

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

(21) **PI0806507-1 A2**

(22) Data de Depósito: 07/01/2008
(43) Data da Publicação: 06/09/2011
(RPI 2122)



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

(54) Título: EXPRESSÃO DIFERENCIAL DE ALELOS SUBGENÔMICOS ESPECÍFICOS NO ALGODÃO E USOS DOS MESMOS

(30) Prioridade Unionista: 11/01/2007 EP 07 000550.9,
11/01/2007 US 60/884,564

(73) Titular(es): Bayer Bioscience N.V.

(72) Inventor(es): Antonio Arioli, Steven Engelen

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

(86) Pedido Internacional: PCT EP2008000161 de
07/01/2008

(87) Publicação Internacional: WO 2008/083969de
17/07/2008

(57) Resumo: EXPRESSÃO DIFERENCIAL DE ALELOS SUBGENÔMICOS ESPECÍFICOS NO ALGODÃO E USOS DOS MESMOS. A presente invenção refere-se a métodos e meios para alterar as propriedades da fibra tais como o comprimento da fibra em plantas que produzem fibra tal como o algodão. Adicionalmente, são providos promotores para plantas com um perfil de expressão preferencial para fibra ou seletiva para fibra.



PI0806507-1

Relatório Descritivo da Patente de Invenção para "EXPRESSÃO DIFERENCIAL DE ALELOS SUBGENÔMICOS ESPECÍFICOS NO ALGODÃO E USOS DOS MESMOS".

Campo da Invenção

5 A presente invenção refere-se ao campo da agricultura, mais especificamente dirigida ao uso de técnicas de biologia molecular para alterar plantas que produzem fibras, particularmente plantas de algodão e/ou para acelerar a reprodução de tais plantas que contêm fibras. São providos métodos e meios para alterar a qualidade das fibras tais como aumentar o

10 comprimento das fibras, particularmente o comprimento das fibras em plumas ou para diminuir o comprimento de fibras distorcidas. Também são providos métodos para identificar marcadores moleculares associados com o comprimento da fibra em uma população de variedades de algodão e plantas parentais relacionadas. Além disso, são fornecidas plantas promotoras

15 com perfil de expressão preferencial para fibra ou seletivo para fibra. Os promotores também são temporariamente regulados.

Antecedentes da Técnica

O algodão fornece a maior parte da fibra de alta qualidade para a indústria têxtil. A modificação das características das fibras de algodão que

20 melhor satisfaz as necessidades da indústria é um empenho mais importante na reprodução tanto por métodos clássicos quanto por alterar geneticamente o genoma das plantas de algodão.

Cerca de 90% do algodão cultivado mundialmente é *Gossypium hirsutum* L., enquanto que *Gossypium barbadense* é responsável por cerca

25 de 8%. Como na maioria das plantas que produzem flores, os genomas de algodão são considerados como tendo incorrido em um ou mais eventos de poliploidização e terem desenvolvido-se pela união de genomas divergentes em um núcleo comum. O comércio do algodão é dominado por formas aperfeiçoadas de duas espécies tetraploides "AD", *Gossypium hirsutum* L. e

30 *Gossypium barbadense* L. Considera-se que algodões tetraploides tenham formado-se a cerca de 1-2 milhões de anos atrás, no Novo Mundo, pela hibridização entre o táxon genômico materno Old World "A" que se assemelha

ao *Gossypium herbaceum* e táxon genômico paterno New World "D" que se assemelha ao *Gossypium raimondii* ou *Gossypium gossypoides*. Os táxons de genoma diploide A selvagem e de *Gossypium* tetraploide AD produzem fibras que podem ser tecidas. Uma espécie de genoma diploide A, *Gossypium arboreum*, permanece intensivamente reproduzida e cultivada na Ásia. Seu ascendente mais próximo e possível, a espécie de genoma diploide A *G. herbaceum* também produz fibra capaz de ser tecida. Embora as sementes de genoma diploide D sejam sexualmente maduras, nenhuma produz fibras capazes de serem tecidas. Nenhum táxon de outros genomas diploides de *Gossypium* reconhecidos (B, C, E, F e G) foi domesticado. A seleção intensa direcionada por seres humanos tem produzido consistentemente algodões tetraploides AD que têm rendimento e/ou características de qualidade superiores comparados aos cultivares de genoma diploide A. A reprodução seletiva de *G. hirsutum* (AADD) tem enfatizado o rendimento máximo, enquanto que *G. barbadense* (AADD) é valorizado por suas fibras de comprimento, resistência e pureza superiores (Jiang *et al.*, 1998 – Proc Natl Acad Sci USA. 14 de Abril de 1998; 95(8): 4419-4424).

Uma fibra de algodão é uma célula única que se inicia na epiderme do tegumento externo dos óvulos, na antese ou pouco antes. Posteriormente, as fibras alongam-se rapidamente por cerca de 3 semanas antes delas mudarem para a síntese intensiva de celulose da parede celular secundária. As células da fibra interligam-se apenas ao revestimento subjacente da semente nas suas extremidades basais e o influxo de solutos, água e outras moléculas ocorre através tanto dos plasmodesmos quanto da membrana plasmática. Ruan *et al.*, 2001 (Plant Cell 13:47-63) demonstraram um fechamento transitório dos plasmodesmos durante o alongamento da fibra. Ruan *et al.*, 2004 (Plant Physiology – Vol 136:pp. 4104-4113) compararam a duração do fechamento de plasmodesmos entre diferentes genótipos de algodão que diferiam no comprimento da fibra e encontraram uma correlação positiva entre a duração do fechamento de plasmodesmos e o comprimento da fibra. Posteriormente, foi apresentada a evidência microscópica que mostrou deposição de calose e degradação da base da fibra, que se correlaciona

com o período de fechamento e reabertura de plasmodesmos. Além disso, a expressão de um gene de β -1,3-endoglucanase (GhGluc 1) nas fibras, permitindo degradar a calose, correlacionou-se com a reabertura dos plasmodesmos na base da fibra.

5 WO2005/017157 descreve métodos e meios para modular o comprimento da fibra em plantas que produzem fibras tais como o algodão, pela alteração da fase de alongamento da fibra como descrito em Ruan et al 2001. A fase de alongamento da fibra pode ser aumentada ou diminuída pela interferência com a deposição de calose nos plasmodesmos na base das
10 células da fibra.

Além disso, poderá ser interessante para a modificação de fibras através de manipulação genética, possuir promotores que são preferencialmente ou especificamente expressos em células de fibra apenas e/ou que são expressos apenas a partir de um estágio particular de desenvolvi-
15 mento da fibra.

WO2004/018620 refere-se a uma molécula de ácido nucleico isolado que codifica uma quitinase endógena de algodão e seu promotor, que são preferencialmente expressos nas fibras durante a deposição da parede secundária. O polipeptídeo codificado pela molécula de ácido nucleico, um
20 construto de DNA que liga a molécula de ácido nucleico isolado com um promotor, o construto de DNA incorporado em um sistema de expressão, uma célula hospedeira, uma planta ou uma semente de planta também são descritos. O documento também refere-se a um construto de DNA que liga o promotor isolado com um segundo DNA assim como sistemas de expressão,
25 células hospedeiras, plantas ou sementes de plantas que contém o construto de DNA. Métodos para conferir resistência a insetos e fungos, regular o conteúdo de celulose da fibra e métodos para expressar um gene preferencialmente nas fibras durante a deposição da parede secundária também são descritos.

30 Poderá ser útil ter promotores alternativos que dirigirão a expressão de gene preferencialmente e/ou acentuadamente nas fibras durante a deposição da parede secundária, isto é, acentuadamente e continuamente

(por exemplo, >50% de sua atividade máxima), por exemplo, a partir do início da deposição da parede secundária até seu término ou, por exemplo, a partir do estágio de maturação. O início da deposição da parede secundária é definido como o momento em que o peso seco/unidade de comprimento de uma fibra de algodão começa a aumentar ou quando o peso seco/unidade de superfície de área de qualquer célula começa a aumentar através da síntese de parede com novo material que contém mais do que 40% (peso/peso) de celulose. No caso de fibra de algodão de *G. hirsutum* L., isso é esperado ocorrer entre 14-17 DPA quando as plantas de algodão são cultivadas sob condições típicas na estufa ou no campo (temperatura diurna de 26 a 34°C, temperatura noturna de 20 a 26°C, intensidade luminosa maior ou igual a 1000 einsteins/m²/s, com água e nutrição mineral adequadas). O final da formação da parede celular secundária e o início da fase de maturação está geralmente em torno de 35 DPA no caso de fibra de algodão de *G. hirsutum* L.

Além disso, poderá ser útil ter promotores alternativos que dirijam a expressão de gene apenas ou preferencialmente nas fibras enquanto excluem ou minimizam a expressão em outros tecidos.

As invenções descritas a seguir nas diferentes modalidades, exemplos, figuras e reivindicações proveem métodos e meios aperfeiçoados para modulação de comprimento de fibra pela diminuição ou aumento da deposição de calose na base da célula da fibra em um momento particular. As invenções descritas a seguir também proveem promotores específicos para fibra e/ou preferenciais para fibra e regiões promotoras.

25 Sumário da Invenção

Em uma modalidade da invenção, é provido um promotor preferencial para célula de fibra compreendendo uma sequência de nucleotídeos selecionada do seguinte grupo de sequências de nucleotídeo:

a) uma sequência de nucleotídeos que compreende a sequência de nucleotídeos de SEQ ID N° 9 a partir do nucleotídeo na posição 2149 até o nucleotídeo na posição 2307;

b) uma sequência de nucleotídeos que compreende a sequência

de nucleotídeos de SEQ ID Nº 10 a partir do nucleotídeo na posição 3109 até o nucleotídeo na posição 3269;

5 c) uma sequência de nucleotídeos que compreende a sequência de nucleotídeos que tem pelo menos 70%, preferivelmente pelo menos 80%, mais preferivelmente pelo menos 90%, particularmente pelo menos 95% de identidade de sequência ou que é, mais particularmente, idêntica a qualquer uma das sequências de nucleotídeos mencionadas sob a) ou b) ou

10 d) uma sequência de nucleotídeos que compreende a sequência de nucleotídeos de um fragmento de DNA com cerca de 150 bp a cerca de 1000 bp que hibridiza sob condições estridentes com um fragmento de DNA que tem a sequência de nucleotídeos mencionada sob a), b) ou c).

O promotor preferencial para célula de fibra pode compreender a sequência de nucleotídeos de SEQ ID Nº 9 da posição 465 até a posição 2307; ou a sequência de nucleotídeos de SEQ ID Nº 9 da posição 1374 até a posição 2307; a sequência de nucleotídeos de SEQ ID Nº 9 da posição 1531 até a posição 2307; ou a sequência de nucleotídeos de SEQ ID Nº 10 da posição 1397 até a posição 3269; ou a sequência de nucleotídeos de SEQ ID Nº 10 da posição 2371 até a posição 3269; ou a sequência de nucleotídeos de SEQ ID Nº 10 da posição 2718 até a posição 3269.

20 Em outra modalidade, a invenção provê uma região promotora preferencial para célula de fibra que compreende um promotor preferencial para célula de fibra como aqui descrito, compreendendo ainda a sequência de nucleotídeos de SEQ ID Nº 9 do nucleotídeo da posição 2308 ao nucleotídeo da posição 2409 ou o nucleotídeo da posição 3270 ao nucleotídeo da posição 3372.

Em outra modalidade da invenção, é provido um gene quimérico que compreende as seguintes regiões de DNA operativamente ligadas:

- 25 a) um promotor preferencial para célula de fibra como aqui descrito;
- 30 b) uma região de DNA heterólogo que codifica um RNA biologicamente ativo de interesse; e
- c) um sinal de término da transcrição e de poliadenilação.

O RNA biologicamente ativo pode codificar uma proteína de interesse, tal como N-acetilglicosamina transferase, preferivelmente uma N-acetilglicosamina transferase do tipo NodC, uma celulose sintase, uma sacarose sintase, preferivelmente uma sacarose sintase do tipo C; uma sacarose fosfato sintase ou uma β -1,3-endoglucanase. O RNA biologicamente ativo também pode ser um RNA inibitório ou RNA de silenciamento tal como uma ribozima, microRNA, RNA em forma de grampo de fita dupla, particularmente marcado para regular negativamente a expressão de um gene endógeno de algodão, tal como β -1,3-endoglucanase xilano sintase (cs1F), xiloglicano sintase (cs1D) ou 1,3-1,4 glicano sintase (cs1C).

Também são providas célula de planta e plantas, tais como plantas de algodão, e sementes dessa que compreendem tais genes quiméricos. Uma semente de uma planta que compreende em suas células um gene quimérico de acordo com qualquer uma das reivindicações 11 a 15.

A invenção provê ainda um método para expressar um RNA biologicamente ativo, preferencialmente em uma célula de fibra de uma planta que produz fibra, tal como uma planta de algodão, o método compreendendo as etapas de fornecimento de células de planta com um gene quimérico de acordo com a invenção e cultivo das plantas.

Em outra modalidade, a invenção é dirigida ao uso de um promotor específico para fibra e/ou preferencial para fibra de acordo com a invenção para expressão preferencial de um RNA biologicamente ativo em células de fibra de uma planta que produz fibra tal como uma planta de algodão.

A invenção ainda provê uma molécula de DNA isolado que compreende uma sequência de nucleotídeos que codificam uma proteína que compreende uma sequência de aminoácidos que tem pelo menos 95% de identidade de sequência, preferivelmente idêntica, a sequência de aminoácidos de SEQ ID Nº 11 ou SEQ ID Nº 12 e que codifica uma β -1,3-endoglucanase; uma molécula de DNA isolado que compreende a sequência de nucleotídeos selecionada do grupo de SEQ ID Nº 1; SEQ ID Nº 2, SEQ ID Nº 3 e SEQ ID Nº 4 ou uma molécula de DNA isolado que compreende uma

sequência de nucleotídeos de SEQ ID Nº 9 da posição 2410 a 2443 e a sequência de nucleotídeos de SEQ ID Nº 9 da posição 2556 a 3499 ou sequência de nucleotídeos de SEQ ID Nº 10 da posição 3373 a 3406 e a sequência de nucleotídeos de SEQ ID Nº 10 da posição 3501 a 4444. Essas moléculas de DNA isolado podem ser usadas para o isolamento de um promotor preferencial ou região promotora para célula de fibra.

A invenção também provê um método para isolar uma região promotora preferencial para célula de fibra, compreendendo as etapas de:

- identificar um fragmento genômico que codifica um transcrito de RNA a partir do qual um cDNA pode ser sintetizado, o cDNA compreendendo a sequência de nucleotídeos de SEQ ID Nº 1, SEQ ID Nº 2, SEQ ID Nº 3 ou SEQ ID Nº 4 ou seus equivalentes funcionais;

- isolar uma região de DNA a montante de uma sequência de nucleotídeos que codifica a proteína com as sequências de aminoácidos de SEQ ID Nº 11 ou SEQ ID Nº 12 ou seus equivalentes funcionais.

Também são providos promotores e regiões promotoras seletivas para fibra por esse método.

De acordo com uma segunda característica da invenção, é provido um método para alterar as propriedades de uma fibra em uma planta que produz fibra, aumentando particularmente o comprimento de uma fibra de uma planta de algodão, compreendendo a etapa de introduzir um gene quimérico em uma célula de planta de algodão, o gene quimérico compreendendo os seguintes elementos de DNA operativamente ligados:

(a) um promotor que pode ser expresso em planta, preferivelmente um promotor que pode ser expresso em planta que controle a transcrição nas células de fibra tais como o promotor beta-específico para tubulina de algodão, o promotor de actina específico para fibra de algodão, um promotor específico para fibra a partir de um gene de proteína de transferência de lipídio de algodão, um promotor de um gene de expansina de algodão ou um promotor de um gene de quitinase no algodão ou um promotor aqui descrito;

(b) uma região de DNA transcrito, que quando transcrito produz

uma molécula de RNA biologicamente ativo capaz de reduzir a expressão de um gene endógeno que codifica β -1,3-endoglucanase para a planta que produz fibra, a β -1,3-endoglucanase que está envolvida na remoção de calose dos plasmodesmos e

5 (c) uma região terminal 3' que compreende sinais de término da transcrição e de poliadenilação que funcionam em células de planta,

caracterizado em que apenas a expressão de um alelo subgenômico específico do gene que codifica β -1,3-endoglucanase endógena é regulado negativamente. Em uma modalidade preferida, o gene que codifica β -1,3-endoglucanase está compreendido dentro do subgenoma D da planta de algodão. O gene que codifica β -1,3-endoglucanase compreendida dentro do subgenoma D pode ser caracterizado pela presença de uma sequência de íntron que tem a sequência de nucleotídeos de SEQ ID N° 10 da posição 3407 a 3500 ou pode ser caracterizado pela presença de uma sequência de nucleotídeos AAGATC cerca de 326 nucleotídeos a jusante do códon de início da tradução (e que não inclui o códon de início da tradução). Convenientemente, o gene que codifica β -1,3-endoglucanase compreendido dentro do subgenoma D pode ser caracterizado pela identificação de um fragmento com cerca de 538 bp e um fragmento com 118 bp após amplificação por PCR com oligonucleotídeos que têm a sequência de nucleotídeos de SEQ ID N° 5 e SEQ ID N° 6 seguida pela digestão com a enzima de restrição A1wl. O gene que codifica β -1,3-endoglucanase também pode estar codificando uma proteína que compreende a sequência de aminoácidos de SEQ ID N° 12 ou o gene compreende a sequência de nucleotídeos de SEQ ID N° 2. A planta de algodão pode ser *Gossypium hirsutum*.

O RNA biologicamente ativo pode ser um RNA antissense que compreende pelo menos 19 nucleotídeos que têm pelo menos 94% de identidade de sequência com o complemento da sequência de nucleotídeos de SEQ ID N° 3 ou um RNA de sense que compreende pelo menos 19 nucleotídeos que têm pelo menos 94% de identidade de sequência com a sequência de nucleotídeos de SEQ ID N° 3 ou uma molécula de RNA de fita dupla que compreende pelo menos 19 nucleotídeos que têm pelo menos 94% de

identidade de sequência com o complemento da sequência de nucleotídeos de SEQ ID N° 3 e uma fita de RNA complementar essencialmente similar ao complemento de pelo menos 19 nucleotídeos. Preferivelmente, os 19 nucleotídeos mencionados compreendem pelo menos um nucleotídeo específico para o gene Ghgluc 1 do subgenoma D como indicado na Figura 2. Em outra modalidade ainda de acordo com a invenção, o RNA de fita dupla é um microRNA processado a partir de um pré-microRNA que compreende a sequência de nucleotídeos de SEQ ID N° 15, SEQ ID N° 16, SEQ ID N° 20 ou SEQ ID N° 21.

10 A invenção também provê um método para diminuir o comprimento de uma fibra de uma planta que produz fibra, compreendendo a etapa de introduzir um gene quimérico em uma célula da planta que produz fibra, cujo gene quimérico compreende os seguintes fragmentos de DNA operativamente ligados:

15 (a) um promotor que pode ser expresso em planta de acordo com as reivindicações 1b, 5, 6 ou 7;

(b) uma região de DNA que codifica uma proteína β -1,3 endoglucanase tal como uma proteína que tem uma sequência de aminoácidos que tem pelo menos 95% de identidade de sequência com a sequência de aminoácidos de SEQ ID N° 11 ou SEQ ID N° 12; e

20 (c) uma região terminal 3' compreendendo sinais de término da transcrição e de poliadenilação que funcionam em células da planta.

Ainda uma outra modalidade da invenção diz respeito a um gene quimérico como descrito aqui ou a uma célula de uma planta que produz fibra ou a uma planta que produz fibra que compreende tal gene quimérico.

