6-19-2007

Cellulose synthase promoter and method for modifying cellulose and lignin biosynthesis in plants

Vincent Lee C. Chiang  
*Michigan Technological University*

Chandrashekhar P. Joshi  
*Michigan Technological University, cpjoshi@mtu.edu*

Luguang Wu

Daniel T. Carraway

Follow this and additional works at: [http://digitalcommons.mtu.edu/patents](http://digitalcommons.mtu.edu/patents)

Part of the *Engineering Commons*

**Recommended Citation**


[http://digitalcommons.mtu.edu/patents/10](http://digitalcommons.mtu.edu/patents/10)

Follow this and additional works at: [http://digitalcommons.mtu.edu/patents](http://digitalcommons.mtu.edu/patents)

Part of the *Engineering Commons*
This invention relates to an isolated cellulose synthase promoter, methods for genetically altering cellulose and lignin biosynthesis, and to methods for improving strength properties of juvenile wood and fiber in trees. The invention further relates to methods for identifying regulatory elements in a cellulose synthase promoter and to methods for augmenting expression of polynucleotides operably linked to a cellulose synthase promoter.

22 Claims, 10 Drawing Sheets
OTHER PUBLICATIONS

Oliphant et al., Gene 44:177, (1986).
Timmell, Compression Wood in Gymnosperms, Springer Verlag, (1986).
Fig. 1: DNA and predicted protein sequence of PtCelA cDNA
Figure 2. Southern blot analysis of aspen genomic DNA at low (a) and high (b) stringency. DNA (25 μg per lane) was digested with Pst I (lane P), HindIII (H), and EcoRI (E), respectively, and probed with a $^{32}$P-labeled 1-kb 5' end PtCelA cDNA fragment. (c) Northern blot analysis of total RNA (40 μg per lane) from aspen stem internodes, probed with $^{32}$P-labeled PtCelA cDNA. 1-2, 1st and 2nd; 3-4, 3rd and 4th; 5-6, 5th and 6th; 9-10, 9th and 10th internodes.
Figure 3. *In situ* localization of *PtCelA* gene transcripts. Transverse sections (10 μm) from 2nd (a), 4th (b) and 6th (c) aspen stem internodes were hybridized with DIG-labeled *PtCelA* antisense RNA probe or with a DIG-labeled *PtCelA* sense RNA probe as the control (d, sections from the 5th internode) according to Seidman with modifications. Positive RNA-RNA hybridization signals are stained purple blue. Arrows indicate the cellular localization of the *PtCelA* transcripts. No hybridization signal is detectable in control section. PX, primary xylem cells; SX, secondary xylem cells. Bars = 100 μm.
Fig. 4: DNA sequence of PtCelAP, the 5' flanking region of PtCelA coding sequence

```
  1  GAATTCGCCCTTTTGAAATTTGCCAGGAGACGATAGTTTCGGTTTGAATTTGGCTTTTCTCA  
  60
  61  CTTCTGGTCTAGCAATTGCAAAGAACAGACTTTCCGGTTCGTTGAATGGCTTTGTTCA  
  120
  121  ATAGTCTCGATTCGAAGTTGCAAACTGCCGTTTCTGGTATTGCAATTATGTAGCCATAAC  
  180
  181  TGTTAAATCTGTAGCTATTAGCGGACCACAACCAGATTACGCGATCGTCGATAAA  
  240
  241  GAGATCTCCATTTCTACGGTTTCTTTCTTTACCTTTGTCAGAGAATTACCTGTGAT  
  300
  301  ACGTACGATGATGATTGATGATTATGGGAACCATTCCGATGTTAGACACGACAGACATCT  
  360
  361  TGTTAAATCTGTAGCTATTAGCGGACCACAACCAGATTACGCGATCGTCGATAAA  
  420
  421  GAGATCTCCATTTCTACGGTTTCTTTCTTTACCTTTGTCAGAGAATTACCTGTGAT  
  480
  481  ACGTACGATGATGATTGATGATTATGGGAACCATTCCGATGTTAGACACGACAGACATCT  
  540
  541  TGTTAAATCTGTAGCTATTAGCGGACCACAACCAGATTACGCGATCGTCGATAAA  
  600
  601  GCTTTGGTTAGGTATTTGATATTTGGAATTTGTAAGTTGCTCGCCCAACAGAAACCTTC  
  660
  661  GAATTCGCCCTTTTGAAATTTGCCAGGAGACGATAGTTTCGGTTTGAATTTGGCTTTTCTCA  
  720
  721  ATAGTCTCGATTCGAAGTTGCAAACTGCCGTTTCTGGTATTGCAATTATGTAGCCATAAC  
  780
  781  TGTTAAATCTGTAGCTATTAGCGGACCACAACCAGATTACGCGATCGTCGATAAA  
  840
  841  GCTTTGGTTAGGTATTTGATATTTGGAATTTGTAAGTTGCTCGCCCAACAGAAACCTTC  
  900
  901  GAATTCGCCCTTTTGAAATTTGCCAGGAGACGATAGTTTCGGTTTGAATTTGGCTTTTCTCA  
  960
  961
```

MMESGAPICHTCQEGVHGDA
NGELFVACHECSYMCKSCF
EFIEKGRKVCLRCGSE
Figure 5. Histochemical analysis of transgenic tobacco for GUS gene expression driven by aspen PtCelA gene promoter. Stem transverse sections from the 3rd (a), 5th (b), 7th (c), and 8th (d and f) internodes were stained for GUS activity. Fluorescence microscopy (e) showing the identical section as in (d). Lignin autofluorescence was visualized after UV irradiation. An entire section from the 8th internode stained for GUS activity is shown in (f). Bars = 100 μm in a to e; bar = 1.5 mm in f.
Figure 6. GUS gene expression driven by aspen PtCelA gene promoter in transgenic tobacco plants under tension stress. Tension stress was induced by bending the transgenic plants. Tangential and longitudinal sections were harvested before bending (a) and 4 (b), 20 (c) and 40 (d) hours after bending and stained for GUS expression. Arrows indicate the bend sites.
<table>
<thead>
<tr>
<th><strong>Arabidopsis thaliana cellulose synthase mRNA</strong></th>
<th><strong>SEQ ID NO:</strong> 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>gcggcgcgcgg ttaatcgcgc gttcttcacaa cacggaatgag tttgtctcca ttaatgcgcg</td>
<td>1</td>
</tr>
<tr>
<td>tcgaagactgc cgaagaagtc gctgtgcttg cggctggatt gccatcgac</td>
<td>61</td>
</tr>
<tr>
<td>gaccttgctgt gcgctgccat gggctgcttg cgggctattc gagatctgtg gtttagactg</td>
<td>121</td>
</tr>
<tr>
<td>cgtgcgtgctg agttgcgttg tctctttgct tgggctgcga</td>
<td>181</td>
</tr>
<tr>
<td>cgtcgtgctgt gcgctgccat gggctgcttg cgggctattc gagatctgtg gtttagactg</td>
<td>241</td>
</tr>
<tr>
<td>gcggccgcgg ttaatcgcgc gttcttcacaa cacggaatgag tttgtctcca ttaatgcgcg</td>
<td>301</td>
</tr>
<tr>
<td>tcgaagactgc cgaagaagtc gctgtgcttg cggctggatt gccatcgac</td>
<td>361</td>
</tr>
<tr>
<td>gaccttgctgt gcgctgccat gggctgcttg cgggctattc gagatctgtg gtttagactg</td>
<td>421</td>
</tr>
<tr>
<td>gcggcgcgcgg ttaatcgcgc gttcttcacaa cacggaatgag tttgtctcca ttaatgcgcg</td>
<td>481</td>
</tr>
<tr>
<td>tcgaagactgc cgaagaagtc gctgtgcttg cggctggatt gccatcgac</td>
<td>541</td>
</tr>
<tr>
<td>gaccttgctgt gcgctgccat gggctgcttg cgggctattc gagatctgtg gtttagactg</td>
<td>601</td>
</tr>
<tr>
<td>gcggcgcgcgg ttaatcgcgc gttcttcacaa cacggaatgag tttgtctcca ttaatgcgcg</td>
<td>661</td>
</tr>
<tr>
<td>tcgaagactgc cgaagaagtc gctgtgcttg cggctggatt gccatcgac</td>
<td>721</td>
</tr>
<tr>
<td>gaccttgctgt gcgctgccat gggctgcttg cgggctattc gagatctgtg gtttagactg</td>
<td>781</td>
</tr>
<tr>
<td>gcggcgcgcgg ttaatcgcgc gttcttcacaa cacggaatgag tttgtctcca ttaatgcgcg</td>
<td>841</td>
</tr>
<tr>
<td>tcgaagactgc cgaagaagtc gctgtgcttg cggctggatt gccatcgac</td>
<td>901</td>
</tr>
<tr>
<td>gaccttgctgt gcgctgccat gggctgcttg cgggctattc gagatctgtg gtttagactg</td>
<td>961</td>
</tr>
<tr>
<td>gcggcgcgcgg ttaatcgcgc gttcttcacaa cacggaatgag tttgtctcca ttaatgcgcg</td>
<td>1021</td>
</tr>
<tr>
<td>tcgaagactgc cgaagaagtc gctgtgcttg cggctggatt gccatcgac</td>
<td>1081</td>
</tr>
<tr>
<td>gaccttgctgt gcgctgccat gggctgcttg cgggctattc gagatctgtg gtttagactg</td>
<td>1141</td>
</tr>
<tr>
<td>gcggcgcgcgg ttaatcgcgc gttcttcacaa cacggaatgag tttgtctcca ttaatgcgcg</td>
<td>1201</td>
</tr>
<tr>
<td>tcgaagactgc cgaagaagtc gctgtgcttg cggctggatt gccatcgac</td>
<td>1261</td>
</tr>
</tbody>
</table>

FIG. 7
FIG. 8

Arabidopsis thaliana cellulose synthase SEQ ID NO: 5

RPRLIAGSHNRNEFVLINADENARIRSVQELSQTQICRDEIE
LTVDGEPFVACNECAFVCRPCYEYERREGNQACPQCKTRFKRLKGSPRGVEGDEEDD
IDDLDNEFEYGNNGIGFDQVSEGMSISRNSGFPQSDLDSDAPPQGIPLLTGDEDEVE
ISSDRHALIVPPSLGHHGHRVHPVSLSDPTVAAHRLMPQKDLAVYGYGVSVAWKRMM
EEWKRKQNEKLQVVRHEGDPDFEDGDDADFPMMDEGRQPLSMKIPIKSSKINPYRMLI
VRLVILGLFFHYRILHPKDAYALWLISVIECIWFAVSVWLDQFPKWPYPIERETYLD
RLSLRVEKEGKPSLSPDVVFSTVDPLKEPPLANTVLSILAVDPVVDKAVCVSD
DGAAMLTPEALSETAEFARKWVFCKKCYCIEPRAPEWYFCHXMDYLNKNKHPPAFVRER
RAMKRDYEEFKVINALVATAQKVPEDGWMTQDGTWPWPGNSVRDHGMIQVFLGSGDV
RDVENNELPRLYVSREKRPFDHDKKAGAMNSLIRVSGVLSNAPLYLLNVCDHYINN
SKALREAMCFNMDPQSGKICICYVQFPQRDFDGDRHDYRSNRNFFVDFDNMKGLDMLQG
PIYVGTGVCVFRRQALYGFDAPKKKGPRKTCNCWFKWCLCFSRKNKARATYAADK
KKNREASKQHLENIEGERHHKVLNEQSTQAMMKLQKKGQSPVFVASRLENGG
MARNSPACLLEAIQVISRGYEDKTEWKEIGWIYGSVEILTGSKMHSHGWRHVV
CTPKLFAFKGSAPINLSDLRHQVLRWALSVEIFLSRHCPITYWGGGLKWLRLSLYI
NSVVPWTSLPLIVYCSLPAICLLTUTGKIVPEISNYASLFMLFSSIAITGILEMOQW
GKVGIIDDWWRNENQFVYGVSCAHNFALFQGLLKVLAGVDTNFTVTSKADDEGFDLY
LFKWTSSLIPMTLLINIVGIVGVGVSADINSNGDSWGPLFCRLFFALWVIHLYPLFL
KGLLGKQDRMTPIIVWSSILLASILTLLWVNVNPVAKGPGPHELICGLDCL
CELLULOSE SYNTHASE PROMOTER AND METHOD FOR MODIFYING CELLULOSE AND LIGNIN BIOSYNTHESIS IN PLANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a divisional of U.S. application Ser. No. 09/980,043 filed Apr. 5, 2002, now U.S. Pat. No. 7,049,481, which is a 371 of International Application No. PCT/US00/13637 filed May 18, 2000, which claims priority to U.S. Provisional Application No. 60/135,280 filed May 21, 1999.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

Not Applicable

FIELD OF THE INVENTION

This invention relates to polynucleotide molecules encoding cellulose synthase, promoters of cellulose synthase and cellulose synthase polypeptides, methods for genetically altering cellulose and lignin biosynthesis, and methods for improving strength properties of juvenile wood and fiber in trees. The invention further relates to methods for identifying regulatory elements in a cellulose synthase promoter and transcription factors that bind to such regulatory elements, and to methods for augmenting expression of polynucleotides operably linked to a cellulose synthase promoter.

BACKGROUND OF THE INVENTION

Lignin and cellulose are the two major building blocks of plant cell walls that provide mechanical strength and rigidity. In plants, and especially in trees, these two organic materials exist in a dynamic equilibrium conferring mechanical strength, water transporting ability and protection from biotic and abiotic environmental stresses. Normally, oven-dry wood contains 30 to 50% cellulose, 20 to 30% lignin and 20 to 30% hemicellulose (Higuchi, 1997).

