4-7-2009

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 Engineering Commons

Recommended Citation
http://digitalcommons.mtu.edu/patents/108

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

1 Claim, 29 Drawing Sheets
OTHER PUBLICATIONS


Walker et al., 1987, *PNAS USA* 84:6624.


Yang et al., 1990, *PNAS USA* 87:4144.


Ebert et al. 1987, *PNAS USA*, 84:5745-5749.


Sambrook et al., 1982.


EMBL Acc#X62096 Bugos et al., 1991; Alignment with SEQ ID No. 6.


* cited by examiner
Phenylalanine $\rightarrow$ Cinnamate $\rightarrow$ C4H $\rightarrow$ 4-Coumarate $\rightarrow$ C3H $\rightarrow$ Caffeate $\rightarrow$ 4CL $\rightarrow$ Caffeoyl-CoA $\rightarrow$ CCoAOMT $\rightarrow$ COA $\rightarrow$ Feruloyl-CoA $\rightarrow$ GCR $\rightarrow$ Coniferaldehyde $\rightarrow$ CAD $\rightarrow$ Coniferyl alcohol $\rightarrow$ Guaiacyl-syringyl lignin
FIG. 2A  SAD cDNA sequence (SEQ ID NO: 1).

```
1   TTTTTTTTTTTTTTTTTTTTTCTAGCC TTCTCTCTCG ACAGATATTT TCTATCTGAA 
51  GCAAGCACCA TGTCAGAGTC ACCAGAAGAA GAACGCCCTG TGAGGCCTTT 
101 CCGGTGGGCT GCTAGGATCA AATCTGTGCA TCTTTCTCCC TTCAACTTCT 
151 CGAGGAGGCTG AACTGGTGA GAGGATGGA GGTTCAAGGT GCCTGATGTC 
201 GGGATATGCC ATTCTGACCT TCAAGATATC AAGAATGACT GGGCTATTC 
251 CATGTACCCT TTGTTGCTCG TGGATGAAAT TTGGGGGTA ATGACAGAG 
301 TTGGGAGCAA CGTGGAAAAAG GTTAATGCTG GAGACAAAGT GGCGTGGGA 
351 TGCTTGTTTG GTGCATGTCA CTCTTGAG AGTTGTGACCA ATGATCTTA 
401 AAATTACTGT CAAAATGTA TCTGACATA CAGCTTCCAT TACCAGTACG 
451 GAACCATCAC TTACGGTGGC TACTCAGATC ACATGGTCGC TAACGAAGC 
501 TACATCATTC GATTCCTGCA ACAAATGCG CCAGGAGGGC AGACTGGTAC 
551 GGAGGCTTT GTGCTTGGTTG GTGCATGTCA CTTGACGGTG GCGCTCCTCT 
601 AAATTACTGT CAAAATGTA TCTGACATA CAGCTTCCAT TACCAGTACG 
651 GTTGGAGCCT TACATCATTC GATTCCTGCA ACAAATGCG CCAGGAGGGC 
701 TACATCATTC GATTCCTGCA ACAAATGCG CCAGGAGGGC AGACTGGTAC 
751 CACATCATTC GATTCCTGCA ACAAATGCG CCAGGAGGGC AGACTGGTAC 
801 TTAGATGCTAA TAGTTAGCTA GAGGATGGA GGTTCAAGGT GCCTGATGTC 
851 TTAGATGCTAA TAGTTAGCTA GAGGATGGA GGTTCAAGGT GCCTGATGTC 
901 AAATTACTGT CAAAATGTA TCTGACATA CAGCTTCCAT TACCAGTACG 
951 GTTGGAGCCT TACATCATTC GATTCCTGCA ACAAATGCG CCAGGAGGGC 
1001 TTTTGAGCCT ATACATATTC TCTGATCTGA TATCGAAGTT ATTTCAACGG 
1051 AAATTACTGT CAAAATGTA TCTGACATA CAGCTTCCAT TACCAGTACG 
1101 GAAGATCCCT TACTCTGCTG GAAGATCCCT TACTCTGCTG GAAGATCCCT 
1151 ATGATCGCTA ATACATATTC TCTGATCTGA TATCGAAGTT ATTTCAACGG 
1201 ATTTATCTAA TACTCTGCTG GAAGATCCCT TACTCTGCTG GAAGATCCCT 
1251 GTTGGAGCCT TACATCATTC GATTCCTGCA ACAAATGCG CCAGGAGGGC 
1301 TTAGATGCTAA TAGTTAGCTA GAGGATGGA GGTTCAAGGT GCCTGATGTC 
1351 GTTGGAGCCT TACATCATTC GATTCCTGCA ACAAATGCG CCAGGAGGGC 
1401 ATGATCGCTA ATACATATTC TCTGATCTGA TATCGAAGTT ATTTCAACGG 
```
FIG. 2B  SAD protein sequence (SEQ ID NO: 2).

1  MSKSPEEHP VKAFFGWADQ QSGHLSPFNFSR RATGEEDVRFKVLYGIC
51  HSDLHSIKNDWGFSMYPLVPCHEIVGEVTEVSKVVKVNGDKVGVGCLV
101  GACHSCESCA NDLENYCPKMI LTYSIYHDMGTITYGYSDDMVANERYII
151  RFPDNMPLDG GAPLLCAGITVYSPLKYGLEDGKHHIGVGLGGLGHVAV
201  KFAKAFGSKVTVISTSPSKKEAALNFGADSFLVSRDLQEMMQAAAGTLDG
251  IIDTVSAYLPLFGLLKSHGKLILVGAP EKPLELFAFLIMGRKIVAG
301  SGIGGMKETQEMIDFAAKHNITADIEVISTDYLNTAIERLAKNDVRYRFV
351  I DVGNLTAAT KP*
FIG. 3A  Aspen (P. tremuloides) PtCAld5H cDNA sequence (SEQ ID NO: 3).

1 TAAAGTCTTG TGGATTACAC AAAATACAGA CTGAAAACAT CCATAGGCAC
51 CAACACATAA ACCATCCATG GATTCTCTTG TCCAATCTTT GCAAGCTTCA
101 CCCATGTCTC TCTTCTGTAT CGTTATCTCT TCACTCTTTCT TCTTCCGTCT
151 CCTCTCTCCG CTTCGCCGAA GATTGCCATA TCCACCAGGG CCTAAAGGGT
201 TGCCACTTGT AGGTAGCAGT CACATGATGG ACCAAATAAC TCACCGTTGG
251 TTAGCTAAAC TAGCTAAGCA ATATGTTGAG CTCTTTCATA TGCCCATGGA
301 GTACTTGCA ATGGTCACTG TTTCATCTCC TGAAATAGCT CGCCAAGTTT
351 TGCAAGGTCCA GGCAAACATT TTCTCCAACA GACCAGCCAA CATAGCCATA
401 AGTTACTTAA CCTATGATCG TGCAGATATG GCCTTTGCCC ACTACGTTCC
451 TTTCCTGGCGA CAGATGGCTA AGCTCTCGGT CATGAAGCTT TTTAGCCCGA
501 AAAGGGCTGA ATCATGGAGG TCTGTGAGAG ATGAGGTGGA CTCTTT CATA
551 TTTTACGCTTG ACCATGAACA TCACTTACAG AGCAGCTTTTC GGGCTAAAAA
601 ATGAGGACA GGATGAGTCC ATCAAGATTT TGCAGGAGTT CTCTAAGGTT
651 TTGGAGCAT TCAAACATGC TGATTTCATT CCTGGCTGGA GCTGGATTGA
701 CCCCCAGGGA CTACGCGCTA GACTTGTCAA GGCTCGCAAG GCTCTTGATA
751 GATTCATCGA CTCTATCATC GATGATCATA TCCAGAAAAG AAAACAGAAT
801 AGTTTCTCTG AAGATGCTCA AAGCGATATG GTCGATGACA TGCTAGCCTT
851 TTATGGTGAA GAAGCAAGGA AAGTAGATGA ATCAGATGA TTACAAAAAG
901 CCATCAGCCT TACTAAAGAC AACATCAAAG CCATAATCAT GGATGTGAG
951 TTTGGTGGGA CAGAGACGGT GGCGTCGGCA ATAGAGTGGG TCATGGCGGA
1001 GCTAATGAAG AGTCCAGAGG ATCAAAAAAG AGTCCAGCGA GAGCTCGCAG
1051 AGGTGGTGGG TTTAGACGCG CCAGTGGGAG AAAGTGATAT TGACAAACTT
1101 ACGTTCCTGA ATAGGCGGCT CAAAAGAAACC TTAAAGGATGC ACCCAACCAAT
1151 CCCACTTCTC TTACATGAAA TTCTTGAGGA TGCTGGTTATT
FIG. 3A (Continued)

1251 TCATTCCAAA GCAACAAGG GTGATGATCA ATGCTTATGC TATTGAGGA
1301 GACAAGAATT CATGGGAAGA TCCCTGATGCT TTTAAGCCTT CAGGTTTTTT
1351 GAAACCAGGG GTGCCCTGAT TTTAAAGGGAA TCACCTTGAG TTTATTCCTT
1401 TCGGGTCTGG TCGGAGGGTCT TGCCCCGGTA TGCAGGGTTG GTTATACACA
1451 CTTGATTCTGG CTGTTGCTCA CTGCTTCA CTGTTTACAT TGTTTACAT GGGAAATTGCC
1501 TGATGGCATG AAACCGAGTG AACTTGACAT GACTGAGTATG TTTGACTCA
1551 CCGGCGCAAAG AGCAACTCGA CTCGTTGCCG TTCCGAGCAA GCGTGTGCTC
1601 TGTCCTCTCT AAGGAAGGGGA AAAAGGTAAG GGATGGAAAT GAATGGGATT
1651 CAGCGTCTCTCT TGGGATTCTA TACAGGATGG AGGCAATGGT GACAAAGGGT
1701 CAGAAGCTGAG GTTTTTTTTT TTATATATAT ATATATATAA TTGGGTAAAA
1751 AAAAAAAAAAAAA AAA
FIG. 3B  Aspen (P. tremuloides) PtCald5H protein sequence (SEQ ID NO: 4).

1  MDSLVQSLQA SPMSLFLIVI SSLFFFGLLS RLRRLPYPP GPKGLPLVGS
51  MMMDQITHR GLAKLAKQYG GLFHMRMGLY HMVTVSSPEI ARQVLQVQDN
101  IFSNRPANIA ISYLYDRAD MAFHYGFDM RQMRKLCVMK LFSRKRAESW
151  ESVRDENVSM LKTVENANIG KVNLGEFILFT LMNITYRAA FGAKNEEQQDE
201  FIKILQEFSDK LFGAFNMDVF IPWLRWIDDPQ GLSARLVKAK KALDFIDSII
251  IDDHIQKRKQ NKFSEDAETD MVDDMLAFYG EEARKVDESID DLQKAISLTK
301  DINAIIMDD MFGGTTESVASE EWEVMAELM KSPEDQKRVQ QELAEVVGLE
351  RRV EEESIDDK LTLKCALKE TLRMHPPPPL LLLHESETAE VAGYFIPKQT
401  RVMINAYAIG RDKNWSEDPD AFKFSRFLKPP GVPDFKGNHF EPIFFGSGRR
451  SCPGMLGLSY TDLAVAHL LHCWTWELPDG MKPSELDMDT MFGLTAPRAT
501  RLVAVPERSKRLCPL*
FIG. 4A  Aspen (P. tremuloides) PtAldOMT cDNA sequence

GenBank accession number: X62096 (SEQ ID NO: 5).