25 Em outra modalidade ainda, a invenção provê fibras produzidas de acordo com os métodos aqui descritos.

Descrição resumida das figuras

Figura 1: alinhamento de regiões promotoras do gene de β -1,3 endoglucanase do subgenoma A e do subgenoma D de *Gossypium hirsutum*. A sequência de nucleotídeos do promotor do subgenoma A (sequência superior) corresponde a sequência de nucleotídeos de SEQ ID N° 9 da posi-

ção 1460 a 2408. A sequência de nucleotídeos do promotor do subgenoma D (sequência inferior) corresponde a sequência de nucleotídeos de SEQ ID N° 10 da posição 2372 a 3372. As diferenças na sequência de nucleotídeos são indicadas por caixas cinza. Nucleotídeos que não têm um nucleotídeo correspondente na outra região promotora são indicados por traços na sequência de nucleotídeos que não tem os nucleotídeos. A homologia total entre as duas regiões promotoras é de cerca de 71% de identidade de sequência.

Figura 2: alinhamento de regiões codificadoras do gene de β -1,3 endoglucanase (sem íntron) do subgenoma A e do subgenoma D de *Gossypium hirsutum*. A sequência de nucleotídeos da região codificadora do alelo do subgenoma A (sequência superior) corresponde a sequência de nucleotídeos de SEQ ID N° 9 da posição 2410 a posição 2443 e da posição 2556 a 3499. A sequência de nucleotídeos da região codificadora do alelo do subgenoma D (sequência inferior) corresponde a sequência de nucleotídeos de SEQ ID N° 10 da posição 3373 a posição 3406 e da posição 3501 a 4444. As diferenças na sequência de nucleotídeos são indicadas por caixas cinza. Nucleotídeos que não têm um nucleotídeo correspondente na outra região promotora são indicados por traços na sequência de nucleotídeos que não tem os nucleotídeos. A homologia total entre as duas regiões promotoras é de cerca de 97% de identidade de sequência.

Figura 3: alinhamento da sequência de íntron do gene de β -1,3 endoglucanase do subgenoma A e do subgenoma D de *Gossypium hirsutum*. A sequência superior corresponde a sequência de nucleotídeos do íntron do alelo do subgenoma A, enquanto que a sequência inferior corresponde a sequência de nucleotídeos do íntron do alelo do subgenoma D.

Figura 4: alinhamento da sequência de aminoácidos de β -1,3 endoglucanase codificada pelo alelo do subgenoma A (sequência superior que corresponde a SEQ ID N° 11) e pelo alelo do subgenoma D (sequência inferior que corresponde a SEQ ID N° 12).

Figura 5: alinhamento da sequência de nucleotídeos de SEQ IDs 1 a 4 da posição do nucleotídeo 201 a 211. O sítio de reconhecimento

da enzima de restrição *Alw1* presente apenas na sequência de nucleotídeos do alelo do subgenoma A do gene de β -1,3 endoglucanase está sublinhado e representado em itálico.

Figura 6: Correlação entre a curva de crescimento da fibra (painel A) e a expressão do gene de β -1,3 endoglucanase (*GhGluc 1*) (painel B). DNA de uma biblioteca de cDNA de fibras (em desenvolvimento) em *Gossypium hirsutum* foi extraído e equalizado. Os fragmentos de PCR foram amplificados usando os iniciadores de oligonucleotídeo SE002 e SE003 (SEQ ID Nº 5 e 6) e digeridos com *Alw1*. Um produto de PCR amplificado para a variante do genoma A rendeu 3 fragmentos (479bp+118bp+59bp), enquanto que para a variante do genoma D rendeu apenas 2 fragmentos (538bp+118bp).

Figura 7: Análise de expressão de *GhGluc 1* em folhas, raízes e caules de algodão por RT-PCR. O RNA extraído das folhas, raízes e caules de algodão foi submetido a análise RT-PCR para detectar a expressão de glicanase 1 ou da proteína fosfatase 2A (pp2A). A síntese da primeira fita de cDNA foi realizada usando SuperScript First-Strand Synthesis System de Invitrogen. Banda 1: marcador biológico de 100 bp; banda 2: amostra de RNA de folha de algodão + iniciadores específicos para Gluc 1 + transcriptase reversa; banda 3: amostra de RNA de folha de algodão + iniciadores específicos para Gluc 1 - transcriptase reversa; banda 4: amostra de RNA de raiz de algodão + iniciadores específicos para Gluc 1 + transcriptase reversa; banda 5: amostra de RNA de raiz de algodão + iniciadores específicos para Gluc 1 - transcriptase reversa; banda 6: amostra de RNA de caule de algodão + iniciadores específicos para Gluc 1 + transcriptase reversa; banda 7: amostra de RNA de caule de algodão + iniciadores específicos para Gluc 1 - transcriptase reversa; banda 8: DNA genômico (Fibermax 400 ng) + iniciadores específicos para Gluc 1; banda 9: marcador biológico de 100 bp; banda 10: amostra de RNA de folha de algodão + iniciadores específicos para pp2A + transcriptase reversa; banda 11: amostra de RNA de folha de algodão + iniciadores específicos para pp2A - transcriptase reversa; banda 12: amostra de RNA de raiz de algodão + iniciadores específicos para pp2A + transcriptase reversa; banda 13: amostra de RNA de raiz de algodão + iniciadores

- específicos para pp2A - transcriptase reversa; banda 14: amostra de RNA de caule de algodão + iniciadores específicos para pp2A + transcriptase reversa; banda 15: amostra de RNA de caule de algodão + iniciadores específicos para pp2A - transcriptase reversa; banda 16: marcador biológico de 100 bp; banda 17: RNA de controle positivo (fornecido com o kit).

Modalidades Detalhadas

A presente invenção é baseada no achado inesperado de que o momento da expressão dos alelos específicos dos subgenomas A e D do gene de β -1,3-endoglucanase específica para fibra (*GhGluc 1*) é diferente pelo menos em *Gossypium hirsutum*. Enquanto o início da expressão do alelo específico do subgenoma D correlaciona-se com o final da fase de alongamento rápido (cerca de 14 a 17 dias pós-antese, daqui por diante "DPA"), o início da expressão do subgenoma A é postergado até o início da fase de maturação tardia da fibra (cerca de 35 a 40 DPA) dependendo das condições de crescimento.

Consequentemente, em um aspecto, a presente invenção refere-se a identificação de novos promotores específicos para fibra que são expressos, preferencialmente, em células de fibra tanto no final da fase de alongamento rápido quanto no início da fase de maturação da fibra. Tais promotores têm uma utilidade, por exemplo, na expressão de genes quiméricos que expressam uma nova biomolécula nos estágios de desenvolvimento tardio da fibra, por exemplo, XET aumentada, Domínio de Ligação de Celulose (CBD), genes mutantes do Inchaço da Raiz (RSW), compostos/propriedades de coloração, plásticas, de digestibilidade de celulose, de parede celular nova. Os promotores também podem ser usados para dirigir a expressão de gene em fibras com 20, 30, 40 DPA; por exemplo, a produção de quitina nos estágios tardios do desenvolvimento da fibra de algodão (20, 30, 40 DPA) ou para expressão de gene em 40 DPA em fibras, por exemplo, produção de quitina em estágios muito tardios de desenvolvimento da fibra (40 DPA). Os promotores de acordo com a invenção também poderiam ser usados para dirigir a expressão de genes Cesa em 40 DPA para biossíntese intensificada de celulose e propriedades aperfeiçoadas da fibra ou para pro-

duzir RNA biologicamente ativo dirigido ao estágio de silenciamento específico de genes endógenos de algodão, tal como o silenciamento de β -1,3-endoglucanase para um crescimento prolongado da fibra. Os promotores poderão ser aplicados no estágio de expressão específica de genes quiméricos que codificam, por exemplo, a calose sintase, para alterar as propriedades e qualidade da fibra.

Em outro aspecto, a presente invenção está dirigida para o silenciamento da expressão de um alelo de subgenoma específico para o gene de β -1,3-endoglucanase específica para fibra, particularmente o silenciamento do alelo específico do subgenoma D, para prevenir a degradação da calose e assim aumentar a fase de alongamento rápido da fibra e o comprimento da fibra, por exemplo, no algodão, particularmente em *Gossypium hirsutum*.

De acordo com uma modalidade do primeiro aspecto da invenção, a invenção provê um fragmento de promotor específico para fibra e/ou preferencial para fibra que pode ser isolado do algodão, particularmente da espécie *Gossypium hirsutum*, e que está localizado a montante das sequências de ácido nucleico que codificam uma β -1,3-endoglucanase específica para fibra, que compreendem a sequência de aminoácidos de SEQ ID 11, 12, 7 ou 8 e suas variantes.

Como usado aqui, a expressão "promotor" denota qualquer DNA que é reconhecido e ligado (diretamente ou indiretamente) por uma RNA polimerase dependente de DNA durante o início da transcrição. Um promotor inclui um sítio de início da transcrição e sítios de ligação para fatores de início da transcrição e RNA polimerase e podem compreender vários outros sítios (por exemplo, intensificadores), nos quais proteínas reguladoras de expressão d gene podem ligar-se.

O termo "região reguladora", como usada aqui, significa qualquer DNA que está envolvido no direcionamento da transcrição e no controle (isto é, regulação) do momento e do nível de transcrição de uma dada sequência de DNA, tal como um DNA que codifica uma proteína ou polipeptídeo. Por exemplo, uma região reguladora 5' (ou "região promotora") é uma

sequência de DNA localizada a montante (isto é, 5') de uma sequência codificadora e que compreende o promotor e a sequência líder 5' não-traduzida. Uma região reguladora 3' é uma sequência de DNA localizada a jusante (isto é, 3') da sequência codificadora e que compreende sinais de formação(e/ou regulação) da transcrição da terminação 3', incluindo um ou mais sinais de poliadenilação.

O termo "gene" significa qualquer fragmento de DNA compreendendo uma região de DNA (a região de DNA transcrito) que é transcrito em uma molécula de RNA (por exemplo, um mRNA) em uma célula sob o controle de regiões reguladoras adequadas, por exemplo, uma região promotora que pode ser expressa em planta. Um gene pode, então, compreender vários fragmentos de DNA operativamente ligados tal como um promotor, uma sequência líder 5' não-traduzida, uma região codificadora e uma região 3' não-traduzida compreendendo um sítio de poliadenilação. Um gene endógeno de planta é um gene que é naturalmente encontrado em uma espécie de planta. Um gene quimérico é qualquer gene que não é normalmente encontrado em uma espécie de planta ou, alternativamente, qualquer gene no qual o promotor não está associado na natureza com parte ou com toda a região de DNA transcrito (uma região de DNA "heterólogo") ou com pelo menos uma outra região reguladora do gene.

O termo "expressão de gene" refere-se ao processo no qual uma região de DNA sob o controle de regiões reguladoras, particularmente, o promotor, é transcrita em um RNA que é biologicamente ativo, isto é, que é capaz de interagir com outra molécula de ácido nucleico ou que é capaz de ser traduzido em um polipeptídeo ou proteína biologicamente ativos. Um gene é dito codificar um RNA quando o produto final da expressão do gene é um RNA biologicamente ativo, tal como um RNA antissenso ou uma ribozima. Um gene é dito codificar uma proteína quando o produto final da expressão do gene é uma proteína ou polipeptídeo funcionalmente ou biologicamente ativos.

O termo "seletivo de fibra" ou "seletivo de célula de fibra" ou "específico de célula de fibra", com relação a expressão de um DNA de a-

cordo com essa invenção, refere-se, para propósitos práticos, a expressão altamente específica de um DNA em células de fibra de planta, tais como plantas de algodão ("seletiva para célula de fibra"). Em outras palavras, os níveis de transcrito de um DNA em tecidos diferentes de células de fibra estão abaixo da detecção ou são muito baixos (menos do que cerca de 0,2 picogramas por micrograma de RNA total).

O termo "preferencial para fibra" ou "preferencial para célula de fibra" com relação a expressão de um DNA de acordo com essa invenção, refere-se a um padrão de expressão através do qual o DNA é expresso predominantemente em células de fibra ou fibras, mas a expressão pode ser identificada em outros tecidos da planta. Preferivelmente, a expressão em células de fibra é cerca de 2 a cerca de 10 vezes maior nas células de fibra do que em outros tecidos.

As sequências de nucleotídeos de SEQ ID N° 9 e SEQ ID N° 10 representam a sequência de nucleotídeos de moléculas de DNA genômico que codificam, respectivamente, uma β -1,3-endoglucanase seletiva para fibra do subgenoma A e do subgenoma D. O sítio de início da transcrição de SEQ ID N° 9 foi determinado na posição 2308, enquanto que o sítio de início da transcrição de SEQ ID N° 10 foi determinado na posição 3270. Assim em uma modalidade da invenção, é provido um promotor seletivo para fibra que tem uma sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo da posição 1 até o nucleotídeo da posição de 2307 ou 2308; em outra modalidade da invenção, é provido um promotor seletivo para fibra que tem uma sequência de nucleotídeos de SEQ ID N° 10 do nucleotídeo da posição 1 até o nucleotídeo da posição de 3269 ou 3270.

O sítio de início da tradução da sequência de nucleotídeos de SEQ ID N° 9 foi determinado como o códon ATG localizado nas posições 2410 a 2412. A região 5' não-traduzida do promotor seletivo para fibra do alelo subgenômico A que codifica β -1,3-endoglucanase tem, consequentemente, uma sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo da posição 2308 ao nucleotídeo da posição 2409. Uma região promotora seletiva para fibra do alelo subgenômico A é, consequentemente, uma região

promotora que compreende os nucleotídeos 1 a 2409 de SEQ ID N° 9.

5 Similarmente, o sítio de início da tradução da sequência de nucleotídeos de SEQ ID N° 10 foi determinado como o códon ATG localizado nas posições 3373 a 3375. A região 5' não-traduzida do promotor seletivo para fibra do alelo subgenômico D que codifica β -1,3-endoglucanase tem, consequentemente, uma sequência de nucleotídeos de SEQ ID N° 10 do nucleotídeo da posição 3270 ao nucleotídeo da posição 3372. Uma região promotora seletiva para fibra do alelo subgenômico D é, consequentemente, uma região promotora que compreende os nucleotídeos 1 a 3372 de SEQ ID
10 N° 10.

O alinhamento das regiões promotoras seletivas para fibra de SEQ ID N° 9 e SEQ ID N° 10 (Figura 1) revelou uma homologia de sequência geral nas sequências de nucleotídeos cerca de 1000 nucleotídeos a montante do códon de início da tradução ATG, particularmente na região a
15 cerca de 150 nucleotídeos a montante do códon de início da tradução ATG. Assim, em outra modalidade da invenção, é provido um promotor seletivo para fibra que compreende a sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo da posição 2149 ao nucleotídeo da posição 2307 ou que compreende a sequência de nucleotídeos de SEQ ID N° 10 do nucleotídeo da
20 posição 3109 ao nucleotídeo da posição 3269.

É óbvio que o promotor ou a região promotora podem estar contidos dentro de sequências de nucleotídeos maiores. Uma região promotora seletiva para célula de fibra pode assim ser determinada como a região a montante (isto é, localizada a 5' de) do códon codificador para o primeiro
25 aminoácido da proteína codificada pelo mRNA indicado na SEQ ID N° 9 ou 10. Tal região promotora pode ter pelo menos cerca de 400 a 500 bp, pelo menos cerca de 1000 bp, pelo menos 1200 bp, pelo menos cerca de 1300 bp ou pelo menos cerca de 1500 a 2000 bp a montante do códon de partida. Por conveniência, é preferido que tal região promotora não se prolongue por
30 mais do que cerca de 3000 a 5000 bp a montante do códon de partida. O tamanho do fragmento pode ser determinado parcialmente pela presença dos sítios de restrição convenientes. Por exemplo, para a região promotora

seletiva para fibra do alelo do subgenoma A que codifica β -1,3-endoglucanase, um fragmento com cerca de 1945 bp (do nucleotídeo na posição 465 ao nucleotídeo na posição 2409 de SEQ ID Nº 9) pode ser convenientemente clonado. A presença de enzimas de restrição adequadas

5 permite gerar convenientemente regiões promotoras com cerca de 1036 bp (do nucleotídeo na posição 1374 ao nucleotídeo na posição 2409 de SEQ ID Nº 9) ou com cerca de 879 np (do nucleotídeo na posição 1531 ao nucleotídeo na posição 2409 de SEQ ID Nº 9). Para a região promotora seletiva para fibra do alelo do subgenoma D que codifica β -1,3-endoglucanase, um

10 fragmento com cerca de 1976 bp (do nucleotídeo na posição 1397 ao nucleotídeo na posição 3372 de SEQ ID Nº 10) pode ser convenientemente clonado. A presença de enzimas de restrição adequadas permite gerar convenientemente regiões promotoras com cerca de 1002 bp (do nucleotídeo na posição 1397 ao nucleotídeo na posição 3372 de SEQ ID Nº 10) ou com cerca

15 de 655 np (do nucleotídeo na posição 2371 ao nucleotídeo na posição 3372 de SEQ ID Nº 10).

Será esclarecido mais tarde que promotores ou regiões promotoras seletivas para fibra equivalentes podem ser isoladas de outras plantas de algodão ou de plantas de algodão parentais. Para essa finalidade, frag-

20 mentos de promotor podem ser isolados de outras plantas de algodão, tal com *G. barbadense*, incluindo as variedades assim chamadas de PIMA, ou de outras variedades que usam um fragmento de promotor como descrito aqui como uma sonda e que identificam sequências de nucleotídeos dessas outras plantas que hibridizam sob condições estringentes de hibridização.

25 "Condições estringentes de hibridização" como usado aqui significa que a hibridização geralmente ocorrerá se houver pelo menos 95% e, preferivelmente, pelo menos 97% de identidade de sequência entre a sonda e a sequência-alvo. Exemplos de condições estringentes de hibridização são a incubação durante uma noite em uma solução que compreende formamida

30 50%, 5xSSD (NaCl 150 mM, citrato trissódico 15 mM), fosfato de sódio 50 mM (pH 7.6), 5 x solução de Denhardt, sulfato de dextran 10% e 20 μ g/ml de DNA transportador desnaturado, tal como DNA de espermatozoides de salmão, se-

guido pela lavagem do suporte de hibridização em 0,1 x SSC em aproximadamente 65°C. Outras condições de hibridização e lavagem são bem conhecidas e estão exemplificadas em Sambrook et al, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, NY (1989), particularmente no capítulo 11.

Tais promotores seletivos para fibra equivalentes isolados de outras plantas de algodão ou de plantas de algodão parentais podem representar promotores variantes nos quais um ou mais nucleotídeos foram substituídos por substituição, deleção ou inserção. Promotor variante seletivo para fibra de acordo com a invenção também pode ser gerado sinteticamente. A Figura 1 representa o alinhamento de promotores seletivos para fibra do subgenoma A e D de *G. hirsutum* e posições variantes, onde a sequência de nucleotídeos pode ser aparentemente alterada sem um efeito significativo sobre a capacidade de início da transcrição eletiva para fibra, foram indicadas pelas caixas em cinza. A partir da Figura 1 e de sua legenda, é claro que a identidade de sequência global entre os dois fragmentos de promotor seletivo para fibra exemplificados pode ser tão baixa quanto 71% de identidade de sequência.

Consequentemente, em outra modalidade da presente invenção, são providos o promotor e as regiões promotoras seletivas para fibra que compreendem uma sequência de nucleotídeos que tem pelo menos 70%, pelo menos 80%, pelo menos 90% ou pelo menos 95% de identidade de sequência com os promotores e regiões promotoras seletivas para fibra aqui descritas.