Proportions of lignin and cellulose are known to change with variation in the natural environment. For example, during the development of compression wood in conifers, the percentage of cellulose increases from 30 to 40%, while lignin content decreases from 30 to 20% (Timmell, 1986). Conversely, in angiosperm tension wood the percentage of cellulose decreases from 30 to 40%, while lignin content decreases from 30 to 20% (Timmell, 1986).

It was recently discovered that the genetic down-regulation of a key tissue-specific enzyme from the lignin biosynthesis pathway, 4CL, results in reduction of lignin content by up to 45% in transgenic aspen trees (Hu et al., 1999). This down-regulation is also associated with a 15% increase in the cellulose content. If the converse were true, i.e., that increasing cellulose content by genetic up-regulation of cellulose biosynthesis results in reduction of lignin content, then the pulp yield could be increased. This would allow tremendous savings in chemical and energy costs during pulping because, for example, lignin must be degraded and removed during the pulping process.

Cellulose is a linear glucan consisting of 3-D-1,4-linked glucose residues. It is formed by a cellulose synthase enzyme which catalyzes assembly of UDP-glucose units in plasma membrane complexes known as "particle rosettes" (Delmer and Amor, 1995). Cellulose synthase is thought to be anchored to the membrane by eight transmembrane binding domains to form the basis of the cellulose biosynthesis machinery in the plant cell wall (Pear et al., 1996).

In higher plants, the glucan chains in cellulose microfibrils of primary and secondary cell walls are different in their degree of polymerization (Brown et al., 1996). For example, secondary cell walls are known to contain cellulose having a high degree of polymerization, while in primary cell walls the degree of polymerization is lower. In another example, woody cell walls suffering from tension stress produce tension wood on the upper side of a bent angiosperm tree in response to the stress. In these cells, there are elevated quantities of cellulose which have very high crystallinity. The formation of highly crystalline cellulose is important to obtain a higher tensile strength of the wood fiber. Woody cell walls located at the under side of the same stem experience a compression stress, but do not produce highly crystalline cellulose. Such variation in the degree of polymerization in cell walls during development is believed to be due to different types of cellulose synthases for organizing glucose units into different nanocrystalline arrays (Hugler and Blanton, 1996). Therefore, it would be advantageous to determine the molecular basis for the synthesis of highly crystalline cellulose so that higher yields of wood pulp having superior strength properties can be obtained from transgenic trees. Production of highly crystalline cellulose in transgenic trees would also markedly improve the mechanical strength properties of juvenile wood formed in normal trees. This would be a great benefit to the industry because juvenile wood is generally undesirable for solid wood applications because it has inferior mechanical properties.

Since the deposition of cellulose and lignin in trees is regulated in a compensatory fashion, genetic augmentation of cellulose biosynthesis might have a repressive effect on lignin deposition. Since the degree of polymerization and crystallinity may depend upon the type of cellulose synthase incorporated in the cellulose biosynthesis machinery, the expression of heterologous cellulose synthase or a UDP-glucose binding region thereof (e.g., sweetgum protein expression in loblolly pine), could increase the quality of cellulose in transgenic plants. Over-expression of a heterologous cellulose synthase may also increase cellulose quantity in transgenic plants. Thus, genetic engineering of cellulose biosynthesis can provide a strategy to augment cellulose quality and quantity, while reducing lignin content in transgenic plants.

A better understanding of the biochemical processes that lead to wood formation would enable the pulp and paper industries to more effectively use genetic engineering as a tool to meet the increasing demands for wood from a decreasing production area. With this objective, many xylem-specific genes, including most lignin biosynthesis genes, have been isolated from developing xylem tissues of various plants including tree species (Ye and Vanmer, 1993; Fukuda, 1996; Whetten et al., 1998). Genes regulating cellulose biosynthesis in crop plants (Pear et al., 1996 and Arioli et al., 1998), versus in trees, have also been isolated. However, isolation of tree genes which are directly involved in cellulose biosynthesis has remained a great challenge.

For more than 30 years, no gene encoding higher plant cellulose synthase (CelA) was identified. Recently, Pear et al. (1996) isolated the first putative higher plant CelA cDNA, GhCelA (GenBank No. GHU58283), by searching for UDP-glucose binding sequences in a cDNA library prepared from cotton fibers having active secondary wall cellulose synthesis. GhCelA was considered to encode a
cellulose synthase catalytic subunit because it is highly expressed in cotton fibers, actively synthesizes secondary wall cellulose, contains eight transmembrane domains, binds UDP-glucose, and contains two other domains unique to plants.

Recently, Arioli et al. (1998) cloned a CeA homolog, RSW1 (radial swelling) (GenBank No. AF027172), from Arabidopsis by chromosome walking to a defective locus of a temperature sensitive cellulose-deficient mutant. Complementation of the RSW1 mutant with a wild type full-length genomic RSW1 clone restored the normal phenotype. This complementation provided the first genetic proof that a plant CeA gene encodes a catalytic subunit of cellulose synthase and functions in the biosynthesis of cellulose microfibrils. The full-length Arabidopsis RSW1 represents the only known, currently available cellulose synthase cDNA available for further elucidating cellulose biosynthesis in transgenic systems (Wu et al., 1998).

The discovery of the RSW1 gene substantiated the belief that the assembly of a cellulose synthase into the plasma membrane is required for functional cellulose biosynthetic machinery and for manufacturing crystalline cellulose microfibrils in plant cell walls. Most significantly, a single CeA gene, e.g. RSW1, is sufficient for the biosynthesis of cellulose microfibrils in plants, e.g. Arabidopsis. Thus, RSW1 is a prime target for engineering augmented cellulose formation in transgenic plants.

Since many of society’s fiber, chemical and energy demands are met through the industrial-scale production of cellulose from wood, genetic engineering of the cellulose biosynthesis machinery in trees could produce higher pulp yields. This would allow greater returns on investment by pulp and paper industries. Therefore, it would be advantageous to isolate and characterize genes from trees that are involved in cellulose biosynthesis in order to improve the properties of wood.

SUMMARY OF THE INVENTION

The present invention relates to polynucleotides comprising a nucleotide sequence that encodes a cellulose synthase, regulatory sequences, including a stress-inducible promoter, of the cellulose synthase, a cellulose synthase protein or a functional domain thereof and methods for augmenting cellulose biosynthesis in plants.

Thus, in one aspect, the invention provides a polynucleotide comprising a sequence that encodes a cellulose synthase, or a polynucleotide fragment thereof, the fragment encoding a functional domain of cellulose synthase, such as a UDP-glucose binding domain. The invention also provides a cellulose synthase or a functional domain or fragment thereof, including a UDP-glucose binding domain and at least one of eight transmembrane domains. The invention further provides a cellulose synthase promoter, or a functional fragment thereof, which fragment contains one or more mechanical stress response elements (MSRE).

In another aspect, the present invention is directed to a method of improving the quality of wood by altering the quantity of cellulose in plant cells, and optionally decreasing the content of lignin in the cell. The invention also relates to a method of altering the growth or the cellulose content of a plant by expressing an exogenous polynucleotide encoding a cellulose synthase or a UDP-glucose binding domain thereof in the plant. The invention further provides a method for causing a stress-induced gene expression in a plant cell by expressing a polynucleotide of choice using a stress-inducible cellulose synthase promoter.

In yet another aspect, the invention relates to a method for determining a mechanical stress responsive element (MSRE) in a cellulose synthase promoters and a method for identifying transcription factors that binds to the MSRE.

In a further aspect, the invention provides a method for altering (increasing or decreasing) i.e., regulating, the expression of a cellulose synthase in a plant by expressing an exogenous polynucleotide encoding a transcription factor having the property of binding a positive MSRE of a cellulose synthase promoter or by expressing an antisense polynucleotide encoding a transcription factor having the property of binding a negative MSRE to block the expression of the transcription factor.

Other aspects of the invention will be appreciated by a consideration of the detailed description of the invention drawings and appended claims.
hybridization of two sequences at high stringency (such as, tertiary sequences; depending on the stringency of hybridization, the two polynucleotides contain substantially complementary single-stranded molecules, i.e., DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as “protein nucleic acids” (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases, for example thio-uracil, thio-guanine and fluorouracil.

An “isolated” nucleic acid molecule or polynucleotide refers to a component that is removed from its original environment (for example, its natural environment if it is naturally occurring). An isolated nucleic acid or polypeptide may contain less than about 50%, preferably less than about 75%, and most preferably less than about 90%, of the cellular components with which it was originally associated.

A polynucleotide amplified using PCR so that it is suitable for hybridization under conditions of maximum homology is defined as “isolated”. The polynucleotides and polypeptides of the invention may be “substantially pure,” i.e., having the highest degree of purity that can be achieved using purification techniques known in the art.

The term “hybridization” refers to a process in which a strand of nucleic acid joins with a complementary strand through base pairing. Polynucleotides are “hybridizable” to each other when at least one strand of one polynucleotide can anneal to a strand of another polynucleotide under defined stringency conditions. Hybridization requires that the two polynucleotides contain substantially complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high stringency (such as, for example, in an aqueous solution of 0.5xSSC at 65EC) requires that the sequences exhibit a high degree of complementarity over their entire sequence. Conditions of intermediate stringency (such as, for example, an aqueous solution of 2xSSC at 65EC) and low stringency (such as, for example, an aqueous solution of 2xSSC at 55EC), require correspondingly less overall complementarity between the hybridizing sequences. (1xSSC is 0.15 M NaCl, 0.015 M Na citrate.) As used herein, the above solutions and temperatures refer to the probe-washing stage of the hybridization procedure. The term “a polynucleotide that hybridizes under stringent (low, intermediate) conditions” is intended to encompass both single and double-stranded polynucleotides although only one strand will hybridize to the complementary strand of another polynucleotide.

A “sequence-conservative variant” is a polynucleotide that contains a change of one or more nucleotides in a given codon position, as compared with another polynucleotide, but the change does not result in any alteration in the amino acid encoded at that position. A “function-conservative variant” is a polypeptide (or a polynucleotide encoding the polypeptide) having a given amino acid residue that has been changed without altering the overall conformation and function of the polypeptide, including, but not limited to, replacement of an amino acid with one having similar physico-chemical properties (such as, for example, acidic, basic, hydrophobic, and the like). Amino acids having similar physico-chemical properties are well known in the art. For example, arginine, histidine and lysine are hydrophilic-basic amino acids and may be interchangeable. Similarly, isoleucine, a hydrophobic amino acid, may be replaced with leucine, methionine or valine. Sequence- and function-conservative variants are discussed in greater detail below with respect to degeneracy of the genetic code.

A “functional domain” or a “functional fragment” refers to any region or portion of a protein or polypeptide or polynucleotide which is a region or portion of a larger protein or polynucleotide, the region or portion having the specific activity or specific function attributable to the larger protein or polynucleotide, e.g., a functional domain of cellulose synthase is the UDP-glucose binding domain.

The term “% identity” refers to the percentage of the nucleotides/amino acids of one polynucleotide/polypeptide that are identical to the nucleotides/amino acids of another sequence of polynucleotide/polypeptide as identified by program GCG from Genetics Computer Group Wisconsin (GCG) package (version 9.0) (Madison, Wis.). GCG 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. The following parameters are used by the GCG program GCG as default values (for polynucleotides): gap creation penalty: 50; gap extension penalty: 5; scoring matrix: nwsgapdna.cm (local data file).

The “% 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. PowerBlast program, implemented by the National Center for Biotechnology Information, can be used to compute optimal, gapped alignments. GCG program from Genetics Computer Group Wisconsin package (version 9.0) (Madison, Wis.) can also be used. GCG 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. The
The term “oligonucleotide” refers to a nucleic acid, generally of at least 10, preferably at least 15, and more preferably at least 20 nucleotides, that is hybridizable to a genomic DNA molecule, a cDNA molecule, or an mRNA molecule encoding a gene, mRNA, cDNA, or other nucleic acid of interest. Oligonucleotides can be labeled, e.g., with $^{32}$P-nucleotides or nucleotides to which a label, such as biotin, has been covalently conjugated. In one embodiment, a labeled oligonucleotide can be used as a probe to detect the presence of a nucleic acid. In another embodiment, oligonucleotides (one or both of which may be labeled) can be used as PCR primers, either for cloning full length or a fragment of CelA, or to detect the presence of nucleic acids encoding CelA. In a further embodiment, an oligonucleotide of the invention can form a triple helix with a CelA DNA molecule. In still another embodiment, a library of oligonucleotides arranged on a solid support, such as a silicon wafer or chip, can be used to detect various polymorphisms of interest. Generally, oligonucleotides are prepared synthetically, preferably on a nucleic acid synthesizer. Accordingly, oligonucleotides can be prepared with non-naturally occurring phosphoester analog bonds, such as thiester bonds, etc.

The term “coding sequence” 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 5N (amino) terminus and a translation stop codon at the 3N (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from eukaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences.

A “promoter” is a polynucleotide containing elements (e.g., a TATA box) which are capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3N direction) coding sequence. For purposes of describing the present invention, the promoter is a sequence that is bound at its 3N terminus by the transcription initiation site and extends upstream (5N direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conventionally defined for example, by mapping with nucleases S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Examples of promoters that can be used in the present invention include: PCElAP, 4CL-1 and 5SS.