1  tcacttcctt tccttacacc tccttcaacc ttttgtttcc ttgtagaatt
51  caatctcctg caagatggtg tcaacaggtg aaactcagat gactcccaact
101  caggtatcag atgaagaggc acaaccttctt gccatgcaac tagccagttgc
151  ttcagtttct caaatgatcc tcaaaccacgc catgaaacct gaccttcttg
201  aatcatggtg taaagctggc cctggtgtct ttctgtccac atctgagata
251  gcttctcacc tccctaccaaa aacccctgat ggcctgtca tgttagacccg
301  tatcctgcgc ctctcggcta gcctactccat tcttcctgaag ttggaggttyc
351  atccttcctga tggaagttt gggagactgt atggctcctgc tcctgtttgt
401  aatctttctga ccaagaacga ggacggtgct cctgtcagcc ctctctgtct
451  catgaaaccg gacaaggtcc tcatggaaag ctggattttat ttgaaagatg
501  caatcctttga tggaggaatt ccacattata caagcattag gatgactgca
551  ttgtaatata tggcaagaga tcaaaacatt gcaaaacttc tcaacaaggg
601  aatgtctgac cacttcacca ttccatgaa gaagatcctt gagacccctca
651  aaggctttaga aggcctcagc tccctggtgg atgttggtgg ttggaactgga
701  gcgcgtctga acaaccatcg ctcttaaatc ccttcaacta agggcatatta
751  cttcgcacgt gcccaccgcac cagaggtgct cccatcctat ccggagaggg
801  agcatgtggtc tggccgactag tttggtagtg tggcaaacgc agatgccggtt
851  ttcagatgac gataatgcga ttggagccagc gacgcccact gctttaaatc
901  tttgaagata tggctagacg cgttgaccgg caacggcaag tgtgatacttg
951  ttgaggtgcat ttctcccggt gcctctgaca caagccttgc cacaagggga
1001  gcctctgcag tttgatgctt catcgctggcg cacaaccccc gttgggaaga
1051  gaggaccggag aaggaatttg agggccttgc taagggagct gcctttcaag
1101  gttttggaagt aatgtgcgtgt gcattcaaca cacatgctcat tgaattccgc
FIG. 4A (Continued)

1151  aagaaggcct  aagccccatg  tccaagctcc  aagttacctg  gggttttgca
1201  gacaacgttg  ctgctgtctc  tgcgtttgat  gtttctgatt  gtttttttttt
1251  atacgaggag  tagctatctc  ttatgaaca  tgtaaggata  agattgcgtt
1301  ttgtatgcct  gattttctca  aataacttca  ctgcctccct  caaaattcctt
1351  aatacatgtg  aaaaagatttc  ctattggcct  tctgcttcaa  acagtaaaga
1401  cttctgtaac  ggaaaagaaa  gcaattcatg  atgtatgtat  cttgcaagat
1451  tatgagtatt  gttctaaagca  ttaagtgtt  gttcaaaaaaaa  aaaaaaaaaaa
1501  aaa
FIG. 4B  Aspen (P. tremuloides) PtAldOMT protein sequence

GenBank accession number: X62096 (SEQ ID NO: 6).

```
  1  MGSTGETQMT PTQVSDEEAH LFAMQLASAS VLPMILKTAI ELDLLEIMAK
  51  AGPCAFLSTS EIASHLPTKN PDAPVMLDRI LRLLASYSIL TCSKDLPDG
 101  KVERLYGLAP VCKFLTKNED GVSVSPLCLM NQDKVLMESW YYLKDAILDG
 151  GIPFNKAYGM TAFEYHGTDP RFNKVFNKGGM SDHSTITM KK IETYKGFEG
 201  LTLVDVGGG TGAVVNTIVS KYPSTIKGINF DLPHVIDEAP SYPGVEHVGG
 251  DMFVSVPKAD AVFMWKICHD WSDAHCLKFL KNCYDALPEN GKVILVECIL
 301  PVAPDTSLAT KGVVHVDMIV LAHNPPGKER TEKEFEGLAK GAGFQQGEVVM
 351  CCAFNTHVIE FRKKA
```
FIG. 5A  4 CL polynucleotide DNA sequence (SEQ ID NO: 7).

Since the image contains a DNA sequence, it is not transcribed to natural text. The sequence is presented in the format of a DNA strand, which typically includes nucleotide bases (A, C, G, T). The sequence is numbered to indicate specific positions, which might correspond to genetic or functional aspects of the DNA sequence. The presence of such a sequence suggests that it could be related to genetic research, biotechnology, or molecular biology.
FIG. 5B 4CL Aspen (P. tremuloides) amino acid sequence (SEQ ID NO: 10).

```
Met Asn Pro Gin 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  40
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 Glu 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 Ala Ser Arg Ala Lys
110 115 120
Leu Leu Ile Thr Gin Ala Cys Tyr Tyr Glu 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 Glu Val Pro
155 160 165 170
Gln Val Asp Ile Ser Pro Ser Pro Asp Val Ala Leu Pro Tyr Ser Ser
175 180 185
Gly Thr Thr Gin Leu Val Pro Lys Gly Val Met Leu Thr His Lys Gly Leu
190 195 200
Ile Thr Ser Val Ala Gin Gin Val Asp Gly Asp Pro Asn Asn Leu 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 Leu Cys Gly Leu 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 Pro Val Val 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 Gly Ala Pro Leu Gly Lys Glu Leu Glu Asp
300 305 310
Thr Val Arg Ala Lys Phe Pro Gin Ala Arg Leu Gly Gin Gly Tyr Gly
315 320 325 330
Met Thr Gly Ala Gly Pro Val Leu Ala Met Cys Leu Ala Phe Ala Lys
335 340 345
Glu Pro Phe Asp Ile Lys Pro Glu Ala Cys Tyr 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 Gln Pro Gly Glu Ile Cys 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 Glu Gly
395 400 405 410
Trp Leu His Thr Gly Ile Gly Tyr Ile Asp Asp Asp Glu Leu
415 420 425
Phe Ile Val Asp Arg Leu Lys Val Leu Ile Lys Tyr Lys Phe Gin
430 435 440
Val Ala Pro Thr Glu Val Leu Ala Leu Ile Ala His Pro Glu Ile
445 450 455
Ser Asp Ala Ala Val Val Gly Leu Lys Asp Gin Arg Ala Gly Glu Val
460 465 470
Pro Val Ala Phe Val Val Lys Ser Lys Ser Gin Ala Thr Glu Asp
475 480 485 490
```
FIG. 5B (Continued)

Glu Ile Lys Gln Tyr Ile Ser Lys Gln 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 Glu Lys Leu Pro Gly Ile
525 530 535
FIG. 6A  Aspen (*P. tremuloides*) PtCAD protein sequence

GenBank accession number: AF217957 (SEQ ID NO: 9).

```
1  MGSLETERKI VGWAATDSTG HLAPYTYSLR DTGPEDVLIK VISCGICHTD
51  IHQIKNDLGM SHYPMVPGHE VVGEVVVEVGS DVTKFKAGDV VGVGVIVGSC
101  KNCHPCKSEL EQYCNKKIWS YNDVYTDGKP TGSGFAESMV VDKFVVRIP
151  DGMSPEAQAP LLCAGLTVYS PLKHFGLKQS GLRGGILGLG GVGHMGVKIA
201  KAMGHHVTVI SSSDKKREAA MEHLGADEYL VSSDVESMQK AADQLDYIID
251  TVPVVHLELP YLSLLKLDBGK LILMGVINTP LQFVSPMVML GRKSITGSFI
301  GSMKETEEML EFCHEKGLAS MIEVIKMDYI NTAFERLEKN DVRYRFVVDV
351  AGSKLIP*
```
FIG. 6B  Aspen (*P. tremuloides*) PtCAD cDNA sequence

GenBank accession number: AF217957 (SEQ ID NO: 8).

```
1  AAACTCCATC CTCCTCTCTT AGGCTCGTTG TTTCAAGAAA ATGGGTAGCC
51  TTGAAACAGA GAGAAAAATT GTAGGATGGG CAGCAACAGA CTCAACTGGG
101  CATCTGCTC CTTACACCTA TAGTCAGAGA GATACGGGTC CAGAAGATGT
151  TCTTATAAGG GTTATCAGCT GTGGAATTGT CCATACCAGT ATCCACCAAA
201  TCAAAAATGAG TCTGGCCATG TCACACTAAC CTATGTGCCC TGGCCATGAA
251  GTGTTTGGTTG AGGTTGTTGGA GGTGGAGTCA GATTGTGCAA AGTTCAAGGC
301  TTGAGATGTT GTTGTGTTGAG GTACATCGT TGGGAAGTCG AAGAATTGTC
351  ATCCATGCAA ATCAGAGCTT GAGCAAGACT GCAACAAGAA AATCTGGCTT
401  TACAATGATGT TCTACACTGA TGGAAACCC ACCAAAAAGG GCTTTGCTGA
451  ATCCATGTTT GTCAAGTCAA AGTTTTGTGGT GAATATTCC GTAAGGGATGT
501  CACCAGAACA AGCAGCGGCG CTGTTGTGCG CTGGATTGAC AGTTTACAGC
551  CCACTCAAAC ACTTTGGAGCT GAAACAGAGT GGGCTAAGAG GAGGGATTGT
601  AGGACTTGGGA GGAGTGGGGT ACATGGGGGT GAAGATGAGA AAGGCAATGG
651  GACACCATGT AACTGTGATT AGTTTCTCTG ACAAGAAGCG GGAGGAGGCT
701  ATGGGAAATC TTGGTTGCGA TGAATACCTG GTCAGCTCGG ATGTGGAAAG
751  CATGCAAAAA GCTGCTGATC TTTCTGGACT TATCATCGTG ATCTGCTGAC
801  TGTTTCACCC TCTCAGACTT TACTTTGGTC TATGGAACCT TGAAGGCAAG
851  CTGATTTGGA TGCTGCCATT TAATACACCT TGGCAGTTTG TTTGCGCAAT
901  GGTATCTGTT GGGGAAAGG TCGACCCCGG GAGTCTCSSA GGGGACGTGA
951  AGGAGACAGA GGAGATGTGT TCTGGTGCA AAGAAAGGG ATGGGCTTCA
1001 ATGATTGGAAG TATCAAATCT GGATTATATC AACACAGCAT TGGAGGGCT
1051 TGAGAAATAT TATGTTTGGAT AGAGTTTGCTT GTGAGATGAT TGTGGTAGCA
1101 AGCTTTACCC CTGACGACCA ATACATTCC TATTCGAAAA AACGCGATAT
1151 ATACTGATAC CGTTTTGAGA CCTGAGCTTTA TTTGCGAGGT ATGTGTTTTG
```
FIG. 6B (Continued)

1201 TGGTTCAAAT GTGACAGTTT GTCTTTGCTT TTAAAATAAA GAAAAAGTTG
1251 AGTTGTTTTT TTATTTTCAT TAATGGGCAT GCGTTACCTT GTAATTGAAT
1301 GCGCTGCATC TGGTGATCTG TCCCATAAAC TAATCTCTTG TGGCAATGAA
1351 AGATGACGAA CTTTCTGAAA AAAAAAAAA AAAAAA AAAA
Fig. 7. pBKPt<sub>4</sub>CL Pt4CL1-a construct

