populus tremuloides
<400> SEQUENCE: 6
Met Gly Ser Thr Gly Glu Thr Gln Met Thr Pro Thr Gln Val Ser Asp 1 5 10 15
Glu Glu Ala His Leu Phe Ala Met Gln Leu Ala Ser Ala Ser Val Leu 20 25 30
Pro Met Ile Leu Lys Thr Ala Ile Glu Leu Asp Leu Leu Glu Ile Met 35 40 45
### Protein Sequence

<table>
<thead>
<tr>
<th>Position</th>
<th>Residue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-50</td>
<td>Ala Lys Ala Gly Pro Gly Ala Phe Leu Ser Thr Ser Glu Ile Ala Ser</td>
</tr>
<tr>
<td>55-60</td>
<td>His Leu Pro Thr Lys Asn Pro Asp Ala Pro Val Met Leu Asp Arg Ile</td>
</tr>
<tr>
<td>65-70</td>
<td>Leu Arg Leu Ala Ser Tyr Ser Ile Leu Thr Cys Ser Leu Lys Asp</td>
</tr>
<tr>
<td>75-80</td>
<td>Leu Pro Asp Gly Lys Val Glu Arg Leu Tyr Gly Leu Ala Pro Val Cys</td>
</tr>
<tr>
<td>85-90</td>
<td>Lys Phe Leu Thr Lys Asp Gly Val Ser Val Ser Pro Leu Cys</td>
</tr>
<tr>
<td>95-100</td>
<td>Leu Met Asn Gln Asp Lys Val Leu Met Glu Ser Trp Tyr Tyr Leu Lys</td>
</tr>
<tr>
<td>105-110</td>
<td>Asp Ala Ile Leu Asp Gly Ile Pro Phe Asn Lys Ala Tyr Gly Met</td>
</tr>
<tr>
<td>115-120</td>
<td>Thr Ala Phe Glu Tyr His Gly Thr Asp Pro Arg Phe Asn Lys Val Phe</td>
</tr>
<tr>
<td>125-130</td>
<td>Asn Lys Gly Met Ser Asp His Ser Thr Ile Thr Met Lys Lys Ile Leu</td>
</tr>
<tr>
<td>135-140</td>
<td>Glu Thr Tyr Lys Gly Phe Glu Gly Leu Thr Ser Leu Val Asp Val Gly</td>
</tr>
<tr>
<td>145-150</td>
<td>Gly Gly Thr Gly Ala Val Asn Thr Ile Val Ser Lys Tyr Pro Ser</td>
</tr>
<tr>
<td>155-160</td>
<td>Ile Lys Gly Ile Asn Phe Asp Leu Pro His Val Ile Glu Asp Ala Pro</td>
</tr>
<tr>
<td>165-170</td>
<td>Ser Tyr Pro Gly Val Glu His Val Gly Gly Asp Met Phe Val Ser Val</td>
</tr>
<tr>
<td>175-180</td>
<td>Pro Lys Ala Asp Ala Val Phe Met Lys Trp Ile Cys His Asp Trp Ser</td>
</tr>
<tr>
<td>185-190</td>
<td>Asp Ala Asp Cys Leu Lys Phe Leu Lys Asn Cys Tyr Asp Ala Leu Pro</td>
</tr>
<tr>
<td>195-200</td>
<td>Glu Asn Gly Lys Val Ile Leu Val Glu Cys Ile Leu Pro Val Ala Pro</td>
</tr>
<tr>
<td>205-210</td>
<td>Asp Thr Ser Leu Ala Thr Lys Gly Val Val His Val Asp Val Ile Met</td>
</tr>
<tr>
<td>215-220</td>
<td>Leu Ala His Asn Pro Gly Gly Lys Glu Arg Thr Glu Lys Glu Phe Glu</td>
</tr>
<tr>
<td>225-230</td>
<td>Gly Leu Ala Asp Gly Ala Gly Phe Glu Gly Phe Val Met Cys Cys</td>
</tr>
<tr>
<td>235-240</td>
<td>Ala Phe Asn Thr His Val Ile Gly Phe Arg Lys Lys Ala</td>
</tr>
</tbody>
</table>

### DNA Sequence

```
cctcccgtca gcgacgaac ctagacgcc cacacagctca cccatccctc ccctgctcc cctgctcc
```

### Misc Information
- **SEQ ID NO**: 7
- **LENGTH**: 1915
- **TYPE**: DNA
- **ORGANISM**: aspen populus tremuloides
- **FEATURE**: misc_feature
- **OTHER INFORMATION**: 4CL

```
<400> SEQUENCE: 7
```

```
cctcccgtca gcgacgaac ctagacgcc cacacagctca cccatccctc ccctgctcc cctgctcc
ttgagcgcgc atggagccga ctagacgcc cacacagctca cccatccctc ccctgctcc cctgctcc
