



República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial

(21) PI 0721081-7 A2



(22) Data de Depósito: 20/12/2007
(43) Data da Publicação: 25/02/2014
(RPI 2251)

(51) *Int.Cl.*:
C12N 15/82
C12N 15/29

(54) Título: MOLÉCULAS DE ÁCIDOS NUCLEÍCOS QUE CODIFICAM PROTEÍNAS DE PLANTAS NA FAMÍLIA C3HC4 E PROCESSOS PARA A ALTERAÇÃO DO CONTEÚDO DE CELULOSE E LIGNINA DA PLANTA

(57) Resumo:

(30) Prioridade Unionista: 20/12/2006 US 60/871,061

(66) Prioridade Interna: 860446

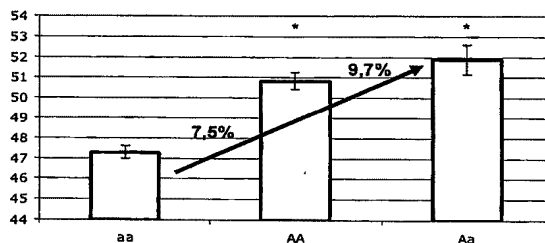
(73) Titular(es): Alellyx S.A.

(72) Inventor(es): ISABEL RODRIGUES GERHARDT, Paulo Arruda

(74) Procurador(es): Dannemann, Siemsen, Bigler & Ipanema Moreira

(86) Pedido Internacional: PCT BR2007000358 de 20/12/2007

(87) Publicação Internacional: WO 2008/074116de 26/06/2008



Relatório Descritivo da Patente de Invenção para **"MOLÉCULAS DE ÁCIDOS NUCLEICOS QUE CODIFICAM PROTEÍNAS DE PLANTAS NA FAMÍLIA C3HC4 E PROCESSOS PARA A ALTERAÇÃO DO CONTEÚDO DE CELULOSE E LIGNINA DA PLANTA"**.

5 REFERÊNCIA CRUZADA A PEDIDO RELACIONADO

Esse pedido reivindica a prioridade do Pedido de Patente Provisório US 60/871.061, depositado em 20 de dezembro de 2006, cujo teor é aqui incorporado a título de referência na íntegra.

CAMPO DA INVENÇÃO

10 A presente invenção refere-se ao campo da biotecnologia vegetal. Mais especificamente, essa invenção se refere à alteração do teor de celulose pela regulação da expressão de genes que codificam proteínas C3HC4.

ANTECEDENTES DA INVENÇÃO

15 Durante a formação de madeira em plantas superiores, a maior parte da glicose produzida durante o metabolismo de carboidratos é canalizada para a celulose para deposição secundária na parede. Djerbi *et al.*, *Cellulose* 11: 301-12 (2004). A celulose é um polímero fibroso que consiste em cadeias lineares de moléculas glicana ligadas em β -(1,4) que cristalizam

20 para formar microfibrilas. As microfibrilas conferem característica de resistência flexível à celulose. A celulose é sintetizada em plantas superiores por grandes complexos multiméricos ligados a membranas de plasma que formam estruturas rosáceas nas extremidades das microfibrilas. Somerville, *Ann. Rev. Dev. Biol.* 22:53-78 (2006).

25 A celulose é valiosa como polpa, como fibra e como um ponto de partida para a síntese de polímeros comercialmente importantes. Espera-se que alterações para aumentar a deposição de celulose tenham um efeito repressivo na deposição de lignina. Hu *et al.*, *Nature Biotech.* 17:808-19 (1999). Uma redução do teor de lignina em vegetais lenhosos é desejável,

30 visto que a produção industrial de celulose e a remoção química da lignina são caras e representam um enorme desafio ambiental.

A via de biossíntese da celulose não é bem compreendida em

nível molecular. Genes da família da sintase de celulose (CESA), e aqueles que codificam proteínas para a síntese e processamento de N-glicana têm sido isolados em um grande número de organismos. Nicol et al., *EMBO J.* 17:5562-76 (1998).

5 Um sistema experimental amplamente utilizado para estudar a formação da madeira, especialmente a síntese e a deposição da celulose, consiste em se curvar uma árvore lenhosa para que a madeira de tensão (TW) seja formada no lado da tensão do embrião. Andersson-Gunners et al., *Plant J.* 45:144-65 (2006). Nas espécies de eucalipto e *Populus*, a madeira
10 de tensão geralmente ocorre no lado superior do embrião curvado e permite que a árvore reoriente seu eixo. A madeira de tensão é caracterizada principalmente pelas fibras do xilema com uma camada secundária gelatinosa extra-grossa (camada G) em seu lúmen. Essa camada G contém quase que exclusivamente microfibras de celulose com alta cristalinidade. Déjardin et
15 al., *Plant Biol.* 6: 55-64 (2004). Como a madeira de tensão é rica em celulose, mas é deficiente em lignina e hemicelulose, ela pode ser usada para detectar e analisar genes envolvidos no controle do fluxo de carbono na lignina, celulose e hemicelulose.

Anderson-Gunners et al. (2006) identificaram genes altamente
20 expressos em TW que estão envolvidos na formação da parede celular, tais como os genes envolvidos no metabolismo de carboidratos e formação do citoesqueleto, bem como genes não destrutivos e dois genes com função desconhecida em *Populus tremula* (L.) x *P. tremuloides* (Michx). Uma proteína de dedo de zinco (RING finger) tipo C3HC4 mostrou expressão diferencial
25 de TW, mas não há dados, neste ou em outros estudos, que impliquem seu papel na biossíntese de celulose.

Domínios de dedo de zinco são fragmentos de proteínas relativamente pequenos que se ligam a um ou mais átomos de zinco. Eles foram primeiramente identificados como um fragmento ligante de DNA na transcri-
30 ção do fator TFIIIA de *Xenopus laevis*, entretanto, sabe-se agora que eles se ligam a DNA, RNA e substratos de lipídios e/ou proteínas. A Ring finger é um tipo especializado de dedo de zinco de 40 a 60 resíduos que se liga a dois

átomos de zinco e está provavelmente envolvida na mediação das interações proteína-proteína. Existem duas variantes diferentes, o tipo C3HC4 e o tipo C3H2C3. O motivo da RING finger tipo C3HC4 é encontrado em várias proteínas celulares e virais, algumas das quais mostraram que contêm atividade de ubiquitina E3 ligase tanto *in vivo* quanto *in vitro*. Laity et al., *Curr. Opin. Struct. Biol.* 11:39-46 (2001).

Considerando-se as dificuldades associadas ao cultivo de árvores florestais tradicionais, tais como o progresso lento devido aos seus longos períodos de geração e a dificuldade de produção de uma planta com uma característica desejável, desenvolvimentos em tecnologia gênica podem reduzir significativamente o tempo requerido para a produção de uma nova variedade de planta e permitir um alvejamento das características consideradas desejáveis pela indústria florestal e de polpa em espécies de árvores específicas.

SUMÁRIO DA INVENÇÃO

Em um aspecto, a invenção proporciona uma sequência de ácido nucleicos, isolada, que compreende uma sequência selecionada do grupo que consiste em: (a) uma sequência de ácido nucleicos revelada em SEQ ID NO:1, ou uma fita complementar desta; (b) uma sequência de ácido nucleicos que codifica a sequência de aminoácidos revelada em SEQ ID NO:2, (c) uma sequência de ácido nucleicos capaz de se hibridizar sob condições estridentes com uma sequência de ácidos nucleicos de (a) ou (b), em que tal sequência hibridizante codifica um polipeptídeo C3HC4; (d) um ácido nucleico que é uma variante alélica ou uma variante *splice* alternativa da sequência de ácido nucleico de (a) ou (b); e (e) uma sequência de ácido nucleico que tem pelo menos 50%, 60%, 70%, 75%, 80%, 85%, 90% ou mais de identidade de sequência com (a) ou (b).

Em outro aspecto, a invenção proporciona uma proteína C3HC4 isolada selecionada do grupo que consiste em: (a) um polipeptídeo revelado em SEQ ID NO:2; (b) um polipeptídeo com uma sequência de aminoácidos que possui pelo menos 85% ou mais de identidade de sequência com a sequência de aminoácidos revelada em SEQ ID NO:2; e (c) uma variante de

um polipetídeo como definido em (a) ou (b).

Em outro aspecto, a invenção proporciona um construto de ácido nucleico que compreende uma sequência de polinucleotídeo C3HC4 isolada operavelmente ligada a um ou mais promotores adequados que acionam a expressão da sequência de polinucleotídeo C3HC4. Em uma modalidade, uma célula de planta compreende a construto de ácido nucleico. Em uma outra modalidade, uma planta transgênica é gerada a partir da célula de planta e a planta tem o teor de celulose e/ou lignina, alterado, em comparação a uma planta não transgênica da mesma espécie. Em ainda outras modalidades, a planta é uma dicotiledônea, monocotiledônea, gimnosperma ou árvore de madeira de lei. Outras modalidades incluem progênie da planta transgênica, incluindo plantas híbridas.

Em outro aspecto, a invenção proporciona um método para alterar o teor de celulose e/ou lignina em uma planta compreendendo (a) introduzir em uma célula de planta isolada, um construto de ácido nucleico que compreende uma sequência de polinucleotídeo C3HC4 isolada, operavelmente ligada a um ou mais promotores adequados que acionam a expressão da sequência de polinucleotídeo C3HC4; e (b) cultivar a dita planta sob condições que promovam o crescimento de uma planta, em que tal planta superexpressa a proteína C3HC4 e tem um teor de celulose e/ou lignina aumentado se comparada a uma planta não transgênica da mesma espécie.

BREVE DESCRIÇÃO DOS DESENHOS

A Figura 1 mostra o perfil de expressão em um conjunto de tecidos de eucalipto de cDNA de C3HC4. Um gene ortólogo foi clonado a partir de mRNA isolado do xilema de *Populus deltoides*.

A Figura 2 retrata esquematicamente o vetor de plasmídeo de expressão de planta pALELLYX-C3HC4 da invenção, o qual compreende um promotor preferido de xilema/câmbio que aciona a expressão da sequência de nucleotídeo C3HC4 da invenção.

A Figura 3 mostra o teor de celulose de várias linhas transgênicas, transformadas com pALELLYX-C3HC4, e as respectivas plantas não transgênicas de controle. O asterisco denota valores médios de celulose

maiores, estatisticamente significantes.

A Figura 4 mostra o teor de celulose de dois genótipos de uma planta transgênica T1 (linha 6B) transformada com pALELLYX-C3HC4. O asterisco denota valores médios de celulose maiores, estatisticamente significantes ($P < 0,05$, t-teste).

A Figura 5 mostra o teor de lignina de dois genótipos de uma planta transgênica T1 (linha 24B) transformada com o vetor plasmidial de expressão de plantas pALELLYX-C3HC4 da invenção. O asterisco denota valores médios de lignina maiores estatisticamente significantes ($P < 0,05$, t-teste).

A Figura 6 mostra o teor de celulose de três genótipos de uma planta transgênica T1 (linha 24B) transformada com o vetor plasmidial de expressão de plantas pALELLYX-C3HC4 da invenção. O asterisco denota valores médios de lignina maiores estatisticamente significantes ($P < 0,05$, t-teste).

A Figura 7 mostra o teor de celulose de três genótipos de uma planta transgênica T1 (linha 25B) transformada com o vetor plasmidial de expressão de plantas pALELLYX-C3HC4 da invenção. O asterisco denota valores médios de lignina maiores estatisticamente significantes ($P < 0,05$, t-teste).

DESCRIÇÃO DETALHADA DA INVENÇÃO

A presente invenção se refere a plantas geneticamente manipuladas que são caracterizadas por um aumento da expressão da proteína C3HC4. Nesse aspecto, a invenção é focada em plantas geneticamente manipuladas que superexpressam uma molécula de ácido nucleico que compreende o gene C3HC4, assim modulando o teor de celulose da dita planta geneticamente manipulada. Aumentar a taxa de transcrição de um gene pode resultar numa melhora do produto de proteína, assim intensificando a taxa do processo metabólico no qual esse gene está envolvido.

Nesse aspecto, os presentes inventores determinaram que plantas geneticamente manipuladas que superexpressam um gene C3HC4 mostram um teor de celulose aumentado. Assim, os inventores determinaram

que a proteína C3HC4 controla, direta ou indiretamente, gene(s) e/ou proteína(s) relacionados com a síntese de celulose.