Fig. 8. pBKPt<sub>4</sub>CL PtCAld5H-s construct
FIG. 9-1  The alignment of plant AldOMT protein sequences

1

1 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS
2 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS
3 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS
4 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTHVSDEEAH LFAMQLASAS
5 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS
6 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTHVSDEEAH LFAMQLASAS
7 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS
8 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTHVSDEEAH LFAMQLASAS
9 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTQVSDEEAH LFAMQLASAS
10 ~~~~~~~~~~ ~~~~~~~~~ MG STG..ETQMT PTHVSDEEAH LFAMQLASAS

11 MESTLAFNSG SNSMNQSFSS SAEFNSPVPE TIPKSEEDTF.VFATLITSAS

51

1 VLPMLKTAI ELDDLLEIMAK A...GPGAFI STSEIASHLTP TKNPDAVPVML
2 VLPMLKTAI ELDDLLEIMAK A...GPGAFI STSEIASHLTP TKNPDAVPVML
3 VLPMLKKAII ELDDLLEIMAK A...GPGVFL SPTDIASQLP TKNPDAVPVML
4 VLPMLKKAII ELDDLLEIMAK A...GPGVFL SPSHLASQLP TKNPEAVPVML
5 VLPMLKSAII ELDDLLEITAK A...GPAQAI SPIEIASQLP TTNPDAVPVML
6 VLPMLKKAII ELDDLLEIMAK A...GPAFL SPSGYAAQLP TQNPDAVPVML
7 VLPMLKKAII ELDDLLEIMAK A...GPGAFI SPSHIAPPQLP TQNPDAVPVML
8 VLPMLKKAII ELDDLLEIMAK A...GPGAFI STSEIASHLTP TKNPDAVPVML
9 VLPMLKSAII ELDDLLEITAK A...GPAHAI STSIAASKLP TKNPDAVPVML
10 VLPMLKSAII ELDDLLEITAK A...GPAHAI STSIAASKLP TKNPDAVPVML
11 VLPMLKSAII ELDDLLEITAK A...GPAHAI STSIAASKLP TKNPDAVPVML

101

1 DRILRLASY SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
2 DRILRLASY SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
3 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
4 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
5 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
6 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
7 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
8 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
9 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
10 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL
11 DRMLRLASS SILTCSLLQL PDGKVERLYG LAPVCKFLTVT NEDGVSAPVL

151

1 CLMNQDKVLM ES..WYLKDA IILGGIPFVNK AYGMTAFYEH GTDPFRNKFVF
2 CLMNQDKVLM ES..WYLKDA IILGGIPFVNK AYGMTAFYEH GTDPFRNKFVF
3 CLMNQDKVLM ES..WYLKDA IILGGIPFVNK AYGMTAFYEH GTDPFRNKFVF
4 CLMNQDKVLM ES..WYLKDA IILGGIPFVNK AYGMTAFYEH GTDPFRNKFVF
5 CLMNQDKVLM ES..WYLKDA IILGGIPFVNK AYGMTAFYEH GTDPFRNKFVF
6 CLMNQDKVLM ES..WYLKDA IILGGIPFVNK AYGMTAFYEH GTDPFRNKFVF

200
FIG. 9-2

1. CLMNQDKVLM ES.WYLLKDA ILDGIPFNK AYGMTAFEHY GTDFRPNKVF
2. SIMNQDKVLM ES.WYHLTEA VLEGGIPFNK AYGMTAFEHY GTDFRNTVF
3. CLMNQDKVLM ES.WYLLKDA ILDGIPFNK AYGMTAFEHY GTDFRPNKVF
4. LLMNQDKVLM ES.WYHLKDA VLDGGIPFNK AYGMTAFEHY GTDFRPNKVF
5. LLMNQDKVPM QSKRYHLKDA VLDGGIPFNK AYGMTDFEHY GTDFRPNKVF

201
1. NKGMSDHSTI TMKKILETYK GFEGLTSLVD VGGGTAVVN TIVSKYPSIK
2. NRMGMSDHSTI TMKKILETYK GFEGLTSLVD VGGGTAVVN TIVSKYPSIK
3. NKGMDHSTI TMKKILETYK GFEGLTSVVD VGGGTAVLN MVSKYPSIK
4. NRGMDHSTI TMKKILETYK GFEGLTSVVD VGGGTAVLN MVSKYPSIK
5. NKGMDHSTI TMKKILEYT GFEGLKSVVD VGGGTAVLN MVSKYPSIK
6. NRGMDHSTI TMKAEKLYT GFEGLKSLVD VGGGTAVLN MVSKYPSIK
7. NRGMDHSTI TMKAEKLYT GFEALPTVVD VGGGTAVLN MVSKYPSIK
8. NNGMSNHSTI TMKAEKLYT GFEGLKSLVD VGGGTAVLN MVSKYPSIK
9. NNGMSNHSTI TMKAEKLYT GFEGLTSLVD VGGGTAVLN MVSKYPSIK
10. NNGMSNHSTI TMKAEKLYT GFEGLTSLVD VGGGTAVLN MVSKYPSIK
11. NNGMSNHSTI TMKAEKLYT GFEGLTSLVD VGGGTAVLN MVSKYPSIK

300
1. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKADAFMKW ICHDWSDAHC
2. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKADAFMKW ICHDWSDAHC
3. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
4. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
5. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
6. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
7. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
8. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
9. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
10. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC
11. GNFDSLPHVI EDAPSYPGVE HVGGDMFSV PKGDAIFMKW ICHDWSDAHC

400
1. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
2. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
3. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
4. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
5. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
6. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
7. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
8. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
9. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
10. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
11. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP

500
1. LKFLKNCYDA LPENGKVLV ECILPVAPDT SLATKGVVHV DVIMLAAHNP
FIG. 9-3

10 LKFLKNCYSA LPANGKVIIL ECILPEAPDT SLATKNTVHV DIVMLAHNPG
11 LKFLENCYQA LPDNGKVIVA ECILPVVPDT SLATKSAVHI DVIMLAYNTG

motif III

351
1 GKERTEKEFE GLAKGAGFQG FEVMCCAFNT HVIEFRKKA
2 GKERTEKEFE GLAKGAGFQG FEVMCCAFNT HVIELRKN~
3 GKERTEQEFQ ALAKGAGFQG FNVACSAFNT YVIEFLKKN
4 GKERTEQEFQ ALAKGSFGFQG IRVCCDAFNT YVIEFLKKI
5 GKERTEQEFQ DLAKGAGFQG FKVHCNAFNT YIMEFLKKV
6 GKERTEQEFQ TLAKGAGFQG FOVMCCAFGT HVMEFLKTA
7 GKERTEKEFE ALAIGAGFKG FKVACAFNT YVMEFLKTA
8 GKERTEKEFE ALAKGAGFQG FRVALCAYNT WIIIEFLKKI
9 GKERTEKEFE ALAKSGFKG IKVVCDAFGV NLIELLKKL
10 GKERTEKEFE ALAKGAGFTG FARLVALTLT GSWNSTSN~
11 GKERTEKEFE ALAKGAGFQG FKVVCCAFNS WIMEFCCKTA

Plant AldOMTs from

1) Aspen, X62096 (SEQ ID NO: 6)
2) Poplar, M73431 (SEQ ID NO: 15)
3) Almond, X83217 (SEQ ID NO: 16)
4) Strawberry, AF220491 (SEQ ID NO: 17)
5) Alfalfa, M63853 (SEQ ID NO: 18)
6) Eucalyptus, X74814 (SEQ ID NO: 19)
7) Clarkia breweri, AF006009 (SEQ ID NO: 20)
8) Sweetgum, AF139533 (SEQ ID NO: 21)
9) Arabidopsis, U70424 (SEQ ID NO: 22)
10) Tobacco, X74452 (SEQ ID NO: 23)
11) Vitis vinifera, AF239740 (SEQ ID NO: 24)
FIG. 10-1 The alignment of full length plant CAD protein sequences available in the GenBank database

```
1 MGSLE.TEKT VTGYAARDDS GHLSPYTYNL RKKGPEDEVIV KVIYCGICHS
2 MGSLE.VEKT VTGYAARDDS GHLSPYTYNL RKKGPEDEVIV KVIYCGICHS
3 MGSLE.SESKT VTGYAARDDS GHLSPYTYNL RKKGPEDEVIV KVIYCGICHS
4 MGSLE.SERT VTGYAARDDS GHLSPYTYLT RKKGPEDEVV KVLYCGICHT
5 MGSLE.ASERK VGWAARDAT GHLPSYTYTL RNTGPEDVVV KVLYGCHT
6 MGSLE.ASERK VVGAARDAT GHLPSYTYLT RNTGPEDVVV KVLYGCHT
7 MGSLE.ASERK VGWAARDAT GLAPYTYTL RSTGPEDVVV KVLYGCHT
8 MGSLE.AERT TVGLAARDPS GILFTYTLT RNTGPDVYI KIHYCVGCH
9 MGSLE.AERT TVGLAARDPS GILFTYTLT RNTGPDVYI KIHYCVGCH
10 MGSLE.ERT TVGAARDPS GVLSPYTLT RNTGPEVYI KVLSCGCH
11 MGSLE.ERT TTGAARDPS GVLSPYTLT RNTGPEVYI KVLSCGCH
12 MGSLE.EKT TVGLAARDPS GVLSPYTTLT RNTGPEVYI KVLSCGCH
13 MGSLE.EK TVGLAARDPS GVLSPYTTLT RNTGPEVYQV KVLYCGCH
14 MGSLE.EK TVGLAARDPS GVLSPYTTLT RNTGPEVYI KVLSCGCH
15 MGSLE.EK TVGLAARDPS GVLSPYTTLT RNTGPEVYI KVLSCGCH
16 MGSLE.EK TVGLAARDPS GVLSPYTYTL RNTGPEVYI KVLSCGCH

51
1 DLVQVMNEMG MSYPVPMQPGH EVGVITIEIG SEVKKFKVGE HVGVGCIVTG
2 DLVQVMNEMG MSYPVPMQPGH EVGVITIEIG SEVKKFKVGE HVGVGCIVTG
3 DLVQVMNEMG MSYPVPMQPGH EVGVITIEIG SEVKKFKVGE HVGVGCIVTG
4 DLVQVMNEMG MSYPVPMQPGH EVGVITIEIG SEVKKFKVGE HVGVGCIVTG
5 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
6 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
7 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
8 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
9 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
10 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
11 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
12 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
13 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
14 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
15 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG
16 DIHQAKHNLG ASYMPVMPGH EVGGEVVEVG PEVAKYGVGD VVGVGITYG

100
1 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
2 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
3 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
4 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
5 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
6 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
7 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
8 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
9 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
10 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
11 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
12 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
13 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
14 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
15 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG
16 DIHQKNDLGN SNSYPMVPQH EVGGEVVEVG SEVTKYRDG RVTGGTVG

150
1 CRSCGCNQTS MEQYCSKRIW TYNVDNHGRT PTQGQFASSM VVQDFMVFRV
2 CRSCGCNQTS MEQYCSKRIW TYNVDNHGRT PTQGQFASSM VVQDFMVFRV
3 CRSCGCNQTS MEQYCSKRIW TYNVDNHGRT PTQGQFASSM VVQDFMVFRV
4 CRSCGCNQTS MEQYCSKRIW TYNVDNHGRT PTQGQFASSM VVQDFMVFRV
5 CRESCPCAN VEQCNKKIQ SYNDVFYDGR PTQGQFASM VVQDFMVVRK
6 CRESCPCAN VEQCNKKIQ SYNDVFYDGR PTQGQFASM VVQDFMVVRK
7 CRESCPCAN VEQCNKKIQ SYNDVFYDGR PTQGQFASM VVQDFMVVRK
8 CKSCRCDCS IEQYCNKKIQ SYNDVFYDGR ITQGQFAEST VVQDFMVVRK
9 CKSCRCDCS IEQYCNKKIQ SYNDVFYDGR ITQGQFAEST VVQDFMVVRK
```
FIG. 10-2