agacactctac atcgacgagaa accttcaccc gcatctgact gcgagagact gcgagagact gcgagagact
acagacgcc tgaagagaggc gcgaagagagc gcgaagagagc gcgaagagagc gcgaagagagc gcgaagagagc
```

```
```
```
tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
300
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
360
tgtgcttca ctttcgacga ctccaccaac gcagacgaag atgaagatgc cttccggtga
cattgctt gatgctcttgc ttcgaatca aggctcaag ggtgccaaa
420
agyggctagt ttcagcataa aaggcataat aaccagttgt gtcaccaagag tagatggaga
480
castctcact ctgtatccttc acagtaaaga gttctgcttg tgggtgtcct tcaattgtgtc
tatctctgt ctgaatcct gcctgctctg aagccgctgc cttctgcgttc aattgcctgc
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
540
tgacgtgatc atgctcttcc taccaagttc acctgaattc(gt)gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
600
agaacctag ttcagcataa aaggcataat aaccagttgt gtcaccaagag tagatggaga
660
castctcact ctgtatccttc acagtaaaga gttctgcttg tgggtgtcct tcaattgtgtc
tatctctgt ctgaatcct gcctgctctg aagccgctgc cttctgcgttc aattgcctgc
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
720
agaacctag ttcagcataa aaggcataat aaccagttgt gtcaccaagag tagatggaga
780
castctcact ctgtatccttc acagtaaaga gttctgcttg tgggtgtcct tcaattgtgtc
tatctctgt ctgaatcct gcctgctctg aagccgctgc cttctgcgttc aattgcctgc
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
840
agaacctag ttcagcataa aaggcataat aaccagttgt gtcaccaagag tagatggaga
900
castctcact ctgtatccttc acagtaaaga gttctgcttg tgggtgtcct tcaattgtgtc
tatctctgt ctgaatcct gcctgctctg aagccgctgc cttctgcgttc aattgcctgc
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
960
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1020
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1080
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1140
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1200
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1260
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1320
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1380
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1440
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1500
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1560
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1620
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1680
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1740
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1800
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1860
tgacgtgatc atgctcttcc taccaagttcacctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
1915

<210> SEQ ID NO 8
<211> LENGTH: 1395
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremuloides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: CAD; GenBank accession number: AF217957
<400> SEQUENCE: 8

aaacctcactcctctctcct acctgcctgtg tttcaagaaa atgggtatccc tggaaacaga
gagaaaaatt gagtaggatggc cagcaacaga ctcaactggg catctcgctc cttccggtga