The term “constitutive promoter” refers to a promoter which typically, does not require positive regulatory proteins to activate expression of an associated coding sequence, i.e., a constitutive promoter maintains some basal level of expression. A constitutive promoter is commonly used in creation of an expression cassette. An example of a constitutive promoter are 3SS CaMV (Cauliflower Mosaic Virus), available from Clonetech, Palo Alto, Calif.

The term “inducible promoter” refers to the promoter which requires a positive regulation to activate expression of an associated coding sequence. An example of such a promoter is a stress-inducible cellulose synthase promoter from aspen described herein.

A coding sequence is “under the control” of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced and translated into the protein encoded by the coding sequence.

A “vector” is a recombinant nucleic acid construct, such as plasmid, phage genome, virus genome, cosmids, or artificial chromosome to which a polynucleotide of 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.

The term “expression cassette” refers to a polynucleotide which contains both a promoter and a protein coding sequence such that expression of a given protein is achieved upon insertion of the expression cassette into a cell.

A cell has been “transfected” by exogenous or heterologous polynucleotide when such polynucleotide has been introduced inside the cell. A cell has been “transformed” by exogenous or heterologous polynucleotide when the transfected polynucleotide effects a phenotypic change. Preferably, the transforming polynucleotide should be integrated (covalently linked) into chromosomal DNA making up the genome of the cell.

“Exogenous” refers to biological material, such as a polynucleotide or protein, that has been isolated from a cell and is then introduced into the same or a different cell. For example, a polynucleotide encoding a cellulose synthase of the invention can be cloned from xylem cells of a particular species of tree, inserted into a plasmid and reintroduced into xylem cells of the same or different species. The species thus contains an exogenous cellulose synthase polynucleotide.

“Heterologous polynucleotide” refers to an exogenous polynucleotide not naturally occurring in the cell into which it is introduced.

“Homologous polynucleotide” refers to an exogenous polynucleotide that naturally exists in the cells into which it is introduced.


The present invention relates to a novel, full-length cellulose synthase gene (CelA), a novel stress inducible promoter of cellulose synthases (CelAP), and cellulose synthase proteins from trees, including UDP-glucose catalytic domains thereof. The invention enables the development of transgenic tree varieties having increased cellulose content, decreased lignin content and, therefore, improved wood fiber characteristics. Production of increased cellulose quantity and quality in multiple varieties of commercially relevant, transgenic forest tree species in operational production scenarios are further contemplated. The invention
further provides a new experimental system for study of CelA gene expression and function in trees.

Polynucleotides Encoding Cellulose Synthase and Fragments Thereof.

The present invention relates to polynucleotides which comprise the nucleotide sequence that encodes cellulose synthase of the invention or a functional fragment thereof. In a preferred embodiment, the polynucleotide comprises the sequence encoding a tree cellulose synthase and most preferably, the sequence encoding a cellulose synthase from aspen. In one embodiment, a polynucleotide of the invention includes the entire cellulose synthase coding region, e.g., nucleotides 69 to 3,005 of SEQ ID NO: 1. In another aspect of the invention, the polynucleotide encoding an Arabidopsis cellulose synthase is provided (see SEQ ID NO: 1 and the translated protein of SEQ ID NO: 5).

Also within the scope of the invention are fragments of the polynucleotides encoding cellulose synthase of the invention, which fragments encode at least one transmembrane domain and/or a UDP-glucose binding domain. For example, a polynucleotide comprising the nucleotides encoding a UDP-glucose binding domain of aspen cellulose synthase (e.g., nucleotides 660 to 2250 of SEQ ID NO: 1) or corresponding nucleotides of SEQ ID NO: 4 are within the scope of the invention. The nucleotides encoding the UDP-glucose binding domain can be determined by, for example, alignment of protein sequences as described below.

The invention further relates to sequence conservative variants of the coding portion of SEQ ID NOS: 1 and 4. Polynucleotides that hybridize under conditions of low, medium, and high stringency to SEQ ID NOS: 1 and 4, and their respective coding portions are also within the scope of the invention. Preferably, the polynucleotide that hybridizes to any of SEQ ID NOS: 1 and 4, or their respective coding portions, is about the same length as that sequence, for example, not more than about 10 to about 20 nucleotides longer or shorter. In another embodiment of the invention, the hybridizable polynucleotide is at least 1500 nucleotides long, preferably at least 2500 nucleotides long and most preferably at least 3000 nucleotides long. In yet another embodiment, the hybridizable polynucleotide comprises the UDP-glucose binding domain as found in SEQ ID NO: 1 or 4, or at least the conserved region QVLRW [SEQ ID NO: 7]. Most preferably, the hybridizable polynucleotide has a UDP-glucose binding activity.

The polynucleotides that occur originally in nature may be isolated from the organisms that contain them using methods described herein or well known in the art. The non-naturally occurring polynucleotides may be prepared using various manipulations known in the field of recombinant DNA. For example, the cloned CelA polynucleotide can be modified according to methods described by Sambrook et al., 1989. The sequence can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated in vitro. In the production of the modified polynucleotides, for example, care should be taken to ensure that the modified polynucleotide remains within the appropriate translational reading frame (if to be expressed) or interrupted by translational stop signals. As a further example, a CelA-encoding nucleic acid sequence can be mutated in vitro or in vivo, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy preexisting ones, to facilitate further in vitro modification. Preferably, such mutations enhance the functional activity of the mutated CelA polynucleotide. Any technique for mutagenesis known in the art can be used, including but not limited to, in vitro site-directed mutagenesis (Hutchinson, C., et al., 1978, J. Biol. Chem. 253: 6551; Zoller and Smith, 1984, DNA 3: 479-488; Oliphant et al., 1986, Gene 44: 177; Hutchinson et al., 1986, Proc. Natl. Acad. Sci. U.S.A. 83: 710), use of Tn5 linkers (Pharmacia), etc. PCR techniques are preferred for site directed mutagenesis (see Higuchi, 1989, “Using PCR to Engineer DNA”, in PCR Technology: Principles and Applications for DNA Amplification, H. Erlich, ed., Stockton Press, Chapter 6, pp. 61–70).

The polynucleotides of the present invention may be introduced into various vectors adapted for plant or non-plant replication. These are well known in the art, thus, choice, construction and use of such vectors is well within the skill of a person skilled in the art. Possible vectors include, but are not limited to, plasmids or modified viruses of plants, but the vector system must be compatible with the host cell used. An example of a suitable vector is Ti plasmid. The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. An expression cassette containing cellulose synthase or recombinant molecules thereof can be introduced into host cells via silicon carbide whiskers, transformed protoplasts, transformation, e.g., Agrobacterium vectors (discussed below), electroporation, infection, etc., so that many copies of the gene sequence are generated. Preferably, the cloned gene is contained on a shuttle vector plasmid, which provides for expansion in a cloning cell, e.g., E. coli, and facile purification for subsequent insertion into an appropriate expression cell line, if such is desired. For example, a shuttle vector, which is a vector that can replicate in more than one type of organism, can be prepared for replication in both E. coli and Saccharomyces cerevisiae by linking sequences from an E. coli plasmid with sequences form the yeast 2m plasmid.

Transgenic plants containing the polynucleotides described herein are also within the scope of the invention. Methods for introducing exogenous polynucleotides into plant cells and regenerating transgenic plants are well known. Some are provided below.

In one embodiment, to introduce a plasmid containing a CelA coding sequence or promoter of the invention into a plant, a 1:1 mixture of plasmid DNA containing a selectable marker expression cassette and plasmid DNA containing a cellulose synthase expression cassette is precipitated with gold to form microprojectiles. The microprojectiles are rinsed in absolute ethanol and aliquots are dried onto a suitable macrocarrier such as the macrocarrier available from BioRad in Hercules, Calif. Prior to bombardment, embryogenic tissue is preferably desiccated under a sterile laminar-flow hood. The desiccated tissue is transferred to a semi-solid proliferation medium. The prepared microprojectiles are accelerated from the macrocarrier into the desiccated target cells using a suitable apparatus such as a BioRad PDS-1000/HE particle gun. In a preferred method, each plate is bombarded once, rotated 180 degrees, and bombarded a second time. Preferred bombardment parameters

1. **US 7,232,941 B2**

The polynucleotides of the present invention may be introduced into various vectors adapted for plant or non-plant replication. These are well known in the art, thus, choice, construction and use of such vectors is well within the skill of a person skilled in the art. Possible vectors include, but are not limited to, plasmids or modified viruses of plants, but the vector system must be compatible with the host cell used. An example of a suitable vector is Ti plasmid. The insertion into a cloning vector can, for example, be accomplished by ligating the DNA fragment into a cloning vector which has complementary cohesive termini. However, if the complementary restriction sites used to fragment the DNA are not present in the cloning vector, the ends of the DNA molecules may be enzymatically modified. Alternatively, any site desired may be produced by ligating nucleotide sequences (linkers) onto the DNA termini; these ligated linkers may comprise specific chemically synthesized oligonucleotides encoding restriction endonuclease recognition sequences. An expression cassette containing cellulose synthase or recombinant molecules thereof can be introduced into host cells via silicon carbide whiskers, transformed protoplasts, transformation, e.g., Agrobacterium vectors (discussed below), electroporation, infection, etc., so that many copies of the gene sequence are generated. Preferably, the cloned gene is contained on a shuttle vector plasmid, which provides for expansion in a cloning cell, e.g., E. coli, and facile purification for subsequent insertion into an appropriate expression cell line, if such is desired. For example, a shuttle vector, which is a vector that can replicate in more than one type of organism, can be prepared for replication in both E. coli and Saccharomyces cerevisiae by linking sequences from an E. coli plasmid with sequences form the yeast 2m plasmid.

Transgenic plants containing the polynucleotides described herein are also within the scope of the invention. Methods for introducing exogenous polynucleotides into plant cells and regenerating transgenic plants are well known. Some are provided below.

In one embodiment, to introduce a plasmid containing a CelA coding sequence or promoter of the invention into a plant, a 1:1 mixture of plasmid DNA containing a selectable marker expression cassette and plasmid DNA containing a cellulose synthase expression cassette is precipitated with gold to form microprojectiles. The microprojectiles are rinsed in absolute ethanol and aliquots are dried onto a suitable macrocarrier such as the macrocarrier available from BioRad in Hercules, Calif. Prior to bombardment, embryogenic tissue is preferably desiccated under a sterile laminar-flow hood. The desiccated tissue is transferred to a semi-solid proliferation medium. The prepared microprojectiles are accelerated from the macrocarrier into the desiccated target cells using a suitable apparatus such as a BioRad PDS-1000/HE particle gun. In a preferred method, each plate is bombarded once, rotated 180 degrees, and bombarded a second time. Preferred bombardment parameters
are 1350 psi rupture disc pressure, 6 mm distance from the rupture disc to macrocarrier (gap distance), 1 cm macrocarrier travel distance, and 10 cm distance from macrocarrier stopping screen to culture plate (macrocarrier travel distance). Tissue is then transferred to semi-solid proliferation medium containing a selection agent, such as hygromycin B, for two days after bombardment.

**Cellulose Synthase Protein and Fragment Thereof**

A cellulose synthase of the invention is a plant protein that contains a catalytic subunit which has UDP-glucose binding activity for the synthesis of glucan from glucose, and eight transmembrane domains for localizing the cellulose synthase to the cell membrane. The cellulose synthase of the invention has eight transmembrane binding domains; two at the amino terminal and six at the carboxyl terminal. The UDP-glucose binding domain is located between transmembrane domains two and three. Examples of this protein structure are seen in the aspen cellulose synthase as well as in those of RSW1 and GhCelA. The location of the transmembrane domain may be identified as described below and as exemplified in the Example. Preferably, the cellulose synthase of the invention has an amino acid sequence of a tree cellulose synthase.

In one embodiment, the cellulose synthase protein of the invention is isolated from aspen. Aspen cellulose synthase contains about 978 amino acids and has a molecular weight of about 110 KDa and a pI of about 6.58. In one embodiment, the aspen cellulose synthase has the amino acid sequence of SEQ ID NO:2 as represented in FIG. 1. In another aspect, the invention relates to cellulose synthase of SEQ ID NO: 5.

The invention further relates to fragments of plant cellulose synthases, such as fragments containing at least one transmembrane region and/or a UDP-glucose binding domain. The transmembrane regions may be identified as described in the Example by using the method of Hoffman and Stoffel (1993). The cellulose synthase fragment containing the UDP-glucose binding domain is functional without the presence of the rest of the protein. This separable activity is as shown in the Example. This result was surprising and unexpected because previously identified UDP-glucose binding domains were not known to be functional when isolated from other portions of the protein. Thus, a fragment of any cellulose synthase (such as PtCelA, RSW1, GhCelA and SEQ ID NO:5) that contains a UDP-glucose binding domain and is independently functional is within the scope of the invention. The function of the UDP-glucose binding domain may be determined using the assay described in the Example. The UDP-glucose binding domain of the invention is located between the second and third transmembrane region of the cellulose synthase and has conserved amino acid sequences for UDP-glucose binding, such as the sequence QVLRW and conserved D residues. The UDP-glucose binding domain and the conserved regions therein may be located in a cellulose synthase using the guidance of the present specification and the general knowledge in the art, for example Brown, 1996. In one embodiment, the UDP-glucose binding domain and the conserved regions therein may be identified by comparing the amino acid sequence of cellulose synthase of interest with the amino acid sequence of aspen cellulose synthase using the algorithms described in the specification or generally known in the art. For example, the UDP-glucose binding domain of SEQ ID NO:2 is in the position amino acids 220 to 749. The conserved QVLRW sequence is located at positions 715—719 of SEQ ID NO:2.