Para o propósito dessa invenção, a "identidade de sequência" de duas sequências de nucleotídeos ou aminoácidos relacionadas, expressa como uma porcentagem, refere-se ao número de posições em duas sequências otimamente alinhadas que têm resíduos idênticos (x100) divididas pelo número de posições comparadas. Um espaço, isto é, uma posição no alinhamento onde um resíduo está presente em uma sequência mas não na outra, é considerado como uma posição com resíduos não idênticos. O alinhamento de duas sequências é realizado pelo algoritmo de Needleman e

Wunsch (Needleman e Wunsch 1970). O alinhamento de sequência com ajuda de computador acima, pode ser convenientemente realizado usando um programa padronizado tal como ESPAÇO que é parte de Wisconsin Package Versão 10.1 (Genetics Computer Group, Madison, Wisconsin, U-
5 SA) usando a matriz de escore de default com uma penalidade de criação de espaço de 50 e uma penalidade de extensão de espaço de 3.

Será esclarecido que sempre que sequências de nucleotídeos de moléculas de RNA são definidas por referência a uma sequência de nucleotídeos de moléculas de DNA correspondentes, a timina (T) na sequência
10 de nucleotídeos deverá ser substituída por uracil (U). Será esclarecido a partir do contexto do pedido se a referência é feita a moléculas de RNA ou DNA.

A capacidade de expressão seletiva para fibra do promotor ou região promotora identificados ou gerados pode ser testada, convenientemente, pela ligação operativa de tais moléculas de DNA com uma região
15 codificadora que codifique um marcador facilmente contável, por exemplo, um gene de β -glicuronidase, introduzindo tal gene quimérico em uma planta que produz fibra e analisando o padrão de expressão do marcador em células de fibra (preferivelmente durante o desenvolvimento da fibra) em comparação com o padrão de expressão do marcador em outras partes das plantas.
20

É óbvio que os promotores e regiões promotoras da invenção podem também compreender elementos adicionais conhecidos por melhorar a eficiência da transcrição tais como intensificadores, íntrons, etc.

25 Além disso, as sequências 5'UTR exemplificadas das regiões promotoras seletivas para fibra exemplificadas são particularmente similares na sequência de nucleotídeos e espera-se que as sequências 5'UTR sejam intercambiáveis.

A invenção inclui ainda moléculas de DNA que compreendem os
30 promotores ou regiões promotoras seletivas para fibra da invenção operativamente ligadas a uma ou mais regiões heterólogas que codificam um RNA biologicamente ativo, peptídeo ou proteína. Os promotores da invenção po-

dem ser usados para expressar qualquer região codificadora heteróloga desejada.

Exemplos de moléculas de DNA que codificam outras proteínas que podem ser expressas usando o promotor da presente invenção incluem, mas não são limitados a celulosas sintases (genes *CesA*) homólogas e heterólogas, ambas na forma normal e mutada (Arioli *et al.*, "Molecular Analysis of Cellulose Biosynthesis in Arabidopsis," *Science*, 279: 717-720 (1998); Holland *et al.*, "A Comparative Analysis of the Plant Cellulose Synthase (*CesA*) Gene Family," *Plant Physiol.*, 123: 1313-1324 (2000)); genes que podem modular o fracionamento de carbono para celulose (Delmer, "Cellulose Biosynthesis in Developing Cotton Fibers" in: A. S. Basra (ed.), *Cotton Fibers: Developmental Biology Quality Improvement, and Textile Processing*, The Haworth Press, New York, pp. 85-112 (1999)) tal como sacarose sintase (Amor *et al.*, "A Membrane-Associated Form of Sucrose Synthase and Its Potential Role Synthesis of Cellulose and Callose in Plants," *Proc. Natl. Acad. Sci. USA*, 92: 9353-9357 (1995), sacarose fosfato sintase (Haigler *et al.*, "Transgenic Cotton Over-Expressing Sucrose Phosphate Synthase Produces Higher Quality Fibers with Increased Cellulose Content and Has Enhanced Seeded Cotton Yield" Abstract 477. In: *Proceedings of Plant Biology 2000*, July 15-19, San Diego, CA. American Society of Plant Physiologists, Rockville, MD, (2000), UDPG-pirofosforilase (Waffler and Meier, "Enzyme Activities in Developing Cotton Fibers," *Plant Physiol. Biochem.* 32: 697-702 (1994), pirofosfatase inorgânica (Geigenberger *et al.*, "Overexpression of Pyrophosphatase Leads to Increased Sucrose Degradation and Starch Synthesis, Increased Activities of Enzymes for Sucrose-Starch Interconversions, and Increased Levels of Nucleotides in Growing Potato Tubers," *Planta*, 205:428-437 (1998)), hexoquinases (Smeekens, "Sugar Regulation of Gene Expression" *Curr. Op. Plant Biol.*, 1: 230-234 (1998), e invertases (Sturm and Tang, "The Sucrose-Cleaving Enzymes of Plants are Crucial for Development, Growth, and Carbon Partitioning," *Trends Plant Sci.*, 4: 401-407 (1999)); genes que podem afetar as propriedades moleculares e biofísicas da celulose incluindo o grau de polimerização, grau de cristalinidade, tamanho da cristalita e orien-

tação de microfibrila (isto é, genes para codificar proteínas, incluindo polímeros de proteínas cocristalizada ou domínios de ligação de celulose e a síntese de polissacarídeos e outras enzimas) (Delmer, "Cellulose Biosynthesis: Exciting Times for a Difficult Field of Study, "Ann. Rev. Plant Physio. Mol. Biol. 50: 245- 276 (1999); Delmer, "Cellulose Biosynthesis in Developing Cotton Fibers. In: A. S. Basra(ed.), Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing, The Haworth Press, New York, pp.85-112 (1999); Hsieh, "Structural Development of Cotton Fibers. In: A. S. Basra(ed.), Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 137-166 (1999)); fatores de transcrição tais como genes MYB que podem prolongar o período de alongamento e/ou alterar o momento ou a extensão da deposição da parede secundária (Wilkins and Jernstedt,"Molecular Genetics of Developing Cotton Fibers. In: A. S. Basra(ed.), Cotton Fibers: Developmental Biology. Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 231-270 (1999)); genes para afetar a síntese de hormônios de planta e alterar as propriedades da fibra (John, "Genetic Engineering Strategies for Cotton Fiber Modification. In: A. S. Basra(ed.), Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 271-289 (1999)); genes para elementos citoesqueléticos ou proteínas associadas ao citoesqueleto que podem afetar as propriedades da fibra (Seagull, "Cytoskeletal Involvement in Cotton Fiber Growth and Development," Micron, 24: 643-660 (1993)); genes para sintetizar lipídeo ou modificar enzimas que podem alterar as propriedades da membrana e melhorar dessa maneira a qualidade da fibra, inclusive sob condições ambientais estressantes (Haigler,"The Crystallinity of Cotton Cellulose in Relation to Cotton Improvement, "Proc. Cotton Fiber Cellulose: Structure. Function and Utilization Conference, National Cotton Council of America: Memphis, TN., p. 211-225 (1992)); enzimas tais como xiloglicano endotransferase, peroxidase, expansina, ou ATPase vacuolar que podem, através da atividade aumentada ou diminuída, prolongar ou aumentar o crescimento da extensão durante a deposição da parede secundária (Wilkins and Jernstedt, "Molecular Genetics

- of Developing Cotton Fibers. In: A. S. Basra(ed.), Cotton Fibers: Developmental Biology Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 231-270 (1999); genes para proteína ou polímeros plásticos que podem estar retidos na luz da fibra ou integrados na parede celular para aumentar a resistência da fibra ou alterar suas propriedades têxteis (John, "Genetic Engineering Strategies for Cotton Fiber Modification, "In: A. S. Basra(ed.), Cotton Fibers: Developmental Biology. Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 271-289 (1999); Guda *et al.*, "Hyperexpression of an Environmentally Friendly Synthetic PolymerGene, "Biotechnology Letters, 17: 745-750 (1995)); genes para enzimas biossintéticas da matriz da parede celular de planta ou suas proteínas reguladoras tal que outros carboidratos possam ser integrados na parede celular e alterar as propriedades da fibra (Haigler, "The Relationship Between Polymerization and Crystallization in CelluloseBiogenesis," in C. H. Haigler and P. Weimer, eds., Biosynthesis and Biodegradation of Cellulose, New York: Marcel Dekker, pp. 99-124 (1991); Andrawis *et al.*, "Cotton Fiber Annexins: A Potential Role in the Regulation of Callose Synthase, "Plant J., 3: 763-772 (1993); genes para moléculas tais como taninas, suberinas ou pigmentos que podem conferir cor valiosa as fibras (Ryser, "Cotton Fiber Initiation and Histodifferentiation, "In: A. S. Basra(ed.), Cotton Fibers:Developmental Biology. Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 1-46 (1999); genes para moléculas tais como cutina, suberina ou cera que podem alterar a absorvência e resistência de fibras de algodão (May, "Genetic Variation in Fiber Quality, "In: A. S. Basra (ed.), Cotton Fibers: Developmental Biology, Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 183-230 (1999); Ryser, "Cotton Fiber Initiation and Histodifferentiation, "In: A. S. Basra(ed.), Cotton Fibers: Developmental Biology. Quality Improvement, and Textile Processing, The Haworth Press, New York, pp. 1--46 (1999); e genes para moléculas de sinal de transdução tal como Rac que pode regular mudanças entre estágios de desenvolvimento da fibra (Delmer *et al.*, "Genes Encoding Small GTP- Binding Proteins Analogous to Mammalian rac are Preferentially Ex-

pressed in Developing Cotton Fibers, "Mol. Gen. Genet., 248: 43-51 (1995).

Regiões codificadoras de proteína particularmente preferidas são as regiões codificadoras de N-acetilglicosamina sintase como descritas, por exemplo, em WO2006/136351 ou genes de sacarose sintase como descritos em WO2002/45485 ou EP06015433.3 (incorporadas aqui por referência).

RNA biologicamente ativo também pode codificar o assim chamado RNA antissenso, RNA sense, RNA de fita dupla como descrito em WO99/53053 ou moléculas sintéticas de microRNA determinadas, de acordo com regras bem conhecidas na técnica, para regular negativamente a expressão de outros genes compreendidos dentro da célula das plantas que produzem fibra ou até de genes compreendidos dentro de um patógeno ou praga que se alimenta da planta que produz fibra.

Também são providos métodos para expressar uma proteína ou RNA biologicamente ativo especificamente em células de fibra de uma planta que produz fibra, compreendendo a etapa de introduzir uma molécula de DNA que compreende um promotor ou região promotora seletivos para fibra como aqui descrito, operativamente ligados a uma região de DNA transcrito que codifica a molécula de RNA biologicamente ativa nas células da fibra de plantas que produzem fibra.

Em um segundo aspecto da presente invenção, é provido um método para alterar o comprimento da fibra em uma planta que produz fibra, pela alteração especificamente da expressão de um dos alelos subgenômicos que codificam uma β -1,3-endoglucanase seletiva para fibra. Como indicado acima, tal alteração da expressão pode prevenir ou intensificar a degradação da calose que obstrui os plasmodesmos na base da célula de fibra e assim aumentar ou diminuir a fase de alongamento rápido da fibra e o comprimento da fibra, por exemplo, no algodão, particularmente em *Gossypium hirsutum*. A modulação do comprimento da fibra também pode impactar a resistência e outras qualidades ou propriedades da fibra. Consequentemente, a invenção também está direcionada para a alteração de propriedades da fibra ou qualidades da fibra pelos métodos aqui descritos.

Em uma modalidade do segundo aspecto da invenção, é provido um método para aumentar o comprimento de uma fibra em uma planta de algodão, particularmente uma planta *Gossypium hirsutum*, pela introdução de um gene quimérico nas células da dita planta de algodão, o gene quimérico compreendendo os seguintes elementos operativamente ligados ao DNA:

- a. um promotor que pode ser expresso em planta
- b. uma região de DNA transcrito que quando transcrita rende uma molécula de RNA biologicamente ativo capaz de reduzir a expressão de uma β -1,3-endoglucanase seletiva para fibra compreendida dentro do genoma D e
- c. uma região 3' terminal compreendendo sinais de término da transcrição e de poliadenilação que funcionam em células da dita planta.

Como foi descoberto pelos presentes inventores, a expressão aumentada de β -1,3-endoglucanase nas células da fibra de uma planta de algodão previamente descrita, está correlacionada com a reabertura dos plasmodesmos e marca o final da fase de alongamento rápido da fibra e é devida, especificamente, ao início da expressão do alelo específico do subgenoma D nesse momento, enquanto que o alelo específico do subgenoma A é expresso apenas nos estágios tardios do desenvolvimento da fibra, pelo menos no *Gossypium hirsutum*. Consequentemente, os métodos como descritos em WO2005/017157 podem ser ainda aperfeiçoados, pelo menos para *Gossypium hirsutum*, pela alteração específica do padrão de expressão de um alelo subgenômico particular que codifica a β -1,3-endoglucanase seletiva para fibra.

Como descrito aqui, um alelo específico do subgenoma D que codifica uma β -1,3-endoglucanase seletiva para fibra pode ser reconhecido de várias maneiras. O isolamento do ácido nucleico (genômico) que codifica uma β -1,3-endoglucanase seletiva para fibra do subgenoma D de diferentes plantas de algodão, assim como de plantas de algodão diploides parentais com um genoma similar a D, revelou que o íntron no alelo subgenômico A tem uma inserção de cerca de 18 nt de comprimento na sequência de íntron

(vide figura 3) que está ausente na sequência de íntron do alelo subgenômico D.

Além disso, a comparação das sequências de nucleotídeos da região codificadora (cDNA; figura 2) sequências de aminoácidos codificadas (figura 4) revelou várias diferenças características entre os alelos subgenômicos A e D e proteínas codificadas, indicadas nas figuras pelas caixas em cinza.

Muito convenientemente, os alelos subgenômicos A e D podem ser distinguidos por um polimorfismo que resulta da presença de um sítio de reconhecimento de enzima de restrição A1wl, cerca de 326 nucleotídeos a jusante do códon de início da tradução no clone genômico do alelo subgenômico A, que está ausente nos alelos subgenômicos D que codificam a β -1,3-endoglucanase seletiva para fibra. Um método possível para distinguir alelos subgenômicos A e D, ambos à nível de DNA genômico ou cDNA, é a amplificação por PCR usando cDNA ou DNA genômico como modelo e os oligonucleotídeos que têm a sequência de nucleotídeos de SEQ ID N° 5 e 6 como iniciadores, seguido pela digestão do produto de PCR amplificado com a enzima de restrição A1wl. O produto de PCR gerado, que tem cerca de 656 bp para ambos os alelos subgenômicos, é clivado em um fragmento de 538 e 118 bp quando o produto de PCR é gerado a partir de um modelo de alelo subgenômico D, enquanto que o produto de PCR gerado amplificado a partir do modelo de alelo subgenômico A é clivado em três fragmentos de 476 bp, 118 bp e 59 bp, respectivamente. A presença de 118 bp serve como um controle interno para o funcionamento apropriado da reação da enzima de restrição.

Vários métodos estão disponíveis na técnica para produzir uma molécula de RNA de silenciamento, isto é, uma molécula de RNA que quando expressa reduz a expressão de um gene particular ou grupo de genes, incluindo as assim chamadas tecnologias de RNA "sense" ou "antissenso".

Assim, em uma modalidade, um gene quimérico que codifica a molécula de RNA inibitório é baseado na assim chamada tecnologia antisenso. Em outras palavras, a região codificadora do gene quimérico com-

preende uma sequência de nucleotídeos de pelo menos 19 ou 20 nucleotídeos consecutivos do complemento da sequência de nucleotídeos da β -1,3-endoglucanase seletiva para fibra. Tal gene quimérico pode ser construído pela ligação operativa de um fragmento de DNA que compreende pelo menos 19 ou 20 nucleotídeos do gene que codifica a β -1,3-endoglucanase seletiva para fibra, isolada ou identificada como descrito nesse pedido em outro lugar, em orientação inversa para um promotor que pode ser expresso em planta e para uma região de formação da terminação 3' envolvida no término da transcrição e na poliadenilação.

10 Em outra modalidade, o gene quimérico que codifica uma molécula de RNA inibitório está baseado na assim chamada tecnologia de cossupressão. Em outras palavras, a região codificadora do gene quimérico compreende uma sequência de nucleotídeos de pelo menos 19 ou 20 nucleotídeos consecutivos da sequência de nucleotídeos do gene da β -1,3-endoglucanase seletiva para fibra. Tal gene quimérico pode ser construído pela ligação operativa de um fragmento de DNA que compreende pelo menos 19 ou 20 nucleotídeos do gene de β -1,3-endoglucanase seletiva para fibra, em orientação direta para um promotor que pode ser expresso em planta e para uma região de formação da terminação 3' envolvida no término da transcrição e na poliadenilação.

20 A eficiência dos genes quiméricos acima mencionados em reduzir a expressão do gene de β -1,3-endoglucanase seletiva para fibra pode ser ainda intensificada pela inclusão de elemento de DNA que resulta em expressão de moléculas de RNA inibitório aberrante, não-poliadenilado ou resulta na retenção das moléculas de RNA inibitório no núcleo das células. Um de tais elementos de DNA adequados para o propósito é uma região de DNA que codifica uma ribozima autocombinada, como descrito em WO 00/01133 (incorporado por referência). Outro de tal elemento de DNA adequado para esse propósito é uma região de DNA que codifica um sinal de localização nuclear ou de retenção de RNA, como descrito em PCT/AU03/00292 publicado como WO03/076619 (incorporado por referência).

30 Um meio conveniente e muito eficiente de regular negativamen-

te a expressão de um gene de interesse usa o assim chamado RNA de fita dupla (dsRNA) ou RNA de interferência (RNAi), como descrito, por exemplo, em WO99/53050 (incorporado por referência). Nessa tecnologia, uma molécula de RNA é introduzida em uma célula de planta e dessa maneira a molécula de RNA é capaz de formar uma região de RNA de fita dupla sobre pelo menos cerca de 19 a cerca de 21 nucleotídeos, e dessa maneira uma das fitas dessa região de RNA de fita dupla é quase idêntica em sequência de nucleotídeos ao gene-alvo ("região sense"), enquanto que a outra fita é quase idêntica em sequência de nucleotídeos ao complemento do gene-alvo ou à região sense ("região antissenso"). É esperado que para o silenciamento da expressão do gene-alvo, a sequência de nucleotídeos das sequências de 19 nucleotídeos consecutivos pode ter um não-pareamento ou a região sense e a região antissenso podem diferir em um nucleotídeo. Para se obter a construção de tais moléculas de RNA ou dos genes quiméricos codificadores, pode ser feito o uso do vetor como descrito em WO 02/059294.

Assim, em uma modalidade da invenção, é provido um método para aumentar o comprimento de uma fibra de uma planta que produz fibra, tal como algodão, compreendendo a etapa de introduzir um gene quimérico em uma célula da planta que produz fibra, cujo gene quimérico compreende os seguintes elementos de DNA operativamente ligados:

(a) um promotor que pode ser expresso em planta, preferivelmente um promotor que pode ser expresso em planta que controle a transcrição preferencialmente nas células de fibra;

(b) uma região de DNA transcrito, que quando transcrita rende uma molécula de RNA de fita dupla capaz de reduzir especificamente a expressão de um alelo subgenômico de β -1,3-endoglucanase seletiva para fibra e a molécula de RNA compreendendo uma primeira e uma segunda região de RNA em que

i) a primeira região de RNA compreende uma sequência de nucleotídeos de pelo menos 19 nucleotídeos consecutivos que tem pelo menos cerca de 94% de identidade de sequência com a sequência de nucleotídeos do alelo subgenômico;

ii) a segunda região de RNA compreende uma sequência de nucleotídeos complementar aos pelo menos 19 nucleotídeos consecutivos da primeira região de RNA;

5 iii) a primeira e a segunda região de RNA são capazes de parear para formar uma molécula de RNA de fita dupla entre pelo menos 19 nucleotídeos consecutivos da primeira e segunda regiões; e

c) a região 3' terminal compreendendo sinais de término da transcrição e poliadenilação que funcionam em células de planta.