Desta maneira, aplicações da invenção incluem, sem limitação, o aperfeiçoamento da produção de fibra de celulose, durante a fabricação de papel, através do teor aumentado de celulose em árvores lenhosas, e o aperfeiçoamento, da extração de fibras de celulose para a produção de têxteis através do aumento do teor de celulose em fibras de algodão. Adicionalmente, a deposição intensificada de celulose provavelmente tem efeito repressivo na deposição de lignina. A produção industrial de celulose a partir de árvores lenhosas requer a remoção química da lignina durante o processo de polpação, que faz uso de grandes quantidades de produtos químicos concentrados. Para a produção de papel de alta qualidade, a lignina residual necessita ser adicionalmente removida por uma etapa de alveijamento adicional que envolve o uso de substâncias extremamente danosas. O processo geral é dispendioso e representa um desafio ambiental enorme. Por essa razão, espera-se que a redução do teor de lignina em plantas lenhosas, tipicamente árvores, diminua as demandas energéticas e químicas desses processos de extração altamente custosos, e também deve diminuir a quantidade de material efluente, um grande poluidor ambiental potencial que é de processamento caro e difícil. Campbel et al., *Plant Physiol.* 110:3-13 (1996). Assim, engenharia genética da biossíntese de celulose pode fornecer uma estratégia para o aumento da quantidade e qualidade da celulose, ao mesmo tempo em que reduz o teor de lignina nas plantas transgênicas.

A terminologia técnica nesta descrição está em conformidade com o uso comum em bioquímica, biologia molecular e agricultura. Este uso e estes termos técnicos são explicados em: *Molecular Cloning: A Laboratory Manual* (Terceira edição), vol. 1-3, Cold Spring Harbor Laboratory Press (2001); *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY*, Greene Publishing Associates and Wiley-Interscience (1988), com atualizações periódicas; *SHORT PROTOCOLS IN MOLECULAR BIOLOGY: A COMPENDIUM OF METHODS FROM CURRENT PROTOCOLS IN MOLECULAR BIOLOGY* (5ª edição), vol. 1-2, John Wiley & Sons, Inc. (2002); *GENOME ANALYSIS: A*

LABORATORY MANUAL, VOL 1-2, Cold spring Harbor Laboratory Press (1997). Técnicas de biologia de plantas adequadas, conforme descritas aqui, são adicionalmente explicadas em tratados de metodologia como METHODS IN PLANT MOLECULAR BIOLOGY: A LABORATORY COURSE MANUAL, Cold Spring Harbor Laboratory Press (1995). Vários métodos que empregam PCR são descritos, por exemplo, em Innis et al., PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, Academy Press (1990), e por Dieffenbach e Dveksler, PCR PRIMER: A LABORATORY MANUAL (2ª edição), Cold Spring Harbor Laboratory Press (2003). Pares de iniciadores para PCR podem ser derivados de sequências conhecidas por técnicas conhecidas tal como utilizar programas de computadores feitos para esse propósito, como Primer, Versão 0,5, 1991 (Whitehead Institute for Biomedical Research, Cambridge, MA). Metodologia ilustrativa para a síntese química de ácidos nucleicos é discutida, por exemplo, em Beaucage e Caruthers, *Tetra. Letts.* 22:1859-62 (1981), e Matteucci e Caruthers, *J. Am. Chem. Soc.* 103:3185 (1981).

Digestões, fosforilações, ligações e transformações de enzimas de restrição foram feitas conforme descritas por Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL (2ª edição), Cold Spring Harbor Laboratory Press (1989). Todos os materiais e reagentes usados para o crescimento e manutenção de células bacterianas foram obtidos de Aldrich Chemicals (Milwaukee, Wisconsin), DIFCO Laboratories (Detroit, Michigan), Invitrogen (Gaithersburg, Maryland) ou Sigma Chemical Company (St. Louis, Missouri), a não ser que seja indicado o contrário.

Os termos "que codificam" e "codificação" se referem a um processo pelo qual um gene, através dos mecanismos de transcrição e tradução, fornece informações a uma célula a partir das quais uma série de aminoácidos pode ser montada em uma sequência de aminoácidos específica para produzir uma enzima ativa. Devido à degeneratividade do código genético, certas mudanças de base na sequência de DNA não mudam a sequência de aminoácidos de uma proteína. Entende-se, assim, que modificações na sequência de DNA que codifica a proteína C3HC4 que não afetem subs-

tancialmente as propriedades funcionais da proteína estão contempladas.

Nesta descrição, "expressão" denota a produção do produto de proteína ou polipeptídeo codificado por um gene. Alternativa ou adicionalmente, "expressão" denota a combinação de processos intracelulares, incluindo transcrição e tradução, pelos quais passa uma molécula de DNA de codificação como um gene estrutural para produzir um polipeptídeo. "Super-expressão" se refere à expressão de uma sequência de genes particular em que a produção de mRNA ou polipeptídeo em um organismo transgênico excede os níveis de produção em um organismo não transgênico.

O termo "ácido nucleico heterólogo" se refere a um ácido nucleico, DNA ou RNA, que foi introduzido em uma célula (ou no ancestral da célula) através de esforços de seres humanos. Tais ácidos nucleicos exógenos podem ser uma cópia de uma sequência que é naturalmente encontrada na célula na qual foi introduzida, ou fragmentos desta.

Por outro lado, o termo "ácido nucleico endógeno" se refere a um ácido nucleico, gene, polinucleotídeo, molécula de DNA, RNA, mRNA ou cDNA presente em uma planta ou organismo que será geneticamente alterado. Uma sequência endógena é "nativa" a, em outras palavras, indígena à planta ou organismo que será geneticamente alterado.

O termo "sequências homólogas" se refere a sequências de polinucleotídeos ou de polipeptídeo que são similares devido a ancestrais comuns e conservação de sequência.

Para os fins desta invenção, "parálogos" são homólogos produzidos por duplicação de genes. Eles representam genes derivados de um gene ancestral comum que é duplicado dentro de um organismo e, subsequentemente, divergido. "Ortólogos" são homólogos produzidos por especiação. Eles representam genes derivados de um ancestral comum que divergiram devido à divergência dos organismos com os quais eles estão associados. Vide Brinkman e Leipe, em: BIOINFORMATICS, A PRACTICAL GUIDE TO THE ANALYSIS OF GENES AND PROTEINS 323-58, Wiley-Interscience (2001).

O termo "homólogo funcional" se refere a sequências de polinu-

cleotídeos ou polipeptídeos que são similares devido à genealogia comum e conservação de sequência e têm função idêntica ou similar nos níveis catalíticos, celulares ou organizmais.

SEQUÊNCIAS DE C3HC4

5 Muitos dos processos biológicos necessários para o metabolismo e desenvolvimento em um organismo são governados por famílias de genes. Esse é o caso da proteína C3HC4 que pertence a uma família de genes que compreende muitos genes. As proteínas do domínio C3HC4 pertencem à família de proteínas chamada "dedo de zinco", caracterizada pelo
10 domínio "RING finger", que compreende oito ligantes de metal formados pelo motivo de consenso, C3HC4. Domínios de RING finger normalmente se ligam a dois íons de zinco em um arranjo firmado cruzado único e podem, basicamente, ser considerados um domínio de interação de proteínas. Jackson et al., *Trends Cell Biol.* 10:429-39 (2000).

15 Proteínas de RING finger foram implicadas em diversos processos biológicos, desde a regulação transcricional e traducional até proteólise almejada e de desenvolvimento. Assim, outros membros da família C3HC4 também intensificariam o teor de celulose. Um gene C3HC4 ilustrativo revelado em SEQ ID NO:1 foi isolado de choupo.

20 Espera-se que qualquer gene de qualquer organismo que codifique um gene de RING finger, que codifica uma proteína com uma estrutura e propriedades biológicas similares ao produto da tradução de um gene C3HC4 terá o mesmo efeito no teor de celulose, conforme os inventores demonstraram para plantas transgênicas descritas nos exemplos acima. Es-
25 ses genes podem ser identificados e funcionalmente anotados por comparação de sequência. Um biólogo molecular com conhecimento pode prontamente identificar uma sequência relacionada à sequência de C3HC4 com a ajuda da metodologia convencional, como o monitoramento de cDNA ou bibliotecas de genoma com sondas de hibridização adequadas ou por buscas
30 em bancos de dados públicos, como o NCBI's Genbank. Sequências homólogas também podem ser isoladas com a ajuda de oligonucleotídeos degenerados, utilizando-se técnicas baseadas em PCR conhecidas.

Também é possível usar programas computacionais que potencializam várias técnicas conhecidas de identificação de genes por comparação de sequência. Exemplos dessas técnicas são descritos, por exemplo, por Innis et al., PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, Academic Press (1990), e em GENOME ANALYSIS: A LABORATORY MANUAL, volumes 1 e 2, Cold Spring Harbor Laboratory Press (1997).

Em suas investigações acerca desta invenção, os presentes inventores induziram TW em plantas de um híbrido interespecífico, *E. grandis* x *E. urophylla*. Diversos genes diferencialmente expressos entre TW e madeira normal foram assim identificados. Dentre os genes que têm suas expressões alteradas em madeira de tensão, determinou-se que muitos codificam um membro da família de dedo de zinco tipo C3HC4, que é altamente expresso em madeira tensão de eucalipto.

Como um membro da família de proteína C3HC4 é altamente expresso em madeira de tensão, um tecido constituído principalmente de celulose altamente cristalina, e os membros C3HC4 parecem estar envolvidos no controle do meristema vascular e no crescimento secundário, pode-se esperar que a transformação genética de árvores lenhosas com construtos de ácido nucleico, que compreendem moléculas que codificam membros da família C3HC4, alterasse a padronização vascular e a síntese e deposição de celulose. Assim, um aumento da síntese e na deposição de celulose ocorreria quando um gene C3HC4 fosse superexpresso. Inversamente, quando um gene C3HC4 é infrarregulado, os inventores antecipam uma diminuição da síntese de celulose. Como a alteração na síntese e na deposição de celulose produz, em geral, uma alteração na síntese e na deposição de lignina, vide Hu et al., *Nature biotech.* 17:808-19 (1999), os inventores entenderam igualmente que ao aumentar a síntese e a deposição de celulose, o teor de lignina da árvore lenhosa seria reduzido. O cenário oposto ocorreria, se a síntese de celulose fosse reduzida.

No contexto da presente invenção, uma sequência pode ser identificada pela metodologia descrita acima, e assim pode ser funcionalmen-

te anotada como pertencente à família de C3HC4. Nesta descrição, as frases "sequência de polinucleotídeo C3HC4" e "sequência de ácido nucleico C3HC4" denotam qualquer ácido nucleico, gene, polinucleotídeo, molécula de DNA, RNA, mRNA, cDNA que codifique um polipeptídeo C3HC4 que, quando superexpresso, resulta em um aumento do teor de celulose e/ou uma diminuição no teor de lignina de uma planta. As frases "sequência de polinucleotídeo C3HC4" e "sequência de ácido nucleico C3HC4" também englobam qualquer molécula de ácido nucleico com uma sequência de nucleotídeos capaz de hibridizar, sob condições estridentes, com quaisquer das sequências aqui descritas e que codifica um polipeptídeo C3HC4 que, quando superexpresso, resulta em um aumento do teor de celulose e/ou uma diminuição do teor de lignina de uma planta. As frases também conotam sequências que se hibridizam de forma cruzada com SEQ ID NO:1, e que preferivelmente possuem pelo menos 40%, ou mais preferivelmente pelo menos 60%, ainda mais preferivelmente pelo menos 80% e, ainda mais preferivelmente, pelo menos 90% de uma homologia ou de identidade com SEQ ID NO:1.

Uma sequência de nucleotídeos da invenção também pode codificar uma proteína que é homóloga ao produto de gene predito revelado em SEQ ID NO:2.

As frases "sequência de polinucleotídeo C3HC4" e "sequência de ácido nucleico C3HC4" também denotam aquelas sequências representadas pelos fragmentos ou variantes de SEQ ID NO:1 que compartilham pelo menos 50%, preferivelmente pelo menos 60%, mais preferivelmente pelo menos 70%, ainda mais preferivelmente pelo menos 80% e, ainda mais preferivelmente 90% de identidade com SEQ ID NO:1 e codificam polipeptídeos C3HC4 que, quando sobre-expressos, resultam em um aumento no teor de celulose e/ou uma diminuição do teor de lignina de uma planta. Como prova do que é dito, as sequências de nucleotídeos da invenção incluem aquelas sequências que codificam polipeptídeos que compreendem uma sequência de aminoácidos de SEQ ID NO:2 ou uma sequência de aminoácidos que é pelo menos 50%, preferivelmente pelo menos 60%, mais preferivelmente

pelo menos 70%, ainda mais preferivelmente pelo menos 80% e, ainda mais preferivelmente 90% idêntica à SEQ ID NO:2, cuja superexpressão resulta em um aumento do teor de celulose e/ou uma diminuição no teor de lignina de uma planta.

5 A frase "condições rigorosas" aqui conota parâmetros com os quais a técnica está familiarizada. Polinucleotídeos de fita simples se hibridizam quando eles se associam baseados em diversas forças físico-químicas bem caracterizadas, como ligação de hidrogênio, exclusão de solvente e empilhamento de bases. A estringência de uma hibridização reflete o grau de
10 identidade de sequência dos ácidos nucleicos envolvidos, de modo tal que quanto maior a estringência, mais similares serão as duas fitas de polinucleotídeo. A estringência é influenciada por diversos fatores, incluindo temperatura, concentração e composição do sal, aditivos orgânicos e não orgânicos, solvente, etc. presentes na solução de lavagem e hibridização e incubações
15 (e número). Um técnico com conhecimento ordinário pode prontamente selecionar tais condições pela variação de temperatura durante a reação de hibridização e processo de lavagem, de concentração do sal durante a reação de hibridização de processo de lavagem e assim por diante.