60
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
120
ttagtctcga gatagcggggt cagaagatgt ttcttatcag tttatatcag gttgaatttg
180
tcctacctctatatctctcctcttgcgttgcctc ttccactctcct actgtgctccc
ttgccatgaa ggtaggttggt aggtggtgtaa aggcatcata gatagtcagaa
ttagttatactcgatgtgatgct ttcgaccgct ttcagtatcgc
ttgacctgct ctgtatccttc acagtaaaga gttctgcttg tgggtgtcct tcaattgtgtc
tatctctgt ctgaatcct gcctgctctg aagccgctgc
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
240
tgcctgctctg aagccgctgc cttctgcgttc aattgcctgc
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
ttcaccaac tgttgcgcgct acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
300
tgacctgtc gatagcggggt cagaagatgt ttcttatcag tttatatcag gttgaatttg
360
<210> SEQ ID NO 9
<211> LENGTH: 357
<212> TYPE: PRT
<213> ORGANISM: aspen populus tremuloides

<400> SEQUENCE: 9
Met Gly Ser Leu Glu Thr Glu Arg Lys Ile Val Gly Trp Ala Ala Thr
1  5  10  15
Asp Ser Thr Gly His Leu Ala Pro Tyr Thr Tyr Ser Leu Arg Aep Thr
20  25  30
Gly Pro Glu Asp Val Leu Ile Lys Val Ile Ser Cys Gly Ile Cys His
35  40  45
Thr Aep Ile His Gln Ile Lys Asn Aep Leu Gly Met Ser His Tyr Pro
50  55  60
Met Val Pro Gly His Glu Val Val Gly Val Glu Val Gly Ser
65  70  75  80
Asp Val Thr Lys Phe Lys Ala Gly Asp Val Val Gly Val Gly Val Ile
85  90  95
Val Gly Ser Cys Lys Asn Cys His Pro Cys Lys Ser Glu Leu Glu Gln
100 105 110
Tyr Cys Asn Lys Ile Trp Ser Tyr Asn Aep Val Tyr Thr Aep Gly
115 120 125
Lys Pro Thr Glu Gly Phe Ala Glu Ser Met Val Val Aep Glu Lys
130 135 140
Phe Val Val Arg Ile Pro Asp Gly Met Ser Pro Glu Glu Ala Ala Pro
145 150 155 160
Leu Leu Cys Ala Gly Leu Thr Val Tyr Ser Pro Leu Lys His Phe Gly
165 170 175
Leu Lys Gln Ser Gly Leu Arg Gly Gly Ile Leu Gly Leu Gly Val
Gly His Met Gly Val Lys Ile Ala Lys Ala Met Gly His His Val Thr
195 200 205
Val Ile Ser Ser Ser Asp Lys Arg Glu Ala Met Glu His Leu
210 215 220
Gly Ala Asp Glu Tyr Leu Val Ser Ser Asp Val Glu Ser Met Glu Lys
225 230 235 240
Ala Ala Asp Glu Leu Asp Tyr Ile Ile Asp Thr Val Pro Val Val His
245 250 255
Pro Leu Glu Pro Tyr Leu Ser Leu Leu Lys Leu Asp Gly Lys Leu Ile
260 265 270
Leu Met Gly Val Ile Asn Thr Pro Leu Gln Phe Val Ser Pro Met Val
275 280 285
Met Leu Gly Arg Lys Ser Ile Thr Gly Ser Phe Ile Gly Ser Met Lys
290 295 300
Glu Thr Glu Met Leu Glu Phe Cys Lys Glu Lys Leu Ala Ser
305 310 315 320
Met Ile Glu Val Ile Lys Met Asp Tyr Ile Asn Thr Ala Phe Glu Arg
325 330 335
Leu Glu Lys Asn Asp Val Arg Asp Phe Val Val Asp Val Ala Gly
340 345 350
Ser Lys Leu Ile Pro
355
<210> SEQ ID NO 10
<211> LENGTH: 535
<212> TYPE: PRT
<213> ORGANISM: aspen populus tremuloides
<400> SEQUENCE: 10
Met Asn Pro Gln Glu Phe Ile Phe Arg Ser Lys Leu Pro Asp Ile Tyr
1 5 10 15
Ile Pro Lys Asn Leu Pro Leu His Ser Tyr Val Leu Glu Asn Leu Ser
20 25 30
Lys His Ser Ser Lys Pro Cys Leu Ile Asn Gly Ala Asn Gly Asp Val