O comprimento da primeira e segunda regiões de RNA (região
10 sense ou antissense) pode variar entre cerca de 19 nucleotídeos (nt) até um comprimento correspondente ao comprimento (em nucleotídeos) do gene endógeno envolvido na remoção da calose. O comprimento total da sequência de nucleotídeos sense e antissense pode ser, então, pelo menos de 25 nt, ou pelo menos de cerca de 50 nt, ou pelo menos cerca de 100 nt, ou pelo
15 menos de cerca de 150 nt, ou pelo menos de cerca de 200 nt, ou pelo menos de cerca de 500 nt. É esperado que não haja limite superior no comprimento total da sequência de nucleotídeos sense ou antissense. Entretanto, por razões práticas (tais como, por exemplo, estabilidade de genes quiméricos) é esperado que o comprimento da sequência de nucleotídeos sense ou
20 antissense não deve exceder 5000 nt, particularmente não deve exceder 2000 nt e pode estar limitado a cerca de 1000 nt.

Será avaliado que quanto mais longo o comprimento total da região sense ou antissense, menos estridentes os requisitos para identidade de sequência entre essas regiões e a sequência correspondente no gene de
25 β -1,3-endoglucanase seletiva para fibra ou seu complemento. Preferivelmente, o ácido nucleico de interesse deve ter uma identidade de sequência de pelo menos cerca de 75% com a sequência-alvo correspondente, particularmente pelo menos cerca de 80%, mais particularmente pelo menos cerca de 85%, muito particularmente cerca de 90%, especialmente cerca de 95%,
30 mais especialmente cerca de 100%, muito especialmente ser idêntica a parte correspondente da sequência-alvo ou seu complemento. Entretanto, é preferido que o ácido nucleico de interesse sempre inclua uma sequência

com cerca de 19 nucleotídeos consecutivos, particularmente cerca de 25 nt, mais particularmente cerca de 50 nt, especialmente cerca de 100 nt, muito especialmente cerca de 150 nt com 100% de identidade de sequência com a parte correspondente do ácido nucleico alvo. Preferivelmente, para calcular a identidade de sequência e determinar a sequência sense e antissenso correspondentes, o número de espaços deve ser minimizado, particularmente para sequências sense mais curtas.

Genes quiméricos codificando dsRNA de acordo com a invenção podem compreender um íntron, tal como um íntron heterólogo, localizado, por exemplo, na sequência espaçadora entre as regiões sense e antissenso de RNA de acordo com a descrição de WO 99/53050 (incorporado aqui por referência).

É preferido para a presente invenção que a sequência do gene alvo específico incluída na molécula de RNA antissenso, sense ou de fita dupla compreenda pelo menos um nucleotídeo e preferivelmente mais de um nucleotídeo, que são específicos para o alelo subgenômico cuja expressão é para ser regulada negativamente. Tais nucleotídeos específicos estão indicados pelo menos na figura 2 por caixas em cinza.

Em uma modalidade preferida, o RNA biologicamente ativo está adaptado especialmente para regular negativamente o alelo subgenômico D do gene que codifica a β -1,3-endoglucanase seletiva para fibra (GhGluc1). Em outra modalidade preferida, o RNA biologicamente ativo está especialmente adaptado para regular negativamente o alelo subgenômico A do gene que codifica a β -1,3-endoglucanase seletiva para fibra (GhGluc 1).

O uso de micro-RNAs sintéticos para regular negativamente a expressão de um gene particular em uma célula de planta, possibilita especificidade de sequência do gene-alvo muito grande e assim permite discriminar convenientemente entre alelos intimamente relacionados como genes-alvo qual expressão deve ser regulada negativamente.

Assim, em outra modalidade da invenção, o RNA biologicamente ativo ou RNA de silenciamento ou molécula de RNA inibitório pode ser uma molécula de microRNA, determinada, sintetizada e/ou modulada para

atingir e causar a clivagem de alelos específicos de subgenoma, preferivelmente o alelo subgenômico D do gene de β -1,3-endoglucanase seletiva para fibra em plantas que produzem fibra, tal como uma planta de algodão. Vários métodos têm sido descritos para gerar e usar miRNAs para um gene-alvo

5 específico (incluindo mas não limitado a Schwab *et al.* (2006, *Plant Cell*, 18(5):1121-1133), WO2006/44322, WO 2005/047505, EP 06009836, incorporadas por referência). Geralmente, um arcabouço de miRNA existente é modificado na porção de reconhecimento do gene-alvo tal que o miRNA gerado guia agora o complexo RISC para clivar as moléculas de RNA transcri-

10 tas a partir do ácido nucleico alvo. Arcabouços de miRNA podem ser modificados ou sintetizados tal que o miRNA compreenda agora 21 nucleotídeos consecutivos de um dos alelos subgenômicos da sequência de nucleotídeos que codifica β -1,3-endoglucanase seletiva para fibra, tal como as sequências representadas na Listagem de Sequências e permitindo não-pareamentos

15 de acordo com as regras aqui descritas abaixo.

Assim, em uma modalidade, a invenção provê um método para regular negativamente a expressão de ou aumentar a resistência de plantas a condições de crescimento adversas, compreendendo as etapas de:

- a. Introduzir um gene quimérico nas células da dita planta, o dito
- 20 gene quimérico compreendendo as seguintes regiões de DNA operativamente ligadas:
- i. Um promotor que pode ser expresso em planta;
 - ii. Uma região de DNA que após a introdução e transcrição em uma célula de planta é processada em um miRNA, de acordo com o que o
- 25 miRNA é capaz de reconhecer e conduzir a clivagem do mRNA de um alelo subgenômico de um gene que codifica β -1,3-endoglucanase seletiva para fibra da planta mas não de outro alelo subgenômico; e
- iii. opcionalmente, uma região 3' de DNA envolvida no término da transcrição e na poliadenilação.

30 A região de DNA mencionada processada em um miRNA pode compreender uma sequência de nucleotídeos que é essencialmente complementar a sequência de nucleotídeos dos pelo menos 21 nucleotídeos

consecutivos de um gene que codifica β -1,3-endoglucanase seletiva para fibra, desde que um ou mais dos seguintes não-pareamentos sejam permitidos:

- 5 a. Um não-pareamento entre o nucleotídeo na terminação 5' do miRNA e a sequência de nucleotídeos correspondente na molécula de RNA;
- b. Um não-pareamento entre qualquer um dos nucleotídeos na posição 1 a posição 9 do miRNA e a sequência de nucleotídeos correspondente na molécula de RNA;
- 10 c. Três não-pareamentos entre qualquer um dos nucleotídeos na posição 12 a posição 21 do miRNA e a sequência de nucleotídeos correspondente na molécula de RNA, desde que não haja mais do que dois não-pareamentos consecutivos.

Como usado aqui, um "miRNA" é uma molécula de RNA com cerca de 20 a 22 nucleotídeos de comprimento que pode ser carregada em
15 um complexo RISC e direcionar a clivagem de outra molécula de RNA, em que a outra molécula de RNA compreenda uma sequência de nucleotídeos essencialmente complementar a sequência de nucleotídeos da molécula de miRNA de acordo com o que um ou mais dos seguintes não-pareamentos pode ocorrer:

- 20 a. Um não-pareamento entre o nucleotídeo na terminação 5' do dito miRNA e a sequência de nucleotídeos correspondente na molécula de RNA-alvo;
- b. Um não-pareamento entre qualquer um dos nucleotídeos na posição 1 a posição 9 do dito miRNA e a sequência de nucleotídeos correspondente na molécula de RNA-alvo;
- 25 c. Três não-pareamentos entre qualquer um dos nucleotídeos na posição 12 a posição 21 do dito miRNA e a sequência de nucleotídeos correspondente na molécula de RNA-alvo, desde que não haja mais do que dois não-pareamentos consecutivos.
- 30 d. Nenhum não-pareamento é permitido nas posições 10 e 11 do miRNA (todas as posições de miRNA são indicadas começando da terminação 5' da molécula de miRNA).

Um miRNA é processado a partir de uma molécula de "pré-miRNA" por proteínas, tais como proteínas DCL, presentes em qualquer célula de planta e carregadas dentro de um complexo RISC onde elas podem conduzir a clivagem das moléculas de RNA-alvo.

5 Como usado aqui, uma molécula de "pré-miRNA" é uma molécula de RNA com cerca de 100 a cerca de 200 nucleotídeos, preferivelmente cerca de 100 a cerca de 130 nucleotídeos que pode adotar uma estrutura secundária que compreende um tronco de RNA de fita dupla e uma alça de RNA de fita simples e que compreende ainda a sequência de nucleotídeos
10 do miRNA (e sua sequência complementar) no tronco de RNA de fita dupla. Preferivelmente, o miRNA e seu complemento estão localizados a cerca de 10 a cerca de 20 nucleotídeos das extremidades livres do tronco de miRNA de fita dupla. O comprimento e a sequência da região de alça da fita simples não são críticos e podem variar consideravelmente, por exemplo, entre 30 e
15 50 nt de comprimento. Um exemplo de um pré-miRNA sintético está representado na Figura 1. Preferivelmente, a diferença de energia livre entre estruturas de RNA não-pareada e pareada está entre -20 a -60 kcal/mol, particularmente em torno de -40 kcal/mol. A complementaridade entre o miRNA e o miRNA* não precisa ser perfeita e cerca de 1 a 3 projeções de nucleotídeos não-pareados podem ser toleradas. A estrutura secundária adotada por
20 uma molécula de RNA pode ser predita por algoritmos de computador convencionais na técnica tal como mFOLD. A fita particular do tronco de RNA de fita dupla do pré-miRNA que é liberada pela atividade de DCL e carregada sobre o complexo RISC é determinada pelo grau de complementaridade na
25 extremidade 5', de acordo com o que a fita que está menos envolvida em sua terminação 5' com pontes de hidrogênio entre os nucleotídeos de fitas diferentes do tronco de dsRNA clivado é carregada sobre o complexo RISC e determinará a especificidade de sequência da degradação da molécula do RNA-alvo. Entretanto, se empiricamente a molécula de miRNA de uma mo-
30 lécula sintética de pré-miRNA particular não é funcional (por que a fita "errada" foi carregada sobre o complexo RISC) ficará imediatamente claro que esse problema pode ser solucionado pela alteração de posição da molécula

de miRNA e seu complemento sobre as respectivas fitas do tronco de dsRNA da molécula de pré-miRNA. Como é conhecido na técnica, a ligação entre A e U que envolve duas ligações de hidrogênio ou entre G e U que envolve duas ligações de hidrogênio são menos fortes do que entre G e C que envolve três ligações de hidrogênio.

Moléculas de miRNA que ocorrem naturalmente podem estar compreendidas dentro de suas moléculas de pré-miRNA que ocorrem naturalmente, mas elas também podem ser introduzidas em arcabouços existentes de molécula de miRNA pela substituição da sequência de nucleotídeos da molécula de miRNA processada normalmente a partir de tal molécula de pré-miRNA existente pela sequência de nucleotídeos de outro miRNA de interesse. Igualmente, moléculas sintéticas de miRNA podem estar compreendidas dentro e processadas a partir de arcabouços de molécula de pré-miRNA existentes ou arcabouços de pré-miRNA sintético.

As moléculas de pré-miRNA (e consequentemente também as moléculas de miRNA) podem ser convenientemente introduzidas em uma célula de planta pelo provimento das células de planta com um gene que compreende um promotor que pode ser expresso em planta operativamente ligado a uma região de DNA, que quando transcrita rende a molécula de pré-miRNA. O promotor que pode ser expresso em planta pode ser o promotor naturalmente associado com a molécula de pré-miRNA ou pode ser um promotor heterólogo.

Moléculas adequadas de miRNA e pré-microRNA para a regulação negativa da expressão do alelo subgenômico D do gene GhGluc 1 estão descritas nas entradas da listagem de sequência de SEQ ID N° 15, 16, 20 e 21.

Moléculas adequadas de miRNA e pré-microRNA para a regulação negativa da expressão do alelo subgenômico A do gene GhGluc 1 estão descritas nas entradas da listagem de sequência de SEQ ID N° 13, 13, 17, 18 e 19.

Como usado aqui, a expressão "promotor que pode ser expresso em planta" significa uma sequência de DNA que é capaz de controlar (ini-

ciar) a transcrição em uma célula de planta. Isso inclui qualquer promotor originado de planta, mas também qualquer promotor não-originado em planta que é capaz de direcionar a transcrição em uma célula de planta, isto é, certos promotores de origem viral ou bacteriana tais como CaMV35S, o promotor N° 4 ou N° 7 do vírus do trevo subterrâneo ou genes-promotores de T-DNA e similares.

Um promotor que pode ser expresso em planta que controle o início e a manutenção da transcrição, preferencialmente em células de fibra, é um promotor que conduz a transcrição da região de DNA operativamente ligado em um nível mais elevado nas células de fibra e nas células da epiderme subjacente do que em outras células ou tecidos da planta. Tais promotores incluem o promotor para algodão do gene de β -tubulina específico para fibra (como descrito em WO0210377), o promotor para algodão do gene de actina específico para fibra (como descrito em WO0210413), o promotor de um gene de proteína de transferência de lipídeo específica para fibra de algodão (descrito em US5792933), o promotor para um gene de expansina de algodão (WO9830698) ou um promotor de um gene de quitinase no algodão (US2003106097) ou os promotores dos genes específicos para fibra descritos em US6259003 ou US6166294. Promotores seletivos para fibra como descritos aqui também podem ser usados.

Para algumas aplicações, pode ser benéfico diminuir o comprimento da fibra em uma planta que produz fibra, tal como uma planta de algodão. Por exemplo, pode ser benéfico diminuir a quantidade de fibra distorcida em uma planta de algodão, deixando assim mais energia e material para incorporação na fibra em plumas em plantas de algodão. Isso pode ser obtido convenientemente pela (super)expressão de uma β -1,3-endoglucanase em células de fibra em um estágio precoce do desenvolvimento de uma célula de fibra, por exemplo, usando um promotor seletivo para fibra como aqui descrito que é expresso no final da fase de alongamento rápido.

A invenção também abrange os genes quiméricos aqui descritos, assim como as plantas, sementes, tecidos que compreendem esses genes quiméricos e fibras produzidas a partir de tais plantas.

O comprimento da fibra ou outras propriedades da fibra e a qualidade da fibra no algodão também podem ser aumentadas ou alteradas pelo isolamento de variantes alélicas do subgenoma D que não codificam mais uma β -1,3-endoglucanase funcional ou que codificam uma proteína variante com atividade enzimática fraca.

Os métodos para transformação de plantas são bem conhecidos na técnica e são de pouca relevância para a presente invenção. Métodos para transformar plantas de algodão também são bem conhecidos na técnica. A transformação do algodão mediada por Agrobactéria foi descrita, por exemplo, na patente U.S. 5.004.863 ou na patente U.S. 6.483.013 e a transformação do algodão pelo bombardeio de partícula está descrito, por exemplo, em WO 92/15675.

Os genes quiméricos de acordo com a invenção podem ser introduzidos nas plantas de uma maneira estável ou de uma maneira transitória usando métodos bem conhecidos na técnica. Os genes quiméricos podem ser introduzidos nas plantas ou podem ser gerados dentro da célula da planta como descrito, por exemplo, em EP 1339859.

Os genes quiméricos podem ser introduzidos pela transformação em plantas de algodão a partir das quais os calos embriogênicos podem ser derivados, tais como Coker 312, Coker310, Coker 5Acala SJ-5, GSC25110, FIBERMAX 819, Siokra 1-3, T25, GSA75, Acala SJ2, Acala SJ4, Acala SJ5, Acala SJ-C1, Acala B1644, Acala B1654-26, Acala B1654-43, Acala B3991, Acala GC356, Acala GC510, Acala GAM1, Acala C1, Acala Royale, Acala Maxxa, Acala Prema, Acala B638, Acala B1810, Acala B2724, Acala B4894, Acala B5002, non Acala "picker" Siokra, variedade "stripper" FC2017, Coker 315, STONEVILLE 506, STONEVILLE 825, DP50, DP61, DP90, DP77, DES119, McN235, HBX87, HBX191, HBX107, FC 3027, CHEMBRED A1, CHEMBRED A2, CHEMBRED A3, CHEMBRED A4, CHEMBRED B1, CHEMBRED B2, CHEMBRED B3, CHEMBRED C1, CHEMBRED C2, CHEMBRED C3, CHEMBRED C4, PAYMASTER 145, HS26, HS46, SICALA, PIMA S6 ORO BLANCO PIMA, FIBERMAX FM5013, FIBERMAX FM5015, FIBERMAX FM5017, FIBERMAX FM989, FIBERMAX

FM832, FIBERMAX FM966, FIBERMAX FM958, FIBERMAX FM989, FIBERMAX FM958, FIBERMAX FM832, FIBERMAX FM991, FIBERMAX FM819, FIBERMAX FM800, FIBERMAX FM960, FIBERMAX FM966, FIBERMAX FM981, FIBERMAX FM5035, FIBERMAX FM5044, FIBERMAX FM5045, FIBERMAX FM5013, FIBERMAX FM5015, FIBERMAX FM5017 ou FIBERMAX FM5024 e plantas com genótipos derivados dessas.

"Algodão" como usado aqui inclui *Gossypium hirsutum* ou *Gossypium barbadense*. "Plantas de algodão parentais" incluem *Gossypium arboreum*, *Gossypium herbaceum* e *Gossypium raimondii* e *Gossypium longicalyx*.

Os métodos e meios da presente invenção também podem ser empregados para outras espécies de planta tais como cânhamo, juta, linho e plantas lenhosas, incluindo mas não limitado a *Pinus spp.*, *Populus spp.*, *Picea spp.*, *Eucalyptus spp.* etc.

A planta transformada obtida pode ser usada em um esquema de reprodução convencional para produzir mais plantas transformadas com as mesmas características ou para introduzir o gene quimérico de acordo com a invenção em outras variedades da mesma espécie ou de espécies de planta relacionadas ou em plantas híbridas. As sementes obtidas das plantas transformadas contém o gene quimérico da invenção como um inserto genômico estável e são também abrangidas pela invenção.

Em outra modalidade da invenção, são providos anticorpos surgidos contra as proteínas β -1,3-endoglucanase, particularmente anticorpos que reconhecem proteínas β -1,3-endoglucanase que têm sequências de aminoácidos de SEQ ID N° 11 e 12.

Como usado aqui, "compreendendo" deve ser interpretado como especificando a presença dos aspectos estabelecidos, números inteiros, etapas ou componentes como referidos, mas não excluindo a presença ou adição de um ou mais aspectos, números inteiros, etapas ou componentes ou seus grupos. Assim, por exemplo, um ácido nucleico ou proteína que compreende uma sequência de nucleotídeos ou aminoácidos, pode compreender mais nucleotídeos ou aminoácidos do que aqueles realmente citados,

isto é, estar embebidos em um ácido nucleico ou proteína maiores. Um gene quimérico que compreende uma região de DNA, que está funcionalmente ou estruturalmente definida, pode compreender regiões de DNA adicionais, etc.

Os seguintes Exemplos não-limitantes descrevem a identificação de duas β -1,3-endoglucanases do subtipo A e D no algodão assim como suas regiões promotoras e a análise do momento de expressão e envolvimento na síntese da parede celular secundária. Também são descritos genes quiméricos para alteração das características da fibra, para expressão preferencial em fibra e específica para fibra em plantas que produzem fibras tais como o algodão e seus usos. A menos que estabelecido de outra maneira nos Exemplos, todas as técnicas de DNA recombinante são realizadas de acordo com protocolos padronizados como descrito em Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, NY e nos Volumes 1 e 2 de Ausubel *et al.* (1994) *Current Protocols in Molecular Biology*, Current Protocols, USA. Materiais e métodos padronizados para o trabalho molecular em planta estão descritos em *Plant Molecular Biology Labfax* (1993) by R.D.D. Croy, conjuntamente publicado por BIOS Scientific Publications Ltd (UK) e Blackwell Scientific Publications, UK.