10 CRSCSPCNSD QEYCNKKIW NYNDVYTDGK PTQGGFAGEI VVGERFVVKI
11 CRSCSPCNSD QEYCNKKIW NYNDVYTDGK PTQGGFAGEI VVGERFVVKI
12 CRNCGPKRD IEQYCNKKIW NCDNYTDGK PTQGGFATM VVDDKQFVKI
13 CRNCGPKRE IEQYCNKKIW NCDNYTDGK PTQGGFAMS N VVDDQFVKI
14 CKNCHPCKSE LEQYCNKKIW SYNDVYTDGK PTQGGFAEILV VVDDQFVKVRI
15 CKNCHPCKSE IEQYCNKKIW SYNDVYTDGK PTQGGFAESM VVHQKFVKRI
16 CKTCRCKAD VEQYCNKKIW NYNDVYTDGK PTQGGFSGHM VVDDQFVKV

Zn2

1 PENLPLEQAA PLLCAGVTVF SPMKHFMTE .PGKKGILG LGGVGHMGVK
2 PENLPLEQAA PLLCAGVTVF SPMKHFMTE .PGKKGILG LGGVGHMGVK
3 PENLPLEQAA PLLCAGVTVF SPMKHFMTE .PGKKGILG LGGVGHMGVK
4 PENLPLEQAA PLLCAGVTVF SPMKHGFMT .PGKKGILG LGGVGHMGVK
5 PAGLAPQEA PLLCAGVTVF SPLKHFGLT TPGLRGGILG LGGVGHMGVK
6 PAGLAPQEA PLLCAGVTVF SPLKHFGLT TPGLRGGILG LGGVGHMGVK
7 PAGLAPQEA PLLCAGVTVF SPLKAFGLT TPGLRGAIGL LGGVGHMGVK
8 PEGLAPQVA PLLCAGVTVF SPLSHFGLK. TPGLRGGILG LGGVGHMGVK
9 PEGLAPQVA PLLCAGVTVF SPLSHFGLK. TPGLRGGILG LGGVGHMGVK
10 PDLLEQEAQ PLLCAGVTVF SPLVRFGLKQ .SGLRGGILG LGGVGHMGVK
11 PDLLEQEAQ PLLCAGVTVF SPLVRFGLKQ .SGLRGGILG LGGVGHMGVK
12 PEGMAPEQAA PLLCAGTVFY SPLNHFGFKQ .SGLRGGILG LGGVGHMGVK
13 PEGMAPEQAA PLLCAGTVFY SPNFHGFSQN .SGLRGGILG LGGVGHMGVK
14 PGMMSPEQAA PLLCAGTVFY SPLKFHLKQ .SGLRGGILG LGGVGHMGVK
15 PGMMSPEQAA PLLCAGTVFY SPLKFHLKQ .SGLRGGILG LGGVGHMGVK
16 PGMMSPEQAA PLLCAGTVFY SPLTHFGLKE ISGLRGGILG LGGVGHMGVK

NADP

200

1 IAKAFGLHVIT VISSSDKKE EAMEVLGADA YLVSKDTEKM MEAAESLDDYI
2 IAKAFGLHVIT VISSSDKKE EAMEVLGADA YLVSKDTEKM MEAAESLDDYI
3 IAKAFGLHVIT VISSSDKKE EAMEVLGADA YLVSKDTEKM MEAAESLDDYI
4 IAKAFGLHVIT VISSSDKKE EAMEVLGADA YLVSKDTEKM MEAAESLDDYI
5 VAAMGHHVIT VISSSDKRA EAMDHLGADA YLVSSDAAM GAPADSLDDYI
6 VAAMGHHVIT VISSSDKRA EAMDHLGADA YLVSSDAAM GAPADSLDDYI
7 VAAMGHHVIT VISSSDKRA EAMDHLGADA YLVSSDAAM GAPADSLDDYI
8 VAAMGHHVIT VISSSDKRA EAMDHLGADA YLVSSDAAM GAPADSLDDYI
9 VAAMGHHVIT VISSSDKRA EAMDHLGADA YLVSSDAAM GAPADSLDDYI
10 IAKAMHHVIT VISSSDKKRT EALEHLGADN YLVSSDVTGM QSAADSLDDYI
11 IAKAMHHVIT VISSSDKKRT EALEHLGADN YLVSSDVTGM QSAADSLDDYI
12 IAKAMHHVIT VISSHKKQR EALEHLGADN YLVSSDVTGM QSAADSLDDYI
13 IAKAMHHVIT VISSHKKQR EALEHLGADN YLVSSDVTGM QSAADSLDDYI
14 IAKAMHHVIT VISSDHKRR EAMEHLGADN YLVSSDVSM QKAADQSLDDYI
15 IAKAMHHVIT VISSDHKRR EAMEHLGADN YLVSSDVSM QKAADQSLDDYI
16 IAKAMHHVIT VISSSDKKE EAIMHLDGADN YLVSSDQTM QSAADSLDDYI
### FIG. 10-3

<table>
<thead>
<tr>
<th>251</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MDTIPVAHPL EPYLALLKTN GKLVMGLGVP EPLHFVTPLL ILGRRSIAGS</td>
</tr>
<tr>
<td>2</td>
<td>MDTIPVAHPL EPYLALLKTN GKLVMGLGVP EPLHFVTPLL ILGRRSIAGS</td>
</tr>
<tr>
<td>3</td>
<td>MDTIPVAHPL EPYLALLKTN GKLVMGLGVP EPLHFVTPLL ILGRRSIAGS</td>
</tr>
<tr>
<td>4</td>
<td>MDTIPVAHPL EPYLALLKTN GKLVMGLGVP EPLHFVTPLL ILGRRSIAGS</td>
</tr>
<tr>
<td>5</td>
<td>IDTVPVHHPD EPYLALLKLD GKLIVLGVIQ EPLSFVSAPMV MLGRKAITGS</td>
</tr>
<tr>
<td>6</td>
<td>IDTVPVHHPD EPYLALLKLD GKLIVLGVIQ EPLSFVSAPMV MLGRKAITGS</td>
</tr>
<tr>
<td>7</td>
<td>IDTVPVHHPD EPYLALLKLD GKLIVLGVIQ EPLSFVSAPMV MLGRKAITGS</td>
</tr>
<tr>
<td>8</td>
<td>IDTVPVHHPD EPYLALLKLD GKLIVLGVIQ EPLSFVSAPMV MLGRKAITGS</td>
</tr>
<tr>
<td>9</td>
<td>IDTVPVHHPD EPYLALLKLD GKLIVLGVIQ EPLSFVSAPMV MLGRKAITGS</td>
</tr>
<tr>
<td>10</td>
<td>FDTIPVVHPL EPYLALLKLD GKLILTGVIN TPLQVFVPD MVLRKAITGS</td>
</tr>
<tr>
<td>11</td>
<td>FDTIPVVHPL EPYLALLKLD GKLILTGVIN TPLQVFVPD MVLRKAITGS</td>
</tr>
<tr>
<td>12</td>
<td>FDTIPVVHPL EPYLALLKLD GKLILTGVIN TPLQVFVPD MVLRKAITGS</td>
</tr>
<tr>
<td>13</td>
<td>FDTIPVVHPL EPYLALLKLD GKLILTGVIN TPLQVFVPD MVLRKAITGS</td>
</tr>
<tr>
<td>14</td>
<td>FDTIPVVHPL EPYLALLKLD GKLILTGVIN TPLQVFVPD MVLRKAITGS</td>
</tr>
<tr>
<td>15</td>
<td>FDTIPVVHPL EPYLALLKLD GKLILTGVIN TPLQVFVPD MVLRKAITGS</td>
</tr>
<tr>
<td>16</td>
<td>FDTIPVVHPL EPYLALLKLD GKLILTGVIN TPLQVFVPD MVLRKAITGS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>301</th>
<th>350</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FIGMEETQE TLDFCAEKKV SSIMSVVGLD YINTAMERLE KNDVRYRFV</td>
</tr>
<tr>
<td>2</td>
<td>FIGMEETQE TLDFCAEKKV SSIMSVVGLD YINTAMERLE KNDVRYRFV</td>
</tr>
<tr>
<td>3</td>
<td>FIGMEETQE TLDFCAEKKV SSIMSVVGLD YINTAMERLE KNDVRYRFV</td>
</tr>
<tr>
<td>4</td>
<td>FIGMEETQE TLDFCAEKKV SSIMSVVGLD YINTAMERLE KNDVRYRFV</td>
</tr>
<tr>
<td>5</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>6</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>7</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>8</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>9</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>10</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>11</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>12</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>13</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>14</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>15</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
<tr>
<td>16</td>
<td>FIGSIDETAE VLQFCVVDGKL TSTQIEVVKMG YVEALERLE KNDVRYRFV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>351</th>
<th>370</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>2</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>3</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>4</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>5</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>6</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>7</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>8</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>9</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
<tr>
<td>10</td>
<td>DVAGSKLND* ~~~~~~~~~~~</td>
</tr>
</tbody>
</table>
FIG. 10-4

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

Full length plant CADs from

1) Radiata pine, U62394 (SEQ ID NO: 25)
2) Loblolly pine, Z37992 (SEQ ID NO: 26)
3) Loblolly pine, Z37991 (SEQ ID NO: 27)
4) Norway spruce, X72675 (SEQ ID NO: 28)
5) Maize, aj005702 (SEQ ID NO: 29)
6) Maize, Y13733 (SEQ ID NO: 30)
7) Sugarcane, AJ231135 (SEQ ID NO: 31)
8) Lucerne, AF083332 (SEQ ID NO: 32)
9) Lucerne, Z19573 (SEQ ID NO: 33)
10) Eucalyptus, AF038561 (SEQ ID NO: 34)
11) Eucalyptus, X65631 (SEQ ID NO: 35)
12) Tobacco, X62343 (SEQ ID NO: 36)
13) Tobacco, X62344 (SEQ ID NO: 37)
14) Aspen, AF217957 (SEQ ID NO: 38)
15) Cottonwood, Z19568 (SEQ ID NO: 39)
16) Udo, D1.3991 (SEQ ID NO: 39)
FIG. 11-1 The alignment of full length plant CAld5H protein sequences

1) Aspen (SEQ ID NO: 4); 2) Poplar, AJ010324 (SEQ ID NO: 40); 3) Sweetgum, AF139532 (SEQ ID NO: 41); 4) Arabidopsis, U38416 (SEQ ID NO: 42). *, Heme-binding signature

1
1 -MDSLVQSLQ AS..PMSLFL IVISSLFPFG LLSSLRRRLP YPPGPKGLPL
2 -MDSLLPLSLQ TL..PMSFFL ITISSLFFFL LGSLRRRRSP YPPGPKGLPL
3 MDSSLHEALQ PL..PMTLFF I.PFLLLLLG LVSRLQRLPL YPPGPKLFLV
4 MESSISQTLK KLSDDPTTSLV IVVSLPTIFIS FITR.RRRRPY YPPGPRGWPI