35 40 45
Tyr Thr Tyr Ala Asp Val Glu Leu Thr Ala Arg Arg Val Ala Ser Gly
50 55 60
Leu Asn Lys Ile Gly Ile Gln Gin Gly Asp Val Ile Met Leu Phe Leu
65 70 75 80
Pro Ser Ser Pro Glu Phe Val Leu Ala Phe Leu Gly Ala Ser His Arg
85 90 95
Gly Ala Met Ile Thr Ala Ala Asn Pro Phe Ser Thr Pro Ala Glu Leu
100 105 110
Ala Lys His Ala Lys Ala Ser Arg Ala Lys Leu Leu Ile Thr Gin Ala
115 120 125
Cys Tyr Tyr Glu Lys Val Lys Asp Phe Ala Arg Glu Ser Asp Val Lys
130 135 140
Val Met Cys Val Asp Ser Ala Pro Asp Gly Ala Ser Leu Phe Arg Ala
145 150 155 160
His Thr Gin Ala Asp Glu Gin Val Pro Gin Val Asp Ile Ser Pro
165 170 175
Asp Asp Val Val Ala Leu Pro Tyr Ser Ser Gly Thr Thr Gin Leu Pro
180 185 190
Lys Gly Val Met Leu Thr His Lys Gin Leu Ile Thr Ser Val Ala Gin
195 200 205
Gln Val Asp Gly Asp Asn Pro Asn Leu Tyr Phe His Ser Glu Asp Val
210 215 220
Ile Leu Cys Val Leu Pro Met Phe His Ile Tyr Ala Leu Asn Ser Met
225 230 235 240
Met Leu Cys Gly Leu Arg Val Gly Ala Ser Ile Leu Ile Met Pro Lys
245 250 255
Phe Glu Ile Gly Ser Leu Leu Gly Leu Ile Glu Lys Tyr Lys Val Ser
260 265 270
Ile Ala Pro Val Val Pro Val Met Met Ala Ile Ala Lys Ser Pro
275 280 285
Asp Leu Asp Lys His Asp Leu Ser Leu Arg Met Ile Lys Ser Gly
290 295 300
Gly Ala Pro Leu Gly Lys Glu Leu Glu Asp Thr Val Arg Ala Lys Phe
305 310 315 320
Pro Gln Ala Arg Leu Gly Gln Gly Tyr Gly Met Thr Glu Ala Gly Pro
325 330 335
Val Leu Ala Met Cys Leu Ala Phe Ala Lys Glu Pro Phe Asp Ile Lys
340 345 350
Pro Gly Ala Cys Gly Thr Val Val Arg Asn Ala Glu Met Lys Ile Val
355 360 365
Asp Pro Glu Thr Gly Val Ser Leu Pro Arg Asn Gln Pro Gly Glu Ile
370 375 380
Cys Ile Arg Gly Asp Gln Ile Met Lys Gly Tyr Leu Asn Asp Pro Glu
385 390 395 400
Ala Thr Ser Arg Thr Ile Asp Glu Gly Trp Leu His Thr Gly Asp
405 410 415
Ile Gly Tyr Ile Asp Asp Asp Glu Leu Phe Ile Val Asp Arg Leu
420 425 430
Lys Glu Leu Ile Lys Tyr Lys Gly Phe Gln Val Ala Pro Thr Glu Leu
435 440 445
Glu Ala Leu Leu Ile Ala His Pro Glu Ile Ser Asp Ala Ala Val Val
450 455 460
Gly Leu Lys Asp Glu Asp Ala Gly Glu Val Pro Val Ala Phe Val Val
465 470 475 480
Lys Ser Glu Lys Ser Gln Ala Thr Glu Asp Glu Ile Lys Gln Tyr Ile
485 490 495
Ser Lys Gln Val Ile Phe Tyr Lys Arg Ile Lys Arg Val Phe Phe Ile
500 505 510
Glu Ala Ile Pro Lys Ala Pro Ser Gly Lys Ile Leu Arg Lys Asn Leu
515 520 525
Lys Glu Lys Leu Pro Gly Ile
530 535
<210> SEQ ID NO 11
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremuloides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: Pt4CLl promoter sense primer
<400> SEQUENCE: 11

caggaatgct ctgcactctg
What is claimed is:

1. A method of producing a transgenic plant comprising:
   (a) introducing into a plant cell at least one polynucleotide sequence encoding at least two of 4CL antisense RNA, aspen CALd5H, aspen AldOMT and SEQ ID NO:2; and (b) regenerating the transformed cell to produce a transgenic plant.