Polypeptides having at least 75%, preferably at least 85% and most preferably at least 95% similarity to the amino acid sequence of SEQ ID NO: 2, amino acids 220—749 of SEQ ID NO:2, SEQ ID NO:5 or its UDP-glucose binding domain using Power Blast or GAP algorithm described above. In a preferred embodiment, these polypeptides are of about the same length as the polypeptide of SEQ ID NO: 2 or amino acids 220—749 of SEQ ID NO:2. For example, the polypeptide may be from about 2—3 to about 5—7 and to about 10—15 amino acids longer or shorter. In another embodiment, the polypeptides described in this paragraph are not originally found (i.e., naturally occurring) in Arabidopsis or cotton. These polypeptides may be prepared by, for example, altering the nucleic acid sequence of a cloned nucleotide encoding the protein of SEQ ID NO:2 or SEQ ID NO:5 using the methods well known in the art.

Function conservative variants of cellulose synthase are also within the scope of the invention and can be prepared by altering the sequence of a cloned nucleotide encoding cellulose synthase or fragments thereof. Conventional methods used in the art can be used to make substitutions, additions or deletions in one or more amino acids, to provide functionally equivalent molecules. For example, a function conservative variant that has substitutions, deletions and/or additions in the amino and/or carboxyl terminus of the protein, outside of the UDP-glucose binding domain is within the scope of the invention. Preferably, variants are made that have enhanced or increased functional activity relative to native cellulose synthase. Methods of directed evolution can be used for this purpose.

The invention also includes function conservative variants which include altered sequences in which functionally equivalent amino acid residues are substituted for residues within the sequence resulting in a conservative amino acid substitution. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity, which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example, the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. Amino acids containing aromatic ring structures are phenylalanine, tryptophan, and tyrosine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. Such alterations will not be expected to affect apparent molecular weight as determined by polyacrylamide gel electrophoresis, or isoelectric point. Particularly preferred substitutions are: (i) Lys for Arg and vice versa such that a positive charge may be maintained; (ii) Glu for Asp and vice versa such that a negative charge may be maintained; (iii) Ser for Thr such that a free —OH can be maintained; and (iv) Gin for Asn such that a free CONH2 can be maintained. Amino acid substitutions may also be introduced to substitute an amino acid with a particularly preferable property. For example, a Cys may be introduced a potential site for disulfide bridges with another Cys. A His may be introduced as a particularly catalytic site (i.e., His can act as an acid or base and is the most common amino acid in biochemical catalysis). Pro may be introduced because of its particularly planar structure, which induces b-turns in the protein’s structure.
The cellulose synthase of the invention can be isolated by expressing a cloned polynucleotide encoding the cellulose synthase as well as using direct protein purification techniques. These methods will be apparent to those of skill in the art.

Polynucleotides Containing Cellulose Synthase Promoter

The present invention further relates to a cellulose synthase promoter. The promoter is a stress-inducible promoter and may be used to synthesize greater quantities of high crystalline cellulose in plant, and preferably in trees. This permits an increase in the proportion of cellulose in transgenic plants, greater strength of juvenile wood and fiber, and acceleration of overall growth rate.

In one embodiment, the promoter of the invention is from aspen and is represented in FIG. 4. The promoter sequence is located within the region of nucleotides 1-840 of SEQ ID NO:3. A person of skill in the art will appreciate that not the entire sequence is required for the promoter function and can easily identify the critical regions by looking for conserve boxes and doing routine deletion analysis. Thus, functional fragments of SEQ ID NO:1 are within the scope of the invention.

Polynucleotides that hybridize under conditions of low, medium, and high stringency to SEQ ID NO:3, and its non-coding portion are also within the scope of the invention. The hybridizable polynucleotide may be about the same length as the sequence to which it hybridizes, for example, not more than about 10 to about 20 nucleotides longer or shorter. In another embodiment, the hybridizable polynucleotide is at least about 200 nucleotides long, preferably at least about 400 nucleotides long and most preferably at least 500 nucleotides long. In yet another embodiment, the hybridizable polynucleotide comprises at least one MSRE element identified according to the method described below.

A cellulose synthase promoter of the invention typically provides tissue-specific gene regulation in xylem, but also permits up-regulation of gene expression in other tissues as well, e.g., phloem under tension stress. Furthermore, expression of cellulose synthase is localized to an area of the plant under stress.

This stress-inducible phenomenon is regulated by positive and negative mechanical stress response elements (MSREs). These MSREs upregulate (positive) or downregulate (negative) the expression of a cellulose synthase polynucleotide under stress conditions through binding of transcription factors. MSRE-regulated expression of cellulose synthase permits synthesis of cellulose with high crystallinity.

The MSREs of cellulose synthase can be modified or employed otherwise in methods to regulate expression of a polynucleotide, including a cellulose synthase, operatively linked to a promoter containing an MSRE in response to mechanical stress (e.g., tension or compression) to a transgenic plant.

Negative MSREs of a cellulose synthase promoter can be modified, removed or blocked to improve expression of a cellulose synthase, and thereby increase cellulose production and improve wood quality. Alternatively, positive MSREs can be removed or blocked to decrease expression of a cellulose synthase, which decreases cellulose production and increases lignin deposition. This is useful for increasing the fuel value of wood because lignin has a higher BTU value than cellulose. Moreover, a modified cellulose synthase promoter can be operatively linked to a polynucleotide of interest to control its expression upon mechanical stress to a plant harboring it.

The location of MSRE elements in the SEQ ID NO:5 may be identified, for example, using promoter deletion analysis, DNase Foot Print Analysis, and Southwestern screening of an expression library for an MSRE. In one embodiment, cellulose synthase promoter that has one or more portions deleted, and is operatively linked to a reporter sequence, is introduced into a plant or a plant cell. A positive MSRE is detected by observing no relative change or increase in the amount of reporter in a transgenic plant or tissue, e.g., phloem after inducing a stress to the plant, and a negative MSRE is detected by observing increases in the amount of reporter in the plant in the absence of any stress to the plant. A positive element is detected when by removing it, GUS expression goes down and by adding it kept at the same level or more. The negative element does not support, or suppress, expression of GUS and by removing it, normal or enhanced GUS expression is observed as compared to when negative element is present.

Manipulation of a MSRE binding sites and/or providing transcription factors that bind thereto, provides a mechanism to continuously produce high crystalline cellulose in woody plant cell walls of transgenic plants. For example, one having ordinary skill in the art can delete or block negative MSRE elements, or provide cDNA encoding protein(s) that bind the positive MSREs, to enable constitutive expression of a cellulose synthase without the requirement of a mechanical stress. The increased cellulose synthase, and therefore, increased cellulose content, can improve the strength properties of juvenile wood and fiber. It is also contemplated that the positive MSREs can be deleted or blocked, or cDNA in an antisense direction, which in the sense direction encodes a protein that binds a positive MSRE, can be provided, to reduce cellulose synthase activity and decrease cellulose production.

Method of Isolating Polynucleotides Encoding Cellulose Synthase

The invention further relates to identifying and isolating polynucleotides encoding cellulose synthase in plants, e.g., trees, (in addition to those polynucleotides provided in the Example, and represented in FIG. 1 and FIG. 7). These polynucleotides may be used to manipulate expression of cellulose synthase with an objective to improve the cellulose content and properties of wood.

The method comprises identifying a nucleic acid fragment containing a sequence encoding cellulose synthase or a portion thereof by using a fragment of SEQ ID NO:1 or 4 as a probe or a primer. Once identified, the nucleic acid fragment containing a sequence encoding cellulose synthase or a portion thereof is isolated.

Polynucleotides encoding cellulose synthases of the invention, whether genomic DNA, cDNA, or fragments thereof, can be isolated from many sources, particularly from cDNA or genomic libraries from plants, preferably trees (e.g. aspen, sweetgum, loblolly pine, eucalyptus, and other angiosperms and gymnosperms). Molecular biology methods for obtaining polynucleotides encoding a cellulose synthase are well known in the art, as described above (see, e.g., Sambrook et al., 1989, supra).

Accordingly, cells from any species of plant can potentially serve as a nucleic acid source for the molecular cloning of a polynucleotide encoding a cellulose synthase of the invention. The DNA may be obtained by standard procedures known in the art from cloned DNA (e.g., a DNA "library"), and preferably is obtained from a cDNA library prepared from tissues with high level expression of a cellulose synthase (e.g., xylem tissue, since cells in this tissue
evidence very high levels of expression of CelA), by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from a desired cell (see, for example, Sambrook et al., 1989, supra; Glover, D. M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II). Clones derived from genomic DNA may contain regulatory and intron DNA regions in addition to coding regions; clones derived from cDNA will not contain intron sequences. Whatever the source, a polynucleotide should be molecularly cloned into a suitable vector for its propagation.

In another embodiment for the molecular cloning of a polynucleotide encoding a cellulose synthase of the invention from genomic DNA, DNA fragments are generated from a genome of interest, such as from a plant, or more particularly a tree genome, part of which will correspond to a desired polynucleotide. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNase I in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing a desired CelA sequence may be accomplished in a number of ways. For example, if an amount of a portion of a CelA sequence or its specific RNA, or a fragment thereof, is available and can be purified and labeled, the generated DNA fragments may be screened by nucleic acid hybridization to a labeled probe (Benton and Davis, 1977, Science 196:180; Gunstein and Hogness, 1975, Proc. Natl. Acad. Sci. U.S.A. 72:3961). For example, a set of oligonucleotides corresponding to the partial amino acid sequence information obtained for a CelA protein from trees can be prepared and used as probes for DNA encoding cellulose synthase, or as primers for cDNA or mRNA (e.g., in combination with a poly-T primer for RT-PCR). Preferably, a fragment is selected that is highly unique to a cellulose synthase of the invention, such as the UDP-glucose binding regions. Those DNA fragments with substantial homology to the probe will hybridize. As noted above, the greater the degree of homology, the more stringent hybridization conditions can be used. In a specific embodiment, stringency hybridization conditions can be used to identify homologous CelA sequences from trees or other plants.

Thus, in one embodiment, a labeled cellulose synthase cDNA from, e.g., *Populus tremuloides* (PtCelA), can be used to probe a library of genes or DNA fragments from various species of plants, especially angiosperm and gymnosperm, to determine whether any bind to a CelA of the invention. Once genes or fragments are identified, they can be amplified using standard PCR techniques, cloned into a vector, e.g., pBluescript vector (Stratagene of LaJolla, Calif.), and transformed into a bacteria, e.g., DH5 E. coli strain (Gibco BRL of Gaithersburg, Md.). Bacterial colonies are typically tested to determine whether any contains a cellulose synthase-encoding nucleic acid. Once a positive clone is identified through binding, it is sequenced from an origin of replication. The elements are collectively termed herein a “promoter.” Thus, a nucleic acid encoding CelA of the invention can be operatively associated with a promoter in an expression vector of the invention.

Further selection can be carried out on the basis of the properties of the gene, e.g., if the gene encodes a protein product having the isoelectric, electrophoretic, hydrophobicity, amino acid composition, or partial amino acid sequence of a cellulose synthase protein of the invention, as described herein. Thus, the presence of the gene may be detected by assays based on the physical, chemical, or immunological properties of its expressed product. For example, cDNA clones or DNA clones which hybrid-select the proper mRNAs can be used to produce a protein that has similar properties known for cellulose synthases of the invention.

Such properties may include, for example, similar or identical electrophoretic migration patterns, isoelectric focusing or non-equilibrium pH gel electrophoresis behavior, proteolytic digestion maps, hydrophobicity plots, or functional properties (such as isolated, functional UDP-glucose binding domains). A cellulose synthase polynucleotide of the invention can also be identified by mRNA selection, i.e., by nucleic acid hybridization followed by in vitro translation. In this procedure, nucleotide fragments are used to isolate complementary mRNAs by hybridization. Such DNA fragments may represent available, purified CelA DNA, or may be synthetic oligonucleotides designed from the partial amino acid sequence information. Functional assays (e.g., UDP-glucose activity) of the in vitro translation products of the products of the isolated mRNAs identifies the mRNA and, therefore, the complementary DNA fragments, that contain the desired sequences.

A radiolabeled CelA cDNA can be synthesized using a selected mRNA as a template. The radiolabeled mRNA or cDNA may then be used as a probe to identify homologous CelA DNA fragments from amongst other genomic DNA fragments. It will be appreciated that other polynucleotides, in addition to a CelA of the invention can be operatively linked to a CelA promoter to control expression of the polynucleotide upon application of a mechanical stress.

Expression of CelA Polypeptides

The nucleotide sequence coding for CelA, or a functional fragment, derivative or analog thereof, including chimeric proteins, can be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted protein-coding sequence. Preferably, an expression vector includes an origin of replication. The elements are collectively termed herein a “promoter.” Thus, a nucleic acid encoding CelA of the invention can be operatively associated with a promoter in an expression vector of the invention. Both cDNA and genomic sequences can be cloned and expressed under control of such regulatory sequences. The necessary transcriptional and translational signals can be provided on a recombinant expression vector, or they may be supplied by the native gene encoding CelA and/or its flanking regions.

In addition to a CelAP, expression of cellulose synthase can be controlled by any promoter/enhancer element known in the art, but these regulatory elements must be functional in the host selected for expression. Promoters which may be used to control CelA polynucleotide expression include, constitutive, development-specific and tissue-specific. Examples of these promoters include 35S Cauliflower Mosaic Virus, terminal flower and 4CL-1. Thus, there are
various ways to alter the growth of a plant using different promoters, depending on the needs of the practitioner.