51
1 VGSNSMDQO THRLGLAKA QYQGLFHMNM GYLSMVTVSS PEIARQVLQV
2 IGSMHMLDQL TDRGLAKA QYQGLFHMNM GYLSMVAVSS PEVARQVLQV
3 IGNMLMMDQL THRLGLAKA KYQGLPHLMK GFLYMVAVST PDARQVLQV
4 IGNMLMMDQL THRLGANLAK KYQGLCHLMK GFLYMYAVSS PEVARQVLQV

101
1 QDNIFSNRPA NIAISLYLTYD RADMAFAYHG PFWFRQMRKLC VMKLFSRKRRA
2 QDNIFSNRPA NIAISLYLTYD RADMAFAYHG PFWFRQMRKLC VMKLFSRKRRA
3 QDNIFSNRPA TIAISLYLTYD RADMAFAYHG PFWFRQMRKLC VMKLFSRKRRA
4 QDVFNSRPA TIAISLYLTYD RADMAFAYHG PFWFRQMRKVC VMKVFSRKRRA

151
1 ESWSVRDEVE DSMVLKTEAN IGKPVNLGEL IFITLMNITY RAAPGGA.KNE
2 ESWSVRDEVE DSMVLKTEAN IGKPVNLGEL IFITLMNITY RAAPGGA.KNE
3 ESWSVRDEVE DSATRVSAN IGSTVNLGEL VFALTJKNIT RAAPGTLISHE
4 ESWSVRDEVE DSMVRVSVSN GVKPINVGQEQ IFALTRNITY RAAPGSACEK

201
1 GQDEFIKILQ BFSKLPAGAFN MSDFIPWLGW IDPQGQSARL VKARKALDRF
2 GQDEFIKILQ BFSKLPAGAFN ISDFIPWLGW IDPQGQTARL VKARKDLKF
3 DQDEFVAILQ BFEQLPAGAFN IADFIPWLKW V.PQGINVRL NKARGALDF
4 GQDEFIRILQ BFSKLPAGAFN VADFIPFYGW IDPQGJNKRKL VKARNDLDF

251
1 IDSIIIDDIQ KRKQNFSDQ ..AEIDMVE DMLAYFGEBA RKVGESDDLQ
2 IDHIIIDDIQ KRKQNFSEEE ..AEIDMVE DMLTNYEET .KVNSDDLQ
3 IDKIIDDIQ KGSGN ..SEE ..VDIDMVE DLLAYFGEBA .KVSSEDDLQ
4 IDDDIDEMHK KKENQNAVDD GDDVVTMDVD DLLAYFGEBA KLVSETALDQ

301
1 KAILSTKDNI KALIMDVDFG GTELVISAIE WVMARLKSP KEDQKRQVQEL
2 NAIKILRDNI KALIMDVDFG GTELVISAIE WAMARLLKSP EDIKRQVQEL
3 NSIKLTKDNI KA.IDMVDFG GTELVISAIE WAMTELKSP EDELQVQEL
4 NSIKLTRDNI KAIIMDVDFG GTELVISAIE WALTBLRSP EDELQVQEL
FIG. 11-2 (Continued)

<table>
<thead>
<tr>
<th>351</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AEVVLERRV EBSDIKLTF LKCALKETLR MHPIPPLLH ETSBDAEVAG</td>
</tr>
<tr>
<td>2</td>
<td>ADVVLERRV EBSDFKLTF FKCTLKETLR LHPPPPLLH ETSBDAEVAG</td>
</tr>
<tr>
<td>3</td>
<td>AVVGLDRRV EKDFEKLY LKCVLKEVLR LHPPPPLLH ETAEBAVEVG</td>
</tr>
<tr>
<td>4</td>
<td>AEVVLDRRRV EBSDIEKLTY LKCTLKETLR MHPIPPLLH ETAEDTSIGD</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>401</th>
<th>450</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>YFIPKQTRVM INAYAIGRDK NSWEDPDFAFK PSRFLKPGVP DFKGNHFEPI</td>
</tr>
<tr>
<td>2</td>
<td>YYVKKKTRVM INAYAIGRDK NSWEDPDSDK PFSFLEPGVP DFKGNHFEPI</td>
</tr>
<tr>
<td>3</td>
<td>YYIPAKSRVM INACAIIGRDK NSWADDPTFR PSRFLKDGVP DFKGNFEPI</td>
</tr>
<tr>
<td>4</td>
<td>FFIPKKSQVM INAFAIIGRD PTSWDPTFTR PSRFLPQVP DFKGNFEPI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>451</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PFGSGRRSCP GMQGLYTLQ LAVAHLLHCF TWELPDGMKP SELEMDTDMFG</td>
</tr>
<tr>
<td>2</td>
<td>PFGSGRRSCP GMQGLYLALD LAVAHLLHCF TWELPDGMKP SELEMDTDMFG</td>
</tr>
<tr>
<td>3</td>
<td>PFGSGRRSCP GMQGLYALE TTVAHLLHCF TWELPDGMKP SELEMDVDFG</td>
</tr>
<tr>
<td>4</td>
<td>PFGSGRRSCP GMQGLYLAD LAVAHLLHCF TWKLPDGMKP SELEMDVDFG</td>
</tr>
</tbody>
</table>

********** *

<table>
<thead>
<tr>
<th>501</th>
<th>523</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LTPRATRLV AVPSKRLCP L*</td>
</tr>
<tr>
<td>2</td>
<td>LTPRATRLV AVPRKRVVCP L--</td>
</tr>
<tr>
<td>3</td>
<td>LTPRAIRLT AVSPRLLCP LY*</td>
</tr>
<tr>
<td>4</td>
<td>LTPKATRLF AVPTTRLICA L--</td>
</tr>
</tbody>
</table>
FIG. 12-1 PLANT 4CL AMINO ACID SEQUENCE ALIGNMENTS

(1) 1:--------------------------MNPDQ-ETIFRSKLPDYYPIKNLPLSYVLENLSKHSSKPCLI 41
(2) 1:--------------------------MDATNMSQEEIFRSKLPDYYPIKNLPLSYVLENLSKSXKCLI 46
(3) 1:--------------------------MGDCVARPELEIFRSKLPDYYPIKNLPLHLYCENISKVRDSCLI 46
(4) 1:--------------------------MMDCTIREFQGQDLIFRSKLPDYYPIKNLPLHLYCENISFERSRCLI 48
(5) 1:--------------------------M-NVQTFQINIVYKSKLPDYYPIKNLPLHRYEFQKWSHTSCKCLI 46
(6) 1:--------------------------MMDCTTMOQSDLIFRSKLPDYYPIKNLPLHLYCENISFERSRCLI 48
(7) 1:--------------------------MEKDTKYH-DIIFRSKLPDYYPIKNLPLHLYCENISFERSRCLI 45
(8) 1:KGSMEQ-Q-QFFES-AAPATEASPIIFRSKLDQIAITWEPLPHYERLPFVAAPRCLI 57
(9) 1:MTYLAQSLDPDQDQVDPSGHWVYFSDPDIPSNISHLNLYCQGQFQAPHRCLI 60
(10) 1:MAPQFQ-Q-QQVES-SHMMHNSDIFRSLPDIYIHSLHSGYIQHQPISFATRCLI 56
(11) 1:--------------------------K-Q--E-----HLYRSKLPDIEISDHLHLYCENISFERSRCLI 44
(12) 1:--------------------------M-A-N-GI-K-EMV---E----HLYRSKLPDIEISDHLHLYCENISFERSRCLI 45
(13) 1:------------------------------------------------------------------LI 2
(14) 1:------------------------------------------------------------------PCLI 4

(1) 42:NG-ANGDVTYAVELITALRVA--SGLMKGIQQQGVIDYMLFSSFSSTLAFGLASHVHGM 99
(2) 47:NG-ANGDVTYAVELITALRVA--SGLMKGIQQQGVIDYMLFSSFSSTLAFGLASHVHGM 104
(3) 47:NG-ATGEYSSQVEILRRSRLA-SGLMKGIQQQGVIDYMLFSSFSSTLAFGLASHVHGM 104
(4) 49:DG-ANDRTYTAELITSRKVA--VRLMKGIQQQGVIDYMLFSSFSSTLAFGLASHVHGM 106
(5) 46:NG-TGQIHYTARPKLARKVA--SGLMKGIEQGSDVLLPLSTYVAFSGLASFCGGM 103
(6) 47:NG-TGQIHYTARPKLARKVA--SGLMKGIEQGSDVLLPLSTYVAFSGLASFCGGM 103
(7) 46:NG-TGQIHYTARPKLARKVA--SGLMKGIEQGSDVLLPLSTYVAFSGLASFCGGM 103
(8) 58:DGATGVEVYAVELRLTARSRVA--SGLMKGIQQQGVIDYMLFSSFSSTLAFGLASHGMR 117
(9) 61:VQ-FAKSTTVADYLTHIISKIA--AGLSMKGLDKDVMWLMQSAEDVFLSFLAIRMIGAV 116
(10) 61:NGTGHVYTVDSWWNHQIAANFV--LQGQNDVQVWMLPPCSTVAFSGLASFCGGM 116
(11) 56:DG-ATGKTHCFAEVELISRKVA--AGLVNLGLQQGQVVMLPLCTEAFVFGASGVRGA 102
(12) 46:DG-ATDRTYCEILRVRKVA--AGLAKLGLQQQVMLLPPCSTEAFVFGASGVRGA 103

(1) 1:--------------------------A------K-------------------------------A 3
(14) 1:--------------------------A------K-------------------------------A 3
(15) 3:DG-STNKYFAEELISRVA--AGLSMKGIQQVHLLNNQICEAFVFVMGASLIGAV 60
(16) 5:DG-STNKYFAEELISRVA--AGLSMKGIQQVHLLNNQICEAFVFVMGASLIGAV 62

(1) 100:ITAANPSTTPAEHLAKHAKASSRAKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GA 155
(2) 105:VYAAPFSTTPAEHLAKHAKASSRAKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GC 160
(3) 105:STAAMPEFTTEAEVKQLKASKSRIKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GD 160
(4) 107:STAAMPFETTEAEVKQLKASKSRIKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GC 162
(5) 107:STAAMPFETTEAEVKQLKASKSRIKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GD 162
(6) 107:STAAMPFETTEAEVKQLKASKSRIKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GD 162
(7) 107:STAAMPFETTEAEVKQLKASKSRIKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GD 162
(8) 118:VTTANMSNEHISHIESQLAGAGATVITESMAADKL-PHSHGALTVV-LID-E-KRMD 171
(9) 113:ATTANPFTTPEALRAAQAIKSKILITQACYEKVK--DFARESDVKVCVIDS-APD-GD 176
(10) 113:STAAMPFETTEAEVKQLKASKSRIKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GD 176
(11) 103:VTTANPFKGEIAQAKLSKRAKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GD 158
(12) 104:VTTANPFKGEIAQAKLSKRAKLLITQACYEKVK--DFARESDVKVCVIDS-APD-GD 159
(13) 4:--------------------------A------K-------------------------------A 3
(14) 1:--------------------------A------K-------------------------------A 3
(15) 5:DG-STNKYFAEELISRVA--AGLSMKGIQQVHLLNNQICEAFVFVMGASLIGAV 60
(16) 5:DG-STNKYFAEELISRVA--AGLSMKGIQQVHLLNNQICEAFVFVMGASLIGAV 62