 Para a hibridização de ácidos nucleicos complementares, que
20 possuem mais de 100 resíduos complementares, em um filtro em um Southern ou Northern blot, condições de hibridização "estringentes" são exemplificadas por uma temperatura que é de cerca de 5°C a 20°C mais baixa que o ponto de fusão térmico (T_m) da sequência específica, em uma força iônica e pH definidos. A T_m é a temperatura, sob força iônica e pH definidos, na qual
25 50% da substância-alvo hibridiza em uma perfeitamente emparelhada. Moléculas de ácido nucleico que hibridizam sob condições estringentes hibridizam em uma sonda baseada ou no cDNA inteiro ou em partes selecionadas. Mais preferivelmente, "condições estringentes" aqui se referem a parâmetros familiares à técnica, como hibridização em 3,5x de SSC, 1x de solução de
30 Denhardt, tampão de fosfato de sódio 25 mM (pH 7,0), 0,5% de SDS e EDTA 2mM por 18 horas a 65°C, seguido por 4 lavagens do filtro a 65°C por 20 minutos, em 2x de SSC, 0,1% de SDS, e uma lavagem final por até 20 minu-

tos em 0,5x de SSC, 0,1% de SDS ou 0,3% de SSC e 0,1% de SDS para maior estringência, e 0,1x de SSC e 0,1% de SDS, para estringência ainda maior. Outras condições podem ser substituídas, desde que o grau de estringência seja substancialmente igual ao aqui fornecido, usando uma lavagem final de 0,5x de SSC.

Conforme indicado, "sequência de ácido nucleico C3HC4" nesta descrição se refere a qualquer molécula de ácido nucleico com uma sequência de nucleotídeos capaz de se hibridizar sob condições estringentes com a sequência aqui revelada, e de codificar um polipeptídeo equivalente à proteína que possui a sequência de aminoácidos aqui revelada em SEQ ID NO:2. A frase também inclui sequências que se hibridizam de forma cruzada com SEQ ID NO:1, e que são preferivelmente pelo menos 55%, mais preferivelmente 65%, mais preferivelmente 75%, mais preferivelmente 85% e ainda mais preferivelmente 90% idênticas para homologia ou identidade com SEQ ID NO:1. A sequência de nucleotídeos da invenção pode codificar uma proteína que é homóloga ao produto de gene predito em SEQ ID NO:2.

Modalidades adicionais incluem qualquer molécula de ácido nucleico que compreenda quaisquer das sequências de bases acima com uma ou mais bases deletadas, substituídas, inseridas ou adicionadas, e que codifica um polipeptídeo que é homólogo à proteína codificada por SEQ ID NO:2. Tais moléculas de ácido nucleico incluem variantes alélicas e variantes *splice* de SEQ ID NO:1.

Assim, o termo "variante" é uma sequência de aminoácidos ou nucleotídeos que se desvia da sequência de aminoácidos ou nucleotídeos padrão, ou de um dado gene ou proteína em particular. A variante pode ter mudanças "conservativas", em que um aminoácido substituído possui propriedades químicas ou estruturais similares, por exemplo, substituição de leucina por isoleucina. Uma variante pode ter mudanças "não conservativas", por exemplo, substituição de leucina por um triptofano. Variações menores análogas podem também incluir inserções ou deleções de aminoácidos, ou ambos. Orientação na determinação de que quais aminoácidos podem ser substituídos, inseridos ou deletados, pode ser encontrada utilizando-

do-se programas computacionais bem conhecidos na técnica como o software Vector NTI Suite (InforMax, Maryland). "Variante" também pode se referir a um "gene embaralhado", como descrito, por exemplo, nas patentes US 6.506.603, US 6.132.970, US 6.165.793, e US 6.117.679.

5 As "sequências de bases acima com uma ou mais bases deletadas, substituídas, inseridas ou adicionadas" referidas aqui, são amplamente conhecidas por aqueles com conhecimento ordinário no assunto como re-
tendo atividade fisiológica mesmo quando a sequência de aminoácidos de
uma proteína que possui geralmente aquela atividade fisiológica tem um ou
10 mais aminoácidos substituídos, deletados, inseridos ou adicionados. Sequências de nucleotídeos que possuem tais modificações e que codificam a proteína C3HC4 estão incluídas no escopo desta invenção. Por exemplo, as regiões de cauda poli A ou de regiões não traduzidas na terminação 5' ou 3' podem ser deletadas, e as bases podem ser deletadas na medida em que os
15 aminoácidos são deletados. As bases também podem ser substituídas, desde que não resulte em deslocamentos de quadros. Bases também podem ser "adicionadas" na medida em que aminoácidos são adicionados. Entretanto, é essencial que tais modificações não resultem na perda da função da proteína C3HC4. Tais ácidos nucleicos modificados podem ser obtidos, por
20 exemplo, pela modificação de sequências de bases da invenção para que aminoácidos em locais específicos sejam substituídos, deletados, inseridos ou adicionados por mutagênese em local específico. Vide Zoller e Smith, Nucleic Acid Res. 10:6487-500 (1982).

Entende-se que sequências de aminoácidos e de ácidos nucleicos
25 podem incluir resíduos adicionais, tais como aminoácidos N ou C-terminal ou sequências 5' ou 3'. Tais adições são adequadas desde que as sequências resultantes codifiquem um polipetídeo que mantém as mesmas atividades biológicas de proteína ou equivalentes.

A presente invenção fornece uma sequência de nucleotídeos
30 que codifica a proteína C3HC4. Esta sequência pode ser derivada de cDNA, tal como cDNA de *Populus deltoides* ou de DNA genômico. Um clone de cDNA típico é revelado em SEQ ID NO:1, acima, que codifica uma proteína

C3HC4. De acordo com um aspecto da invenção, modifica-se o teor de celulose em tecidos de planta, tais como células fibrosas do xilema lenhoso ou sementes de algodão pelo controle da expressão da proteína C3HC4. Assim, células de plantas ou plantas inteiras são geneticamente modificadas, por exemplo, com a sequência que codifica a proteína C3HC4 de, por exemplo, *Populus deltoides* que é expressa preferivelmente em células de fibra e causam um aumento na síntese e deposição da celulose.

Adicionalmente, a invenção fornece uma molécula de ácido nucleico que compreende uma sequência de nucleotídeo selecionada de (a) SEQ ID NO:1, ou uma parte desta, ou um complemento desta; (b) uma sequência de nucleotídeos que hibridiza na dita sequência de nucleotídeo de (a) sob uma estringência de lavagem equivalente a 0,1X de SSC a 1,0X de SSC, 0,1% de SDS, a 50-65°C, (c) uma sequência de nucleotídeos que codifica uma proteína que possui a mesma sequência de aminoácidos como a que é codificada pela sequência de nucleotídeos de (a), mas é degenerada de acordo com a degenerescência do código genético; e (d) uma sequência de nucleotídeos que codifica a mesma sequência de aminoácidos que aquela sequência de nucleotídeos de (b), mas é degenerada de acordo com a degenerescência do código genético.

Outra característica da invenção são proteínas e polipeptídeos codificados pela molécula de ácido nucleico da invenção, exemplificados por, sem limitação, o polipeptídeo que tem as sequências de aminoácidos compreendidas em SEQ ID NO:2. Preferivelmente, os polipeptídeos da invenção possuem sequências de aminoácidos que contêm regiões que são pelo menos 60% idênticas às sequências referidas acima. Identidade superior a 70% é preferível, enquanto identidades superiores a 80%, 90% ou mesmo 95% são ainda mais preferidas.

A molécula de ácido nucleico da invenção pode ser utilizada "pura" ou, preferivelmente, em construtos de vetores de expressão, para introdução em células, como células de plantas. Técnicas padrão de biologia molecular, bem conhecidas por um técnico no assunto, podem ser utilizadas.

Construtos de Ácidos Nucleicos

Construtos de ácidos nucleicos recombinantes podem ser feitos utilizando técnicas padrão. Por exemplo, uma sequência de nucleotídeos para transcrição pode ser obtida pelo tratamento de um vetor contendo tal sequência com enzimas de restrição para cortar o segmento apropriado. A

5 sequência de nucleotídeos para transcrição pode também ser gerada por oligonucleotídeos de anelamento ou ligação de oligonucleotídeos sintéticos ou pelo uso de oligonucleotídeos sintéticos em uma reação em cadeia de polimerase (PCR) para que obtenção de locais de restrição apropriados em cada extremidade. A sequência de nucleotídeos é então clonada em um ve-

10 tor contendo os elementos reguladores adequados, como sequências de promotores a montante e de terminadores a jusante. Tipicamente, os vetores de transformação de plantas incluem uma ou mais sequências de codificação de plantas clonadas (genômica ou cDNA) sob o controle transcricional de sequências reguladoras 5' e 3', e um marcador selecionável. Tais vetores

15 de transformação de plantas tipicamente contêm também um promotor, um sítio de partida de iniciação da transcrição, um sinal de processamento de RNA (tais como sequências de sinal *splicing*), um sítio de terminação de transcrição e/ou um sinal de poliadenilação. Intensificadores e sequências de alvejamento também podem estar presentes.

20 Promotores de plantas constitutivos adequados, que podem ser úteis para a expressão de sequências de proteína C3HC4 incluem, sem limitação, o promotor do vírus mosaico da couve-flor (CaMV) 35S, os promotores de poliubiquitina de milho e de *Populus*, que conferem expressão constitutiva e de alto nível na maioria dos tecidos de planta (ver, por exemplo, WO

25 2007/00611, patente US 5.510.474; Odell et al., *Nature*, 1985, 313:810-812); o promotor de nopalina sintase (An et al., 1988, *Plant Physiol.* 88: 547-552); o promotor de FMV do vírus do mosaico da escrofulária (Patente US 5.378.619); e promotor de octopina sintase (Fromm et al., 1989, *Plant Cell* 1:977-984).

30 O vetor pode conter também sequências de terminação, que são posicionadas a jusante das moléculas de ácido nucleico da invenção, de tal modo que a transcrição de mRNA é terminada e sequências de poli A são

adicionadas. Terminadores típicos são o terminador do vírus mosaico da couve-flor (CaMV) 35S e o terminador do gene de nopalina sintase (NOS).

Vetores de expressão podem conter também um marcador de seleção com o qual células transformadas podem ser identificadas em uma cultura. O marcador pode estar associado a uma molécula de ácido nucleico heterólogo, ou seja, com o gene ligado operavelmente a um promotor. Como utilizado aqui, "marcador" se refere a um gene que codifica uma característica ou fenótipo que permite a seleção ou o monitoramento de uma planta ou de uma célula contendo o marcador. Em plantas, por exemplo, o gene marcador codificará resistência a antibiótico e a herbicida. Isso permite a seleção de células transformadas dentre células que não foram transformadas ou transfectadas.

Exemplos de marcadores selecionáveis adequados incluem adenosina desaminase, diidrofolato redutase, higromicina-B-fosfotransferase, timidina cinase, xantina-guanina fosfotransferase (resistência à canamicina, neomicina e a G148). Esses marcadores podem incluir resistência a G148, higromicina, bleomicina, canamicina e gentamicina. O construto também contém o gene marcador selecionável *Bar*, que confere resistência aos análogos de fosfinotricina herbicida como o glifosinato de amônio. Thompson et al., EMBO J. 6: 2519-23 (1987). Outros marcadores de seleção adequados são bem conhecidos.

Marcadores visíveis, tal como a proteína fluorescente verde (GFP), podem ser utilizados. Métodos para identificar ou selecionar plantas transformadas baseados no controle de divisão celular também têm sido descritos. Vide John e Van Mellaert, WO 2000/052168, e Fabijansk et al., WO 2001/059086.

Sequências de replicação, de origem bacteriana ou viral, também podem ser incluídas para permitir que o vetor seja clonado em um hospedeiro bacteriano ou de fagos. Preferivelmente, um hospedeiro de amplo espectro de origem procariótica de replicação é utilizado. Um marcador selecionável para bactérias pode ser incluído para permitir a seleção de células bacterianas que contêm a construto desejada. Marcadores selecionáveis

procarióticos adequados também incluem resistência a antibióticos tal como a canamicina ou a tetraciclina.

Outras sequências de ácido nucleico que codificam funções adicionais também podem estar presentes no vetor, conforme conhecido da técnica. Por exemplo, quando a agrobactéria é a hospedeira, sequências de T-DNA podem ser incluídas para facilitar a transferência e incorporação subsequente aos cromossomos da planta.