Através da descrição e dos Exemplos, é feita referência as seguintes sequências representadas na listagem de sequências:

SEQ ID Nº 1: fragmento genômico amplificado de β -1,3-endoglucanase de *Gossypium hirsutum* Fiber Max966, subtipo A

SEQ ID Nº 2: fragmento genômico amplificado de β -1,3-endoglucanase de *Gossypium hirsutum* Fiber Max966, subtipo D

SEQ ID Nº 3: fragmento genômico amplificado de β -1,3-endoglucanase de *Gossypium arboreum*

SEQ ID Nº 4: fragmento genômico amplificado de β -1,3-endoglucanase de *Gossypium raimondii*

SEQ ID Nº 5: iniciador em sentido reverso (SE002) para amplificação de fragmento genômico de β -1,3-endoglucanase

SEQ ID Nº 6: iniciador em sentido direto (SE003) para amplifi-

cação de fragmento genômico de β -1,3-endoglucanase

SEQ ID N° 7: sequência de aminoácidos codificada pelo fragmento genômico amplificado de β -1,3-endoglucanase de *Gossypium hirsutum* Fiber Max966, subtipo A

5 SEQ ID N° 8: sequência de aminoácidos codificada pelo fragmento genômico amplificado de β -1,3-endoglucanase de *Gossypium hirsutum* Fiber Max966, subtipo D

SEQ ID N° 9: sequência de nucleotídeos do clone genômico completo de β -1,3-endoglucanase de *Gossypium hirsutum* Fiber Max966, subtipo A, incluindo a sequência promotora

10 SEQ ID N° 10: sequência de nucleotídeos do clone genômico completo de β -1,3-endoglucanase de *Gossypium hirsutum* Fiber Max966, subtipo D, incluindo a sequência promotora

SEQ ID N° 11: sequência de aminoácidos de β -1,3-endoglucanase completa de *Gossypium hirsutum* Fiber Max966, subtipo A

SEQ ID N° 12: sequência de aminoácidos de β -1,3-endoglucanase completa de *Gossypium hirsutum* Fiber Max966, subtipo D

SEQ ID N° 13: sequência de nucleotídeos de pré-miRNA1

20 SEQ ID N° 14: sequência de nucleotídeos de pré-miRNA1 com sítios de clonagem

SEQ ID N° 15: sequência de nucleotídeos de pré-miRD1

SEQ ID N° 16: sequência de nucleotídeos de pré-miRD1 com sítios de clonagem

SEQ ID N° 17: sequência de nucleotídeos de pré-miRA2

25 SEQ ID N° 18: sequência de nucleotídeos de pré-miRA2aa

SEQ ID N° 19: sequência de nucleotídeos de pré-miRA2aa com sítios de clonagem

SEQ ID N° 20: sequência de nucleotídeos de pré-miRD2

30 SEQ ID N° 21: sequência de nucleotídeos de pré-miRD2 com sítios de clonagem

Exemplo 1: Identificação de alelos específicos subgenômicos A e D que codificam uma β -1,3-endoglucanase seletiva para fibra em plantas de algodão

e plantas de algodão parentais.

Foi proposto que a calose está envolvida no processo de manutenção da turgência em células de fibra em crescimento. A remoção da calose na base das estruturas de ligação entre células de fibra e não fibra de algodão, conhecidas como plasmodesmos, dissipa a pressão de turgência. Isso resulta no término do alongamento rápido da fibra. Portanto, pelo retardamento da remoção da calose, o comprimento da fibra deve ser aumentado.

A enzima β -1,3-endoglucanase catalisa a hidrólise das ligações β -1,3-D-glicosídicas em β -1,3-D-glicano (calose). Um gene de β -1,3-endoglucanase específica para fibra codificado por *GhGluc 1* não foi detectado quando a calose foi depositada na base da fibra, mas tornou-se evidente no momento da degradação da calose.

Em resumo, o fechamento dos plasmodesmos parece desempenhar um papel importante no alongamento das fibras de algodão. A deposição e a degradação da calose devem estar envolvidas no fechamento e reabertura dos plasmodesmos, respectivamente. A expressão de *GhGluc 1* pode desempenhar um papel nesse processo pela degradação da calose, reabrindo assim os plasmodesmos. (Ruan *et al.*, 2004, Plant Physiology – 136:4104-4113).

Baseado na sequência de nucleotídeos de *GhGluc 1* (número de acesso EMBL D88416) descrita em Ruan *et al.*, 2004, Plant Physiology – 136:4104-4113, 2 iniciadores (SE002: ggccgaagccgatcttatctagg (iniciador em sentido reverso; SEQ ID Nº 5) e SE003: cggcaacaatcttccatctccag (iniciador em sentido direto, SEQ ID Nº 6)) foram determinados para amplificar fragmentos genômicos de DNA de *G. hirsutum* (genoma AD), *G. arboreum* (genoma A) e *G.raimondii* (genoma D). Esses fragmentos foram seqüenciados (vide SEQ ID Nº 1 a 4). Para *G. hirsutum* foram derivadas 2 sequências de consenso, para *G. arboreum* foi derivada 1 sequência de consenso e para *G.raimondii* foi derivada 1 sequência de consenso.

Síntese dos polimorfismos entre as duas sequências de *G. hirsutum* e as duas sequências diploides:

Consequentemente, um tipo da sequência corresponde a sequência do genoma A e o outro tipo corresponde ao genoma D.

Usando um digesto *AlwI* (sítio de reconhecimento = GGATC) sobre o fragmento amplificado por PCR, ambas as variantes subgenômicas dentro de *hirsutum* ou do genoma podem ser distinguidas (vide figura 5). Assim o seguinte protocolo pode ser usado para distinguir os alelos do subgenoma A e D de GhGluc 1. Os iniciadores usados são:

a. SE002: GGCCGAAGCCGATCTTATCTAGG (SEQ ID Nº 5)

b. SE003: CGGCAACAATCTTCCATCTCCAG (SEQ ID Nº 6)

O comprimento do produto de PCR esperado é de 655 bp.

Condições de PCR:

	1 x
DNA Padrão (200ng/μl)	1 μl
Tampão 5 x GreenGoTaq	5 μl
SE002 (10μM)	0,75 μl
SE003 (10μM)	0,75 μl
dNTP's (20 mM)	0,5 μl
GoTaq polYmerase	0,25 μl
Água MQ	16,75 μl
	25 μl

Perfil de PCR

	95°C – 5 min	
	95°C – 1 min	5 x
15	58°C – 1 min	
	72°C – 2 min	
	93°C – 30 s	25 x
	58°C – 30 s	
	72°C – 1 min	
20	72°C – 10 min	

Após a amplificação por PCR, o fragmento de PCR é digerido com o digesto *AlwI* (3 h de incubação @ 37°C) usando 10 μl de padrão; 1 μl de enzima *AlwI*; 2 μl de tampão de restrição NEB 4, 7 μl de água MQ. Os

fragmentos resultantes são analisados sobre gel TAE 1,5% corado com EtBr.

Os tamanhos esperados das bandas para o fragmento de PCR específicos para o alelo do subgenoma A são: 479+118+59 bp. Os tamanhos esperados das bandas para o fragmento de PCR específicos para o alelo do subgenoma D são: 53+118 bp.

Exemplo 2: Expressão diferencial de alelos específicos dos subgenomas A e D que codificam uma β -1,3-endoglucanase seletiva para fibra e correlação com os estágios de desenvolvimento da fibra em *Gossypium hirsutum*.

A expressão alélica específica através do perfil de crescimento/desenvolvimento da fibra foi analisada para *G. hirsutum*. DNA de bibliotecas de cDNA criadas a partir de células de fibra e semente com 5DPA e a partir de células de fibra com 10, 15, 20 e 40 DPA foi extraído, a concentração foi equalizada e a amplificação por PCR mencionada acima foi realizada. As diferenças nas intensidades das bandas correspondem a diferenças relativas na expressão (Figura 6, banda 2, 4, 6, 8 e 10. Um controle positivo (DNA genômico) foi incluído (Figura 6, banda 12). Uma digestão com AlwI foi realizada como descrito no Exemplo 1 para distinguir entre os 2 alelos subgenômicos dos genes GhGluc 1 (Figura 6, bandas 3, 5, 7, 9, 11 e 13).

O perfil de expressão pode ser resumido como se segue:

5 dpa	10 dpa	15 dpa	20 dpa	30 dpa	40 dpa
/	/	D	D	A e D	A e D

Durante a fase de alongamento rápido no desenvolvimento da fibra (5 e 10 dpa), não há expressão de *GhGluc 1*. Portanto, o tampão de calose não é degradado; a turgência na célula não é liberada e o alongamento continua. Acredita-SE que a banda mais fraca com 5 dpa indica alguma expressão na semente (não na fibra). Na transição durante a fase de alongamento e a fase de formação da parede secundária (15 DPA), *GhGluc 1* é expresso. Isso libera a turgência na célula e o alongamento rápido cessa. Apenas a variante similar ao genoma D do gene *GhGluc 1* é expressa. Com 20 DPA há mais expressão do que com 15 DPA (banda mais intensa). A variante similar ao genoma A não é expressa com 15 e 20 DPA. A variante

similar ao genoma A, assim como a variante similar ao genoma D, é expressa com 30 e 40 DPA e possivelmente desempenha um papel na maturação ou outras propriedades da fibra.

Exemplo 3: Isolamento e identificação de região promotora de alelos específicos dos subgenomas A e D que codificam uma β -1,3-endoglucanase seletiva para fibra.

Um fragmento de PCR que compreende a sequência de nucleotídeos dos alelos específicos subgenômicos A e D de GhGluc 1 foi usado para rastrear uma biblioteca BAC que contém clones de DNA genômico da variedade de *Gossypium hirsutum*. 4 clones diferentes foram identificados, 2 para cada variante subgenômica. A sequência de nucleotídeos dos fragmentos genômicos para cada uma das variantes alélicas foi identificada e está representada na SEQ ID N° 9 (genoma A) e SEQ ID N° 10 (genoma D).

Para a variante do genoma A, um TATA box pode ser identificado nas posições 2278 a 2281; um sítio de início da transcrição na posição 2308. A líder 5' não-traduzida se estende do nucleotídeo 2308 a 2409; o códon de início da tradução está localizado nas posições 2410 a 2412. A sequência codificadora consiste em dois éxons (nt 2410 a 2443 e nt 2556 a 3499) separada por uma sequência de íntron (2444 a 2555). O códon de parada da tradução está localizado na posição 3497 a 3499 e o sítio de poliadenilação está localizado na posição 3624.

Para a variante do genoma D, um TATA box pode ser identificado nas posições 3242 a 3245; um sítio de início da transcrição na posição 3270. A líder 5' não-traduzida se estende do nucleotídeo 3270 a 3372; o códon de início da tradução está localizado nas posições 3373 a 3375. A sequência codificadora consiste em dois éxons (nt 3373 a 3406 e nt 3501 a 4444) separada por uma sequência de íntron (nt 3407 a 3500). O códon de parada da tradução está localizado na posição 4442 a 4444 e o sítio de poliadenilação está localizado na posição 4566.

A sequência de nucleotídeos das regiões codificadoras fundidas foi alinhada (Figura 2) e as diferenças estão indicadas; similarmente, as sequências de nucleotídeos dos íntrons foram alinhadas (Figura 3). A Figura 4

mostra um alinhamento das proteínas codificadas, enquanto que a figura 1 mostra um alinhamento das sequências de nucleotídeos localizadas a montante da região codificadora.

Exemplo 4: Construtos de gene quimérico que compreendem regiões promotoras seletivas para fibra diferentes operativamente ligadas a um gene marcador.

Os seguintes fragmentos de DNA foram operativamente ligados usando técnicas recombinantes padronizadas:

- a. A sequência de nucleotídeos do promotor específico do alelo A compreendendo a sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo 465 ao nucleotídeo 2409.
- b. Uma região codificadora de β -glicuronidase (GUS)
- c. Um fragmento que compreende um sinal de término da transcrição e de poliadenilação (CaMV) 3' 35S (indicado daqui por diante como Gluc1-SGA (A1.9)) e
- d. A sequência de nucleotídeos do promotor específico para o alelo A que compreende a sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo 1374 ao nucleotídeo 2409.
- e. Uma região codificadora de β -glicuronidase (GUS)
- f. Um fragmento que compreende um sinal de término da transcrição e de poliadenilação (CaMV) 3' 35S
- e
- g. A sequência de nucleotídeos do promotor específico para o alelo A que compreende a sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo 1531 ao nucleotídeo 2409.
- h. Uma região codificadora de β -glicuronidase (GUS)
- i. Um fragmento que compreende um sinal de término da transcrição e de poliadenilação (CaMV) 3' 35S
- e
- j. A sequência de nucleotídeos do promotor específico para o alelo D que compreende a sequência de nucleotídeos de SEQ ID N° 10 do nucleotídeo 1397 ao nucleotídeo 3372. (indicado daqui por diante como

Gluc1-SGD (D2.0))

k. Uma região codificadora de β -glicuronidase (GUS)

l. Um fragmento que compreende um sinal de término da transcrição e de poliadenilação (CaMV) 3' 35S

5 e

m. A sequência de nucleotídeos do promotor específico para o alelo D que compreende a sequência de nucleotídeos de SEQ ID Nº 10 do nucleotídeo 2371 ao nucleotídeo 3372.

n. Uma região codificadora de β -glicuronidase (GUS)

10 Um fragmento que compreende um sinal de término da transcrição e de poliadenilação (CaMV) 3' 35S

ou

o. A sequência de nucleotídeos do promotor específico para o alelo D que compreende a sequência de nucleotídeos de SEQ ID Nº 10 do nucleotídeo 2718 ao nucleotídeo 3372.

15 p. Uma região codificadora de β -glicuronidase (GUS)

q. Um fragmento que compreende um sinal de término da transcrição e de poliadenilação (CaMV) 3' 35S

Os genes quiméricos acima foram clonados separadamente em um vetor de T-DNA, na presença de um gene marcador quimérico selecionado e introduzidos na cepa de *Agrobacterium tumefaciens* que compreende um plasmídeo Ti auxiliar desarmado. As cepas de *Agrobacterium* foram usadas para gerar plantas de algodão transgênicas de acordo com o método descrito em WO00/71733. As células de fibra e outros tecidos dessas plantas foram corados histoquimicamente.

20 A expressão dos construtos quiméricos diferentes é predominantemente observada nas células de fibra e fibras.

Exemplo 5: Construtos de gene quimérico compreendendo regiões promotoras seletivas para fibra diferentes operativamente ligadas a região codificadora de N-acetilglicosamina transferase

30 Construtos similares aos construtos do Exemplo 4, mas em que a região codificadora de GUS foi substituída por uma região codificadora de

NodC (vide WO2006/136531, especificamente as SEQ ID 1 a 9, incorporadas por referência) precedida por uma sequência líder não-traduzida Cab22, que foram reunidas usando técnicas comuns de DNA recombinante.

Esses genes quiméricos são clonados separadamente em um vetor de T-DNA na presença de um gene marcador quimérico selecionado e introduzidos na cepa de *Agrobacterium tumefaciens* que compreende um plasmídeo Ti auxiliar desarmado. As cepas de *Agrobacterium* foram usadas para gerar plantas de algodão transgênicas de acordo com o método descrito em WO00/71733. As células de fibra e outros tecidos dessas plantas são analisados quanto a presença de oligossacarídeos positivamente carregados como descrito em WO2006/136531.

Exemplo 6: Aumento do comprimento da fibra em *Gossypium hirsutum* pela regulação negativa de alelos específicos do subgenoma D que codificam uma β -1,3-glicanase seletiva para fibra.

Pré-microRNAs sintéticos específicos para a variante alélica subgenômica A ou D de GhGluc 1 foram planejados usando regras de planejamento disponíveis na técnica.

Moléculas adequadas de miRNA e pré-microRNA para a regulação negativa específica da expressão do alelo subgenômico D de GhGluc 1 estão descritas nas entradas da listagem de sequência de SEQ ID Nº 15, 16, 20 e 21.

Moléculas adequadas de miRNA e pré-microRNA para a regulação negativa específica da expressão do alelo subgenômico A de GhGluc 1 estão descritas nas entradas da listagem de sequência de SEQ ID Nº 13, 14, 17, 18 e 19.

As sequências de nucleotídeos que codificam o pré-microRNA são clonadas sob o controle de um promotor CaMV 35S ou sob o controle de um promotor específico para fibra e/ou preferencial para fibra como descrito no Exemplo 4.

Esses genes quiméricos são clonados separadamente em um vetor de T-DNA na presença de um gene marcador quimérico selecionado e introduzidos na cepa de *Agrobacterium tumefaciens* que compreende um

plasmídeo Ti auxiliar desarmado. As cepas de *Agrobacterium* foram usadas para gerar plantas de algodão transgênicas de acordo com o método descrito em WO00/71733. As fibras de tais plantas ou de sua progênie são analisadas quanto ao comprimento da fibra aumentado.

5 Exemplo 7: Análise da especificidade de promotores gluc1 para fibras.

RNA foi isolado de folhas, raízes e caules de algodão e analisado para a expressão de genes gluc1 e para o gene da proteína fosfatase 2A (uma Serina/Treonina fosfatase abundante e conservada com ampla especificidade pelo substrato e funções celulares diversas). A síntese da primeira
10 fita de cDNA foi realizada usando o sistema SuperScript First Strand Synthesis para RT-PCR (Invitrogen), após o que os iniciadores específicos para gluc1 ou pp2A foram adicionados e a reação de RT-PCR foi realizada. Os resultados são visualizados na Figura 7. O resultado positivo na banda 17 indica que a síntese da primeira fita foi obtida com sucesso. Os resultados
15 positivos nas bandas 10, 12, 14 usando iniciadores específicos para pp2A em amostras de RNA de folha, raiz e caule indicaram especificamente que a síntese da primeira fita na amostra de folha, raiz e caule funcionou. As bandas 11, 13 e 15 são controles negativos que omitem a transcriptase reversa da reação. A banda 8 é um controle positivo que consiste em DNA genômico
20 para verificar a exatidão da reação de PCR específica para gluc1. Nas bandas 2, 4, 6 nenhum sinal pode ser detectado indicando a ausência da expressão de Gluc1 em folhas, raízes e caules de algodão. As bandas 3, 5 e 7 são controles negativos que omitem a transcriptase reversa.

Em conclusão, nenhuma expressão de Gluc1 pode ser detectada na folha, raiz e tecido do caule de algodão.
25

Exemplo 8: Transformação de óvulo de algodão com cepas de *Agrobacterium* contendo

Óvulos de flores de algodão foram isolados com 0DPA, 1DPA e 2DPA e usados para experimentos de cultura de óvulo (como descrito, por exemplo, por Feng e Brown, 2000, In vitro Cellular & Developmental Biology, Volume 36, Número 4, páginas 293 a 299) após a transformação com *Agrobacterium tumefaciens* que carrega o promotor quimérico diferente de
30

- gluc1::genes de fusão de GUS indicados no Exemplo 4 como Gluc1-SGA(A1.9) e Gluc1-SGD(D2.0). As células de fibra/fibras iniciais foram analisadas histoquimicamente para a expressão de GUS após 1 semana, 2 semanas, 3 semanas, 4 semanas ou 5 semanas. Como um controle positivo, 5 óvulos cultivados foram incluídos que tinham sido infectados com uma cepa de *Agrobacterium* abrigando um promotor quimérico CaMV35::gene de fusão de GUS. Observe que a região codificadora de GUS usada continha uma sequência de íntron para evitar a expressão acidental de GUS nas *Agrobacterias* infectantes. Os resultados estão resumidos na Tabela 1.