(1) 156:SLFRAGHTQADENVQFV-----DISPDGVVALPSSGTLPGVVLTHKLGLTSVA 207
(2) 156:SLFRAGHTQADENVQFV-----DISPDGVVALPSSGTLPGVVLTHKLGLTSVA 212
(3) 156:SLFRAGHTQADENVQFV-----DISPDGVVALPSSGTLPGVVLTHKLGLTSVA 212
(4) 156:SLFRAGHTQADENVQFV-----DISPDGVVALPSSGTLPGVVLTHKLGLTSVA 212
(5) 156:SLFRAGHTQADENVQFV-----DISPDGVVALPSSGTLPGVVLTHKLGLTSVA 212
FIG. 12-3

(10) 353: GYGMTEAGPVLMALGFAKEFPPVKSAGCCTVVRNAEMKIREDFTGDSLRSNQPGECIR 412
(11) 331: GYGMTEAGPVLMALAFKNFPSVFKSGCTTVRNQIKLDGTPHSGNQAEGECIR 390
(12) 332: GYGMTEAGPVLMALAFKNFPSVFKSGCTTVRNQIKLDGTPHSGNQAEGECIR 391
(13) 213: GYGMTEAGPV---------------------------------223
(14) 221: GYGMTEAGP --------------------------------- 230
(15) 289: GYGMTEAGPVLMALFAKEFPPVKSAGCCTVVRNAEMKIREDFTGDSLRSNQPGECIR 317
(16) 291: GYGMTEAGPVLMALFAKEFPPVKSAGCCTVVRNAEMKIREDFTGDSLRSNQPGECIR 318

------------

(1) 388: GDQIMKGYLNDPEATSRTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 447
(2) 393: GDQIMKGYLNDPEATSRTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 452
(3) 393: GDQIMKGYLNDPEATSRTTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 452
(4) 393: GDQIMKGYLNDPEATSRTTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 454
(5) 395: GDQIMKGYLNDPEATSRTTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 454
(6) 395: GDQIMKGYLNDPEATSRTTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 454
(7) 392: GDQIMKGYLNDPEATSRTTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 451
(8) 412: GDQIMKGYLNDPEATSRTTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 471
(9) 410: GDQIMKGYLNDPEATSRTTIDKEGWLHTGDIGYIDDDDELFIVDRLKELIKYKGFQVAPE 472
(10) 413: GPEIMKGYLNNPAATAETIDKEGWLHTGDIGLIDDDDELFIVDRLKELIKYKGFQVAPE 472
(11) 393: GPEIMKGYLNNPAATAETIDKEGWLHTGDIGLIDDDDELFIVDRLKELIKYKGFQVAPE 450
(12) 392: GPEIMKGYLNNPAATAETIDKEGWLHTGDIGLIDDDDELFIVDRLKELIKYKGFQVAPE 451

------------

(1) 448: LEALLIAHPEISDAAVVGLKDEDAGEVPVAFVKSEKSQATEDEIKQYISKQVIFYKRIK 507
(2) 453: LEALLIAHPEISDAAVVGMKDEDAGEVPVAFVKSEKSQATEDEIKQYISKQVIFYKRIK 512
(3) 453: LEALLIAHPEISDAAVVPMIDEKAGEVPVAFVVRTNGFTTETEREKQPSQKVVFYIKRIK 512
(4) 455: LEALLIAHPEISDAAVVPMIDEQAGEVPVAFVVRTNGFTTETEREKQPSQKVVFYIKRIK 514
(5) 453: LEALLIAHPEISDAAVVPMIDEKAGEVPVAFVVRTNGFTTETEREKQPSQKVVFYIKRIK 514
(6) 455: LEALLIAHPEISDAAVVPMIDEQAGEVPVAFVVRTNGFTTETEREKQPSQKVVFYIKRIK 514
(7) 452: LEALLIAHPEISDAAVVPMIDEQAGEVPVAFVVRTNGFTTETEREKQPSQKVVFYIKRIK 511
(8) 472: LEALLIAHPEISDAAVVGLK----FGEIPVFAKTEGELSEEDVQVAKKEYIVYIKRIK 528
(9) 470: LEALLIAHPEISDAAVVGLK----FGEIPVFAKTEGELSEEDVQVAKKEYIVYIKRIK 529
(10) 473: LEALLIAHPEISDAAVVGLK----FGEIPVFAKTEGELSEEDVQVAKKEYIVYIKRIK 529
(11) 451: LEALLIAHPEISDAAVVGLK----FGEIPVFAKTEGELSEEDVQVAKKEYIVYIKRIK 532
(12) 451: LEALLIAHPEISDAAVVGLK----FGEIPVFAKTEGELSEEDVQVAKKEYIVYIKRIK 532

------------

(1) 508: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 535
(2) 513: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 540
(3) 513: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 544
(4) 513: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 545
(5) 513: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 543
(6) 513: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 545
(7) 513: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 545
(8) 529: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 563
(9) 530: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 562
(10) 533: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 561
(11) 509: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 536
(12) 510: RVYFVDAIPKSPGIKLRKDLRAKTEATQTP-- 537
FIG. 12-4

1: aspen AF041049 (SEQ ID NO: 10)
2: Hybrid populus AF283552 (SEQ ID NO: 43)
3: Parsley X13324 (SEQ ID NO: 44)
4: potato M62755 (SEQ ID NO: 45)
5: Rubus idaeus AF239687 (SEQ ID NO: 46)
6: solanum AF150686 (SEQ ID NO: 47)
7: Tobacco D43773 (SEQ ID NO: 48)
8: rice x52623 (SEQ ID NO: 49)
9: soybean x69955 (SEQ ID NO: 50)
10: Arabidopsis AF106084 (SEQ ID NO: 51)
11: FinusteadaU12012 (SEQ ID NO: 52)
12: Finus teada U12013 (SEQ ID NO: 53)
13: Larix AF144513 (SEQ ID NO: 54)
14: Pseudolarix AF144528 (SEQ ID NO: 55)
15: Pseudotsuga AF144511 (SEQ ID NO: 56)
16: Tsuga AF144526 (SEQ ID NO: 57)
METHODS FOR SIMULTANEOUS CONTROL OF LIGNIN CONTENT AND COMPOSITION, AND CELLULOSE CONTENT IN PLANTS

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 transformation 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 transformation unless appropriate antibiotic or herbicide selection is used.

Chen et al. (1998) recently reported the genetic transformation of rice with multiple genes by co-bombardment of several gene constructs into embryogenic suspension tissues. However, particle bombardment-mediated gene transfer into embryogenic tissues is highly species-dependent, and regeneration of whole plants from embryogenic cells cannot be achieved for a variety of plant species (Horsch et al., 1985).

In contrast, Agrobacterium-mediated gene transfer and whole plant regeneration through organogenesis is a simple process and a less species-dependent system than bombardment-mediated transformation and regeneration via embryogenesis. However, the introduction of more than one gene in a single 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 copending, 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 Sarkane, 1973; Chiang and Funakoka, 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 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 indicated 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 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 pulp production. 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 this purpose, increased lignin content and/or S/G lignin ratio are desired to provide extra sturdiness to plants in order to prevent the loss of socially and economically important food crops due to dislodging 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 (CALD5H) (Osakabe et al., 1999) and S-adenosyl-L-methionine (SAM)-dependent 5-hydroxycinnamaldehyde 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 signature sequence motifs and the high sequence homology of these proteins to PtAldOMT attest to their function as an AldOMT specific for converting 5-hydroxycinnamaldehyde into sinapaldehyde (Li et al., 2000). J. Biol. Chem., 275:6537-6545). This AldOMT, like CALD5H, 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 NADPH 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 PtCAD attest to these CAD 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, C5H1 is the sole enzyme catalyzing specifically the conversion of coniferaldehyde into 5-hydroxycinnamaldehyde. 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 (PF-GXGXXCXGX, SEQ ID NO: 58) that is conserved in all P450 proteins (Nelson et al. 1996, Pharmacogenetics, 6:1-41). These signature sequence characteristics and high sequence homology of these proteins to PtCAD1H their function as a 5-hydroxylase that is specific for converting coniferaldehyde into 5-hydroxycinnamaldehyde (Osakabe et al., 1999, Proc. Natl. Acad. Sci. USA, 96:8955-8960).

The protein sequence alignment of plant 4CLs are 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 (SSGTTGLPKG, SEQ ID NO: 59) and a catalytic motif (GEICIRG, SEQ ID NO: 60). 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 angiosperm plants with increased 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 angiosperm 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 gymnosperm 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 CAD5H polynucleotide DNA sequence (SEQ ID NO: 3) and the CAD5H 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: 10) respectively FIGS. 5A and 5B.

FIG. 6 is the CAD5H polynucleotide DNA sequence (SEQ ID NO: 8) and the CAD amino acid sequence (SEQ ID NO: 9) respectively FIGS. 6A and 6B.

FIG. 7 is a map of the DNA construct, pBKPtr4CLl-a, positioned in a plant transformation binary vector.

FIG. 8 is a map of the DNA construct, pBKPtr4CLl- PCAld5H-s, positioned in a plant transformation binary vector.

FIG. 9 is the protein sequence alignment of AldOMTs for representative species of Plants, including Aspen X62096 (SEQ ID NO: 6).

FIG. 10 is the protein sequence alignment of CADs for representative species of plants, including Aspen AF217957 (SEQ ID NO: 9).

FIG. 11 is the protein sequence alignment of CAD5Hs for representative species of plants, including Aspen (SEQ ID NO: 4).

FIG. 12 is the protein sequence alignment of 4CLs for representative species of plants, including Aspen AF041049 (SEQ ID NO: 10).

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' non-coding) and following (3' non-coding) the coding region or coding sequence (See, 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 involve 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 reflect 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 iden-
vital 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.).

"Angiosperms" refers to plants that produce seeds encased in an ovary. A specific example of an angiosperm is *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.) [lobolly pine].

As used herein, the terms "isolated and/or purified" with reference to a nucleic acid molecule or polypeptide refer to isolation in vitro 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-hydroxylation or "Cald5H", 5-hydroxyconiferaldehyde O-methyltransferase or "AldOMT", and 4-coumarate-CoA ligase (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 Phenylpropanoid Biosynthetic Pathway

Reference is made to FIG. 1 which shows various 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 cinnamoyl-CoA (3) which in turn is converted to feruloyl-CoA (4) by cinnamoyl-CoA O-methyltransferase (CoAOMT); cinnamyl-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 (Cald5H); 5-hydroxyconiferaldehyde O-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, Cald5H, AldOMT, 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 the transgene (e.g., 4CL alone or together with SAD, Cald5H, and AldOMT), and combinations thereof, or 4CL and CAD alone, or together with Cald5H, SAD, 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 Cald5H, 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: Cald5H, 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 particular 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. Any promoter in the construct in accordance with the present invention suitably provides for expression of the linked DNA segment. The promoter can 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, GPA1.2, GPA1.3 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) may 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 (Hudspeth et al., 1989) or those associated with the R gene complex (Chandler 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. Suitably, tissue-specific promoters, such as those which could confine the expression of the transgenes 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., pBluescript II or pBluescript II). 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/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 the 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. For example, expression of some genes occurs in the leafy shoot system and may be achieved by joining a DNA sequence encoding a transit or signal peptide sequence to the coding sequence of the 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., Agro-
bacterium-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) coculturing Agrobacterium in low-pFl 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-pFl 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., Altmorbe 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 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 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 radio-labelled 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 reflectometry 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

pBKPt.sub.4CL Pt4CLl-a: Aspen 4CLl xylem specific promoter (Pt4CLl. 1.1 kb, Genbank AF041051 was prepared and linked to aspen 4CLl cDNA (Pt4CLl. Genbank AF041049, SEQ ID NO: 10) which was orientated in the antisense direction. Then the cassette containing aspen 4CLl promoter and antisense aspen 4CLl cDNA was positioned in a plant transformation binary vector, as shown in FIG. 1. (pBKPt.sub.4CL Pt4CLl-a construct).

pBKPt4d PtCAld5H-s: From pBKPt4d Pt4CLl-a construct, the antisense Pt4CLl was replaced with PtCAld5H cDNA (SEQ ID NO: 3) in a sense orientation, yielding a pBKPt4d PtCAld5H-s transformation binary construct, as shown in FIG. 2.