De acordo com outro aspecto da invenção, construtos de ácido nucleico são fornecidos que compreendem uma sequência de DNA de C3HC4, como descrito acima, sob o controle de uma região de iniciação transcricional operativa em plantas, de tal forma que o construto pode gerar RNA em células de plantas. Preferivelmente, a região de iniciação transcricional é parte de um promotor de órgão ou tecido específico de planta, conforme quaisquer dos descritos no WO 2005/096805. Mais preferivelmente, o promotor específico de tecido, quando ligado operavelmente à sequência de DNA de C3HC4, garante a transcrição em tipos de células, tecidos ou órgãos específicos de modo que a síntese de celulose pode ser especificamente almejada sem se afetar outras funções da planta.

As plantas transgênicas da invenção podem ser caracterizadas pelo aumento do teor de celulose e/ou pela diminuição do teor de lignina. O teor de celulose aumentado na planta geneticamente modificada é preferivelmente conseguido por um aumento na expressão de C3HC4 em tecidos da planta em que a deposição de celulose ocorre. Assim, em uma modalidade preferida, as plantas transgênicas da invenção contêm um construto de ácido nucleico que compreende um promotor preferido de câmbio/xilema como aqueles descritos na patente citada acima, operavelmente ligada a um gene que codifica uma proteína C3HC4, levando a um aumento da expressão no sistema vascular da planta do gene C3HC4, que por sua vez resulta em um aumento da síntese e da deposição de celulose nesses tecidos sem afetar outras funções da planta.

Conforme foi mencionado, o teor de celulose e outras características relacionadas de partes da planta podem ser modificados geneticamen-

te com um construto de ácido nucleico em conformidade com a invenção. A invenção também fornece células de planta que contêm ou que foram geneticamente alteradas com tais construtos, plantas derivadas destas que têm expressão do gene C3HC4 modificada, e sementes de tais plantas.

5 Construtos de ácidos nucleicos de acordo com a invenção podem compreender uma sequência de bases de comprimento mínimo para gerar mRNA e, conseqüentemente, um polipeptídeo que retém a função de C3HC4. Por conveniência, geralmente se achará adequado o uso de se-
10 quências entre 100 e 1.000 bases de comprimento, mas não há limite superior teórico do comprimento de sequência de bases. A preparação de tais construtos é descrita a seguir em mais detalhes.

 As moléculas de ácido nucleico isoladas da invenção podem ser incorporadas em construtos de ácido nucleico, de forma que elas estejam ligadas operavelmente a um promotor. Preferivelmente, o promotor é um que
15 se saiba que opera em células de planta e, mais preferivelmente, operável em células de órgãos ou em tecidos de planta específicos, tais como raízes, brotos, folhas, xilema, etc. As moléculas de ácido nucleico da invenção po-
20 dem ser colocadas em ligação operável com promotores constitutivos ou induzíveis. Alternativamente, as moléculas de ácido nucleico da invenção
25 podem ser postas em ligação operável com promotores, os quais direcionam a expressão do gene a jusante preferivelmente, ou especificamente a um órgão ou tecido da planta, tais como o xilema e o câmbio.

 Adicionalmente, promotores específicos do sistema vascular, específicos do xilema ou preferidos do xilema podem ser úteis para promo-
25 ver a expressão de moléculas de ácido nucleico da invenção especificamen-
30 te em tecidos vasculares, especialmente em tecido do xilema. Em geral, o uso de um promotor constitutivo afeta os níveis e funções protéicas em todas as partes da planta, enquanto o uso de um promotor preferido de um tecido permite que se foque na expressão do gene modificado em partes
35 específicas da planta, o que leva a fenótipos mais controláveis. Assim, na aplicação da invenção, pode se achar conveniente o uso de um promotor que dará a expressão durante o desenvolvimento do xilema, por meio do

que as proteínas da invenção somente seriam superproduzidas no(s) órgão(s) ou tecido(s) ou tipo(s) celular(es) nos quais sua ação fosse requerida para usos aqui descritos. Promotores específicos de tecido vascular, específicos de xilema, preferidos de tecido vascular e preferidos de xilema que poderiam ser utilizados incluem, sem limitação, o promotor de gene de cumarato-4-hidroxilase (C4H) preferido de xilema, promotor do gene de tubulina (TUB) preferido de xilema e o promotor de gene de proteína de transferência lipídica (LTP) preferido de xilema descritos no pedido internacional publicado mencionado acima '805'. O promotor selecionado em particular deve ser capaz de provocar expressão suficiente para que resulte em superexpressão da proteína da invenção, para modificar o tamanho do xilema ou para modificar a composição química do xilema de uma planta ou ainda uma combinação destes efeitos.

Embora a taxa de expressão dos genes seja modulada principalmente pelo promotor, um aperfeiçoamento na expressão também pode ser obtido pela identificação e uso de sequências de intensificadores, como porções intrônicas de genes, que elevam os níveis de expressão de genes proximamente localizados em uma orientação de maneira independente. Em plantas, a inclusão de alguns íntrons em uma construção de gene em uma posição entre o promotor e a sequência que codifica o gene leva a um aumento da acumulação de mRNA e proteína. Introns conhecidos por elevarem a expressão em plantas foram identificados em genes de milho, por exemplo, *hsp70*, *tubA1*, *Adh1*, *Sh1*, *UbH* (Brown e Santino, Patentes US 5.424.412 e US 5.859.347; Jeon et al., 2000, *Plant Physiol.* 123: 1005-1014; Callis et al., 1987, *Genes Dev.* 1:1183-1200; Vasil et al., 1989, *Plant Physiol.* 91:1575-1579), e em genes de plantas dicotiledôneas como *rbcS* de petúnia (Dean et al., 1989, *Plant Cell* 1: 201-208); ST-LS1 da batata (Leon et al., 1991, *Plant Physiol.* 95:968-972) e UBQ3 (Norris et al., 1993, *Plant Mol. Biol.* 21:895-906) and PAT1 de *Arabidopsis thaliana* (Rose and Last, 1997, *Plant J.* 11:455-464).

Adicionalmente, o vetor de expressão recombinante compreende um promotor funcional em uma célula de planta, uma molécula de ácido nu-

cleico que é homóloga das sequências de nucleotídeos descritas acima, e que codifica um polipeptídeo cuja sequência de aminoácidos contém regiões que são pelo menos 60% idênticas à sequência revelada em SEQ ID NO:2, conforme descrição acima. Mais preferivelmente, a molécula de ácido nucleico codifica um polipeptídeo cuja sequência de aminoácidos contém regiões que são pelo menos 70%, 80% ou mesmo 90% idênticas à sequência acima.

Construtos de acordo com a invenção podem ser usados para modificar geneticamente qualquer planta utilizando-se qualquer técnica adequada. Ambas as células de plantas de angiospermas ou gimnospermas de monocotiledôneas e de dicotiledôneas podem ser geneticamente modificadas de várias maneiras conhecidas na técnica. Klein et al., *biotechnology* 4:583-90 (1993); Bechtold et al., *C. R. Acad. Sci. Paris* 316:1194-99 (1993); Bent et al., *Mol. Gen. Genet.* 204: 383-96 (1986); Paszowski et al., *EMBO J.* 3:2717-2722 (1984); Sagi et al., *Plant Cell Rep.* 13:262-66 (1994).

15 Plantas para Engenharia Genética

A invenção se refere em geral a plantas transgênicas que expressam genes ou segmentos de genes que codificam as composições originais de polipeptídeos descritas aqui. Como utilizado aqui, o termo "plantas transgênicas" pretende se referir a plantas que incorporaram sequências de ácidos nucleicos incluindo, sem limitação, genes que talvez não estejam normalmente presentes, sequências de ácidos nucleicos não normalmente transcritas em RNA ou traduzidas em uma proteína ("expressas"), ou quaisquer outras sequências de ácidos nucleicos ou genes que se deseje introduzir na planta, tais como genes que talvez estejam presentes na planta, mas que se deseja ou modificar geneticamente ou alterar a expressão. Contempla-se que em alguns exemplos, o genoma das plantas transgênicas da presente invenção terá sido aumentado pela introdução estável de um transgene. Em outros exemplos, entretanto, a sequência ou gene introduzido vai substituir uma sequência endógena. Um gene preferido, que pode ser introduzido, inclui sem limitação a sequência de ácidos nucleicos C3HC4 de *Populus deltoides*.

Plantas que podem ser modificadas de acordo com a presente

invenção incluem, mas sem limitação, árvores como as da espécie de eucalipto (*E. alba*, *E. albens*, *E. amygdalina*, *E. aromaphloia*, *E. baileyana*, *E. balladoniensis*, *E. bicostata*, *E. botryoides*, *E. brachyandra*, *E. brassiana*, *E. hrevistylis*, *E. brockwayi*, *E. camaldulensis*, *E. ceracea*, *E. cloeziana*, *E. coccifera*, *E. cordata*, *E. cornuta*, *E. corticosa*, *E. crebra*, *E. croajingolensis*, *E. curtisii*, *E. dalrympleana*, *E. deglupta*, *E. delegatensis*, *E. delicata*, *E. diversicolor*, *E. diversifolia*, *E. dives*, *E. dolichocarpa*, *E. dundasii*, *E. dunnii*, *E. elata*, *E. erythrocorys*, *E. erythrophloia*, *E. eudesmoides*, *E. falcata*, *E. gamophylla*, *E. glaucina*, *E. globulus*, *E. globulus* subsp. *bicostata*, *E. globulus* subsp. *globulus*, *E. gongylocarpa*, *E. grandis*, *E. grandis* x *urophylla*, *E. guilfoylei*, *E. gunnii*, *E. hallii*, *E. houseana*, *E. jacksonii*, *E. lansdowneana*, *E. latisinensis*, *E. leucophloia*, *E. leucoxylon*, *E. lockyeri*, *E. lucasii*, *E. maidenii*, *E. marginata*, *E. megacarpa*, *E. melliodora*, *E. michaeliana*, *E. microcorys*, *E. microtheca*, *E. muelleriana*, *E. nitens*, *E. nitida*, *E. obliqua*, *E. obtusiflora*, *E. occidentalis*, *E. optima*, *E. ovata*, *E. pachyphylla*, *E. pauciflora*, *E. pellita*, *E. perriniana*, *E. petiolaris*, *E. pilularis*, *E. piperita*, *E. platyphylla*, *E. polyanthemos*, *E. populnea*, *E. preissiana*, *E. pseudo globulus*, *E. pulchella*, *E. radiata*, *E. radiata* subsp. *radiata*, *E. regnans*, *E. risdonii*, *E. robertsonii*, *E. rodwayi*, *E. rubida*, *E. rubiginosa*, *E. saligna*, *E. salmonophloia*, *E. scoparia*, *E. sieberi*, *E. spathulata*, *E. staeri*, *E. stoatei*, *E. tenuipes*, *E. tenuiramis*, *E. tereticornis*, *E. tetragona*, *E. tetrodonta*, *E. tindaliae*, *E. torquata*, *E. umbra*, *E. urophylla*, *E. vernicosa*, *E. viminalis*, *E. wandoo*, *E. wetarensis*, *E. willisii*, *E. willisii* subsp. *falciformis*, *E. willisii* subsp. *willisii*, *E. woodwardii*), *Populus* spp. (*P. alba*, *P. alba* x *P. grandidentata*, *P. alba* x *P. tremula*, *P. alba* x *P. tremula* var. *glandulosa*, *P. alba* x *P. tremuloides*, *P. balsamifera*, *P. balsamifera* subsp. *trichocarpa*, *P. balsamifera* subsp. *trichocarpa* x *P. deltoides*, *P. ciliata*, *P. deltoides*, *P. euphratica*, *P. euramericana*, *P. kitakamiensis*, *P. lasiocarpa*, *P. laurifolia*, *P. maximowiczii*, *P. maximowiczii* x *P. balsamifera* subsp. *trichocarpa*, *P. nigra*, *P. sieboldii* x *P. grandidentata*, *P. suaveolens*, *P. szechuanica*, *P. tomentosa*, *P. tremula*, *P. tremula* x *P. tremuloides*, *P. tremuloides*, *P. wilsonii*, *P. canadensis*, *P. yunnanensis*), coníferas tais como pinheiro teda (*Pinus taeda*), pinheiro da Flórida (*Pinus elliottii*), pinheiro ponderosa (*Pinus ponderosa*),

pinheiro *lodgepole* (*Pinus contorta*), e pinheiro Monterey (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); cicuta do Ocidente (*Tsuga canadensis*); espruce Sitka (*Picea glauca*); madeira vermelha (*Sequoia sempervirens*); abetos verdadeiros tais como abeto prata (*Abies amabilis*) e abeto de bálsamo (*Abies balsamea*); e cedros tais como cedro vermelho do Oeste (*Thuja plicata*) e cedro amarelo do Alasca (*Chamaecyparis nootkatensis*).