The nucleotide sequence may be inserted in a sense or antisense direction depending on the needs of the practitioner. For example, if augmentation of cellulose biosynthesis is desired then polynucleotides encoding, e.g., cellulose synthase, can be inserted into the expression vector in the sense direction to increase cellulose synthase production and thus cellulose biosynthesis. Alternatively, if it is desired that cellulose biosynthesis is reduced or lignin content is increased, then polynucleotides encoding, e.g., cellulose synthase, can be inserted in the antisense direction so that upon transcription the antisense mRNA hybridizes to other complementary transcripts in the sense orientation to prevent translation. In other embodiments, the polynucleotide encodes a UDP-glucose binding domain and is used in a similar manner as described.

A recombinant CelA protein of the invention, or functional fragment, derivative, chimeric construct, or analog thereof, may be expressed chromosomally, after integration of the coding sequence by recombination. In this regard, any of a number of amplification systems for plants may be used to achieve high levels of stable gene expression, as discussed above. Any of the methods previously described for the insertion of DNA fragments into a cloning vector may be used to construct expression vectors containing a gene consisting of appropriate transcriptional/translational control signals and the protein coding sequences. These methods may include in vitro recombinant DNA and synthetic techniques and in vivo recombination (genetic recombination).

Expression vectors containing a nucleic acid encoding a CelA of the invention can be identified by four general approaches: (a) PCR amplification of the desired plasmid DNA or specific mRNA, (b) nucleic acid hybridization, (c) presence or absence of selection marker gene functions, (d) analyses with appropriate restriction endonucleases, and (e) expression of inserted sequences. In the first approach, the nucleic acids can be amplified by PCR to provide for detection of the amplified product. In the second approach, the presence of a foreign gene inserted in an expression vector can be detected by nucleic acid hybridization using probes comprising sequences that are homologous to an inserted marker gene. In the third approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain “selection marker” gene functions (e.g., ß-glucuronidase activity, resistance to antibiotics, transformation phenotype, etc.) caused by the insertion of foreign genes in the vector. In another example, if the nucleic acid encoding CelA is inserted within the “selection marker” gene sequence of the vector, recombinants containing the CelA insert can be identified by the absence of the CelA gene function. In the fourth approach, recombinant expression vectors are identified by digestion with appropriate restriction enzymes. In the fifth approach, recombinant expression vectors can be identified by assaying for the activity, biochemical, or immunological characteristics of the gene product expressed by the recombinant, provided that the expressed protein assumes a functionally active conformation.

After a particular recombinant DNA molecule is identified and isolated, several methods known in the art may be used to propagate it. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As previously explained, the expression vectors which can be used include, but are not limited to those vectors or their derivatives described above.

Vectors are introduced into the desired host cells by methods known in the art, e.g., Agrobacterium-mediated transformation (described in greater detail below), transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (liysosome fusion), use of a gene gun, or a DNA vector transporter (see, e.g., Wu et al., 1992, J. Biol. Chem. 267:963–967; Wu and Wu, 1988, J. Biol. Chem. 263:14621–14624; Hartmut et al., Canadian Patent Application No. 2,012,311, filed Mar. 15, 1990).

The cell into which the recombinant vector comprising the nucleic acid encoding CelA is cultured in an appropriate cell culture medium under conditions that provide for expression of CelA by the cell. In addition, a host cell strain may be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific fashion desired. Different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (such as glycosylation, cleavage, e.g., of a signal sequence) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed.

Agrobacterium-Mediated Transformation and Inducing Somatic Embryos

The culture media used in the invention, and for transforming Agrobacterium, contain an effective amount of each of the medium components (e.g. basal medium, growth regulator, carbon source) described above. As used in describing the present invention, an “effective amount” of a given medium component is the amount necessary to cause a desired effect. For example, an effective amount of a growth hormone in the primary callus growth medium is the amount of the growth hormone that induces callus formation when combined with other medium components. Other compounds known to be useful for tissue culture media, such as vitamins and gelling agents, may also be used as optional components of the culture media of the invention.

Transformation of cells from plants, e.g., trees, and the subsequent production of transgenic plants using Agrobacterium-mediated transformation procedures known in the art, and further described herein, is one example of a method for introducing a foreign gene into trees. Transgenic plants may be produced by various methods, such as by the following steps: (i) culturing Agrobacterium in low-pH induction medium at low temperature and preconditioning, i.e., coculturing bacteria with wounded tobacco leaf extract in order to induce a high level of expression of the Agrobacterium vir genes whose products are involved in the Ti-DNA transfer; (ii) coculturing a 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 (v) converting the embryos into plantlets.

Any non-tumorigenic A. tumefaciens strain harboring a disarmed Ti plasmid may be used in the method of 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 resis-
The well-defined critical factors for inducing virulence activ­
most preferably about 5.2, and at low temperature such as
with tobacco extract and the vir genes involved in the
bacteria vir gene expression and infectivity. Preconditioning
wounded leaf tissues of tobacco plants grown in vitro. To
particular needs.

To increase the infectivity of the bacteria, Agrobacterium
is cultured in low-pH induction medium, i.e., any bacterium
culture media with a pH value adjusted to from 4.5 to 6.0,
most preferably about 5.2, and at low temperature such as
for example about 19–30EC, preferably about 21–26EC.
The conditions of low-pH and low temperature are among
the well-defined critical factors for inducing virulence activity
in Agrobacterium (e.g., Altunbor et al., Mol. Plant-
Microbe. Interac. 2: 301, 1989; Fullner et al., Science 273:
1107, 1996; Fullner and Nester, J. Bacteriol. 178: 1498,
1996).

The bacteria is preconditioned by coculturing with
wounded tobacco leaf extract (prepared according to meth­
ods known generally known 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 Agrobac­
terium 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 cocul­
tured 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 invention also relates to methods of altering the
growth of a plant by expressing the polynucleotide of the
invention, which as a result alters the growth of the plant.
The polynucleotide used in the method may be a homolo-
gous polynucleotide or a heterologous polynucleotide and
are described in detail above. For example, both full-length
and UDP-glucose binding region containing fragments may
be expressed. Additionally, depending on the aim of the
method, the polynucleotide may be introduced into the plant
in the sense or in the antisense orientation. Any suitable
promoter may be used to provide expression. The promoter
or a functional fragment thereof is operatively linked to the
polynucleotide. The promoter may be a constitutive pro-
moter, a tissue-specific promoter or a development-specific
plant promoter. Examples of suitable promoters are Cauli-
flower Mosaic Virus 35S, 4CL, cellulose synthase promoter,
PtCelAP and terminal flower promoter.
EXAMPLE

Molecular Cloning of Cellulose Synthase

This Example describes the first tree cellulose synthase cDNA (PtCelA, GenBank No. AF072131) cloned from developing secondary xylem of aspen trees using RSW1 cDNA.

Prior to the present invention, only partial clones of cellulose synthases from crop species and cotton GhCelA have been discovered, which have significant homology to each other. The present inventors have discovered and cloned a new full-length cellulose synthase cDNA, AraxCelA (GenBank No. AF062485) (FIG. 7, [SEQ ID NO: 4]), from an Arabidopsis primary library. AraxCelA is a new member of cellulose synthase and shows 63–85% identity and 72–90% similarity in amino acid sequence with other Arabidopsis CelA members.

Another cellulose synthase was cloned in aspen using a 32P-labeled 1.65 kb long EcoRI fragment of Arabidopsis CelA cDNA, which encodes a centrally located UDP-glucose binding domain, was used as a probe to screen about 500,000 pfu of a developing xylem cDNA library from aspen (Populus tremuloides) (Ge and Chiang, 1996). Four positive clones were obtained after three rounds of plaque purification. Sequencing the 3' ends of these four cDNAs showed that they were identical clones. The longest cDNA clone was fully sequenced and determined to be a full-length cDNA having a 3232 bp nucleotide sequence (FIG. 1) [SEQ ID NO: 1], which encodes a protein of 978 amino acids [SEQ ID NO: 2].

Characterization of a Cellulose Synthase from Aspen

The first AUG codon of PtCelA was in the optimum context for initiation of transcription on the basis of optimal context sequence described by Joshi (1987a) and Joshi et al. (1997). A putative polyadenylation signal (AATATA) was found 16 bp upstream of a polyadenylated tail of 28 bp, which is similar to the proposed plant structure (Joshi, 1987b). The 5' untranslated leader was determined to have 68 bp and the 3' untranslated trailer was 227 bp. Both of these regions have a typical length observed in many plant genes (Joshi, 1987a and Joshi, 1987b). This cDNA clone exhibited 90% amino acid sequence similarity with cellulose synthase from cotton (GhCelA) and 71% with cellulose synthase from Arabidopsis (RSW1), suggesting that this particular tree homolog also encodes a cellulose synthase.

The full length cDNA was designated PtCelA, and encodes a 110,278 Da polypeptide having an isoelectric point (pI) of 6.58 and 8 charged molecules. The hydropathy curve indicated that this particular cellulose synthase has eight transmembrane binding domains; two at the amino terminal and six at the carboxyl terminal, using the method of Hoffman and Stoffel (1993). This protein structure is analogous to those of RSW1 and GhCelA. All of the conserved domains for UDP-glucose binding, such as QVI,RW and conserved D residues, are also present in a cellulose synthase of the invention, e.g., PtCelA (Brown et al., 1996). Thus, based on sequence and molecular analyses, it was concluded that PtCelA encodes a catalytic subunit which, like RSW1 in Arabidopsis, is essential for the cellulose biosynthesis machinery in aspen.

In situ localization of PtCelA mRNA transcripts along the developmental gradient defined by stem primary and secondary growth demonstrated that cellulose synthase expression is confined exclusively to developing xylem cells undergoing secondary wall thickening. This cell-type-specific nature of PtCelA gene expression was also consistent with xylem-specific activity of cellulose synthase promoter (PtCelAP) based on heterologous promoter-β-glucuronidase (GUS) fusion analysis. Overall, the results provide several lines of evidence that cellulose synthase is the gene primarily responsible for cellulose biosynthesis during secondary wall formation in woody xylem of trees, such as aspen. Previous results by the inventors (Hu et al., 1999) showed that cellulose and lignin are deposited in a compensatory fashion in wood. The discovery of a cellulose synthase in trees, such as aspen, permits the up-regulation of the protein to elevate cellulose production. Surprisingly, expression of CelA in trees suppressed lignin biosynthesis to further improve wood properties of trees.

Preparation of Transgenic Plants

The UDP-glucose binding sequence was subcloned into pBI121, which was used to prepare transgenic tobacco plants (Hu et al., 1998). The expression of a heterologous UDP-glucose binding sequence resulted in a remarkable growth-accelerating effect. This was surprising because current knowledge of the function of plant cellulose synthases teaches that a UDP-glucose sequence must remain intact with other functional domains in CelA, e.g., the transmembrane domains, in order for cellulose synthase to initiate cellulose biosynthesis. The remarkable growth and tremendous increase in plant biomass observed in transgenic tobacco was due likely to an augmented deposition of cellulose, indicating that the UDP-glucose domain alone is sufficient for genetic augmentation of cellulose biosynthesis in plants.

Genome Organization and Expression of a Novel Cellulose Synthase

To confirm that the cDNA clone of FIG. 1 [SEQ ID NO: 1] was a cellulose synthase, genomic Southern blot analysis was performed under both high and low stringency conditions using the cDNA. Genomic DNA from aspen was digested with PstI (lane P), HindIII (lane H) and EcoRI (lane E), and probed using a 1 kb 32P-labeled fragment from the 5' end of a cellulose synthase of FIG. 1. The Southern blot suggested the presence of a small family of cellulose synthase genes in aspen genome (FIG. 2, panels a and b). Repeated screening of the aspen xylem cDNA library with various plant CelA gene-related probes always resulted in the isolation of the same cellulose synthase cDNA clone. This suggested that the cellulose synthase cDNA cloned (FIG. 1) [SEQ ID NO: 1], represents the primary and most abundant cellulose synthase-encoding gene in developing xylem of trees, such as aspen, where active cellulose deposition takes place. It also indicates that manipulation of cellulose synthase gene expression can have a profound influence on cellulose biosynthesis in trees.

In Situ Hybridization

Northern blot analysis of total RNA from the internodes of aspen seedling stems (FIG. 2, panel c) using the labeled probe (as described above) revealed the near absence of cellulose synthase transcripts in tissues undergoing primary growth (internodes 1 to 4), and that the presence of cellulose synthase transcripts occurs during the secondary growth of stem tissues (internodes 5 to 11). However, weak northern
signals in primary growth may only suggest that cellulose synthase gene expression is specific to xylem, of which there is little in primary growth tissue.

Xylogenesis in higher plants offers a unique model that involves sequential execution of cambium cell division, commitment to xylem cell differentiation, and culmination in xylem cell death (Fukuda, 1996). Although primary and secondary xylem cells originate from different types of cambia, namely procambium and inter/intraxylary cambium, both exhibit conspicuous secondary wall development with massive cellulose and lignin deposition (Easu, K., 1960, Anatomy of Seed Plants, New York: John Wiley and Sons). To further investigate spatial and temporal cellulose synthase gene expression patterns at the cellular level, in situ hybridization was used to localize cellulose synthase mRNA along the developmental gradient defined by stem primary and secondary growth.