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 genes from the phenylpropanoid pathway (i.e., 4CL, AEOMT, 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 pg of binary vector DNA was added to 200 nL of 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 nL 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, pBKPt4d Pt4CLl-a and pBKPt4d PtCAld5H Agrobacterium clones were cultured in LB medium at 28° C. overnight separately. The Agrobacterium strains were subcultured individually by a 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 immersed 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 days. Then, the pre-cultured leaf discs were rinsed with sterile water several times to remove the Agrobacterium cells and washed in 1 mg/ml clorox 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 clorox for regeneration shoots. After shoots were grown to about 0.5 cm high, they were excised and planted to rooting media (WPM with 2 mg/l BAP, 0.2 mg/l NAA and 0.2 mg/l ticarcillin) for 2 to 3 subcultures (10 days/subculture), the callus grown on the leaf discs was excised and transferred onto shoot induction medium (WPM+TDZ: N-phenyl-N'-1,2,3-thiadiazol-5-yl-urea) containing 50 μg/ml kanamycin and 300 μg/ml clorox for selection of transformed cells.

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% hexadecyltrimethylammonium 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-isooctyl alcohol (24:1) was added and mixed gently. The two phases were separated by centrifugation at 10,000xg 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 mL of isopropanol (-20° C.) and kept at -20° C. for 20 minutes. Following the centrifugation, the precipitated DNA was dissolved in 2 mL TE buffer (10 mM Tris-HCl/0.1 mM EDTA, pH 8) and treated with 2 μg RNase A at 37° C. After removing excess RNA by a brief digestion, the DNA was dissolved in 2 mL TE buffer and kept at -20° C. for 2 days. Then, the precipitated 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 pHKPp1ac2, Pt4CL1-a construct, two specific primers were synthesized that amplify a 4CL cDNA fragment. Pt4CL1 promoter sense primer (5'GAGGAATCCGTGGTGCACCTCTG3') (SEQ ID NO:11) and Pt4CL1 promoter antisense primer (5'ACAGATATGCTCTGACACATG3') (SEQ ID NO:13) and PtCalD5H antisense primer (5'TTAGAGAGGAGCACAGAGCACAG3') (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 PtCalD5H or can vary between cDNA templates used) according to different gene checked) and 72° C. for 2 minute, for 40 cycles, with 5 minutes at 72° C. extension. The PCR products were electrophoresized 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, CalD5H 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 1**

<table>
<thead>
<tr>
<th>Protein sequence homology (%) of plant Coniferyl Aldehyde 5-hydroxylase (CalD5H) from:</th>
<th>1) Aspen, SEQ ID NO: 4; 2) Poplar, AJ010324, SEQ ID NO: 40; 3) Sweetgum, AF139532, SEQ ID NO: 41; 4) Arabidopsis, U38416, SEQ ID NO: 42 (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></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 2**

<table>
<thead>
<tr>
<th>Protein sequence homology (%) of plant AldOMTs from: 1) Aspen, X62956, SEQ ID NO: 6; 2) Poplar, M73431, SEQ ID NO: 15; 3) Almond, X83217, SEQ ID NO: 16; 4) Strawberry, AF220491, SEQ ID NO: 17; 5) Alfalfa, M83853, SEQ ID NO: 18; 6) Eucalyptus, T74814, SEQ ID NO: 19; 7) Clarkia breweri, AVJ06599, SEQ ID NO: 20</th>
<th>Tobacco, T74452, SEQ ID NO: 23; 11) Vitis vinifera, AF239740, SEQ ID NO: 24</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>92</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>89</td>
</tr>
<tr>
<td>5</td>
<td>88</td>
</tr>
<tr>
<td>6</td>
<td>88</td>
</tr>
<tr>
<td>7</td>
<td>84</td>
</tr>
</tbody>
</table>
TABLE 2-continued

| Protein sequence homology (%) of plant AldOMTs from 1) Aspen, X62096, SEQ ID NO: 6 | 2) Poplar, M73431, SEQ ID NO: 15 | 3) Almond, X83217, SEQ ID NO: 16 | 4) Strawberry, AF220491, SEQ ID NO: 17 | 5) Aspens, M63853, SEQ ID NO: 18 | 6) Eucalyptus, X74814, SEQ ID NO: 19 | 7) Celastrus brevifolius, AF065009, SEQ ID NO: 20 | 8) Swartgum, AF139533, SEQ ID NO: 21 | 9) Arabidopsis, U70424, SEQ ID NO: 22 | 10) Tobacco, X74452, SEQ ID NO: 23 | 11) Pits vinifera, AF239740, SEQ ID NO: 24 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 10 | 83 | 83 | 83 | 82 | 81 | 82 | 80 | 83 | 77 |
| 11 | 80 | 80 | 78 | 77 | 77 | 78 | 78 | 80 | 76 | 77 |

TABLE 3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>97</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>2</td>
<td>85</td>
<td>84</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>3</td>
<td>82</td>
<td>82</td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>80</td>
<td>80</td>
<td>81</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>81</td>
<td>82</td>
<td>80</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>81</td>
<td>82</td>
<td>80</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>81</td>
<td>80</td>
<td>81</td>
<td>80</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>79</td>
<td>79</td>
<td>80</td>
<td>79</td>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>79</td>
<td>80</td>
<td>79</td>
<td>78</td>
<td>79</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td>77</td>
<td>79</td>
<td>76</td>
<td>74</td>
<td>76</td>
<td>77</td>
<td>73</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>78</td>
<td>78</td>
<td>79</td>
<td>77</td>
<td>74</td>
<td>76</td>
<td>75</td>
<td>73</td>
<td>72</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>77</td>
<td>76</td>
<td>78</td>
<td>74</td>
<td>73</td>
<td>75</td>
<td>74</td>
<td>73</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>77</td>
<td>77</td>
<td>76</td>
<td>74</td>
<td>74</td>
<td>75</td>
<td>74</td>
<td>73</td>
<td>73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>70</td>
<td>69</td>
<td>70</td>
<td>69</td>
<td>69</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>68</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>69</td>
<td>69</td>
<td>68</td>
<td>69</td>
<td>69</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>99</td>
</tr>
<tr>
<td>16</td>
<td>69</td>
<td>69</td>
<td>70</td>
<td>69</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
<td>67</td>
</tr>
</tbody>
</table>

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 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 manipulating xylem-specific expression of 4CL (SEQ ID NO: 10) and CAd5H (SEQ ID NO: 4) in transgenic aspen. After DNA constructs were 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 Pt4CL1 cDNA (SEQ ID NO: 7) in the antisense orientation (Table 4). Group II plants (#32, 84, 93, and 94) harbored only sense PtcAd5H cDNA (SEQ ID NO: 3), whereas Group III plants (#71, 72, 74, and 141) contained both antisense Pt4CL1 and sense PtcAd5H 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 Pt4CL1 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 CAd5H with sense PtcAd5H transgene had pronounced increases in their S/G ratio, but their lignin and cellulose contents remained essentially unaffected. When engineered for the simultaneous suppression of 4CL gene and overexpression of CAd5H 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 50% 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.
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 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 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 and CAD genes 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.

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 and CAD genes in the 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 angiosperm plants, angiosperm 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 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 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 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 claims.

REFERENCES

Brasiliero et al., 1992, Transgenic Res. 1:133.
Chandler et al., 1989.
Danekar et al., 1987, Bio/Technology 5:587.
Ebert et al. 1987.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 14
<210> SBQ ID NO 1
<211> LENGTH: 1446
<220> FEATURE:
<223> NAME/KEY: misc_feature
<224> OTHER INFORMATION: SAD
<400> SEQUENCE: 1
ttttttttttt tttctcagct ctcctctcg acgatatcc tctatctgaa gcagcagcaca 60
<210> SEQ ID NO 2
<211> LENGTH: 362
<212> TYPE: PRT
<213> ORGANISM: aspen populus tremuloides

<400> SEQUENCE: 2

Met Ser Lys Ser Pro Glu Glu His Pro Val Lys Ala Phe Gly Trp 1  5  10 15
Ala Ala Arg Asp Gln Ser Gly His Leu Ser Pro Phe Asn Phe Ser Arg 20  25  30
Arg Ala Thr Gly 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 Val Asn Val Gly Asp Lys Val Gly Val 85  90  95
Gly Cys Leu Val Gly Ala Cys His Ser Cys Glu Ser Cys Ala Asn 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
Asn  Glu  Arg  Tyr  Ile  Ile  Arg  Phe  Pro  Asp  Asn  Met  Pro  Leu  Asp  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  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  Glu  Ala  Leu 
210  215  220
Lys  Asn  Phe  Gly  Ala  Asp  Ser  Phe  Leu  Val  Ser  Arg  Asp  Gln  Glu  Gln 
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  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  Gln  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  Ann  Thr  Ala 
325  330  335
Ile  Glu  Arg  Leu  Ala  Lys  Asm  Asp  Val  Arg  Tyr  Arg  Phe  Val  Ile  Asp 
340  345  350
Val  Gly  Ann  Thr  Leu  Ala  Thr  Lys  Pro 
355  360

<211> LENGTH: 1764
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremuloides
<215> NAME/KEY: misc_feature
<221> OTHER INFORMATION: CALD5H
<223> SEQUENCE: 3

`taaagtcttg tggattacac aaatacaga ctagaaacat ctaggacc cacacataaa` 60
`acccatctag gattctcttg tccaatcttt gcaagttcca cccatgtcct tttccttgat` 120
`agttcatctc ttaaccttcct tcttctgtc ctctcttcgc ctctggcga gattggcata` 180
`tccacaggg cotaagggg tgcacattg aggtacatt cagcatatgg accaaatgac` 240
`tcacgctggg ttagctaaac tagtcaagca atagtggagg ctctttctca tgcagcatgg` 300
`gtcttctgct atgttcagtt tttcattcgc tgaatagct cgccagttc tcgaggtccaa` 360
`ggcagcactt ttctccaaac gaccagccaa ctagacaaaa gttcactaatt ctttgagatc` 420
`tgccagcttt gccttcgcgc actacggtcc ttcctggcga cagatgctga agcttgctgt` 480
`ctagctgg ttagcggga aaggggtag atcagggag tttgtgagag atgaggtgga` 540
`ctcataagggtt cacagacatg ttgcaagtcttg ggcagatgtt ggtactttc` 600
`tctttccttt accatcaca tcaaacagc agcaggtctc ggggtctaaa atgaagggca` 660
ggatgagttc atcaagattt tgccagagtt ctctaaagct tttggagcat tcacagttgc
720