A presente invenção também contempla a modificação de plantas produtoras de fibras, tais como algodão (*Gossipium* spp.), linho (*Linum usitatissimum*), urtiga ferrão (*Urtica dioica*), lúpulo (*Humulus lupulus*), árvores de limeira (*Tilia cordata*, T. x. *europaea* e *T. platyphyllus*), sorgo espanhol (*Spartium junceum*), rami (*Boehmeria nivea*), amoreira de papel (*Broussonetia papyrifera*), linho da Nova Zelândia (*Phormium tenax*), apócino (*Apocynum cannabinum*), espécie de Iris (*I. douglasiana*, *I. macrosiphon* e *I. purdyi*), asclépias (espécies de *Asclépias*), abacaxi, banana e outros. São também contempladas colheitas de forragem, tais como alfafa, lólio, festuca e trevo.

Nesta descrição, "planta" indica, de modo amplo, qualquer material de planta contendo celulose que pode ser geneticamente manipulado, incluindo, mas sem limitação, células de plantas diferenciadas ou indiferenciadas, protoplastos, plantas inteiras, tecidos de plantas e órgãos de plantas, bem como qualquer componente de uma planta tais como folha, caule, raiz, broto, tubérculo, fruta, rizoma e similares.

Na presente descrição, "planta transgênica" refere-se a uma planta na qual foi incorporada uma sequência de ácido nucleico, incluindo, mas sem limitação, genes que não estão normalmente presentes em um genoma de planta hospedeira, sequências de ácidos nucleicos não normalmente transcritos em RNA ou traduzidos em uma proteína, ou quaisquer outros genes ou sequências de ácidos nucleicos que se deseja introduzir na planta tipo selvagem, tais como genes que normalmente podem estar presentes na planta do tipo selvagem, mas que se deseja ou engenheirar geneticamente ou ter a expressão alterada. A categoria de "planta transgênica" inclui tanto um transformante primário como uma planta que inclui um trans-

formante em sua linhagem, por exemplo, a título de introgressão padrão ou qualquer outro procedimento de multiplicação. Em contraste, uma planta que não é geneticamente manipulada é uma planta de controle e é referida como uma planta "não transgênica" ou de "controle". Planta não transgênica pode

5 ser uma planta cujo genoma não está modificado pela introdução de um construção compreendendo as sequências de polinucleotídeos ou fragmento destas da presente invenção. Pode ser também uma planta gerada a partir de células ou tecidos cultivados sem modificação anterior pela introdução de um construção que compreende as sequências de polinucleotídeos da in-

10 venção, ou pode compreender progênie recessiva de homozigoto (isto é, não tem nenhuma cópia do transgene) resultante de autofertilização de uma planta transgênica. Como utilizado aqui, uma "planta híbrida" se refere a uma planta ou a uma parte desta resultante de um cruzamento entre duas plantas parentes, em que um parente é a planta geneticamente alterada da

15 invenção. Isto pode ocorrer naturalmente, por exemplo, por reprodução sexual, ou artificialmente, por exemplo, por fusão nuclear *in vitro*.

Uma planta transgênica da presente invenção contém uma sequência de ácidos nucleicos, conforme descrição aqui, que é expressa sob o controle de um promotor operativo em plantas, de modo que a planta é ca-

20 racterizada, por exemplo, por um teor reduzido de lignina e um teor aumentado de celulose.

Métodos para Engenharia Genética

Os construtos de acordo com a invenção podem ser introduzidos em qualquer célula da planta, usando técnica de engenharia adequada. Tan-

25 to as células de planta do angiosperma como do gimnosperma de monocotiledôneas e dicotiledôneas podem ser geneticamente engenheiradas de vários modos conhecidos da técnica. Por exemplo, vide Klein et al., 1993, *Bio-technology* 4: 583-590; Bechtold et al., 1993, *C. R. Acad. Sci. Paris* 316:1194-1199; Koncz e Schell, 1986, *Mol. Gen. Genet.* 204: 383-396; Paszkowski et al., 1984, *EMBO J.* 3: 2717-2722; Sagi et al., 1994, *Plant Cell Rep.*

30 13: 262-266.

Espécies de *Agrobactéria* tais como *A. tumefaciens* e *A. rhiso-*

genes podem ser usadas, por exemplo, de acordo com Nagel et al., 1990, *Microbiol Lett* 67: 325. Em resumo, a *Agrobactéria* pode ser usada com um vetor de expressão de planta, por exemplo, eletroporação, depois da qual a *Agrobactéria* é introduzida nas células da planta via, por exemplo, o método bem conhecido de discos foliares.

Os métodos adicionais para obtenção disto incluem, mas sem limitação, a transformação por *Rhizobium*, *Sinorhizobium* ou *Mesorhizobium* (Broothaerts et al., 2005, *Nature* 433: 629-633), eletroporação, bombardeio com pistola de partículas, precipitação, e fusão de glicol, transferências nos grãos de pólen germinantes, transformação direta (Lorz et al., 1985, *Mol. Genet.* 199: 179-182), e outros métodos conhecidos da técnica. Se um marcador de seleção, tal como resistência à canamicina é empregada, torna mais fácil determinar as células que foram transformadas com êxito.

Os métodos de transformação de *Agrobactéria* discutidos acima são conhecidos por serem úteis para dicotiledôneas transformadas. Adicionalmente, de La Pena et al., 1987, *Nature* 325: 274-276; Rhodes et al, 1988, *Science* 240: 204-207; e Shimamoto et al, 1989, *Nature* 328: 274-276, todos os quais são incorporados a título de referência, têm monocotiledôneas cereais transformadas usando *Agrobactéria*. Vide Bechtold e Pelletier, 1998, *Methods Mol. Biol.* 82: 259-266, mostrando o uso de infiltração a vácuo para transformação mediada por *Agrobactéria*.

A presença de uma proteína, polipeptídeo, ou molécula de ácido nucleico em uma célula particular pode ser medida para determinar se, por exemplo, uma célula foi transformada ou transfectada com sucesso. A capacidade de executar tal ensaio é bem conhecida e não precisa ser reiterada aqui.

Quantificação do Comprimento da Fibra ou Altura da Planta

A frase "teor de celulose aumentado", empregado aqui para descrever uma planta da invenção, refere-se ao aumento quantitativo de celulose na planta quando comparado à quantidade de celulose em uma planta tipo selvagem. Um aumento quantitativo de celulose pode ser ensaiado por diversos métodos, como, por exemplo, por quantificação baseado nos açú-

cares totais depois da hidrólise ácida de polissacarídeos em madeira moída de embrião. Chiang e Sarkanen, *Wood Sci. Technol.* 17:217-26 (1983); Davis, *J. Wood Chem. Technol.* 18: 235-52 (1988).

O teor de celulose na planta engenheirada da invenção pode ser
 5 aumentado até níveis de cerca de 30% a cerca de 50%, mais preferivelmen-
 te de cerca de 25% a cerca de 45%, ainda mais preferivelmente de cerca de
 20% a cerca de 40% do teor de celulose de uma planta tipo selvagem. Uma
 modalidade mais preferível da planta da invenção possui um teor de celulose
 de cerca de 10% a cerca de 15% do teor de celulose de uma planta tipo sel-
 10 vagem.

As frases "teor de lignina reduzido" e "teor diminuído", utilizadas
 aqui para descrever um aspecto de uma planta da presente invenção, se
 referem respectivamente a uma redução quantitativa da quantidade de ligni-
 na em uma planta quando comparada à quantidade de lignina em uma plan-
 15 ta tipo selvagem ou não transformada. Uma redução quantitativa de lignina
 pode ser ensaiada por metodologia convencional ilustrada no ensaio de lig-
 nina Klason (Kirk et al., *Method in Enzymol.* 161: 87-101 (1988)) e o ensaio
 de brometo de acetila de lignina (Iiyama et al., *Wood Sci. Technol.* 22: 271-
 80 (1988)).

O teor de lignina em uma planta engenheirada da invenção pode
 20 ser reduzido a níveis de cerca de 5% a cerca de 90%, preferivelmente de
 cerca de 10% a cerca de 75%, ainda mais preferivelmente de cerca de 15%
 a cerca de 65% em peso seco do teor de lignina de uma planta tipo selva-
 gem. Uma modalidade mais preferível da planta da invenção possui um teor
 25 de lignina de cerca de 20% a cerca de 60% do teor da lignina tipo selvagem.

Exemplos são fornecidos abaixo sobre a metodologia para a ob-
 tenção de um gene C3HC4 de *Populus deltoides*, bem como técnicas para o
 uso de Agrobactéria na introdução do gene-alvo para produzir as plantas
 transformantes são dadas abaixo. Eles pretendem ser apenas meros exem-
 30 plos e não uma limitação da presente invenção.

EXEMPLO 1

Perfil de expressão de genes preferivelmente expressos em madeira de ten-

são, madeira de reação e madeira normal

Marcadores de Sequências Expressas (ESTs) de *Eucalyptus grandis* x *Eucalyptus urophylla* foram aglomerados usando o programa CAP3, conforme descrito por Huang e Madan, Genome Res. 9: 868-77 (1999), que é incorporado a título de referência. Um grupo de 53.522 ESTs foi obtido de bibliotecas representando os seguintes tecidos: madeira de tensão, madeira de reação, e madeira normal de árvores de eucalipto crescidas no campo (*Eucalyptus grandis* x *Eucalyptus urophylla*), de 6,5 m de altura). O conjunto de aglomeração assim gerados foi investigado quanto a aglomerações de compostos de pelo menos 90% de leituras de EST de bibliotecas representando tecido de madeira de tensão. Adicionalmente, o conjunto de aglomerações foi investigado quando a aglomerações de compostos de pelo menos três EST de tecido de tensão de madeira e, preferivelmente, menos que duas leituras de outras bibliotecas.

Uma aglomeração assim selecionada, compostos de 14 leituras de EST da biblioteca de cDNA de madeira de tensão e 0 leituras de outras bibliotecas (madeira de reação e normal), representa um membro da família de proteína C3HCA (Figura 1).

A aglomeração selecionada usando-se esses parâmetros foi então alinhada usando o algoritmo Blast-X com um valor e-valor de cutoff $\leq 1e-5$, vide Altschul et al., Nucleic Acids Res. 25: 3389-402 (1997), para sequências de uma base de dados da espécie *Populus* curada composta de sequências obtidas da base de dados v.1 de JGI *Populus trichocarpa* (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>). Os resultados de comparação foram armazenados em uma base de dados local de sequências de *Populus*. Por esse procedimento, uma codificação de cluster foi corrigida para a proteína C3HC4 que é ortóloga a uma escolhida das bibliotecas de eucalipto. A sequência da leitura mais longo nessa aglomeração é estabelecida aqui como a SEQ ID NO:1, que codifica o polipeptídeo descrito aqui sob SEQ ID NO:2.

EXEMPLO 2

Isolamento de uma sequência de DNA de C3HC4 de *Populus deltoides*

(a) Preparação de RNA de câmbio/xilema de *Populus deltoides* e síntese de cDNA

A casca foi removida a partir de talhos no caule em árvores de *Populus deltoides* de um ano de idade. A parte interna do caule, contendo câmbio, xilema e medula, foi cortada em pedaços pequenos, congelados em nitrogênio líquido e usados para extração de RNA usando o método de extração com brometo de cetiltrimetil-amônio (CTAB) (Aldrich e Cullis, Plant Mol. Biol. Report., 11;128-141 (1993)). Uma reunião de cDNA foi usada em experimentos de RT-PCR em que o RNA total isolado foi usado como modelo, e transcriptase reversa Superscript II (Invitrogen) e iniciador oligo (dT) foram usados para sintetizar o cDNA de primeira fita. O cDNA de fita dupla foi obtido pela reação de polimerase subsequente usando-se iniciadores específicos de gene, conforme descrição abaixo.

(b) Desenho de Iniciadores de PCR e reação de RT-PCR.

Oligômeros com base em SEQ ID NO:1 foram sintetizados como iniciadores para PCR, incluindo tanto a região em torno do primeiro códon ATG como em torno do códon de terminação do principal ORF que codifica o polipeptídeo para amplificar a região de codificação inteira do ORF. As sequências dos iniciadores são:

C3HC4NDE: Comprimento: 30

Catatgaata cgcggtaccc cttccaatg (SEQ ID NO: 3)

C3HC4XBA: Comprimento: 31

Tctagactat ctctccaatc ctgtttaca g (SEQ ID No:4)

A reunião de cDNA obtida em (a) foi usada como modelo em uma reação de PCR com os iniciadores das SEQ ID NOs: 5, 6, 7, e 8. A PCR envolveu 40 ciclos de 1 minuto a 94°C, 1 minuto a 51°C, e 2 minutos a 72°C seguidos por uma etapa extra de alongamento a 72°C, por 7 minutos. Os produtos de PCT foram isolados por eletroforese em gel em agarose a 1,0% seguido por brometo de etídio colorindo o gel submetido à eletroforese e detecção de bandas amplificadas em um trans-iluminador de UV. As bandas amplificadas detectadas foram verificadas e cortadas do gel de agarose com um barbeador. Os pedaços de gel foram transferidos para microtubos de 1,5

mL, e os fragmentos de DNA foram isolados e purificados usando-se um kit de limpeza de PCR e de purificação de banda de gel GFX (Amersham). Os fragmentos de DNA recuperados foram subclonados em um vetor de clonagem comercialmente disponível, transformados em *E. coli*, e então usados para preparar DNA plasmidial, que foi então sequenciado pelo método de didesóxi (Messing, Methods in Enzymol, 101, 2-78 (1983)) usando métodos padrão, resultando na sequência de nucleotídeos SEQ ID NO:1, que codifica o polipeptídeo descrito aqui sob SEQ ID NO:2.