10 Tabela 1: Análise histoquímica de GUS dos óvulos de algodão cultivados e transfectados

	7DPA	14DPA	21DPA	28DPA	35DPA
CaMV35S-gus	+	+	+	+	+
Gluc1-SGA-gus	-	-	-	-	+
Gluc1-SGD-gus	-	+	+	+	-

Esses dados corroboram os perfis de expressão deduzidos para os promotores dos diferentes alelos subgenômicos para GhGluc 1 a partir dos dados do Exemplo 2.

LISTAGEM DE SEQUÊNCIAS

	<110>	Bayer BioScience N.V. Engelen, Steven Arioli, Tony	
5	<120>	EXPRESSÃO DIFERENCIAL DE ALELOS SUBGENÔMICOS ESPECÍ- FICOS NO ALGODÃO E USOS DOS MESMOS	
	<130>	BCS 07-2001	
	<150>	EP07000550.9	
	<151>	2007-01-11	
10	<150>	US60/884,564	
	<151>	2007-01-11	
	<160>	21	
	<170>	PatentIn versão 3.3	
	<210>	1	
15	<211>	689	
	<212>	DNA	
	<213>	Gossypium hirsutum	
	<400>	1	
		tggccgtgat attggtgttt gctatggttt gaacggcaac aatcttccat ctccaggaga	60
20		tggtattaat cttttcaaaa ctagtggtcat aaacaatatc aggtcttacc agccttacct	120
		tgaagtgtc gaagcagcaa ggggatcggg aatatccctc tcgatgagta cgacaaacga	180
		ggacatacaa agcctcgcaa cggatcaaag tgcagccgat gcatgggtta acaccaacat	240
		cgtcccttat aaggaagatg ttcaattcag gttcatcatc attgggaatg aagccattcc	300
		aggacagtca agctcttaca ttctggtgc catgaacaac ataatgaact cgctggcctc	360
25		at ttgggcta ggcacgacga aggttacgac cgtgggtccc atgaatgccc taagtacctc	420
		gtaccctcct tcagacggcg cttttggaag cgatataaca tcgatcatga ctagtatcat	480
		ggccattctg gttcgacagg attcgcccct cctgatcaat gtgtaccctt attttgcccta	540
		tgccctcagac cccactcata tttccctcaa ctacgccttg ttcacctcga ccgcaccggt	600
		ggtggtcgac caaggcttgg aatactacaa cctctttgac ggcatgggtc atgctttcaa	660
30		tgccgcccta gataagatcg gcttcggcc	689
	<210>	2	
	<211>	689	

<212> DNA
 <213> *Gossypium hirsutum*
 <400> 2
 tggccgtgat attggtgttt gctatggttt gaacggcaac aatcttccat ctccaggaga 60
 5 tgttattaat ctttataaaa ctagtggcat aaacaatatc aggctctacc agccttacct 120
 tgaagtgtc gaagcagcaa ggggatcggg aatatccctc tcgatgggtc cgagaaacga 180
 ggacatacaa agcctcgcaa aagatcaaag tgcagccgat gcatgggtta acaccaacat 240
 cgtcccttat aaggacgatg ttcagttcaa gttgatcact attgggaatg aagccatttc 300
 aggacaatca agctcttaca ttcctgatgc catgaacaac ataatgaact cgctcgctt 360
 10 atttgggtta ggcacgacga aggttacgac cgtgggtccg atgaatgcc taagtacctc 420
 gtaccctcct tcagacggcg cttttggaag cgatataaca tcgatcatga ctagtatcat 480
 ggccattctg gctgtacagg attcgcccct cctgatcaat gtgtaccctt attttgctta 540
 tgcctcagac cccactcata tttccctcga ttacgccttg ttcacctcga cgcaccgggt 600
 ggtggtcgac caaggcttg aataactaaa cctctttgac ggcattggctg atgctttcaa 660
 15 tgccgcccta gataagatcg gcttcggcc 689
 <210> 3
 <211> 655
 <212> DNA
 <213> *Gossypium arboreum*
 20 <400> 3
 cggcaacaat ctccatctc caggagatgt tattaatctt taaaaacta gtggcataaa 60
 caatatcagg ctctaccagc cttacctgaa gtgtcgaag gagcaagggg atcggaata 120
 tccctctcga tgagtacgac aaacgaggac atacaaagcc tcgcaacgga tcaaagtga 180
 gccgatgcat gggttaacac caacatcgtc cttataagg acgatgttca attcaggttc 240
 25 atcatcattg ggaatgaagc cattccagga cagtcaagct cttacattcc tggtgccatg 300
 aacaacataa tgaactcgct cgcctcattt gggctaggca cgacgaagggt tacgaccgtg 360
 gtcccgatga atgccctaag tacctcgtac cctccttcag acggcgcttt tggaagcgat 420
 ataacatcga tcatgactag tatcatggcc attctgggtc gacaggattc gccctcctg 480
 atcaatgtgt acccttattt tgcctatgcc tcagaccca ctcattttc cctcaactac 540
 30 gccttggtca cctcgaccgc accggtggtg gtcgaccaag gcttgaata ctacaacctc 600
 tttgacggca tggtcgatgc tttcaatgcc gccctagata agatcggctt cgcc 655
 <210> 4

	<211>	656	
	<212>	DNA	
	<213>	Gossypium raimondii	
	<400>	4	
5	cggcaacaat cttccatctc caggagatgt tattaatctt tacaaaacta gtggcataaa	60	
	caatatcagg ctctaccagc cttaccctga agtgctcgaa gcagcaaggg gatcgggaat	120	
	atccctctcg atgggtccga gaaacgagga catacaaagc ctcgcaaaag atcaaagtgc	180	
	agccgatgca tgggttaaca ccaacatcgt cccttataag gacgatgttc agttcaaatt	240	
	gatcactatt gggaatgaag ccatttcagg acaatcaagc tcttacattc ctgatgccat	300	
10	gaacaacata atgaactcgc tcgcctcatt tgggttaggc acaacgaagg ttacgaccgt	360	
	gggtcccgatg aatgccctaa gtacctcgta ccttccttca gacggcgctt ttggaagcga	420	
	tataacatcg atcatgacta gtatcatggc cattctggct gtacaggatt cgccccctct	480	
	gatcaatgtg tacccttatt ttgcctatgc ctcagacccc actcatatct ccctcgatta	540	
	cgcccttgctt acctcgaccg caccgggtggg ggtcgaccaa ggcttggaat actacaacct	600	
15	ctttgacggc atgggtcgatg ctttcaatgc cgccctagat aagatcggct tcggcc	656	
	<210>	5	
	<211>	23	
	<212>	DNA	
	<213>	Artificial	
20	<220>		
	<223>	Iniciador de oligonucleotídeo para amplificação do fragmento genômico de endoglicanase	
	<400>	5	
	ggccgaagcc gatcttatct agg	23	
25	<210>	6	
	<211>	23	
	<212>	DNA	
	<213>	Artificial	
	<220>		
30	<223>	Iniciador de oligonucleotídeo para amplificação do fragmento genômico de endoglicanase	
	<400>	6	

* cggcaacaat cttccatctc cag 23
 <210> 7
 <211> 229
 <212> PRT
 5 <213> Gossypium hirsutum
 <400> 7
 Gly Arg Asp Ile Gly Val Cys Tyr Gly Leu Asn Gly Asn Asn Leu Pro
 1 5 10 15
 Ser Pro Gly Asp Val Ile Asn Leu Phe Lys Thr Ser Gly Ile Asn Asn
 10 20 25 30
 Ile Arg Leu Tyr Gln Pro Tyr Pro Glu Val Leu Glu Ala Ala Arg Gly
 35 40 45
 Ser Gly Ile Ser Leu Ser Met Ser Thr Thr Asn Glu Asp Ile Gln Ser
 50 55 60
 15 Leu Ala Thr Asp Gln Ser Ala Ala Asp Ala Trp Val Asn Thr Asn Ile
 65 70 75 80
 Val Pro Tyr Lys Glu Asp Val Gln Phe Arg Phe Ile Ile Ile Gly Asn
 85 90 95
 Glu Ala Ile Pro Gly Gln Ser Ser Ser Tyr Ile Pro Gly Ala Met Asn
 20 100 105 110
 Asn Ile Met Asn Ser Leu Ala Ser Phe Gly Leu Gly Thr Thr Lys Val
 115 120 125
 Thr Thr Val Val Pro Met Asn Ala Leu Ser Thr Ser Tyr Pro Pro Ser
 130 135 140
 25 Asp Gly Ala Phe Gly Ser Asp Ile Thr Ser Ile Met Thr Ser Ile Met
 145 150 155 160
 Ala Ile Leu Val Arg Gln Asp Ser Pro Leu Leu Ile Asn Val Tyr Pro
 165 170 175
 Tyr Phe Ala Tyr Ala Ser Asp Pro Thr His Ile Ser Leu Asn Tyr Ala
 30 180 185 190
 Leu Phe Thr Ser Thr Ala Pro Val Val Val Asp Gln Gly Leu Glu Tyr
 195 200 205

Tyr Asn Leu Phe Asp Gly Met Val Asp Ala Phe Asn Ala Ala Leu Asp
 210 215 220
 Lys Ile Gly Phe Gly
 225
 5 <210> 8
 <211> 229
 <212> PRT
 <213> Gossypium hirsutum
 <400> 8
 10 Gly Arg Asp Ile Gly Val Cys Tyr Gly Leu Asn Gly Asn Asn Leu Pro
 1 5 10 15
 Ser Pro Gly Asp Val Ile Asn Leu Tyr Lys Thr Ser Gly Ile Asn Asn
 20 25 30
 Ile Arg Leu Tyr Gln Pro Tyr Pro Glu Val Leu Glu Ala Ala Arg Gly
 15 35 40 45
 Ser Gly Ile Ser Leu Ser Met Gly Pro Arg Asn Glu Asp Ile Gln Ser
 50 55 60
 Leu Ala Lys Asp Gln Ser Ala Ala Asp Ala Trp Val Asn Thr Asn Ile
 65 70 75 80
 20 Val Pro Tyr Lys Asp Asp Val Gln Phe Lys Leu Ile Thr Ile Gly Asn
 85 90 95
 Glu Ala Ile Ser Gly Gln Ser Ser Ser Tyr Ile Pro Asp Ala Met Asn
 100 105 110
 Asn Ile Met Asn Ser Leu Ala Leu Phe Gly Leu Gly Thr Thr Lys Val
 25 115 120 125
 Thr Thr Val Val Pro Met Asn Ala Leu Ser Thr Ser Tyr Pro Pro Ser
 130 135 140
 Asp Gly Ala Phe Gly Ser Asp Ile Thr Ser Ile Met Thr Ser Ile Met
 145 150 155 160
 30 Ala Ile Leu Ala Val Gln Asp Ser Pro Leu Leu Ile Asn Val Tyr Pro
 165 170 175
 Tyr Phe Ala Tyr Ala Ser Asp Pro Thr His Ile Ser Leu Asp Tyr Ala

		180		185		190	
		Leu Phe Thr Ser Thr Ala Pro Val Val Val Asp Gln Gly Leu Glu Tyr					
		195		200		205	
		Tyr Asn Leu Phe Asp Gly Met Val Asp Ala Phe Asn Ala Ala Leu Asp					
5		210		215		220	
		Lys Ile Gly Phe Gly					
		225					
		<210>	9				
		<211>	6009				
10		<212>	DNA				
		<213>	Gossypium hirsutum				
		<220>					
		<221>	Região de interesse biológico				
		<222>	(2278)..(2281)				
15		<223>	TATA box				
		<220>					
		<221>	Região de interesse biológico				
		<222>	(2308)..(2308)				
		<223>	Sítio de início da transcrição				
20		<220>					
		<221>	Região de interesse biológico				
		<222>	(2308)..(2409)				
		<223>	Sequência líder 5' não traduzida				
		<220>					
25		<221>	Região de interesse biológico				
		<222>	(2410)..(2412)				
		<223>	Códon de início da tradução (ATG)				
		<220>					
		<221>	Região de interesse biológico				
30		<222>	(2410)..(2443)				
		<223>	Sequência codificante parte I				
		<220>					

<221> Região de interesse biológico
 <222> (2444)..(2555)
 <223> Sequência do íntron
 <220>
 5 <221> Região de interesse biológico
 <222> (2556)..(3499)
 <223> Sequência codificante parte II
 <220>
 <221> Região de interesse biológico
 10 <222> (3497)..(3499)
 <223> Códon de parada da tradução
 <220>
 <221> Região de interesse biológico
 <222> (3624)..(3624)
 15 <223> Sítio de poliadenilação
 <400> 9
 cgcgccatat aattttatgt gtgtaatttg ttgggttaat tacttaaaat agtatatttt 60
 taattgctgt aattaatgta agataatttt tattatttga atcattgcac aaaattaaaa 120
 tagaataatt tatttaacaa ttcaaataata ataataatcc aaattataat tatagtattt 180
 20 ttacaatatt caatatacaa tatagtttta cttcatacaa ttaatatataaaa aaaatatttat 240
 tcaaaataat aactaataaaa cataattacc atatattaat tattttgata tttcgaacat 300
 aacgctaata aaaaatttcc taatcattat taaatcattt gtataaacta taaagaaatt 360
 gatatattgt aaattaaact ttattcattt tttttcttaa tactcaataa attaatcata 420
 ataactcata aataatatat aattaaaata atcataacat ggtagattat ataaataggg 480
 25 ggcgaatcta gggagctggc atgaccctta aaatagaatt ttctattttg acctatcaaa 540
 atttttaaaa ttttaaatta gtaaaggtaa atttgactt tgacctctta aatgataaaa 600
 attttacttt aatcctttta aatttacatt ttactatca taaaattac aatttgattt 660
 tgcccctaaa atttttttct agcttagccc tgtatataaa tatattattt ataattttta 720
 tattttaaata ataaagtttt taattataca aataattaaa atctgatatt taaaactaaa 780
 30 gtaatttctt ttttcttttt actttttttt aattgcaaca taatggttta aatatctata 840
 taacgtatga agtaatttga tataaatttt attttaattt attattatat aaattcattt 900
 agtaaaaact tttaatagaa tcaaaatttt tatttgtaaa ttcgataact tttcttatca 960

agtatatttg tgagaaccaa atatttagta aaattaatat tcttatttat aaatatgata 1020
 aatcttataa aaaaatattt aaaatgaaaa aaattgtaca aatattataa aaaaatattt 1080
 aaaatgaaaa acattgtaca aaggctatat aagaagttca aaagtttctt cgaccatgta 1140
 ctcttataga gattatagat agattataaa actatatgta gtttctctta acttttaaat 1200
 5 aagaggataa atgtatttta atgtactcaa acttatatat ttttatattg acaataatat 1260
 caatatcaac ctaattaaga ttcattctaa cattaatggt gaagattttt aataaaagaa 1320
 aaggttaata aattaattag aacacaaaca aacacaaatt taagtgggtat gtaaggctct 1380
 tgacccaaag gaaaaatttg ttacgtcgat taaattataa attaatttaa agtaaaatta 1440
 cattttaacc taaaaaaga gaaaagtata tctaatttct tcgaaaatgg aaagaaaatt 1500
 10 ataaatttat ggcatttcta aaaaaattct gaattcgcta ctaaaagatg aaattataaa 1560
 atccgaagca ttaccagaag atggatcacc aaatcacaaa caatcaatga aaagtaatga 1620
 taattaattg aaagtgaagca ttttaatttg atagccatat acttcctgct gaatttatag 1680
 gttctcatta atgcaattaa attatattcg acaccttttg aatgaaataa aatgacacaa 1740
 gaggaagac gggtcatcta ttttttcttt caatcgccca tcaaaatacc aaaaatgtaa 1800
 15 ctacatgcaa aaaatcaa atgaaaaata ttcattttt gatattttta tatattgtgt 1860
 gttcaaacg taaatgtatt gaaaaattat gatgggtgtg ttgctgtatg tccataaaat 1920
 tcaatgtact cacatttatc aaatgtatac tttgagagaa gttattttga taatactcaa 1980
 gtttttttta tagatgggaa aattttttta attatttttt gattttgatg aaatgtatat 2040
 ataaatttta attcgatata tataaatata tatgtaaatt ttaaatttaa atttaataat 2100
 20 atacaattaa gaaaataatt tataaatatt ttccgattaa aaataaatct ggaaagaaga 2160
 aatgtcaaca ctttttcatt aaatacaatt aggatgggac acgatacctt catgcattga 2220
 tatctcaggt ggtccaaaaa ctcggaatcc tttttgaaaa aaaacttcca gagagagtat 2280
 ataaatccag cagtaggcac aagaaacgag caccagttat tgactttcct ttgtaaaaaa 2340
 aaaaagtgtc gagatcaaga aatatagtga aatatgggtc caagattttc tgggttttta 2400
 25 atctaagcaa tgctgttttt aactcaactc ctctctctaa caggtaaaac aaacttctct 2460
 acagtgattt tacagtaa atggctttga aaaatataca acaaacatt tatcttcaat 2520
 ccattttaat tactgatcta ctatatatgt tgcagatggc cgtgatattg gtgtttgcta 2580
 tgggttgaac ggcaacaatc ttccatctcc aggagatggtt attaattctt tcaaaactag 2640
 tggcataaac aatatcaggc tctaccagcc ttaccctgaa gtgctcgaag cagcaagggg 2700
 30 atcgggaata tccctctcga tgagtacgac aaacgaggac atacaaagcc tcgcaacgga 2760
 tcaaagtgca gccgatgcat gggttaacac caacatcgct ccttataagg aagatgttca 2820
 attcaggttc atcatcattg ggaatgaagc cattccagga cagtcaagct cttacattcc 2880

tgggtgccatg aacaacataa tgaactcgct ggcctcattt gggctaggca cgacgaaggt 2940
 tacgaccgtg gtcccgatga atgccctaag tacctcgtac cctccttcag acggcgcttt 3000
 tggaagcgat ataacatcga tcatgactag tatcatggcc attctgggtc gacaggattc 3060
 gcccctcctg atcaatgtgt acccttattt tgcctatgcc tcagacccca ctcattttt 3120
 5 cctcaactac gccttggttca cctcgaccgc accggtgggtg gtcgaccaag gcttggaata 3180
 ctacaacctc tttagcggca tggtcgatgc tttaatgcc gccctagata agatcggctt 3240
 cggccaaatt actctcattg tagccgaaac tggatggccg accgccggtg acgagcctta 3300
 cacgagtgtc gcgaacgctc aaacttataa caagaacttg ttgaatcatg tgacgcagaa 3360
 agggactccg aaaagacctg aatatataat gccgacgttt ttcttcgaga tgttcaacga 3420
 10 gaacttgaag caaccacag ttgagcagaa ttctggattc ttcttcccca atatgaacct 3480
 tgtttatcca ttttggtgaa cttgaaatgt tattgttggc tatttaaadc ttttgccaga 3540
 gacgcttcat atagtttctg catattttga aagtggaaaa tcaatctaaa tataaataag 3600
 ttttatttgt tgttttttaa ttaaataaaa ttttaaataat tttaaaaaca tctttatttg 3660
 taattaaata ttaaataaaa agtttaatat tcaaatttta tcaattcaaa aataaaataa 3720
 15 aaatatatta aattttattt tacgaataaa ttgattttct attaatgcag attttaaata 3780
 atttgatata aattttcaat tcaacaatag taattttgat cacatcaaag gagaaagggg 3840
 aagatttaac ttttaattggt gacctaatat aacacgttga aaacggagtt cccaataagg 3900
 caaatgact tgtaatgacg aaagagatgt ccaagtgaag tctgctttta agtgaaagaa 3960
 gcataaaagg ataactaat aactcatgat cttaaattgaa gttctataaa atgcaacttt 4020
 20 catctagaaa caaggtatgt cttaaattgat gttttatgaa tttgtcttaa ttgggtttta 4080
 tgcaatgaat tcatggatag cacatctcta attatacgtt gctgggttat atgagagtgg 4140
 tgcagaagtt aattgtgctt taaatacttg cttagtgttt atgaaatttg aaaagtgtta 4200
 tatacttata ataaaaataa ttcgattcgg aatccaattc aggggttcgac tcaatataat 4260
 aaaattttac agatatcttg aaggggatct tcttcttctc tacttctcga gcagtgttat 4320
 25 atatttaca taaagataac tcaattcgag atccgacctg atataataaa attctacaga 4380
 catatcaaag agggagatct tcttcttccc tacatcttga ccttcttgat caaatgacc 4440
 ttccttatat ttttacatac gttgattata tgaatcaaaa gaaagatacc aaaaagtgtt 4500
 taaaaataaa caacgggggtt cttatgtaga gatgcttatg ggccgggccc gactcaacta 4560
 aaaatttagg cacattcatt gggcccaggt cgggcctaac ccaaaaatgg gcctaaaatt 4620
 30 ttgccaagc ttgactcaaa taaaatgct aaaattcggg cctgaccccg tattaatttt 4680
 atatttttt atataacttt taaatatata taatatataa aaaatactaa aaaaattaaa 4740
 ataaatattt cccaactaaa ctaaaattat taagaaaaat aattcatatt agcgtataaa 4800