780

tgattcatt ccctcggtgg gctgtgattga ccoccaaggg ccoccaagcta gctttgcca

740

ggcctcagcag gccttcacag ctctatcatc gctgtatcag tcacagaaag

800

860

aatcagag aagttctctg aagatgctga ttatggagaa gaagcaagga aagtagatga

920

980

ggctcagcag gccttcacag ctctatcatc gctgtatcag tcacagaaag

1040

1100

tgacaaagctgtaagtgaaag atgccagagcag ccacagctga

1160

1220

tgaaacaagcag gccttcacag ctctatcatc gctgtatcag tcacagaaag

1280

1340

tcgtcagctgtaagtgaaag atgccagagcag ccacagctga

1400

1460

tggaaccagcag cccttcacag ctctatcatc gctgtatcag tcacagaaag

1520

1580

agccctcagcag cccttcacag ctctatcatc gctgtatcag tcacagaaag

1640

1700

ttggaaccagcag cccttcacag ctctatcatc gctgtatcag tcacagaaag

1764
-continued

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 Phe Ile Lys Ile Leu Gln Glu Phe
195 200 205
Ser Lys 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 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 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 Gly Thr Glu Thr Val Ala Ser
305 310 315 320
Ala Ile Glu Trp Val Met Ala Leu Met Lys Ser Pro Glu Asp Gln
325 330 335
Lys Arg Val Glu Gin Gin Glu Leu Ala Glu Val Val Gly Leu Glu Arg Arg
340 345 350
Val Glu Glu Ser Asp Ile 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 Gin 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 Asp 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
tcactctcc tcttcaacc ttcttaacc tttgtttcc ttgtgatt caacacgctatc 60
cacaattg gtcaccggct caacacgctatc gactcaact caggtcatcag atgaagaggc 120
aacaacctttt gcacatgcaac tagcagtgct ttcagtttta ccaatgatcc ccacaaacagc
33
ctgtgagttgcc cttcagttcct atcagttccttat ctcagttcct 180
ttcagttcct tatcagttcct 240
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 300
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 360
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 420
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 480
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 540
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 600
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 660
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 720
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 780
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 840
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 900
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 960
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 1020
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 1080
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 1140
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 1200
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 1260
gacagttgcc cttcagttcct atcagttccttat ctcagttcct 1320
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 1380
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 1440
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 1500
ttcagttcct gagagacctgttcctagttcctaatgatcc ccacaaacagc 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 Gly 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 Gln Leu Asp Leu Leu Glu Ile Met
35 40 45
Ala Lys Ala Gly Pro Gly Ala Phe Leu Ser Thr Ser Gln Ile Ala Ser
50 55 60
His Leu Pro Thr Lys Asn Pro Asp Ala Pro Val Met Leu Asp Arg Ile
65 70 75 80
Leu Arg Leu Ala Ser Tyr Ser Ile Leu Thr Cys Ser Leu Lys Asp
85 90 95
Leu Pro Asp Gly Lys Val Glu Arg Leu Tyr Gly Leu Ala Pro Val Cys
100 105 110
Lys Phe Leu Thr Lys Asn Glu Asp Gly Val Ser Val Ser Pro Leu Cys

-continued
Leu Met Asn Gin Asp Lys Val Leu Met Glu Ser Thr Tyr Tyr Leu Lys
Leu Met Asn Gin Asp Lys Val Leu Met Glu Ser Thr Tyr Tyr Leu Lys
Asp Ala Ile Leu Asp Gly Gly Ile Pro Phe Asn Lys Ala Tyr Gly Met
Thr Ala Phe Glu Tyr His Gly Thr Asp Pro Arg Phe Asn Lys Val Phe
Asn Lys Gly Met Ser Asp His Ser Thr Ile Thr Met Lys Ile Leu
Glu Thr Tyr Lys Gly Phe Glu Gly Leu Thr Ser Leu Val Asp Val Gly
Gly Gly Thr Gly Ala Val Asn Thr Ile Val Ser Lys Tyr Pro Ser
Ile Lys Gly Ile Asn Phe Asp Leu Pro His Val Ile Glu Asp Ala Pro
Ser Tyr Pro Gly Val Glu His Val Gly Gly Asp Met Phe Val Ser Val
Pro Lys Ala Asp Ala Val Phe Met Lys Thr Ile Cys His Asp Thr Ser
Asp Ala His Cys Leu Lys Phe Leu Lys Asn Cys Tyr Asp Ala Leu Pro
Glu Asn Gly Lys Val Ile Leu Val Glu Cys Ile Leu Pro Val Ala Pro
Asp Thr Ser Leu Ala Thr Lys Gly Val Val His Val Asp Val Ile Met
Leu Ala His Asn Pro Gly Gly Lys Glu Arg Thr Glu Lys Glu Phe Glu
Gly Leu Ala Lys Gly Ala Gly Phe Gin Gly Phe Glu Val Met Cys Cys
 Ala Phe Asn Thr His Val Ile Glu Phe Arg Lys Lys Ala

<210> SEQ ID NO 7
<211> LENGTH: 1915
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremuloides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> OTHER INFORMATION: 4CL
<223> OTHER INFORMATION: 4CL
<240> SEQUENCE: 7

ccctcgcgaa actccgaaaa cagagagcac ctaaaactca ccattctccc ctctgcattc
60
ttagcccgca atggacgcca caaatgaatcc acaagaattc atctttcgct caaaattacc
120
agacatctac atcccgaaaa accttccct gcaacctac gttcttgaga actttgctaa
180
acattcaca aaaccggcag tgataaatgg cggaattga gatgtcctca cctatgctga
240
tgttgagctc acagcaagaa gagttgcttc tggtctgaac aagattggta ttcaacaagg
300
tgacgtgatc atgctcttcc taccaagttc acctgaattc gtgcttgctt tcctaggcgc
360
ttcacacaga ggtgcctcgc tcacgctccc tcaccccotc cagagctgc
420
aaaacctgcag aaggtcctga gagaagagct cggataaaca cagccgtttgc actacgagaa
480
ggtaaagat tttgcgcag aagatggatg taaggctcatg tgcgtggaact ctgccccgga
540
cgtgtcctca ctttcgcag ttacacacca ggcagacgaag aatgaagtgc ctccaggtca
600
cattgctct gatgtcctgc tagcatgccc ttacatcaca ggacactacag ggtgcacag
660
<table>
<thead>
<tr>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>aggggtctag ttaacgcaac aagggcta aaccagttg gctcaacag tagagt gaga</code></td>
</tr>
<tr>
<td><code>caatccatac ctgtatttc acagtagaga tgtgattcct tgtgtcct ctcgatt cca</code></td>
</tr>
<tr>
<td><code>ttatcctgac tgtatctctg tggctctgaa gtctgtgct cgtattt ggtg</code></td>
</tr>
<tr>
<td><code>aagctcaag gttggatgc tgtagagttg gaggattgaa gagaagaca aagtagt ttaa</code></td>
</tr>
<tr>
<td><code>agcatctgc acatatgacct tagcttgact cagggatag gaaagacca</code></td>
</tr>
<tr>
<td><code>ggcaagacct tcgctggtgc atttgccaaag gaaaaagttag gctctggtt cgtg</code></td>
</tr>
<tr>
<td><code>agagctaga gctgctgtgc tccacgctc ttcgcttgtt gacaccaga aacctgaa gagaagact</code></td>
</tr>
<tr>
<td><code>gaaactcctg tttctggacg tttctggagct ctgtctctg gggagacc ctgc</code></td>
</tr>
<tr>
<td><code>aaactccatc ctotctcttc agctctgttg tttcagaaa atgggtagcc ttgaacaga 60</code></td>
</tr>
<tr>
<td><code>gagaaaaatt tgtcgttgcc cagcagacga acctgactctt caaactct gttccacacta 120</code></td>
</tr>
<tr>
<td><code>tagtccgcca gatacggccc cagagagtct ctttataaagt ggatacctg ggttaaattg 180</code></td>
</tr>
<tr>
<td><code>ccatacagat atccacacac tcaaaaaagct cttggtgct acatcactct tctttgcc 240</code></td>
</tr>
<tr>
<td><code>tggccttgaagttctcctg ggattgtgca ggttgggtg cattcg aatcagcacc caccag</code></td>
</tr>
<tr>
<td><code>ttgagagttgtg ggtcgttcag tggagtgcct cgaagagtt ctgttgctgt </code></td>
</tr>
<tr>
<td><code>agcgactcgt ggtggtgagttggc gggtgtctgc agttgggtatctccaggtg caggtctgag</code></td>
</tr>
<tr>
<td><code>ttggaagttgctgttg gggagctg gtagttcgc ttttcagagtctg tccctgagtt tcgag</code></td>
</tr>
<tr>
<td><code>ggcaactaga gctgctgtgc tccacgctc ttcgcttgtt gacaccaga aacctgaa gagaagact</code></td>
</tr>
<tr>
<td><code>aggtatgtctg agttttttttt ttctctgtcc ttcagagagtt cactcactact acag</code></td>
</tr>
</tbody>
</table>

**SEQ ID NO 8**

**LENGTH: 1395**

**FEATURE: aspen populus tremuloides**

**NAME/KEY: misc_feature**

**OTHER INFORMATION: CAD; GenBank accession number: AF217957**

**SEQUENCE: 8**
<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 Asp Thr
20  25  30
Gly Pro Glu Asp Val Leu Ile Lys Val Ile Ser Cys Gly Ile Cys His
35  40  45
Thr Asp Ile His Glu Ile Ser Cys Gly Ile Cys His
50  55  60
Met Val Pro Gly His Glu Val Val Gly Val Val Val Gly Ser
65  70  75  80
Asp Val Thr Lys Phe Lys Ala Gly Asp Val Val Val Gly Val Ile
85  90
Val Gly Ser Cys Lys Asn Cys His Pro Cys Lys Ser Glu Leu Glu Glu
100 105 110
Tyr Cys Asn Lys Ile Trp Ser Tyr Asn Asp Val Tyr Thr Asp Gly
115 120 125
Lys Pro Thr Glu Gly Gly Phe Ala Glu Ser Met Val Val Asp 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 Glu Ser Gly Leu Arg Gly Gly Ile Leu Gly Leu Gly Gly Val
180 185 190
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 Lys Arg Glu 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
-continued

Pro Leu Glu Pro Tyr Leu Ser Leu Leu Lys Leu 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 Glu Met Leu Glu Phe Cys Lys Gly 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 Val Arg Arg 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 Gin 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 Asn Glu Val Pro Gin Val Asp Ile Ser Pro 165 170 175
Asp Asp Val Ala Leu Pro Tyr Ser Ser Gly Thr Thr Gly Leu Pro 180 185 190
Lys Gly Val Met Leu Thr His Lys Gly 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 Tyr Lys Val Ser 260 265 270
Ile Ala Pro Val Val Pro Pro Val Met Met Ala Ile Ala Lys Ser Pro
275 280 285
Asp Leu Asp Lys His Asp Leu Ser 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 Gly Lys Gly Trp Leu His Thr Gly Asp
405 410 415
Ile Gly Tyr Ile Asp Asp Asp Asp Asp Asp Gly 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 Gln Ile Lys Gln Tyr Ile
485 490 495
Ser Lys Gln Val Ile Phe Tyr 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 20

<210> SEQ ID NO 12
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: aspen populus tremuloides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: Pt4CLl sense primer
<400> SEQUENCE: 12
atgaatccac aagaattcat 20
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, aspen CAld5H,
       aspen AldOMT, SEQ ID NO: 9, and SEQ ID NO: 2; and
   (b) regenerating the plant cell to produce a transgenic plant.

* * * * *