EXEMPLO 3

10 Preparação das plantas transgênicas de *Nicotiana benthamiana*

As moléculas de ácido nucleico da *Populus deltoides* obtida no Exemplo 2 acima foram introduzidas em um hospedeiro de planta para produzir plantas de tabaco transgênico.

15 As moléculas de ácido nucleico isoladas da *Populus deltoides* e obtidas no Exemplo 2 foram clonadas em um vetor de expressão a jusante de um promotor de gene (C4H) de cumarato-4-hidroxalase preferido de xilema (Figura 2). Os contrutos de expressão resultantes foram amplificados em *E. coli*, e então transformados por transformação química em cepa de *A. tumefaciens* LBA4404.

20 Transformação mediada por agrobactéria de *Nicotiana benthamiana* foi obtida usando-se o método de disco foliar de Horsch et al., Science 227: 1229 (1985). Em resumo, a cepa de agrobactéria LBA4404 foi desenvolvida durante a noite até alcançar o crescimento de fase mid-log. As culturas foram diluídas 1:10 em água estéril e cocultivadas por 20 minutos
25 com discos foliares de plantas jovens *Nicotiana benthamiana* crescida estéril. Esses discos foram incubados em meio de Murashige-Skoog, no escuro. Depois de 48 horas, os discos foliares foram colocados de cabeça para baixo em placas frescas do mesmo meio de crescimento suplementado com 0,4 mg/L de ácido indolacético (IAA), 2 mg/L de benzil-aminopurina (hoBAP), 1
30 mg/L de Finale e 500 mg/L de carbenicilina. Quando os brotos se formaram, eles foram removidos do disco foliar e colocados em meio fresco, suplementado com somente 1 mg/L de Finale. Brotos de transformantes primários de

Nicotiana benthamiana, heterozigoto para o transgene, foram deixados enraizar em meio de Murashige e Skoog, e, subsequentemente, transferidos para o solo e crescidos na estufa. As condições (~50pM; m²/s de luz branca, 27°C) foram suficientes para identificar aquelas plantas transgênicas, que exibiram

5 estrutura de xilema e/ou composição química do xilema alteradas, ou uma combinação desses efeitos, de acordo com as descrições proporcionadas aqui.

EXEMPLO 4

Inserção de gene estranho de verificação de PCR no genoma da planta hospedeira

10

PCR foi usada para verificar a integração do construto de gene no genoma das plantas transgênicas. Um par de iniciadores foi sintetizado para amplificar uma sequência de DNA de 400 bp do gene *Bar* de marcador selecionável. Adicionalmente, um outro par foi sintetizado para amplificar o

15 gene de chalcona sintase (CHS) da *Nicotiana benthamiana*. Esses conjuntos de iniciadores estão descritos no Pedido Internacional Publicado, WO 2006/096951.

Bar 35: Comprimento: 20

tctaccatga gcccgaaacg

20 Bar 36: Comprimento: 23

aattcgggg atctggattt tag

CHS 150: Comprimento: 24

gccagcccaa atccaagatt actc

CHS 151: Comprimento: 23

25 aatgttagcc caacttcacg gag

Os iniciadores de *Bar* foram usados para amplificar por PCR parte da porção de T-DNA do construto de expressão contendo uma molécula de ácido nucleico da invenção, isto é, do DNA genômico de transformantes de *Nicotiana benthamiana*.

30 A mistura reacional da PCR continha 100 ng de DNA genômico de planta transformada, preparada usando-se o método de extração com brometo de cetiltrimetil-amônio (CTAB) (Aldrich e Cullis, *Plant Mol. Biol. Re-*

port 11: 128-41 (1993)), 0,2 μ M de cada iniciador para o gene de controle *CHS* endógeno, 100 μ M de cada trifosfato de desorribonucleotídeo, 1x de tampão de PCR e 2,5 Unidades de AmpliTaq DNA polimerase (Applied Biosystems) em um volume total de 50 μ L. Os parâmetros de ciclagem foram como a seguir: 94°C, por 1 minuto, 57°C, por 1 minuto e 72°C por 1 minuto, para 40 ciclos, com 5 minutos a 72°C, na extensão. Os produtos de PCR foram submetidos à eletroforese em um gel de agarose a 1%.

EXEMPLO 5

Determinação dos níveis de expressão de transgene em plantas transgênicas

RT-PCR semiquantitativa foi usada para detectar o acúmulo de transcritos C3HC4 de *Populus deltoides* no tecido do caule das plantas transgênicas. O RNA total foi isolado dos cortes de caule de plantas *Nicotiana* T0 transgênicas de 4 meses de idade usando o método de CTAB (Aldrich e Cullis, *Plant Mol. Biol. Report.* 11:128-141 (1993)). cDNA foi sintetizado de 500 ng de RNA total usando RNase H Superscript II-RT (Invitrogen, E.U.A.). Os iniciadores descritos acima foram usados juntamente com iniciadores para o gene constitutivo que codifica a chalcone sintase (CHS) como controle interno para normalizar a quantidade de RNA total usado em cada amostra. A PCR foi feita com uma diluição de 12,5 vezes do cDNA de primeira fita sob as seguintes condições: 94°C por 3 minutos e 27 ciclos de 94°C por 1 minuto, 51°C por 1 minuto, e 72°C por 1 minuto e 30 segundos.

A descrição acima e os exemplos descrevem várias características da invenção, que essencialmente levam ao isolamento e à clonagem de moléculas de ácido nucleico que codificam um membro da família de proteínas C3HC4 e que são úteis na produção de plantas geneticamente engenhiradas. As plantas recombinantes que foram transformadas ou transfectadas com tal molécula de ácido nucleico isolada podem exibir alteração quantitativa do teor de celulose e/ou do teor de lignina.

EXEMPLO 6

Análise histoquímica de plantas transgênicas

Caules de *Nicotiana* transgênica e plantas não transgênicas de

controle foram seccionadas e fixadas em paraformaldeído a 4% por 24 horas. Tecidos fixados foram então seccionados em um micrótomo (Leica RM2255) e subsequentemente coloridos com azul astra/safranina. As secções coloridas histologicamente foram observadas em um microscópio invertido Leica DM1I usando iluminação de campo, clara e escura.

EXEMPLO 7

Aumento do teor de celulose em plantas transgênicas que superexpressam C3HC4 em tecidos vasculares

Os caules principais de eventos transgênicos de *Nicotiana* transformada com construções que compreendem o gene de C3HC4 de *Populus deltoides* sob o controle do promotor C4H de *Populus deltoides* preferido do xilema e plantas de controle não transgênicas foram coletados e secados a ar por duas semanas. Os caules secos foram cortados em pedaços e formados em pó em um moinho de facas usando-se uma peneira de 30 malhas.

As amostras em pó de caule foram então submetidas a análises químicas para determinar o teor de celulose e de lignina. Em resumo, os teores de celulose e hemicelulose foram determinados com base nos açúcares totais depois da hidrólise ácida destes polissacarídeos extraídos do caule. Os caules moídos foram secados sob vácuo a 45°C e hidrolisados com H₂SO₄. Seguinte a cromatografia de troca iônica em pH alto, glicana e outros polissacarídeos (hemiceluloses) foram quantificados como base na composição do hidrolisado. Chiang e Sarkanen (1983) e Davis (1988), acima. Três dos eventos transgênicos de C3HC4, conhecidos por expressarem o transgene de acordo com o procedimento detalhado no Exemplo 5, mostram um aumento estatisticamente significativo de teor de celulose (Figura 3). O evento transgênico 6B exibe 54,09% de celulose em comparação com 50,00% das plantas não transgênicas de controle, representando um aumento significativo de 8,18% de teor de celulose ($P \leq 0,05$, t-teste). O evento transgênico 24B exibe 53,90% de celulose, em comparação com 50,00% nas plantas de controle, representando 7,80% de aumento no teor de celulose (Figura 3. $P \leq 0,05$, t-teste). O evento transgênico 4B exibe 53,26% de teor de celulose em comparação com 50,00% em plantas não transgênicas de controle, representan-

do 6,52% de aumento do teor de celulose (Figura 3; $P \leq 0,05$, t-teste).

Depois de crescer até a maturidade, os eventos T0 foram autopolinizadas para geração de linhagens T1. Aqui, são apresentados os resultados referentes a três eventos que têm a sua geração T1 analisada.

5 A análise da população T1 a partir do evento 6B indicou que a condição dominante do homozigoto para o gene de C3HC4 é letal, já que não foi detectada nenhuma planta homozigótica na população segregante. O desenvolvimento da planta foi provavelmente afetado. No entanto, as plantas que são hemizigóticas para o gene de CrHC4 apresentaram um aumento
10 significante de 8,4% do teor de celulose ($P < 0,05$, t-test), quando comparadas com plantas recessivas homozigóticas (Figura 4). Elas mostraram também uma redução de 18% de seu teor de lignina ($P < 0,05$, t-teste), em comparação com plantas recessivas homozigóticas (Figura 5).

Na população segregante de eventos 24B e 25B, foi possível
15 identificar plantas dominantes homozigóticas. Mas o aumento maior de teor de celulose foi observado no grupo de plantas hemizigóticas, em comparação com as plantas recessivas homozigóticas. As plantas homozigóticas do evento 24B mostraram um aumento de 9,7% do teor de celulose e as plantas dominantes homozigóticas apresentaram um aumento de 7,5% de teor
20 de celulose em comparação com plantas recessivas homozigóticas (Figura 6; $P \leq 0,05$, t-teste). As plantas hemizigóticas do evento 25B mostraram um aumento de 10,4%, em comparação com o grupo de plantas recessivas homozigóticas (Figura 7; $P \leq 0,05$, t-teste). Nenhuma alteração significativa do teor de lignina foi observada nesses dois eventos.

LISTAGEM DE SEQUÊNCIA

<110> GERHARDT, Isabel Rodrigues
ARRUDA, Paulo

<120> MOLÉCULAS DE ÁCIDOS NUCLEICOS QUE CODIFICAM PROTEÍNAS DE PLANTAS NA FAMÍLIA C3HC4 E
PROCESSOS PARA A ALTERAÇÃO DO CONTEÚDO DE CELULOSE E LIGNINA DA PLANTA

<140>

<141>

<150> 4

<210> 1

<211> 1029

<212> DNA

<213> Populus deltoides

<221> DNA

<220>

<221> CDs

<222> {1}...(1029)

<223> Populus deltoides C3HC4 proteína, cDNA clone

<400> 1

```

atg aat acg cgg tac ccc ttt cca atg aat tca ttc tgt ggt tca agt   48
Met Asn Thr Arg Tyr Pro Phe Pro Met Asn Ser Phe Cys Gly Ser Ser
          5                      10                      15

cct gca gca tct cca gca gga gaa gag cgt gga act gtg tct ctc aga   96
Pro Ala Ala Ser Pro Ala Gly Glu Glu Arg Gly Thr Val Ser Leu Arg
          20                      25                      30

aac agg tot gct cgc act ccc act tgg tcc tta ttg gta aga aca gcg   144
Asn Arg Ser Ala Arg Thr Pro Thr Ser Ser Leu Leu Val Arg Thr Ala
          35                      40                      45

atg aga ata tct aga gca aga tgg ttc acc ttc ttg aga aga gta ttc   192
Met Arg Ile Ser Arg Ala Arg Trp Phe Thr Phe Leu Arg Arg Val Phe
          50                      55                      60

cac tac cag aat ggc tca agg tct aat ctt ggg tot aat cct ttc aac   240
His Tyr Gln Asn Gly Ser Arg Ser Asn Leu Gly Ser Asn Pro Phe Asn
          65                      70                      75                      80

tcc agc act tgg atg atg ctg gag ttt gta gct ttg gtc ctt caa ata   288
Ser Ser Thr Trp Met Met Leu Glu Phe Val Ala Leu Val Leu Gln Ile
          85                      90                      95

agt att aac atg ttc aac ctg gct att tcc aag gca gag aag cca gtt   336
Ser Ile Thr Met Phe Thr Leu Ala Ile Ser Lys Ala Glu Lys Pro Val
          100                     105                     110

tgg cct gtg aga ata tgg att att gcc tat aat att ggt tgt gtc ctt   384
Trp Pro Val Arg Ile Trp Ile Ile Gly Tyr Asn Ile Gly Cys Val Leu
          115                     120                     125

agt ctg ctg ctg ttc tat ggg cga tac cgg caa att aac acg act caa   432
Ser Leu Leu Leu Phe Tyr Gly Arg Tyr Arg Gln Ile Asn Thr Thr Gln
          130                     135                     140

ggc gat ggt ttt ggc cta cct gat ttg gaa caa cag agg ggc agt gag   480
Gly Asp Gly Phe Gly Leu Pro Asp Leu Glu Gln Gln Arg Gly Ser Glu
          145                     150                     155                     160

gaa tcc agt gta tgc agg tgc tca att ttg atg cac aag tgc cgg act   528
Glu Ser Ser Val Cys Arg Cys Ser Ile Leu Met His Lys Cys Arg Thr
          165                     170                     175