Localization of cellulose synthase gene transcripts (RNA) in stem at various growth stages was also observed. FIG. 3 shows transverse sections from 2nd, 4th and 6th internodes hybridized with digoxigenin (DIG)-labeled cellulose synthase antisense or sense (control) RNA probes, as described. PtCelA transcripts were detected in young aspen stem sections by in situ hybridization with transcripts of highly variable 5N region of PtCelA cDNA (a 771 bp long fragment generated from Pst1 and Sac1). This region was first subcloned in the plasmid vector, pGEM,-3Zf (+) (Promega) for the production of digoxigenin (DIG)-labeled transcripts using T7 (for antisense transcripts) and SP6 (for sense transcripts) RNA polymerase (DIG system: Boehringer Mannheim). Probes were subjected to mild alkaline hydrolysis by incubation in 100 mM NaHCO3, pH 10.2 at 60° C., which produced approximately 200 bp fragments.

Aspen young stems were prepared for sectioning by fixation in 4% (w/v) paraformaldehyde in 100 mM phosphate buffer (pH 7.0) at 4° C. overnight, dehydrated through an ethanol series on ice, and embedded in Paraplast medium (Sigma). Ten μm sections were mounted on Superfrost/plus (Fisher) slides at 42° C. overnight, dewaxed and then rehydrated through a descending ethanol series. The sections were incubated with proteinase K (10 μg/ml in 100 mM Tris-HCl, 50 mM EDTA, pH 7.5) for 30 min and were post-fixed with FAA. The sections were acetylated with 0.33% (v/v) acetic anhydride in 0.1 M triethanolamine-HCl (pH 8.0) prior to hybridization. The sections were then incubated in a hybridization mixture (approximately 2 μg/ml DIG-labeled probes, 50% (v/v) formamide, 2xSSPE, 10% (w/v) dextran sulfate, 125 μg/ml tRNA, pH 7.5) at 45° C. for 12–16 hrs. Nonhybridized single-stranded RNA probe was removed by treatment with 20 μg/ml RNase A in TE buffer with 500 mM NaCl. The sections were washed at 50° C.

Hybridized DIG-labeled probe was detected on sections using anti-digoxigenin antiserum at a 1:1500 dilution, as described in the manufacturer’s instruction (DIG system: Boehringer Mannheim). Sections were examined by Eclipse 400 light microscope (Nikon) and photographed.

During the primary growth stage (FIG. 3, panels a and b), strong expression of cellulose synthase was found localized exclusively to primary xylem (PX) cells. At this stage, young internodes are elongating, resulting in thickening of primary xylem cells through formation of secondary walls (Easu, K., 1960, Anatomy of Seed Plants, New York: John Wiley and Sons). The concurrence of shoot elongation with high expression of cellulose synthase strongly suggests the association of cellulose synthase protein with secondary cell wall cellulose synthesis. Later stages of primary growth (FIG. 3, panel b) are characterized by the appearance of an orderly alignment of primary xylem cells. Active cellulose biosynthesis accompanies cell elongation-induced wall thickening, as indicated by the strong expression of cellulose synthase in these primary xylem cells.

At the beginning of secondary growth in older internodes, it was observed that expression of cellulose synthase is also exclusively localized to xylem cells (FIG. 3, panel c). Instead of elongation in internodes distal to the meristematic activity, growth at this stage is mainly radial due to thickening in secondary cell walls of secondary xylem. At the same time, expression of PtCelA gene becomes localized to the secondary developing xylem cells (SX in FIG. 3, panel c), which is again consistent with the idea that PtCelA encodes a secondary cell wall cellulose synthase. At this stage, secondary xylem cells cover the elongated and differentiated primary xylem cells in which PtCelA gene expression is no longer detectable (FIG. 3, panel c). These results demonstrate that expression of PtCelA gene is xylem-specific and the cellulose synthase of FIG. 1 [SEQ ID NO: 1] encodes a cellulose synthase associated with cellulose biosynthesis in secondary walls of xylem cells. To further confirm xylem-specific expression of cellulose synthase, a cellulose synthase gene promoter sequence was cloned and characterized for regulatory activities.

Characterization of Expression Regulated by Cellulose Synthase Promoter

A 5N 1,200 bp cDNA fragment of a cellulose synthase of FIG. 1 [SEQ ID NO: 1] was used as a probe to screen an aspen genomic library for 5N regulatory sequences of a novel cellulose synthase gene, PtCelA. The library was constructed by cloning aspen genomic DNA fragments, generated from an Sau3AI partial-digest and sucrose gradient-separated, into the BamHI site of a Lambda DASH II vector (Stratagene, La Jolla, Calif.). Five positive clones were obtained from about 150,000 pfu and Lambda DNA was purified. One clone having about a 20 kb DNA insert size was selected for restriction mapping and partial sequencing. This resulted in the identification of a 5N flanking region of PtCelA gene of approximately 1 kb. This genomic fragment, designated PtCelAP (FIG. 4) [SEQ ID NO: 3], contained about 800 bp of promoter sequence, 68 bp of 5N end untranslated region and 160 bp of coding sequence. To investigate regulation of tissue-specific cellulose synthase expression at the cellular level, promoter activity was analyzed in transgenic tobacco plants by histochemical staining of a GUS protein. A PtCelAP-GUS fusion binary vector was constructed in pBI121 with the 35S promoter replaced with PtCelAP [SEQ ID NO: 3] and introduced into tobacco (Nicotiana tabacum) as per Hu et al. (1998).

Eleven independent transgenic lines harboring a CelAP-GUS fusion were generated. FIG. 5 shows a histochemical analysis of GUS expression driven by a cellulose synthase promoter of the invention in transgenic tobacco plants. Transverse sections from the 3rd (panel a), 5th (panel b), 7th (panel c), and 8th (panels d and f) internodes were stained from GUS activity, and fluorescence microscopy was used to visualize expression under UV radiation.

GUS staining was detected exclusively in xylem tissue of stems, roots and petioles. In stems, strong GUS activity was found localized to xylem cells undergoing primary (FIG. 5, panel a) and secondary growth (FIG. 5 panels b-d and f). GUS expression was confined to xylem cells in the primary growth stage and became more localized in developing secondary xylem cells during secondary growth. An entire
section from the 8th internode stained for GUS activity (FIG. 5, panel f). These results are consistent with the in vivo expression patterns of cellulose synthase in aspen stems. Lignin autofluorescence was visualized after UV radiation. Phloem fibers, which are also active in cellulose and lignin biosynthesis (FIG. 5, panels d and e), did not show GUS activity, suggesting that cellulose synthase gene expression is not associated with cellulose biosynthesis in cell types other than xylem. Examination of GUS activity in roots, stems, leaves, anthers and fruit also showed GUS expression in xylem tissue of all these organs suggesting that cellulose synthases of the invention are xylem-specific cellulose and expressed in all plant organs.

Characterization of promoter activity and cellular expression of a cellulose synthase of the invention from one particular source (aspen) indicated that expression produces a protein that encodes a secondary cell wall-specific cellulose synthase and is specifically compartmentalized in developing xylem cells. Characterization of the cellulose synthase gene promoter sequence not only confirms cell type-specific expression of cellulose synthase, but also provides a method for over-expressing cellulose synthase in a tissue-specific manner to augment cellulose production in xylem.

Expression of Cellulose Synthase Under Tension Stress

As described earlier, a cellulose synthase promoter of the invention is involved in a novel gene regulatory phenomenon of cellulose synthase. To further characterize a cellulose synthase of the invention, GUS expression driven by an aspen cellulose synthase promoter (PtCelAP) was observed in transgenic tobacco plants without or under tension stress. The stress was induced by bending and affixing the plants to maintain the bent position (e.g., tying) over a 40 hour period. Tangential and longitudinal sections were taken before bending, and 4 hrs, 20 hrs and 40 hrs after bending (panels a—d, respectively).

The cellulose synthase promoter-GUS fusion binary constructs showed exclusive xylem-specific expression of GUS without any tension stress (FIG. 6, panel a). However, under tension stress conditions endured by angiosperms in nature, the transgenic tobacco plants induced xylem and phloem-specific expression on the upper side of the stem within the first four hours of stress (FIG. 6, panel b).

This observation was surprising because during tension wood development fibers produce highly crystalline cellulose in order to provide essential mechanical strength to a bending stem. The present observation was the first showing of transcriptional up-regulation of a cellulose synthase, mediated through a cellulose synthase promoter that is directly responsible for development of highly crystalline cellulose in trees. Furthermore, after 20 hrs of tension stress, both xylem and phloem exhibited GUS expression, but only on the upper side of the stem that was under tensile stress, i.e., GUS expression on the lower side was inhibited (FIG. 6, panel c). With extended stress (up to 40 hrs), GUS expression was restricted to only one small region on the upper side of the stem where maximum tension stress was present (FIG. 6, panel d). Based on the observation of GUS signal in woody cells upon tension stress and the absence of GUS under compression or no stress, it was concluded that a cellulose synthase promoter of the invention has mechanical stress responsive elements (MSREs) that turn cellulose synthase genes on and off depending on the presence and type of stress to the stem.

The results indicate that positive MSREs exist in a cellulose synthase promoter of the invention to bind transcription factors in response to tension stress for regulating the expression of cellulose synthase and increasing biosynthesis of higher crystalline cellulose. This is evident based on the expression of GUS in xylem and phloem tissue at the upper side of the stem subjected to tension stress, but not when tissue on the lower side was subjected to compression or no stress. Furthermore, the tissue at the lower side of the stem, which was subjected to compression stress, showed no GUS expression, i.e., expression was turned off. This indicated the presence of negative MSREs, which bind transcription factors to turn off expression of cellulose synthase at the lower side of the stem. Negative MSREs likely suppress development of highly crystalline cellulose in normal wood.

These results provide a mechanism for genetically engineering synthesis of highly crystalline cellulose in juvenile wood for enhancing strength properties, and for synthesizing a higher percentage of cellulose in reaction wood. The positive MSREs and their cognate transcription factors are important in the synthesis of highly crystalline cellulose of high tensile strength, as are the negative MSREs and inhibition of cognate transcription factors thereof. The present invention thus provides a starting point for cloning cDNAs for the transcription factors that bind to positive and negative MSREs according to methods known in the art. Constitutive expression of cDNAs for positive MSRE transcription factors allows the continuous production of highly crystalline cellulose in transgenic trees, while expression of antisense cDNAs for negative MSRE transcription factors inhibits those transcription factors so that cellulose synthase cannot turn off. This combination will assure continuous production of highly crystalline cellulose in trees.

Genetic Engineering of Cellulose Synthase in Transgenic Plants

As discussed above, the nucleotide sequence of a cellulose synthase of the invention, e.g., PtCelA cDNA from aspen, shows significant homology with other polynucleotides encoding cellulose synthase proteins that have been suggested as authentic cellulose synthase clones. To further characterize the activity of a cellulose synthase, four constructs were prepared in a PBI121 plasmid.

1) A constitutive plant promoter Cauliflower mosaic Virus 35S was operatively linked to PtCelA (35SP-PtCelA-s) and overexpressed in transgenic plants. This causes excess production of cellulose, resulting in a reduction in lignin content. Tobacco and aspen have been transformed with this construct.

2) Cauliflower mosaic Virus 35S was operatively linked to antisense RNA from PtCelA (35S-PtCelA-a) and constitutively expressed to reduce production of cellulose and increase lignin content in transgenic plants. This negative control construct may not result in healthy plants since cellulose is essential for plant growth and development. Aspen plants have been transformed with this construct.

3) Aspen 4CL-1 promoter (Hu et al., 1998) was operatively linked to PtCelA (4CLP-PtCelA) (the 35S promoter of PBI121 was removed in this construct) and expressed in a tissue-specific manner in developing secondary xylem of transgenic aspen. This expression augments the native cellulose production and reduces lignin content of angiosperm tissues. Tobacco and aspen have transformed with this construct.
4) The cytoplasmic domain of \( \text{PtCelA} \) which contains three conserved regions thought to be involved in UDP-glucose binding during cellulose biosynthesis, was linked to a 35S promoter to produce binary constructs (35S-PtCelA UDP-glucose). Expression by this promoter permits constitutive expression of a UDP-glucose binding domain of \( \text{PtCelA} \) in transgenic plants. Tobacco and aspen have been transformed with this construct.

35S-GUS constructs (pBI121, ClonTech, CA) were used as controls for each experiment with the constructs. Transgenic tobacco plants were transformed with the constructs.

The following table shows the general growth measurements of the T0 tobacco plants. Plants carrying a \( \text{PtCelA} \) construct grew much faster than control plants carrying a pBI121 (control) construct. In comparing developmental 4CL and constitutive 35S promoter control of \( \text{PtCelA} \) expression, the 35S was more effective, permitting faster growth of transgenic tobacco plants. The fastest growth was seen in transgenic plants carrying a 35S promoter driven UDP-G domain from \( \text{PtCelA} \).

It is noted that T0 generation plants can have carry over effects from their tissue culture treatments. Therefore, seeds were collected for testing this growth phenomenon in T1 generations. The transgenic tobacco plants were analyzed for presence of the transferred genes and all tested positive for the respective gene constructs.

### TABLE

<table>
<thead>
<tr>
<th>Construct</th>
<th>Height</th>
<th>Diameter</th>
<th>Internode length</th>
<th>No. of leaves</th>
<th>Longest leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>35S-GUS</td>
<td>17</td>
<td>0.5</td>
<td>1</td>
<td>11</td>
<td>17</td>
</tr>
<tr>
<td>35S-PtCelA</td>
<td>77</td>
<td>1.0</td>
<td>6</td>
<td>13</td>
<td>37</td>
</tr>
<tr>
<td>35S-UDPG</td>
<td>83</td>
<td>1.0</td>
<td>6</td>
<td>13</td>
<td>37</td>
</tr>
<tr>
<td>4CLP-PtCelA</td>
<td>41</td>
<td>0.8</td>
<td>5</td>
<td>10</td>
<td>29</td>
</tr>
</tbody>
</table>

Note:
All values were measured in centimeters, excluding number of leaves.