ttggaaattg accaaaatta aaattattgt atagttaatc tatattaaaa ggacatgtaa 4860
 ttaaaaaacca ttaaaaactat tataacaataa attaaatctt cattgtatac atagaaaggc 4920
 attaataatt aaaaaactat attaagatat aaactaaatt caaaattatt aaaaacaaga 4980
 actaaataaa aaagcaattg aaaattacga attaatgtta aaatcaaagc ttaaaatcaa 5040
 5 gggacttaaa taaaaaatatc ccaaaataca aaacattagc ttcctttccc atccacgtga 5100
 atgcaaagtt tacatggtgt ttcctagtgt ttgtgcgact ccaacctttt atttacctct 5160
 ttttttcttt atttgaacaa ttatttgata atgattagaa ttttgggatt gttgctcatc 5220
 gtacgtgcaa cacttaaaat cactatgatt tttcataatt tatataacct atatcgtttt 5280
 ggaaattaat tttatttttt atattatttt aataaaaata ccatctacct tttttaattt 5340
 10 atgatccctt tcatatttaa aaattcaaat tgacaattgt ctaactaaac accgtcacac 5400
 tccaataaga ttgtaatttc ctccatcttg atattacact caaaagcatg ttgccacaa 5460
 acaaatcaac tagccttttt ctaccactat tcatcatctt cttaagagtg tgtttatgtc 5520
 atgtgccgag attttaggta tggtcacgtt gtggctttta actcaaactt attgcccag 5580
 agtctaagtt agcctccgat cctcactaaa gagaggcttg gcacacttta cctagccaag 5640
 15 tacacaagga atagagctat tagaaagcat taaagagtta ggagaatgtg gaagtgtttt 5700
 tattactcaa agctaacttg gatacaaata aaggagggag cctctccttt aggcaagctt 5760
 cttttgatct gatggttaca attaatctcg aataggaggg gtcaaacttc tctcagtt 5820
 tcatattatc tcttggtgct tggttggcct ccgccttgag acaactttag ataacaccta 5880
 gtcttaacac ttttagcttc acattgtacg catccttcat tactcaaagc ccacaaagcc 5940
 20 tccttactta aggtcttgg tcgctccac taccttcggc tttagactca tctaagatct 6000
 tccaatcg 6009
 <210> 10
 <211> 6877
 <212> DNA
 25 <213> *Gossypium hirsutum*
 <220>
 <221> Região de interesse biológico
 <222> (3242)..(3245)
 <223> TATA box
 30 <220>
 <221> Região de interesse biológico
 <222> (3270)..(3270)

	<223>	sítio de início da transcrição
	<220>	
	<221>	Região de interesse biológico
	<222>	(3270)..(3372)
5	<223>	Sequência líder 5' não-traduzida
	<220>	
	<221>	Região de interesse biológico
	<222>	(3373)..(3375)
	<223>	Códon de início da tradução (ATG)
10	<220>	
	<221>	Região de interesse biológico
	<222>	(3373)..(3406)
	<223>	Sequência codificante parte I
	<220>	
15	<221>	Região de interesse biológico
	<222>	(3407)..(3500)
	<223>	Sequência do intron
	<220>	
	<221>	Região de interesse biológico
20	<222>	(3501)..(4444)
	<223>	Sequência codificante parte II
	<220>	
	<221>	Região de interesse biológico
	<222>	(4442)..(4444)
25	<223>	códon de parada
	<220>	
	<221>	Região de interesse biológico
	<222>	(4566)..(4566)
	<223>	Sítio de poliadenilação
30	<400>	10
ttcaaactta ctcgcttgca caaaaataat ttataaaaag tatttaaaat ataaaatttt 60		
tatgttgata atatttttat atacatttta tattttaaga aataataatt ttttaggaat 120		

tagaaaaaaa atgtagaata atatcattga tattttaatt tttcaaaaaa ttaaaaataa 180
 gttcacgtag tctaatttta tctattttta tttttatact ttcaaattga gataaatatc 240
 aaagaacttt tgggttcaata tgcaatttga tacttaaatt ttaatttgat gtaattatta 300
 catgaaactt ggcttgtggt ttatacgtat acatgaaatt tttattttga ttcaattgta 360
 5 cgcattttaa gaaatgaaaa tgggttcta tcaataatat tattagtga ttgtgaaatt 420
 taaaactttt atgcattaaa ccacacaaaa tcagagttta tgtatgatat tgcacattgg 480
 actatagttc atgcatattt tttatatttt atccatgtca aattttgaaa tttcattcct 540
 aacttatatg atagcagtta aatttggtta gtcaaaactct agtattagtt atatactata 600
 cataacttgt agagtttagt ttaagttcac taatttgatt attttttatc tgtttatttt 660
 10 ttcaatttca agatttaagt tttaagctta acttaaaca tagtcattaa atttattaac 720
 taaaatgtcc tggggttttt tgtaagtatt ataatatgtt tgccacgtga gattttggta 780
 aaagtagagt ttaacttaac aaatttaatg gctactactt agtaaggatt agaatttcaa 840
 aattaaaaaa aaaattatag aggctaaaga tgatcaaatt agaggtttaa attaagtcaa 900
 attaaaaatag ttctggatat taactattta aattaattaa tgtcattata aaattagagg 960
 15 tctaaattat gtaaaattaa aatataaaaa ctaaatctcg aatgtgagta tagtataagg 1020
 atcaaaagtg attttggtca ttttctttta tttacaaata ttcttagaga tgctctttta 1080
 tatataatgg tttctaattg gatatgcgcg gcatataatt ttatgtgttt aatttatttt 1140
 attaattatt taaataatat atttttaatt actataatta atgtaaaata tttttatta 1200
 tttgaatcac tgcacaaaat taaaatatac taacttaatt aacaattcaa atataataat 1260
 20 aatctaaatt ataattaaag cttttttaca atattcagta tataatatag ttttacttta 1320
 tataattaat acaaagaaat attattcaaa ataataacta atcaacataa ttactatata 1380
 tcaattattt tgatatttcg aacataatgc taataaaaaa tttcctaattc attattaaat 1440
 catttgata aactataaag aaattgatat attgtaaatt aaacttttaa ctattcaatt 1500
 ttttcttaat agtcaataaa ttaatcataa taattcataa ttaatatata attaacataa 1560
 25 ccataacata gaatttttta ttttgccca ttaaaatttt taaaatttta aattagtaaa 1620
 ggaaaaatta cactttgacc ccttaaaaat gataaaattt tattttaatc ctttaaaatt 1680
 gacattttta ctattgtaa aattacaatt taattttgcc cccctaaaaa atttttctag 1740
 cttcgccctt gtgtataaat atattaatta caatttttat atttgaatta tataaataat 1800
 taaattttga tatttaaaac taaagtaatc tctttttttt ttactttttt ttaattgaaa 1860
 30 cataatgggt taaatatcta tattacgtat gaagtaattt aatataaatt ttattttaat 1920
 ttattattat ataaattcat ttagtaaaaa cttttaatag aatcaaaatt tttatttgta 1980
 aattcgataa cttttcttat caagtaaatt tgttgaatta aatatttagt aaaattaata 2040

tttttattta taaatatgat aaatcttata aaaaataaaa aaatatttaa aatgaaaaac 2100
 attgtacaaa ggctatataa gaagttcaaa agtttcttcg accctgtact ctaatagaga 2160
 ttatagatag attatagaac tattcatagt ttctcttaac ctttaaataa gaatttttagt 2220
 gtactcaaac ttacatattt ttatattgat aataatgtca ataccagccg agttaagatt 2280
 5 cactcgacat taatgttgaa aatttttaaat aaaagaaaat gttgataagt taattagaac 2340
 acaagcaagc acaaatttaa gtggttaagta aggtccttga ccctaattgga aaaattgtta 2400
 tgttgattaa attataaatt aatttaaggt aaaattatat ttgacctaa aaaaatgaaa 2460
 aaaatatatc tagtttcttc gaaaatgaaa agaaaataat aaattgatac attataaaat 2520
 ttatggcatt tctaaaaaaa ttctgaattt gatgaaatta taataaaaaa aaagttttaa 2580
 10 aacatataga tttcaagaat agtgggaaaa ttatatttga acaacactga agaaatccaa 2640
 agcattagca gaaaatggat caccaaatca caacaatca gtgaaaagta atgataatta 2700
 attgaaagtg agcattttaa tttgatagcc atatacttcc tgctgaattt ataggttctc 2760
 attaattgca ttaaattata tttgtcactt tttgaatgaa ataaatgaca cagttcatct 2820
 attttttttc tttcaatcgc ccatcaaaat accgaaaatg taactacatt aaaaaagatc 2880
 15 gaaaaatatt catattttga tattttaata gattgtgtgt tcaaggcgta atgtactaaa 2940
 aaattatgat ggtgttgctg ctgtatgtcc ataaaattca atgtattcgc atgtatcaaa 3000
 tgtaaatttt gacacaagtt attctaataa taatcaagtt atttttatac atgagataca 3060
 tctcaaaatt atttttatat atccgaaaaa tcataacgta cgatcaaact agaaagagga 3120
 agtgtcaaaa cctattcatt atatgcaaat atgatgggac acgataccct catgcattga 3180
 20 tatctcatat tgtccaaaaa ctcagaatcc tttttgaaaa aaaaaaattc cagagagagt 3240
 gtataaatcc agcagtgtgc acaagaaacg agcaccagtt attgacattc ctttgtaaaa 3300
 aaaaaagaa gctgagatca agaaatatag tgaaatatgg gtccaacatt ttctggggtt 3360
 ttaatctcag caatgggtgtt tttaactcaa ctctctctc taacaggtaa aacaaacttc 3420
 tctacagtga ttttacggta agtatggctt tgaaaaatat acaacaaaac atttatactg 3480
 25 atctaccata tatgttgag atggccgtga tattggtgtt tgctatgggt tgaacggcaa 3540
 caatcttcca tctccaggag atgttattaa tctttacaaa actagtggca taaacaatat 3600
 caggctctac cagccttacc ctgaagtgtc cgaagcagca aggggatcgg gaatatccct 3660
 ctcgatgggt ccgagaaacg aggacataca aagcctcgca aaagatcaaa gtgcagccga 3720
 tgcagtgggt aacaccaaca tcgtccctta taaggacgat gttcagttca agttgatcac 3780
 30 tattgggaat gaagccattt caggacaatc aagctcttac attcctgatg ccatgaacaa 3840
 cataatgaac tcgctcgcct tatttgggtt aggacgacg aagggttacga ccgtgggtccc 3900
 gatgaatgcc ctaagtacct cgtaccctcc ttcagacggc gcttttggaa gcgatataac 3960

atcgatcatg actagtatca tggccattct ggctgtacag gattcgcccc tcctgatcaa 4020
 tgtgtaccct tattttgcct atgcctcaga cccactcat atttccctcg attacgcctt 4080
 gttcacctcg accgcaccgg tgggtggtcga ccaaggcttg gaatactaca acctctttga 4140
 cggcatggtc gatgctttca atgccgcctt agataagatc ggcttcggcc aaattactct 4200
 5 cattgtagcc gaaactggat ggccgaccgc cggtaacgag ccttacacga gtgtcgcgaa 4260
 cgctcaaaact tataacaaga acttggttaa tcatgtgacg cagaagggga ctccgaaaag 4320
 acctgaatat ataatgccga cgtttttctt cgagatgttc aacgaggatt tgaagcaacc 4380
 cacagttgag cagaatttcg gattcttctt cccaatatg aaccctgttt atccattttg 4440
 gtgaagttga aatgttggtg gctattttaa tcttttgcca gagacgcttc atatagtttc 4500
 10 tgcataat tttt gaaagtggaa aatcaatcta aatattaata agttttatgt gttgtttttt 4560
 aattaaataa aattttaaat attttaaaaa tatctttatt ggtaattaaa tattaataa 4620
 aaagtttaat attcaaattt tatcaattca aaaataaaaat aaaaatatat taaattttatt 4680
 tttacgaata aattgatttt ctattaatac agattttgaa taatttgata taaattttaa 4740
 attcaacaat agtaattttg atcacatcaa aggagaaaagg gaaagattta actttaattg 4800
 15 gtgacctaat ataacacgtt gaaaacggag ctcccaggaa ggcaaaatga cttgtaatga 4860
 cgaaagagat gtccaagtag aatctgcatt aaagtgaaaa aagcataaaa ggataagtaa 4920
 actcatgatc tgacataaat tgaagttcta taaaatgcaa ctttcatcta gaaacaaggt 4980
 atgtcttaaa tgatgtttta tgaatttgct ttaactgggt tttatgcaat gaattcatgg 5040
 atagcacctc actaattata cgttgctggt ttatatgaga gtgggtgcaga agttaattgt 5100
 20 gctttaaata cttgcttagt gttcaagaaa tttgaaaagt attatatatt tataataaaa 5160
 ataattcaga tccgactcaa tctagtaaaa ttttacaac attctaaagg ggatcttctt 5220
 ttttctctac ttattgatca gtgttatata cttataataa agacaacctg atttgagatc 5280
 cggcctaata taataaaatt ctacagacat ctcaagggag agatcttctt ctccctaca 5340
 tcttgacctt tttgatcaaa atttctctcc ctctatttcc acattgggtg atcatatgaa 5400
 25 tcaacagaaa ggtaccaaaa agttttttaa aataaacaaa ggggttctta tgaaattcat 5460
 atgatatatatt ggggtctaatt attagaatca attttaagtt taaacaaatt taaaattcaa 5520
 aactcaattc catttttggt tgaacggaaa gttactaatt gttaagaaaa ataattcata 5580
 ttagcgtata aattggaaat tgaccaaacc taaaattatt gtatagttaa tctatattaa 5640
 aaggacatgt aattaaaaac cattaaaact attatagaat aaattaaatc ttcattctat 5700
 30 acatacaaag tcattaataa ttaaaaaact atattaagat ataaactata ttcaaaaaat 5760
 attaaaaaca ataactaat aaaaaaaca attgaaaatt acgaattaat gttaaaatca 5820
 agggacttaa ataaaaatat cccaaaatac aaacattag cttccttcc catccacgtg 5880

attgcaaagt ttacatggtg tttcctagtg cttgtgcgac tccaaccttt tatttacttt 5940
 tttcttttct ttatttgaac aattatttga taatgattag aattttggga ttgttgctca 6000
 tcgtacgtgc aacacttaaa atcactatga tttttcataa tttatataac ctatatcggt 6060
 ttggaaatta atgttattat ttatattggt ttaataaaaa taccatctac ctcttttaat 6120
 5 ttatgatcca tttcttattt gaaaattcaa attgacagtt gtctaactaa acaccatcgc 6180
 actccaataa aattgtaatt ttttctatcg tgaatagtag actcaaaagt atgttggttaa 6240
 caaacaaatc aattagcctt tttctacctc tattcatcat cttcttaata gcgtgtttat 6300
 gtcacgtggt gagatttttag ttccggtcac gtgtggcctt aaaccgaat ttcttacgca 6360
 tgagtctaag ttagcctctg atcctcgcta tggagatgct tggcacagtt tacctaggta 6420
 10 agtaaacaag gaatagagct attagaaagc atcagagagt taggagaatg tggaagtgtt 6480
 tctattactc aaagctaact tggatacaaa taaaagaggg agcctctcct ttaggcaagc 6540
 ctattttgat ctgacggttg caattaatct cgaataggag gggtcgaact tctcactcag 6600
 tttcacatta tctcttggtg cttagttggc ctccgccttg agacacattc aaataacacc 6660
 tagtcttaac acttttggct tcttattgtg cgtatccttc attactcaaa tgccacaaag 6720
 15 cctcattact taagactctc ggtcgctccc actaccttcg acttttagact catctaagat 6780
 cttcccaatc gtagacaact tggccttggg ggggaaatct tgcaccctac ggggccttac 6840
 ataagaagca attaaatggc tttctctcac ccacctt 6877
 <210> 11
 <211> 325
 20 <212> PRT
 <213> Gossypium hirsutum
 <400> 11
 Met Leu Phe Leu Thr Gln Leu Leu Ser Leu Thr Asp Gly Arg Asp Ile
 1 5 10 15
 25 Gly Val Cys Tyr Gly Leu Asn Gly Asn Asn Leu Pro Ser Pro Gly Asp
 20 25 30
 Val Ile Asn Leu Phe Lys Thr Ser Gly Ile Asn Asn Ile Arg Leu Tyr
 35 40 45
 Gln Pro Tyr Pro Glu Val Leu Glu Ala Ala Arg Gly Ser Gly Ile Ser
 30 50 55 60
 Leu Ser Met Ser Thr Thr Asn Glu Asp Ile Gln Ser Leu Ala Thr Asp
 65 70 75 80

Gln Ser Ala Ala Asp Ala Trp Val Asn Thr Asn Ile Val Pro Tyr Lys
 85 90 95
 Glu Asp Val Gln Phe Arg Phe Ile Ile Ile Gly Asn Glu Ala Ile Pro
 100 105 110
 5 Gly Gln Ser Ser Ser Tyr Ile Pro Gly Ala Met Asn Asn Ile Met Asn
 115 120 125
 Ser Leu Ala Ser Phe Gly Leu Gly Thr Thr Lys Val Thr Thr Val Val
 130 135 140
 Pro Met Asn Ala Leu Ser Thr Ser Tyr Pro Pro Ser Asp Gly Ala Phe
 10 145 150 155 160
 Gly Ser Asp Ile Thr Ser Ile Met Thr Ser Ile Met Ala Ile Leu Val
 165 170 175
 Arg Gln Asp Ser Pro Leu Leu Ile Asn Val Tyr Pro Tyr Phe Ala Tyr
 180 185 190
 15 Ala Ser Asp Pro Thr His Ile Ser Leu Asn Tyr Ala Leu Phe Thr Ser
 195 200 205
 Thr Ala Pro Val Val Val Asp Gln Gly Leu Glu Tyr Tyr Asn Leu Phe
 210 215 220
 Asp Gly Met Val Asp Ala Phe Asn Ala Ala Leu Asp Lys Ile Gly Phe
 20 225 230 235 240
 Gly Gln Ile Thr Leu Ile Val Ala Glu Thr Gly Trp Pro Thr Ala Gly
 245 250 255
 Asn Glu Pro Tyr Thr Ser Val Ala Asn Ala Gln Thr Tyr Asn Lys Asn
 260 265 270
 25 Leu Leu Asn His Val Thr Gln Lys Gly Thr Pro Lys Arg Pro Glu Tyr
 275 280 285
 Ile Met Pro Thr Phe Phe Phe Glu Met Phe Asn Glu Asn Leu Lys Gln
 290 295 300
 Pro Thr Val Glu Gln Asn Phe Gly Phe Phe Phe Pro Asn Met Asn Pro
 30 305 310 315 320
 Val Tyr Pro Phe Trp
 325