2. The method of claim 1, wherein the polynucleotide sequence comprises a sequence encoding 4CL antisense RNA and sequences encoding CAld5H, AldOMT, and SEQ ID NO:2.

3. The method of claim 2 wherein the transgenic plant comprises decreased lignin content, increased syringyl/guaiacyl (S/G) lignin ratio, or increased cellulose content compared to a non-transformed plant.

4. The method of claim 2 wherein the plant is an angiosperm or a gymnosperm.

5. The method of claim 1 wherein the lignin content of the transgenic plant is decreased relative to the lignin content of a non-transformed plant.

6. The method of claim 1 wherein the transgenic plant comprises an increased syringyl/guaiacyl (S/G) lignin ratio relative to a non-transformed plant.

7. The method of claim 1 wherein the transgenic plant comprises increased cellulose content relative to a non-transformed plant.

8. The method of claim 1 wherein the plant is an angiosperm.

9. The method of claim 8 wherein the angiosperm is a Populus tremuloides.

10. The method of claim 1 wherein the plant is a gymnosperm.

11. The method of claim 1, wherein the polynucleotide comprises 4CL in an antisense orientation; and wherein the transgenic plant comprises a characteristic selected from the group consisting of altered lignin content, increased syringyl/guaiacyl (S/G) lignin ratio, altered cellulose content and combinations thereof compared to a non-transformed control plant.

12. The method of claim 11 wherein the transgenic plant comprises decreased lignin content.

13. The method of claim 11 wherein the transgenic plant comprises increased syringyl/guaiacyl (S/G) lignin ratio.

14. The method of claim 11 wherein the transgenic plant comprises increased cellulose content.

15. The method of claim 11 wherein the plant is an angiosperm.

16. The method of claim 15 wherein the angiosperm is a Populus tremuloides.

17. The method of claim 11 wherein the plant is a gymnosperm.

18. The method of claim 1 wherein each polynucleotide coding sequence is operably linked to a promoter sequence functional in the plant and a termination sequence; and wherein the plant comprises a characteristic selected from the group consisting of altered lignin content, increased syringyl/guaiacyl (S/G) lignin ratio, altered cellulose content, altered agronomic traits, and combinations thereof compared to a control plant that is not transformed with the polynucleotide sequence.

19. The method of claim 18 wherein the promoter is tissue-specific.

20. The method of claim 18 wherein the plant is an angiosperm.

21. The method of claim 20 wherein the angiosperm is a Populus tremuloides.

22. The method of claim 18 wherein the plant is a gymnosperm.
23. The method of claim 18 wherein the plant comprises an altered agronomic trait selected from the group consisting of growth, wood quality, stress resistance, sterility, grain yield or nutritional value.

24. The method of claim 1, wherein the polynucleotide sequence encodes 4CL antisense RNA and CAldSH.

25. The method of claim 24, wherein the transgenic plant comprises decreased lignin content, increased syringyl/guaiacyl (S/G) lignin ratio, or increased cellulose content compared to a non-transformed plant.

26. The method of claim 24, wherein the CAldSH comprises SEQ ID NO:4.

27. The method of claim 1, wherein the CAldSH comprises SEQ ID NO:4.

28. The method of claim 27, wherein the polynucleotide sequence comprises SEQ ID NO:3.

29. The method of claim 1, wherein the AldOMT comprises SEQ ID NO:6.

30. The method of claim 29, wherein the polynucleotide sequence comprises SEQ ID NO:5.
It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Beginning at Column 1, line 11, delete the following paragraph:

"STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with United States government support awarded by the Energy Biosciences Program, United States Department of Energy, and the United States Department of Agriculture research grant numbers USDA 99-35103-7986, USDA 01-03749, and DOE DE-FG02-01ER15179. The United States government has certain rights in this invention."

Signed and Sealed this

Twelfth Day of December, 2006

[Signature]

JON W. DUDAS
Director of the United States Patent and Trademark Office