```

tca ctc gag ctc ttc ttt gca ata tgg ttt gtg atg ggt aat gtt tgg 576
 Ser Leu Glu Leu Phe Phe Ala Ile Trp Phe Val Met Gly Asn Val Trp
 180 185 190

gtc ttt gat tct cgc ttt gga tcc tac cac cgt gct cca aaa ctc cat 624
 Val Phe Asp Ser Arg Phe Gly Ser Tyr His Arg Ala Pro Lys Leu His
 195 200 205

gtg cta tgc atc tct cta ctt gcc tgg aat gct ctc agt tac tcc ttc 672
 Val Leu Cys Ile Ser Leu Leu Ala Trp Asn Ala Leu Ser Tyr Ser Phe
 210 215 220

cca ttt ctc ttg ttt cta ctc tta tgt tgt tgc gtg cca ctc att agc 720
 Pro Phe Leu Leu Phe Leu Leu Leu Cys Cys Cys Val Pro Leu Ile Ser
 225 230 235 240

act gtc ctt ggt tac aac atg aac atg ggt tct gcc gag aga ggg gct 768
 Thr Val Leu Gly Tyr Asn Met Asn Met Gly Ser Ala Glu Arg Gly Ala
 245 250 255

tcc gac gat caa atc tcc agt ctn ccc agt tgg agg tac aaa gca gct 816
 Ser Asp Asp Gln Ile Ser Ser Leu Pro Ser Trp Arg Tyr Lys Ala Ala
 260 265 270

gac acc aac tgc gag ttt cgc aat aat gct gac tgc aat tca acc att 864
 Asp Thr Asn Ser Glu Phe Arg Asn Asn Ala Asp Cys Asn Ser Thr Ile
 275 280 285

gca agt gaa gat ctg gaa tgt tgt atc tgc cta gcc aag tat aaa gac 912
 Ala Ser Glu Asp Leu Glu Cys Cys Ile Cys Leu Ala Lys Tyr Lys Asp
 290 295 300

aaa gaa gaa gtt agg aaa ttg cca tgt tcc cat atg ttc cac ctc aag 960
 Lys Glu Glu Val Arg Lys Leu Pro Cys Ser His Met Phe His Leu Lys
 305 310 315 320

tgt gta gat caa tgg ctt cga att ata tcc tgc tgc cct ctc tgt aaa 1008
 Cys Val Asp Gln Trp Leu Arg Ile Ile Ser Cys Cys Pro Leu Cys Lys
 325 330 335

caa gga ttg gag aga tag 1029
 Gln Gly Leu Glu Arg

<210> 2
 <211> 342
 <212> FRT
 <213> Populus deltoides
 <220>
 <400> 2

Met Asn Thr Arg Tyr Pro Phe Pro Met Asn Ser Phe Cys Gly Ser Ser
 5 10 15

Pro Ala Ala Ser Pro Ala Gly Glu Glu Arg Gly Thr Val Ser Leu Arg
 20 25 30

Asn Arg Ser Ala Arg Thr Pro Thr Ser Ser Leu Leu Val Arg Thr Ala
 35 40 45

Met Arg Ile Ser Arg Ala Arg Trp Phe Thr Phe Leu Arg Arg Val Phe
 50 55 60

His Tyr Gln Asn Gly Ser Arg Ser Asn Leu Gly Ser Asn Pro Phe Asn
 65 70 75 80

Ser Ser Thr Trp Met Met Leu Glu Phe Val Ala Leu Val Leu Gln Ile
 85 90 95

Ser Ile Thr Met Phe Thr Leu Ala Ile Ser Lys Ala Glu Lys Pro Val
 100 105 110
 Trp Pro Val Arg Ile Trp Ile Ile Gly Tyr Asn Ile Gly Cys Val Leu
 115 120 125
 Ser Leu Leu Leu Phe Tyr Gly Arg Tyr Arg Gln Ile Asn Thr Thr Gln
 130 135 140
 Gly Asp Gly Phe Gly Leu Pro Asp Leu Glu Gln Gln Arg Gly Ser Glu
 145 150 155 160
 Glu Ser Ser Val Cys Arg Cys Ser Ile Leu Met His Lys Cys Arg Thr
 165 170 175
 Ser Leu Glu Leu Phe Phe Ala Ile Trp Phe Val Met Gly Asn Val Trp
 180 185 190
 Val Phe Asp Ser Arg Phe Gly Ser Tyr His Arg Ala Pro Lys Leu His
 195 200 205
 Val Leu Cys Ile Ser Leu Leu Ala Trp Asn Ala Leu Ser Tyr Ser Phe
 210 215 220
 Pro Phe Leu Leu Phe Leu Leu Leu Cys Cys Cys Val Pro Leu Ile Ser
 225 230 235 240
 Thr Val Leu Gly Tyr Asn Met Asn Met Gly Ser Ala Glu Arg Gly Ala
 245 250 255
 Ser Asp Asp Gln Ile Ser Ser Leu Pro Ser Trp Arg Tyr Lys Ala Ala
 260 265 270
 Asp Thr Asn Ser Glu Phe Arg Asn Asn Ala Asp Cys Asn Ser Thr Ile
 275 280 285
 Ala Ser Glu Asp Leu Glu Cys Cys Ile Cys Leu Ala Lys Tyr Lys Asp
 290 295 300
 Lys Glu Glu Val Arg Lys Leu Pro Cys Ser His Met Phe His Leu Lys
 305 310 315 320
 Cys Val Asp Gln Trp Leu Arg Ile Ile Ser Cys Cys Pro Leu Cys Lys
 325 330 335
 Gln Gly Leu Glu Arg
 340

<210> 3
 <211> 30
 <212> DNA
 <213> Sequência artificial
 <220> oligonucleotídeo sintético (iniciador C3HC4NDE)
 <223> 3
 catatgaata cgggtaccc ctttccaatg

30

<210> 4
 <211> 31
 <212> DNA
 <213> Sequência artificial
 <220> oligonucleotídeo sintético (iniciador C3HC4NDE)
 <223> 4
 totagactat ctctccaatc cttgattaca g

31

REIVINDICAÇÕES

1. Sequência de ácido nucleico isolada que compreende uma sequência selecionada do grupo que consiste em:

5 (a) uma sequência de ácido nucleico apresentada na SEQ ID NO: 1 ou o filamento complementar da mesma;

(b) uma sequência de ácido nucleico que codifica a sequência de aminoácidos apresentada na SEQ ID NO: 2;

10 (c) uma sequência de ácido nucleico capaz de se hibridizar sob condições estridentes com uma sequência de ácido nucleico de (a) ou (b), em que a dita sequência de hibridização codifica um polipeptídeo C3HC4;

(d) um ácido nucleico que é uma variação alélica ou uma variação de processamento alternativo da sequência de ácido nucleico de (a) ou (b); e

15 (e) uma sequência de ácido nucleico que possui pelo menos 50%, 60%, 70%, 75%, 80%, 85%, 90% ou mais de identidade de sequência em relação à sequência de (a) ou (b).

2. Proteína C3HC4 isolada selecionada do grupo que consiste em:

20 (a) um polipeptídeo apresentado na SEQ ID NO 2;

(b) um polipeptídeo com uma sequência de aminoácidos que possui pelo menos 85% ou mais de identidade de sequência em relação à sequência de aminoácidos apresentada na SEQ ID NO 2; e

(c) uma variação de um polipeptídeo que é definido em (a) ou (b).

25 3. Construto de ácido nucleico que compreende uma sequência de polinucleotídeo de C3HC4 isolada ligada de forma operacional a um ou mais promotores adequados que controlam a expressão da sequência de polinucleotídeo de C3HC4.

30 4. Construto de ácido nucleico de acordo com a reivindicação 3, em que o promotor é um promotor preferido pelo xilema.

5. Construto de ácido nucleico de acordo com a reivindicação 4, em que o dito promotor preferido pelo xilema é selecionado do grupo que

consiste no promotor do gene TUB, do promotor do gene SuSy, do promotor do gene COMT e do promotor do gene C4H.

6. Célula vegetal que compreende o construto de ácido nucleico como definido na reivindicação 3.

5 7. Planta transgênica gerada partindo da célula vegetal como definida na reivindicação 6, em que a dita planta possui conteúdo alterado de celulose e/ou lignina comparado com a de uma planta não transgênica da mesma espécie.

10 8. Célula vegetal de acordo com a reivindicação 6, em que o dito promotor é um promotor preferido pelo xilema.

9. Célula vegetal de acordo com a reivindicação 8, em que o dito promotor preferido pelo xilema é selecionado do grupo que consiste do promotor do gene TUB, do promotor do gene SuSy, do promotor do gene COMT e do promotor do gene C4H.

15 10. Planta transgênica de acordo com a reivindicação 7, em que a dita planta é uma planta dicotiledônea.

11. Planta transgênica de acordo com a reivindicação 7, em que a dita planta é uma planta monocotiledônea.

20 12. Planta transgênica de acordo com a reivindicação 7, em que a dita planta é uma gimnosperma.

13. Planta transgênica de acordo com a reivindicação 7, em que a dita planta é uma árvore de madeira de lei.

14. Planta transgênica de acordo com a reivindicação 13, em que a dita árvore de madeira de lei é uma planta de *Eucalyptus*.

25 15. Planta transgênica de acordo com a reivindicação 13, em que a dita árvore de madeira de lei é uma planta de *Populus*.

16. Planta transgênica de acordo com a reivindicação 11, em que a dita árvore de madeira mole é uma planta de *Pinus*.

30 17. Parte da planta transgênica de acordo com a reivindicação 7, selecionada do grupo que consiste em uma folha, um caule, uma flor, um ovário, um fruto, uma semente e um calo.

18. Progenie da planta transgênica como definida na reivindica-

ção 7.

19. Progenie de acordo com a reivindicação 18, em que a dita progenie é uma planta híbrida.

5 20. Processo para alterar o conteúdo de celulose e/ou de lignina em uma planta, que compreende

(a) a introdução em uma célula vegetal isolada de uma construto de ácido nucleico que compreende uma sequência de polinucleotídeo de C3HC4 isolada, ligada de forma operacional a um ou mais promotores adequados que controlam a expressão da sequência de polinucleotídeo de C3HC4; e
10

(b) o cultivo da dita célula vegetal sob condições que promovem o crescimento de uma planta, em que a dita planta superexpressa a proteína C3HC4 e conteúdo de celulose maior e/ou de lignina reduzido comparado ao de uma planta não transgênica da mesma espécie.

15 21. Processo de acordo com a reivindicação 20, em que o promotor é um promotor preferido pelo xilema.

22. Processo de acordo com a reivindicação 21, em que o promotor preferido pelo xilema é selecionado do grupo que consiste do promotor do gene TUB, do promotor do gene SuSy, do promotor do gene COMT e
20 do promotor do gene C4H.