It will be appreciated by persons of ordinary skill in the art that the examples and preferred embodiments herein are illustrative, and that the invention may be practiced in a variety of embodiments which share the same inventive concept.

## BIBLIOGRAPHY

Arioli et al., 1998, Science, 279: 717-720
Wu et al., 1998, Pl Physiol, 117: 1125
Hu et al., 1998, PNAS, 95: 5407-5412
Joshi et al., 1997, PMB, 35: 993-1001
Pear et al., 1996, PNAS, 93: 12637-12642
Haigler and Blanton, 1996, PNAS, 93: 12082-12085
Ge and Chiang, 1996, Pl Physiol, 112: 861
Delmer and Amor, 1995, Pl Cell, 7: 987-1000
Joshi, 1987, NAR, 15: 6643-6653
Joshi, 1987, NAR, 15: 9627-9640
Timmell, 1986, Compression Wood in Gymnopserms, Springer Verlag
Ensou, K., 1960, Anatomy of Seed Plants, New York: John Wiley and Sons
Higuchi, 1997, Biochemistry and Molecular Biology of Wood, Springer Verlag

### SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 7

<210> SEQ ID NO 1
<213> ORGANISM: Populus tremuloides
<220> FEATURES:
<221> NAME/KEY: CDS
<222> LOCATION: (69) . . (202)

<400> SEQUENCE: 1

```
GTCGACCCAC GCCTCCGTCT TGAAAGATA TGAAGTTGTA AAGAGCTGGT AAAGTGTTAA
TAAGCAAG ATG ATG GLU SER GLY ALA PRO LIE CYC HIS THR CYC GLY GLU
1  5
CAG GTG GGG CAT GAT GCA AAT GGG GAG ETA TTT GTG GET TGC CAT GAG
15 20 25 30
TGT AGT CCA CAT GAG CAT GGA CCA AAT GGG GAC CAT TTT GTG GTC TGC CAT GAG
45 50 55 60
```

```
taacgaag atg atg gaa ttc gta tcc ata tgg cat acc tgt gtt gaa
Net Met Glu Ser Gly Ala Pro Ile Cys His Thr Cys Gly Glu
1  5
CAG GTG GGG CAT GAT GCA AAT GGG GAC CAT TTT GTG GTC TGC CAT GAG
15 20 25 30
TGTT AGC TAT CCC ATG TGC AAC CTC TGT TGT GAC TTT GAA ATC AAT GAG
Cys Ser Tyr Pro Met Cys Tyr Ser Cys Phe Glu Phe Glu Ile Asn Glu
35 40 45
```
31 32

---continued

gcc tct gta gca aag gct cag aas aca cct gaa gaa tgg act atg Ala Leu Val Ala Lys Ala Gin Lys Thr Pro Glu Glu Gly Thr Thr Met 370 375 380 1214
caa gat gga aca ctt tgg cct ggg aat aac aca cct gat cac cct ggg Gln Asp Gly Thr Pro Thr Pro Gly Asn Asn Thr Arg His Pro Gly 385 390 395 1262
cat gat tca ggt ctt cct tgg gaa atc ctg gga gtt cgt gcg att gaa His Asp Ser Gly Leu Pro Thr Glu Ile Leu Gly Ala Arg Asp Ile Glu 400 405 410 1310
ggc tcc cag cac cac aas aag gct ggt gca gaa aat gct ctg ctg gaa Gly Tyr Gin His His Lys Asp Ala Gin Ala Leu Val Arg 435 440 445 1406
gtg tct gca gta ctc aca aat gct ccc tac atc ctc aat gtt gat tgt Val Ser Ala Val Leu Thr Ala Thr Ala Thr Leu Ala Asp Cys 450 455 460 1454
gat cac tat gta aac aat agc aag gct gt tgg cag gca atg tgg atc Asp His Tyr Val Glu Ala Ala Val Arg Met Cys Ile 465 470 475 1502
ctg atg gc cca aag gta ggt cga gta tgc tat gtc cag ttc Leu Met Ala Val Leu Thr Ala Thr Thr Leu Ala Asp Cys Tyr Val Gin Phe Pro 480 485 490 1550
cag agg ttc cat ggc ata gat aag agt gat cgc tac gcc aat cgt aas Gin Arg Phe Asp Gly Ile Asp Aep Arg Tyr Ala Ala Arg Aen 495 500 505 510 1598
gta gtt tcc ttt gat gtt aac atg aas ggg tgg tga gtt gac att cca gga Val Val Phe Asp Arg Val Asp Met Leu Asp Gly Ile Gin Gly 515 520 525 1646
cca gta tac gta gga act gtt tgt ttc aac agg cca gca ctt tac Pro Val Tyr Val Val Gly Thr Gly Cys Val Phe Asn Arg Gin Ala Leu Tyr 530 535 540 1694
ggc tac ggg cct cct tct atg ccc aag aac aag agt tct Gly Tyr Gly Pro Ser Met Pro Ser Leu Arg Lys Arg Aep Ser 545 550 555 1742
tca tcc tgc ttc tca tgc tgc ttc ccc tca aag aag aag cct gct ccc Ser Ser Phe Ser Cys Cys Cys Pro Ser Leu Lys Lys Pso Pro Ala Gin 560 565 570 1790
gac cgg cgt gaa tgg gta ago aas ago aag gag gat ctc tct catt gat Aep Pro Ala Val Tyr Arg Asp Ala Lys Glu Asp Leu Aen Ala 575 580 585 590 1838
ggc sta ttt aat ctt cca gag att gat aat tga gcg gag cct gga agg Ala Ile Phe Asp Leu Thr Glu Ile Asp Aep Tyr Asp His Glu Arg 595 600 605 1886
tca atg cgg cta atc cag tgg agc aag agg cct gaa att cca gtc Act Ser Met Leu Ile Ser Gin Leu Ser Phe Glu Lys Thr Phe Gly Leu Ser 610 615 620 1934
tct gtc tcc aat cag tga tgc cgc atg ctt gag cca ctt gag att cga gtt ctt Ser Val Phe Ile Glu Ser Thr Leu Met Asp Gly Leu Pro Glu 625 630 635 1982
tct ggc acc tca ccc ctc ttc gcc gag cag aat gaa gga cgg aat cca gcc Pro Arg Ser Pro Pro Phe Gin Ala Gin Val Ile Gin Gin Val Val Glu 640 645 650 2030
tgt ggc tac gta gca aag aag cgt gaa aag cag aat ggt gtc gca gct Cys Gly Tyr Glu Glu Thr Thr Leu Met Arg Aen Arg Arg Thr Thr 655 660 665 670 2078
tat ggg tca gta act gct gat ato tta act cgt gcc tcc aag atg ccc tgc
acaaaacctt ttgggaattgg aataatgatcc tcgttgtagt ttccctcaag aaagcacata

<210> SEQ ID NO 2
<211> LENGTH: 978
<212> TYPE: PRT
<213> ORGANISM: Populus tremuloides

<400> SEQUENCE: 2

Met Met Glu Ser Gly Ala Pro Ile Cys His Thr Cys Gly Glu Gln Val
1 5 10 15
Gly His Asp Ala Asn Gly Leu Phe Val Ala Cys His Gln Cys Ser
20 25 30
Tyr Pro Met Cys Lys Ser Cys Phe Glu Phe Glu Ile Asn Glu Gly Arg
35 40 45
Lys Val Cys Leu Arg Cys Gly Ser Pro Tyr Asp Glu Asn Leu Leu Asp
50 55 60
Asp Val Glu Lys Lys Gly Ser Gly Asn Gln Ser Thr Met Ala Ser His
65 70 75 80
Leu Asn Asp Ser Gln Asp Val Gly Ile His Ala Arg His Ile Ser Ser
85 90 95
Val Ser Thr Val Asp Ser Glu Met Asn Asp Glu Tyr Gly Asn Pro Ile
100 105 110
Trp Lys Asn Arg Val Lys Ser Cys Lys Asp Lys Glu Asn Lys Lys Lys
115 120 125
Lys Arg Ser Pro Lys Ala Glu Thr Glu Pro Ala Glu Val Pro Thr Glu
130 135 140
Gln Glu Met Glu Lys Pro Ser Ala Glu Ala Ser Glu Pro Leu Ser
145 150 155 160
Ile Val Tyr Pro Ile Pro Arg Asn Lys Leu Thr Pro Tyr Arg Ala Val
165 170 175
Ile Ile Met Arg Leu Val Ile Leu Gly Leu Phe Phe His Phe Arg Ile
180 185 190
Thr Asn Pro Val Asp Ser Ala Phe Gly Leu Trp Leu Thr Ser Val Ile
195 200 205
Cys Glu Ile Trp Phe Ala Phe Ser Trp Val Leu Asp Glu Phe Pro Lys
210 215 220
Trp Asn Pro Val Asn Arg Glu Thr Tyr Ile Glu Arg Leu Ser Ala Arg
225 230 235 240
Tyr Glu Arg Glu Gly Glu Pro Ser Gln Leu Ala Gly Val Asp Phe Phe
245 250 255
Val Ser Thr Val Asp Pro Leu Lys Glu Pro Leu Ile Thr Ala Asn
260 265 270
Thr Val Leu Ser Ile Leu Ala Val Asp Tyr Pro Val Asp Lys Val Ser
275 280 285
Cys Tyr Val Ser Asp Gly Ala Ala Met Leu Ser Phe Glu Ser Leu
290 295 300
Val Glu Thr Ala Glu Phe Ala Arg Lys Trp Val Pro Phe Cys Lys Lys
305 310 315 320
Phe Ser Ile Glu Pro Arg Ala Pro Glu Phe Tyr Phe Ser Glu Lys Ile
325 330 335
Thr Ile Pro Ala Val Cys Leu Leu Thr Gly Lys Phe Ile Ile Pro Thr
770 775 780
Leu Ser Asn Leu Ala Ser Met Leu Phe Leu Gly Leu Phe Ile Ser Ile
785 790 795 800
Ile Val Thr Ala Val Leu Glu Leu Arg Trp Ser Gly Val Ser Ile Glu
805 810 815
Asp Leu Trp Arg Asn Glu Phe Trp Val Ile Gly Gly Val Ser Ala
820 825 830
His Leu Phe Ala Val Phe Gly Phe Leu Lys Met Leu Ala Gly Ile
835 840 845
Asp Thr Asn Phe Thr Val Thr Ala Lys Ala Ala Glu Asp Ala Glu Phe
850 855 860
Gly Glu Leu Tyr Met Val Lys Trp Thr Thr Leu Leu Ile Pro Pro Thr
865 870 875 880
Thr Leu Leu Ile Ile Asn Met Ser Gly Cys Ala Gly Phe Ser Asp Ala
885 890 895
Leu Asn Lys Gly Tyr Glu Ala Trp Gly Pro Leu Phe Gly Lys Val Phe
900 905 910
Phe Ala Phe Trp Val Ile Leu His Leu Tyr Pro Phe Leu Lys Gly Leu
915 920 925
Met Gly Arg Glu Asn Leu Thr Pro Thr Ile Val Val Leu Trp Ser Val
930 935 940
Leu Leu Ala Ser Val Phe Ser Val Leu Val Trp Lys Ile Asn Pro Phe
945 950 955 960
Val Asn Lys Val Asp Asn Thr Leu Val Ala Glu Thr Cys Ile Ser Ile
965 970 975
Asp Cys

<210> SEQ ID NO 3
<211> LENGTH: 1010
<212> TYPE: DNA
<213> ORGANISM: Populus tremuloides
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (841) . . (1008)
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: 5' flanking region of PtCelA coding sequence

<400> SEQUENCE: 3

gaattcgcct tttggaattc agggagagt agtttccgt tcggtgatc gcggtttca
60
ccttcggtct aagcaatttg ccaagagatt acaaaaaa tgcatttatg tgaatattta
120
cacagtctg ttcttatgtg cactattta tcggcatct tagttctgg gttactcttt
180
atagtctgat tcgagattgc aacticccgt ttcggtcatt gaaattgatg acgataaaac
240
tgttaaattc gtagctatta gcggaccaac aaccagatg acggtacag cggctgaaaca
300
gagattcctg ttcggtcatt cttcttaatt ttcggttctg aagtgagagaa ttcctgat
360
acattccggt gatgatgat gatgatgag accatctgga tccagagaa accatctgga gacagctct
420
ggatcctgcc gttcctggtgc tcaagtctgca tcgtcgccca aatgtcgtcgt gttatagcc
480
taattctctt tttgtagcag ttgtaaattg gtagatgatg tcgtcgccca agccaccttac
540
cctaaacctc cagcaaatcc ttcaacacttg ctccaacacc accaataag acgtgctctc
600
acggtggctt tgtgaaacca tcagcaccac gocatacacc accaataacc acccctccac
**US 7,232,941 B2**  

---continued---

ctctctctg ccattacaa aastgcagta ccaccottctg aagacaccca aaccacacta 720
gcttggatta gggtagttta tataaaacca aaggggaacc aaaaatagg aaggaagcag 780
agggacacc ttctgaaga attgtagttg taatcagttg ttaaaaagct ttaataacgag 840
atg atg gaa tac ggg gct cct ata tcc gct ggt gat gaa cag gtt Met Met Glu Ser Gly Ala Pro ile Cys His Thr Cys Gly Glu Glu Val 1 5 10
ggg cat gat gca aat ggg gag cta ttt gtt gct gac cat gac gag agc Gyl His Ala Ala Gly Glu Val Ala Ala Cys His Glu Cys Ser 30 25 20

tat ccc atg tgg tct tgg cag gaa gac tgg gac gat ggc tag gaa ggc arg Tyr Pro Met Cys Lys Ser Cys Phe Val Glu Glu Ile Lys Gly Arg 936 35 40 45

aaa gtt tgc cgg tgt ggc tcg ag 1010
lys Val Cys Leu Arg Cys Gly Ser

---

<210> SEQ ID NO 4  
<211> LENGTH: 3444  
<212> TYPE: DNA  
<213> ORGANISM: Arabidopsis thaliana  

<400> SEQUENCE: 4

gcgccgcccgt ttaatgccc ggcttccttg ccatcagga aaggaatgag tttgtcctca 60
tgagatcgc cagatatgc cagcagcag gttgaggtg cagcagtagc aaaaatcgcag 120
agagagacgc gacattcag cgtcagcag caggttgggt gacattcagc aagcagctat 180
cctctgtgt agagtttatt cttgagagc aagagacggc gcaatcagc attcttcagc 240
gttgaaaaacc cttatacaag tctttaaagg gtaacagagg gtagag 300
gagacctt gtagaatag ctaatcattg tgaatcagtt aataattgg gattagttta 360
tcagtttttt cagtctattg cacatcctggc ctagccctaa caccctccag cgcagaagtt 420
gttgatcgc cccattgtct cttttaaacc atggcagctc cttcttgctc 480
gatatttttt gagataaag gctttccttg cttttatttttttcttgctctctctctctctcctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc
Arg Pro Arg Leu Ile Ala Gly Ser His Asn Arg Asn Glu Phe Val Leu
<table>
<thead>
<tr>
<th>No.</th>
<th></th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ile Asn Ala Asp Glu Asn Ala Arg Ile Arg Ser Val Gin Glu Leu Ser</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>Gly Glu Thr Cys Gln Ile Cys Arg Asp Glu Ile Glu Leu Thr Val Asp</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>Gly Glu Pro Phe Val Ala Cys Asn Glu Cys Ala Phe Pro Val Cys Arg</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>Pro Cys Tyr Glu Tyr Glu Arg Arg Glu Gly Asn Gin Ala Cys Pro Gin</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>5</td>
<td>Cys Lys Thr Arg Phe Lys Arg Lys Gin Ser Pro Arg Val Glu Gly</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>Asp Glu Glu Glu Asp Ile Asp Asp Leu Asp Asn Glu Phe Glu Tyr</td>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td>7</td>
<td>Gly Asn Asp Gly Ile Gly Phe Gin Val Ser Glu Gly Met Ser Ile</td>
<td>115</td>
<td>120</td>
</tr>
<tr>
<td>8</td>
<td>Ser Arg Arg Asp Ser Gin Val Ser Arg Leu Asp Ser Ala Pro</td>
<td>130</td>
<td>135</td>
</tr>
<tr>
<td>9</td>
<td>Pro Gly Ser Gin Ile Pro Leu Leu Thr Tyr Gly Asp Glu Asp Val Glu</td>
<td>145</td>
<td>150</td>
</tr>
<tr>
<td>10</td>
<td>Ile Ser Ser Asp Arg His Ala Leu Ile Val Pro Ser Leu Gly Gly</td>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td>11</td>
<td>His Gly Asn Arg Val His Pro Val Ser Leu Ser Asp Pro Thr Val Ala</td>
<td>180</td>
<td>185</td>
</tr>
<tr>
<td>12</td>
<td>Ala His Arg Arg Leu Met Val Pro Gin Lys Gin Gin Ala Val Tyr Gly</td>
<td>195</td>
<td>200</td>
</tr>
<tr>
<td>13</td>
<td>Tyr Gly Ser Val Al Glu Gin Gin Gin Val Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>210</td>
<td>215</td>
</tr>
<tr>
<td>14</td>
<td>Gin Asn Glu Lys Leu Gin Val Arg His Glu Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>225</td>
<td>230</td>
</tr>
<tr>
<td>15</td>
<td>Glu Asp Gly Asp Gin Ile Gin Val Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>240</td>
<td>245</td>
</tr>
<tr>
<td>16</td>
<td>Glu Leu Ser Met Lys Gin Val Val Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>17</td>
<td>Arg Met Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>260</td>
<td>265</td>
</tr>
<tr>
<td>18</td>
<td>Tyr Arg Ile Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>270</td>
<td>275</td>
</tr>
<tr>
<td>19</td>
<td>Ser Val Ile Cys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>20</td>
<td>Phe Pro Lys Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td>21</td>
<td>Gin Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>300</td>
<td>305</td>
</tr>
<tr>
<td>22</td>
<td>Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>23</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>320</td>
<td>325</td>
</tr>
<tr>
<td>24</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>25</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>340</td>
<td>345</td>
</tr>
<tr>
<td>26</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>350</td>
<td>355</td>
</tr>
<tr>
<td>27</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>28</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>370</td>
<td>375</td>
</tr>
<tr>
<td>29</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>380</td>
<td>385</td>
</tr>
<tr>
<td>30</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>31</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>400</td>
<td>405</td>
</tr>
<tr>
<td>32</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>33</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>34</td>
<td>Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>430</td>
<td>435</td>
</tr>
</tbody>
</table>
His Lys Met Asp Tyr Leu Lys Asn Val His Pro Ala Phe Val Arg
435 440 445
Glu Arg Arg Ala Met Lys Arg Asp Tyr Glu Phe Lys Val Lys Ile
450 455 460
Asn Ala Leu Val Ala Thr Ala Gln Lys Val Pro Glu Asp Gly Trp Thr
465 470 475 480
Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Ser Val Arg Asp His Pro
485 490 495 499
Gly Met Ile Gln Val Phe Leu Gly Ser Asp Gly Val Arg Asp Val Glu
500 505 510
Asn Asn Glu Leu Pro Arg Leu Val Tyr Val Ser Arg Glu Lys Arg Pro
515 520 525
Gly Phe Asp His His Lys Ala Gly Ala Met Asn Ser Leu Ile Arg
530 535 540
Val Ser Gly Val Leu Ser Asn Ala Pro Tyr Leu Leu Asn Val Asp Cys
545 550 555 560
Asp His Tyr Ile Asn Asn Ser Lys Ala Leu Arg Glu Ala Met Cys Phe
565 570 575
Met Met Asp Pro Gln Ser Gly Lys Ile Cys Tyr Val Gln Phe Pro
580 585 590
Gln Arg Phe Asp Gly Ile Asp Arg His Asp Tyr Ser Asn Arg Asn
595 600 605
Val Val Phe Phe Asp Ile Asn Met Lys Gly Leu Asp Gly Leu Gln Gly
610 615 620
Pro Ile Tyr Val Thr Gly Cys Val Phe Arg Arg Gln Ala Leu Tyr Gly
625 630 635 640
Phe Asp Ala Pro Lys Lys Lys Gly Pro Arg Lys Thr Cys Asn Cys
645 650 655
Trp Pro Lys Trp Cys Leu Leu Cys Phe Gly Ser Arg Lys Asn Arg Lys
660 665 670
Ala Lys Thr Val Ala Ala Asp Lys Lys Lys Asn Arg Glu Ala Ser
675 680 685
Lys Gln Ile His Ala Leu Gln Ile Glu Glu Gly Arg Gly His Lys
690 695 700
Val Leu Asn Val Glu Gln Ser Thr Glu Ala Met Gln Met Lys Leu Gln
705 710 715 720
Lys Lys Tyr Gly Gln Ser Pro Val Phe Val Ala Ser Ala Arg Leu Glu
725 730 735
Asn Gly Gly Met Ala Arg Asn Ala Ser Pro Ala Cys Leu Leu Lys Glu
740 745 750
Ala Ile Gln Val Ile Ser Arg Gly Tyr Glu Asp Lys Thr Glu Trp Gly
755 760 765
Lys Gln Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr
770 775 780
Gly Ser Lys Met His Ser His Gly Trp Arg His Val Tyr Cys Thr Pro
785 790 795 800
Lys Leu Ala Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg
805 810 815
Leu His Gln Val Leu Arg Trp Ala Leu Gly Ser Val Glu Ile Phe Leu
820 825 830
Ser Arg His Cys Pro Ile Trp Tyr Gly Tyr Gly Gly Leu Lys Trp
835 840 845
Leu Glu Arg Leu Ser Tyr Ile Asn Ser Val Val Tyr Pro Trp Thr Ser
850 855 860
Leu Pro Leu Ile Val Tyr Cys Ser Leu Pro Ala Ile Cys Leu Leu Thr
865 870 875 880
Gly Lys Phe Ile Val Pro Glu Ile Ser Asn Tyr Ala Ser Ile Leu Phe
885 890 895
Met Ala Leu Phe Ser Ser Ser Ile Ala Ile Thr Gly Ile Leu Glu Met Gln
900 905 910
Trp Gly Lys Val Gly Ile Asp Asp Trp Trp Arg Asn Glu Gln Phe Trp
915 920 925
Val Ile Gly Gly Val Ser Ala His Leu Phe Ala Leu Phe Gln Gly Leu
930 935 940
Leu Lys Val Leu Ala Gly Val Asp Thr Asn Phe Thr Val Thr Ser Lys
945 950 955 960
Ala Ala Asp Gly Glu Phe Ser Asp Leu Tyr Leu Phe Lys Trp Thr
965 970 975
Ser Leu Leu Ile Pro Pro Met Thr Leu Leu Ile Aen Val Ile Gly
980 985 990
Val Ile Val Gly Val Ser Asp Ala Ile Ser Aen Gly Tyr Asp Ser Trp
995 1000 1005
Gly Pro Leu Phe Gly Arg Leu Phe Ala Leu Trp Val Ile Ile His
1010 1015 1020 1025
Leu Tyr Pro Phe Leu Lys Gly Leu Leu Gly Lys Gin Asp Arg Met Pro
1030 1035 1040
Thr Ile Ile Val Val Trp Ser Ile Leu Ala Ser Ile Leu Thr Leu
1045 1050 1055
Leu Trp Val Arg Val Asn Pro Phe Val Ala Lys Gly Gly Pro Ile Leu
1060 1065 1070
Glu Ile Cys Gly Leu Asp Cys Leu
1075 1080

<210> SEQ ID NO 6
<211> LENGTH: 56
<212> TYPE: PRT
<213> ORGANISM: Populus tremuloides

<400> SEQUENCE: 6
Met Met Glu Ser Gly Ala Pro Ile Cys His Thr Cys Gly Glu Gln Val
1 5 10 15
Gly His Asp Ala Asn Gly Glu Leu Phe Val Ala Cys His Gly Cys Ser
20 25 30
Tyr Pro Met Cys Lys Ser Cys Phe Glu Phe Glu Ile Gly Lys Gly Arg
35 40 45
Lys Val Cys Leu Arg Cys Gly Ser
50 55

<210> SEQ ID NO 7
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Unknown

<400> SEQUENCE: 7
Gln Val Leu Arg Trp
1 5
What is claimed is:
1. An isolated cellulose synthase promoter comprising nucleotides from position 1 to position 840 of SEQ ID NO:3.
2. A polynucleotide comprising the promoter of claim 1, operatively linked to a polynucleotide encoding a cellulose synthase or a polynucleotide encoding a UDP-glucose binding domain.
3. A vector comprising the promoter of claim 1.
4. The vector of claim 3, wherein the cellulose synthase promoter is operatively linked to a polynucleotide encoding a cellulose synthase or a polynucleotide encoding a UDP-glucose binding domain.
5. A transgenic plant cell comprising the promoter of claim 1.
6. The plant cell of claim 5, wherein the plant cell is a tree cell.
7. The promoter of claim 1, wherein the promoter is a stress-inducible promoter.
8. A method of inducing expression of a polynucleotide in a plant comprising the polynucleotide operatively linked to the cellulose synthase promoter of claim 1, the method comprising exposing the plant to mechanical stress thereby inducing expression of the polynucleotide.
9. The method of claim 8, wherein the polynucleotide encodes a cellulose synthase, and wherein expression of the polynucleotide alters a characteristic of a plant selected from the group consisting of growth, cellulose content, lignin content, juvenile wood strength, fiber strength, and a combination thereof, as compared to a control plant.
10. The method of claim 9, wherein the altered characteristic is selected from the group consisting of accelerated growth, increased cellulose content, decreased lignin content, improved strength of juvenile wood or fiber, and combinations thereof.
12. The plant of claim 11, wherein the plant is a tree.
13. A method of altering a characteristic of a plant cell comprising the promoter of claim 1 operatively linked to a polynucleotide encoding a polypeptide, wherein the promoter comprises a negative mechanical stress response element or a positive mechanical stress response element, the method comprising exposing the plant cell to a stress to alter expression of the polypeptide, thereby altering the characteristic of the plant cell, as compared to a control plant cell.
15. The method of claim 13, wherein the promoter comprises a negative mechanical stress response element and the polypeptide comprises a cellulose synthase, and wherein the altered characteristic is selected from the group consisting of accelerated growth, increased cellulose content, decreased lignin content, improved strength of juvenile wood or fiber, and combinations thereof.
17. The plant cell of claim 16, which is from a tree.
18. The method of claim 13, wherein the promoter comprises a positive mechanical stress response element and the polypeptide comprises a cellulose synthase, and wherein the altered characteristic is selected from the group consisting of decreased cellulose content, increased lignin content, and combinations thereof.
19. A transformed plant cell produced by the method of claim 18.
20. The plant cell of claim 19, which is from a tree.
21. A transformed plant comprising the plant cell of claim 19.
22. The plant of claim 21, wherein the plant is a tree.