FIG. 1

Perfil de Expressão em um conjunto de Tecidos de Eucalipto de cDNA de C3HC4

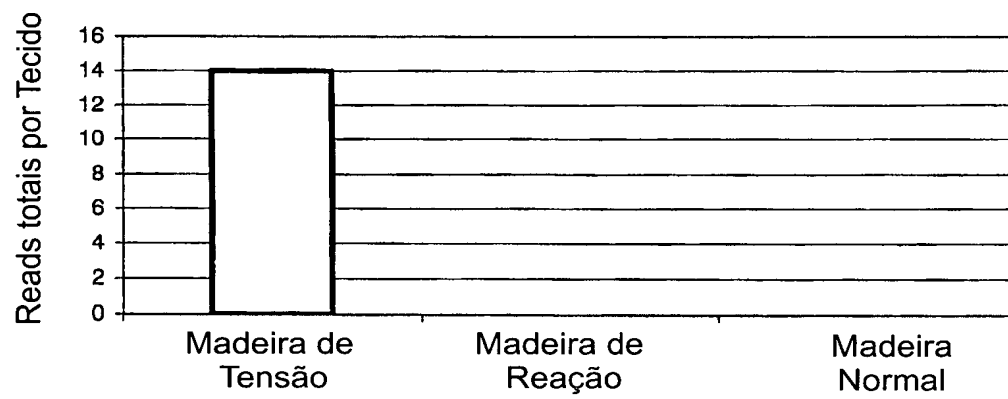


FIG. 2

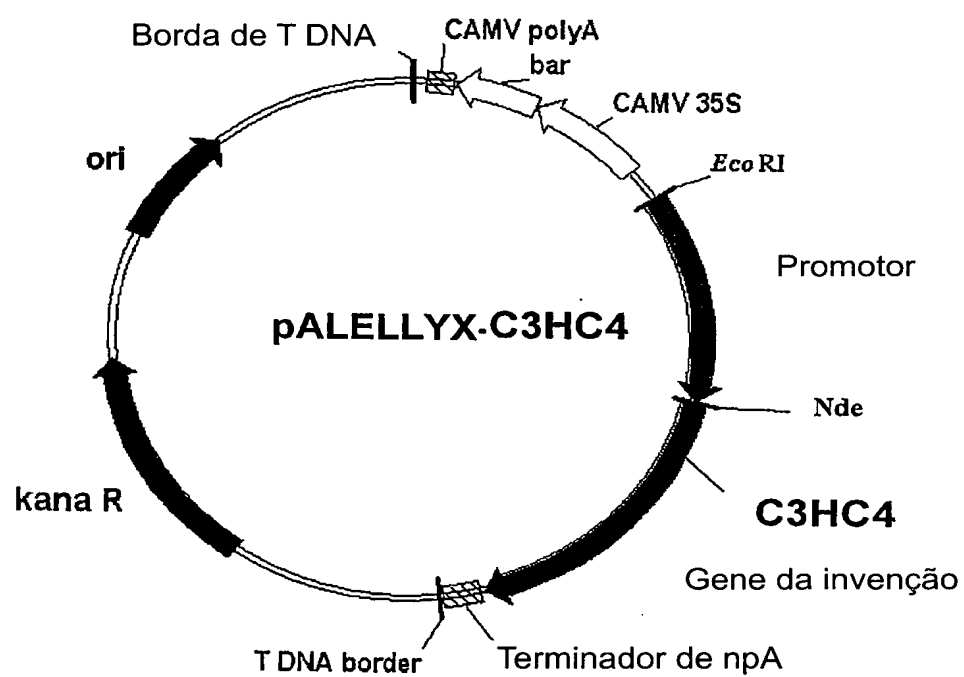


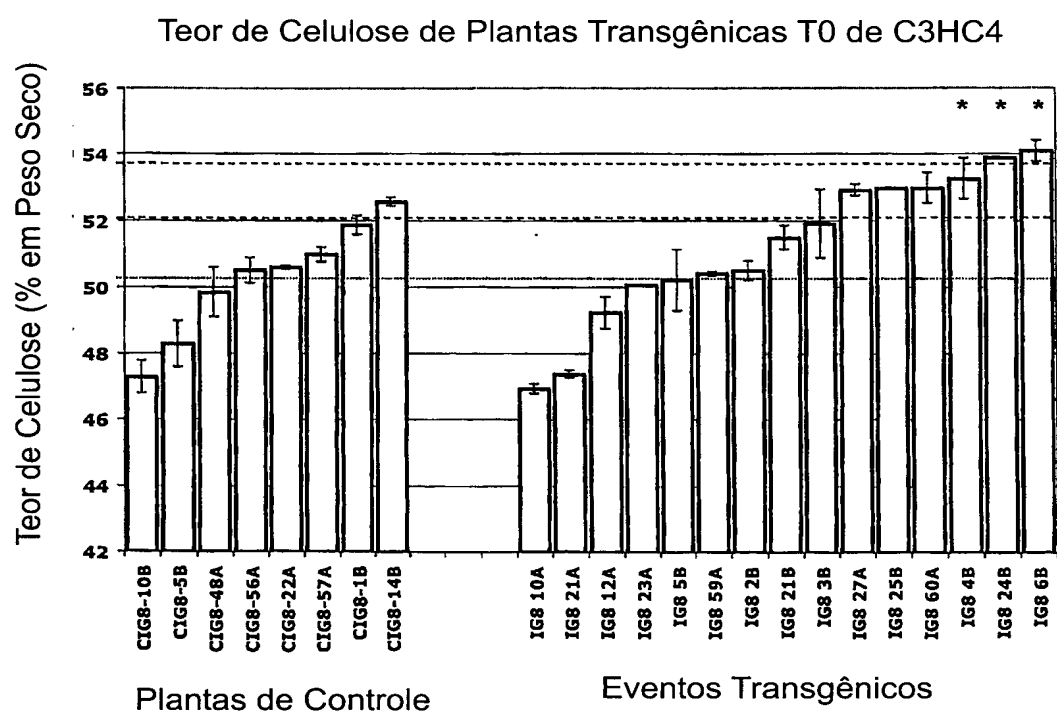
FIG. 3

FIG. 4

Teor de Celulose do Evento 6B (Plantas Transgênicas T1)

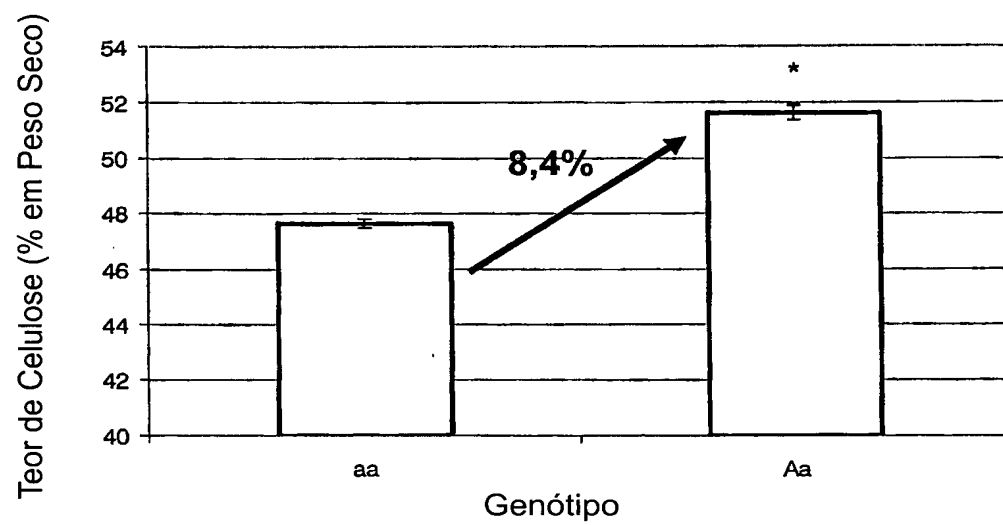


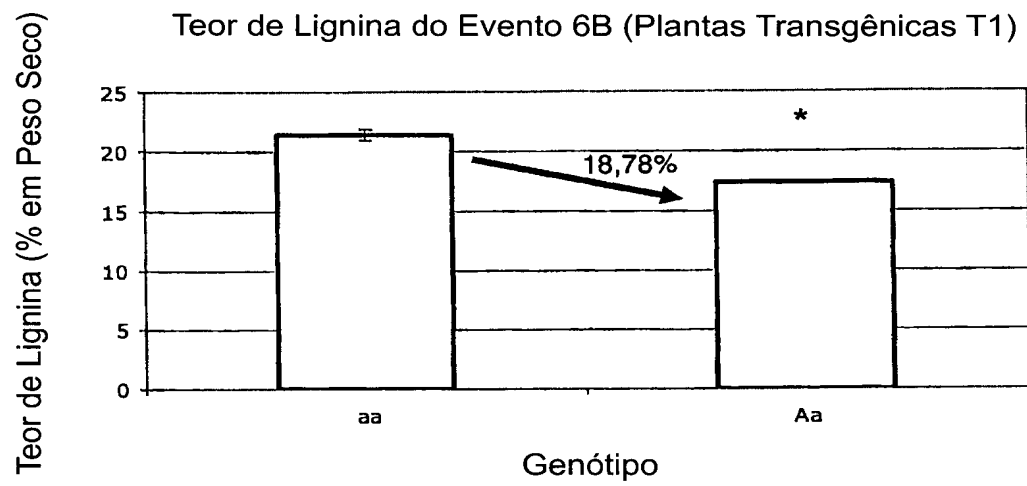
FIG. 5

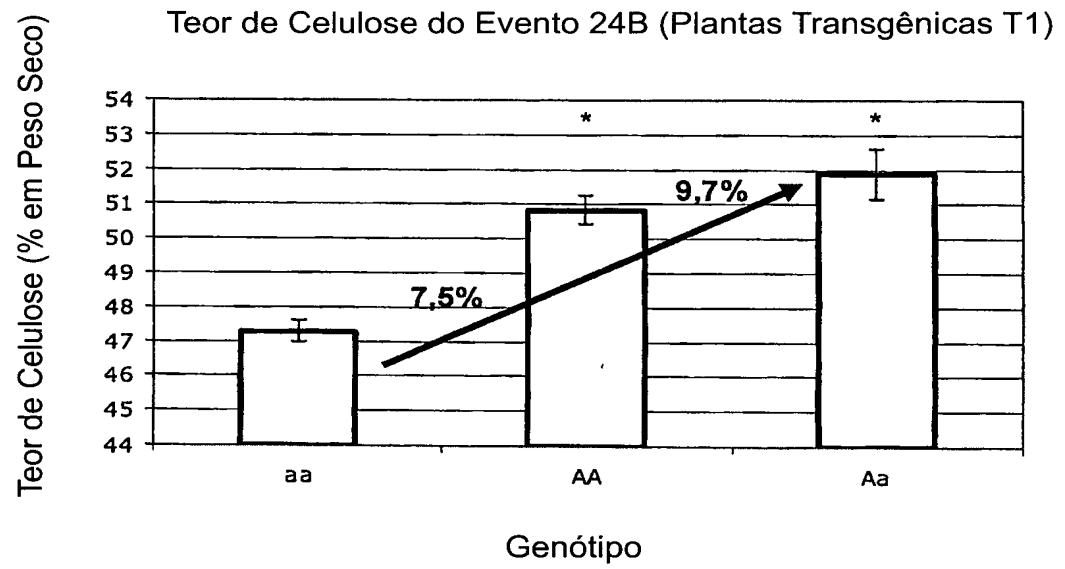
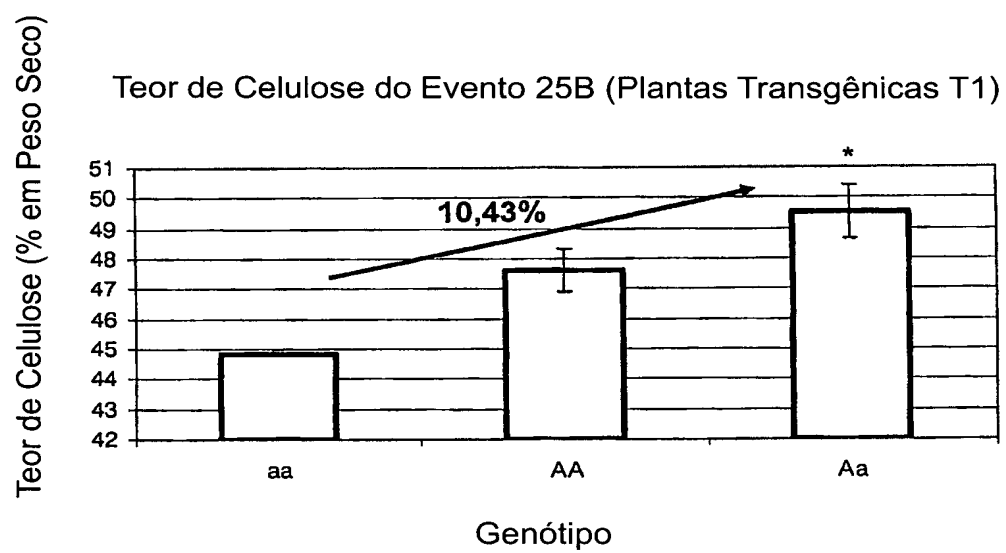
FIG. 6

FIG. 7

RESUMO

Patente de Invenção: **"MOLÉCULAS DE ÁCIDOS NUCLEICOS QUE CODIFICAM PROTEÍNAS DE PLANTAS NA FAMÍLIA C3HC4 E PROCESSOS PARA A ALTERAÇÃO DO CONTEÚDO DE CELULOSE E LIGNINA DA PLANTA"**.

5 A presente invenção refere-se a polinucleotídeos, construtos de ácidos nucleicos e processos para a modificação do conteúdo de celulose e/ou de lignina nos tecidos vegetais. As plantas são engenheiradas geneticamente com um gene que codifica uma proteína C3HC4, que leva a um maior conteúdo de celulose quando superexpressa no sistema vascular da planta. Os transformantes de plantas que carregam o gene da proteína C3HC4 exibem maior conteúdo de celulose e/ou menor conteúdo de lignina, características que se acredita que melhorem as árvores de madeira de lei em relação à extração de celulose durante polpação e a produção de papel.

10