<210> 12
 <211> 325
 <212> PRT
 <213> Gossypium hirsutum
 5 <400> 12
 Met Val Phe Leu Thr Gln Leu Leu Ser Leu Thr Asp Gly Arg Asp Ile
 1 5 10 15
 Gly Val Cys Tyr Gly Leu Asn Gly Asn Asn Leu Pro Ser Pro Gly Asp
 20 25 30
 10 Val Ile Asn Leu Tyr Lys Thr Ser Gly Ile Asn Asn Ile Arg Leu Tyr
 35 40 45
 Gln Pro Tyr Pro Glu Val Leu Glu Ala Ala Arg Gly Ser Gly Ile Ser
 50 55 60
 Leu Ser Met Gly Pro Arg Asn Glu Asp Ile Gln Ser Leu Ala Lys Asp
 15 65 70 75 80
 Gln Ser Ala Ala Asp Ala Trp Val Asn Thr Asn Ile Val Pro Tyr Lys
 85 90 95
 Asp Asp Val Gln Phe Lys Leu Ile Thr Ile Gly Asn Glu Ala Ile Ser
 100 105 110
 20 Gly Gln Ser Ser Ser Tyr Ile Pro Asp Ala Met Asn Asn Ile Met Asn
 115 120 125
 Ser Leu Ala Leu Phe Gly Leu Gly Thr Thr Lys Val Thr Thr Val Val
 130 135 140
 Pro Met Asn Ala Leu Ser Thr Ser Tyr Pro Pro Ser Asp Gly Ala Phe
 25 145 150 155 160
 Gly Ser Asp Ile Thr Ser Ile Met Thr Ser Ile Met Ala Ile Leu Ala
 165 170 175
 Val Gln Asp Ser Pro Leu Leu Ile Asn Val Tyr Pro Tyr Phe Ala Tyr
 180 185 190
 30 Ala Ser Asp Pro Thr His Ile Ser Leu Asp Tyr Ala Leu Phe Thr Ser
 195 200 205
 Thr Ala Pro Val Val Val Asp Gln Gly Leu Glu Tyr Tyr Asn Leu Phe

	210	215	220	
	Asp Gly Met Val	Asp Ala Phe Asn Ala Ala	Leu Asp Lys Ile Gly Phe	
	225	230	235	240
	Gly Gln Ile Thr	Leu Ile Val Ala Glu Thr	Gly Trp Pro Thr	Ala Gly
5	245	250	255	
	Asn Glu Pro Tyr Thr	Ser Val Ala Asn Ala Gln Thr	Tyr Asn Lys Asn	
	260	265	270	
	Leu Leu Asn His Val	Thr Gln Lys Gly Thr	Pro Lys Arg Pro	Glu Tyr
	275	280	285	
10	Ile Met Pro Thr	Phe Phe Phe Glu Met	Phe Asn Glu Asp	Leu Lys Gln
	290	295	300	
	Pro Thr Val Glu Gln	Asn Phe Gly Phe Phe	Phe Pro Asn Met	Asn Pro
	305	310	315	320
	Val Tyr Pro Phe	Trp		
15	325			
	<210>	13		
	<211>	134		
	<212>	DNA		
	<213>	Artificial		
20	<220>			
	<223>	Pré-miRA1		
	<220>			
	<221>	Região de interesse biológico		
	<222>	(17)..(36)		
25	<223>	Fita I de miRNA		
	<220>			
	<221>	Região de interesse biológico		
	<222>	(87)..(107)		
	<223>	Fita II de miRNA I		
30	<400>	13		
	tctagggtaa	gggaggtcct	gtcgaaccag	aatggcacag ggtactttct tgcattgcttg 60
	agcctttcat	gcttgaagct	ctgcgccatt	ctcgttcaga caggatccac ttttctctct 120

	ctctctcact cact	134
	<210> 14	
	<211> 146	
	<212> DNA	
5	<213> Artificial	
	<220>	
	<223> Pré-miRA1 com sítios de clonagem	
	<400> 14	
	ccatgggtcta gggtaaggga ggtcctgtcg aaccagaatg gcacagggtta ctttcttgca	60
10	tgcttgagcc tttcatgctt gaagctctgc gccattctcg ttcagacagg atccactttt	120
	ctctctctct ctctctcact gctagc	146
	<210> 15	
	<211> 134	
	<212> DNA	
15	<213> Artificial	
	<220>	
	<223> Pré-miRD1	
	<220>	
	<221> Região de interesse biológico	
20	<222> (17)..(36)	
	<223> Fita I de miRNA	
	<220>	
	<221> Região de interesse biológico	
	<222> (87)..(107)	
25	<223> Fita II de miRNA	
	<400> 15	
	tctagggtaa gggaggatcc tgtacagcca gaatggacag ggtactttct tgcattgcttg	60
	agcctttcat gcttgaagct ctgcccattc tgcctgtaac aggattccac ttttctctct	120
	ctctctcact cact	134
30	<210> 16	
	<211> 146	
	<212> DNA	

	<213>	Artificial	
	<220>		
	<223>	Pré-miRD1 com sítios de clonagem	
	<400>	16	
5	ccatgggtcta	gggtaaggga ggatcctgta cagccagaat ggacaggga ctttcttgca	60
	tgcttgagcc	tttcatgctt gaagctctgc ccattctgcc tgtaacagga ttccactttt	120
	ctctctctct	ctcactcact gctagc	146
	<210>	17	
	<211>	134	
10	<212>	DNA	
	<213>	Artificial	
	<220>		
	<223>	Pré-miRA2	
	<220>		
15	<221>	Região de interesse biológico	
	<222>	(17)..(36)	
	<223>	Fita I de microRNA	
	<220>		
	<221>	Região de interesse biológico	
20	<222>	(87)..(107)	
	<223>	Fita II de microRNA	
	<400>	17	
	tctagggttaa	gggagggttcc caatgatgat gaacctacag ggtactttct tgcattgcttg	60
	agcctttcat	gcttgaagct ctgcagggttc atcatcaatt gggaatccac ttttctctct	120
25	ctctctcact	cact	134
	<210>	18	
	<211>	134	
	<212>	DNA	
	<213>	Artificial	
30	<220>		
	<223>	Pré-miRA2aa	
	<220>		

<221> Região de interesse biológico
 <222> (17)..(36)
 <223> Fita I de microRNA
 <220>

5 <221> Região de interesse biológico
 <222> (87)..(107)
 <223> Fita II de microRNA
 <400> 18

tctagggtaa gggagggtcc caatgatgat gaacctacag ggtactttct tgcattgcttg 60
 10 agcctttcat gcttgaagct ctgcagggtc ataactcatt gggaatccac ttttctctct 120
 ctctctcact cact 134

<210> 19
 <211> 146
 <212> DNA

15 <213> Artificial
 <220>

<223> Pré-miRA2aa com sítios de clonagem
 <400> 19

ccatgggtcta gggtaaggga gggtcccaat gatgatgaac ctacagggta ctttcttgca 60
 20 tgcttgagcc tttcatgctt gaagctctgc aggttcataa tcacttggga atccactttt 120
 ctctctctct ctactcact gctagc 146

<210> 20
 <211> 134
 <212> DNA

25 <213> Artificial
 <220>

<223> Pré-miRD2
 <220>

<221> Região de interesse biológico
 30 <222> (17)..(36)
 <223> Fita I de microRNA
 <220>

```

<221>      Região de interesse biológico
<222>      (87)..(107)
<223>      Fita II de microRNA
<400>      20
5  tctagggtaa gggaggattc ccaatagtga tcaactacag ggtactttct tgcattgcttg  60
   agcctttcat gcttgaagct ctgcagttga tccctatatg ggaattccac ttttctctct 120
   ctctctcact cact 134
<210>      21
<211>      146
10 <212>      DNA
    <213>      Artificial
    <220>
    <223>      Pré-miRD2 com sítios de clonagem
    <400>      21
15 ccatgggtcta gggtaaggga ggattcccaa tagtgatcaa ctacagggta ctttcttgca  60
   tgcttgagcc tttcatgctt gaagctctgc agttgatccc tatatgggaa ttccactttt 120
   ctctctctct ctactcact gctagc 146

```

REIVINDICAÇÕES

1. Promotor preferencial para célula de fibra compreendendo, uma sequência de nucleotídeos selecionada do seguinte grupo de sequências de nucleotídeos:

5 a) uma sequência de nucleotídeos que compreende a sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo na posição 2149 ao nucleotídeo da posição 2307;

 b) uma sequência de nucleotídeos que compreende a sequência de nucleotídeos de SEQ ID N° 10 do nucleotídeo na posição 3109 ao nucleotídeo da posição 3269;

 c) uma sequência de nucleotídeos que compreende a sequência de nucleotídeos que tem pelo menos 70%, preferivelmente pelo menos 80%, mais preferivelmente pelo menos 90%, particularmente pelo menos 90%, particularmente pelo menos 95% de identidade de sequência com ou, mais
15 particularmente que é idêntica a qualquer uma das ditas sequências de nucleotídeos mencionadas sob a) ou b) ou

 d) uma sequência de nucleotídeos que compreende a sequência de nucleotídeos de um fragmento de DNA com cerca de 150 bp a cerca de 1000 bp que hibridiza sob condições estridentes com um fragmento de
20 DNA que tem a dita sequência de nucleotídeos mencionada sob a), b) ou c).

2. Promotor preferencial para célula de fibra de acordo com a reivindicação 1, compreendendo a sequência de nucleotídeos de SEQ ID N° 9 da posição 465 a posição 2307.

3. Promotor preferencial para célula de fibra de acordo com a
25 reivindicação 1, compreendendo a sequência de nucleotídeos de SEQ ID N° 9 da posição 1374 a posição 2307.

4. Promotor preferencial para célula de fibra de acordo com a reivindicação 1, compreendendo a sequência de nucleotídeos de SEQ ID N° 9 da posição 1531 a posição 2307.

30 5. Promotor preferencial para célula de fibra de acordo com a reivindicação 1, compreendendo a sequência de nucleotídeos de SEQ ID N° 10 da posição 1397 a posição 3269.

6. Promotor preferencial para célula de fibra de acordo com a reivindicação 1, compreendendo a sequência de nucleotídeos de SEQ ID N° 10 da posição 2371 a posição 3269.

7. Promotor preferencial para célula de fibra de acordo com a reivindicação 1, compreendendo a sequência de nucleotídeos de SEQ ID N° 10 da posição 2718 a posição 3269.

8. Região promotora preferencial para célula de fibra, compreendendo um promotor preferencial para célula de fibra como definido em qualquer uma das reivindicações 1 a 7.

9. Região promotora preferencial para célula de fibra de acordo com a reivindicação 8, compreendendo ainda a sequência de nucleotídeos de SEQ ID N° 9 do nucleotídeo da posição 2308 ao nucleotídeo da posição 2409.

10. Região promotora preferencial para célula de fibra de acordo com a reivindicação 8, compreendendo ainda a sequência de nucleotídeos de SEQ ID N° 10 do nucleotídeo da posição 3270 ao nucleotídeo da posição 3372.

11. Gene quimérico, compreendendo as seguintes regiões de DNA operativamente ligadas:

a) um promotor específico para célula de fibra como definido em qualquer uma das reivindicações 1 a 7;

b) uma região de DNA heterólogo codificando um RNA biologicamente ativo de interesse; e

c) um sinal de término da transcrição e de poliadenilação.

12. Gene quimérico de acordo com a reivindicação 11, em que o dito RNA biologicamente ativo codifica uma proteína de interesse.

13. Gene quimérico de acordo com a reivindicação 12, em que a dita região de DNA heterólogo codifique uma molécula selecionada entre uma N-acetilglicosamina transferase, preferivelmente uma N-acetilglicosamina transferase do tipo NodC, uma celulose sintase, uma sacarose sintase, preferivelmente uma sacarose sintase do tipo C; uma sacarose fosfato sintase; uma β -1,3-endoglucanase, xilano sintase (cs1F), xiloglicano sintase

(cs1D) e 1,3-1,4 glicano sintase (cs1C).

14. Gene quimérico de acordo com a reivindicação 11, em que o dito RNA biologicamente ativo é uma ribozima, microRNA, RNA em forma de grampo de fita dupla.

15. Gene quimérico de acordo com a reivindicação 12, em que o dito RNA biologicamente ativo regula negativamente a expressão de um gene endógeno de algodão, tal como β -1,3-endoglucanase.

16. Célula de planta, compreendendo um gene quimérico como definido em qualquer uma das reivindicações 11 a 15.

17. Planta, compreendendo em suas células um gene quimérico como definido em qualquer uma das reivindicações 11 a 15.

18. Planta de acordo com a reivindicação 17, que é uma planta de algodão.

19. Semente de uma planta, compreendendo em suas células um gene quimérico como definido em qualquer uma das reivindicações 11 a 15.

20. Método para expressar um RNA biologicamente ativo preferencialmente em uma célula de fibra de uma planta que produz fibra, tal como uma planta de algodão, o dito método, compreendendo:

a) fornecer as células da dita planta com um gene quimérico como definido em qualquer uma das reivindicações 11 a 15; e

b) cultivar as ditas plantas.

21. Método de acordo com a reivindicação 20, em que a dita planta é uma planta de algodão.

22. Uso de um promotor específico para fibra e/ou preferencial para fibra como definido em qualquer uma das reivindicações 1 a 7, para expressão preferencial de um RNA biologicamente ativo em células de fibra de uma planta que produz fibra.

23. Uso de acordo com a reivindicação 22, em que a dita planta é uma planta de algodão.

24. Molécula de DNA isolado, compreendendo uma sequência de nucleotídeos que codifica uma proteína que compreende uma sequência

de aminoácidos que tem pelo menos 95% de identidade de sequência, preferivelmente é idêntica a sequência de aminoácidos de SEQ ID N° 11 ou SEQ ID N° 12 e que codifica uma β -1,3-endoglucanase.

25. Molécula de DNA isolado compreendendo uma sequência
5 de nucleotídeos selecionada do grupo de SEQ ID N° 1, SEQ ID N° 2, SEQ ID N° 3 ou SEQ ID N° 4.

26. Molécula de DNA isolado compreendendo uma sequência
de nucleotídeos selecionada a partir da sequência de nucleotídeos de SEQ
ID N° 9 da posição 2410 a 2443 e a sequência de nucleotídeos de SEQ ID
10 N° 9 da posição 2556 a 3499 ou a sequência de nucleotídeos de SEQ ID N°
10 da posição 3373 a 3406 e a sequência de nucleotídeos de SEQ ID N° 10
da posição 3501 a 4444.

27. Uso da molécula de DNA isolado de acordo com qualquer
uma das reivindicações 24 a 27 para isolamento de um promotor ou região
15 promotora preferencial para célula de fibra.

28. Método para isolamento de região promotora preferencial
para célula de fibra, compreendendo as etapas de:

a) identificar um fragmento genômico que codifique um transcri-
to de RNA a partir do qual um cDNA pode ser sintetizado, o dito cDNA, com-
20 preendendo a sequência de nucleotídeos de SEQ ID N° 1, SEQ ID N° 2,
SEQ ID N° 3 ou SEQ ID N° 4 ou seus equivalentes funcionais;

b) isolar uma região de DNA a montante de uma sequência de
nucleotídeos que codifique a proteína com a sequência de aminoácidos de
SEQ ID N° 11 ou SEQ ID N° 12 ou seus equivalentes funcionais.

25 29. Promotor ou região promotora preferencial para célula de fi-
bra obtida pelo método como definido na reivindicação 28.

30. Método para aumentar o comprimento de uma fibra de uma
planta de algodão, compreendendo a etapa de introduzir um gene quimérico
em uma célula da dita planta de algodão, o dito gene quimérico compreen-
30 dendo os seguintes elementos de DNA operativamente ligados:

(d) um promotor que pode ser expresso em planta, preferivel-
mente um promotor que pode ser expresso em planta que controle a trans-

crição preferencialmente nas ditas células de fibra;

(e) uma região de DNA transcrito, que quando transcrita produz uma molécula de RNA biologicamente ativa capaz de reduzir a expressão de um gene endógeno que codifica β -1,3-endoglucanase para a dita planta que produz fibra, a dita β -1,3-endoglucanase estando envolvida na remoção de calose dos ditos plasmodesmos e

(f) uma região terminal 3' que compreende sinais de término da transcrição e de poliadenilação que funcionam em células de planta,

caracterizado em que o dito gene que codifica β -1,3-endoglucanase endógena é um gene endógeno que codifica β -1,3-endoglucanase compreendido dentro do subgenoma D da dita planta de algodão.

31. Método de acordo com a reivindicação 30, em que o dito promotor é um promotor de beta-tubulina específico para fibra de algodão, um promotor de actina específico para fibra de algodão, um promotor específico para fibra de algodão de um gene de proteína de transferência de lipídeo, um promotor de um gene de expansina de algodão ou um promotor de um gene de quitinase em algodão ou um promotor como definido em qualquer uma das reivindicações 1 a 7.

32. Método de acordo com a reivindicação 31, em que o dito gene que codifica β -1,3-endoglucanase compreendido dentro do subgenoma D é caracterizado pela presença de uma sequência de íntron que tem a sequência de nucleotídeos de SEQ ID N° 10 da posição 3407 a 3500.

33. Método de acordo com a reivindicação 31, em que o dito gene que codifica β -1,3-endoglucanase compreendido dentro do subgenoma D é caracterizado pela presença de uma sequência de nucleotídeos AAGATC cerca de 326 nucleotídeos a jusante do códon de início da tradução (e que não inclui o códon de início da tradução).

34. Método de acordo com a reivindicação 31, em que o dito gene que codifica β -1,3-endoglucanase compreendido dentro do subgenoma D é caracterizado pela identificação de um fragmento com cerca de 538 bp e um fragmento com cerca de 118 bp após a amplificação por PCR com oligonucleotídeos que têm a sequência de nucleotídeos de SEQ ID N° 5 e SEQ

ID Nº 6 seguida pela digestão com a enzima de restrição A1wl.

35. Método de acordo com a reivindicação 31, em que o dito gene endógeno codifica uma proteína que compreende a sequência de aminoácidos de SEQ ID Nº 12 ou no qual o dito gene endógeno compreende a
5 sequência de nucleotídeos de SEQ ID Nº 2.

36. Método de acordo com qualquer uma das reivindicações 30 a 35, em que a dita planta de algodão é *Gossypium hirsutum*.

37. Método de acordo com qualquer uma das reivindicações 30 a 35, em que o dito RNA biologicamente ativo é um RNA antissenso que
10 compreende pelo menos 19 nucleotídeos que têm pelo menos 94% de identidade de sequência com o complemento da sequência de nucleotídeos de SEQ ID Nº 3.

38. Método de acordo com qualquer uma das reivindicações 30 a 35, em que o dito RNA biologicamente ativo é um RNA sense que com-
15 preende pelo menos 19 nucleotídeos que têm pelo menos 94% de identidade de sequência com o complemento da sequência de nucleotídeos de SEQ ID Nº 3.

39. Método de acordo com qualquer uma das reivindicações 30 a 35, em que o dito RNA biologicamente ativo é uma molécula de RNA de
20 fita dupla que compreende pelo menos 19 nucleotídeos que têm pelo menos 94% de identidade de sequência com o complemento da sequência de nucleotídeos de SEQ ID Nº 3 e uma fita de RNA complementar essencialmente similar ao complemento dos ditos pelo menos 19 nucleotídeos.

40. Método de acordo com qualquer uma das reivindicações 37 a 39, em que os ditos 19 nucleotídeos compreendem pelo menos um nucleotídeo específico para o gene gHgluc 1 do subgenoma D como indicado na
25 Figura 2.

41. Método de acordo com a reivindicação 39, em que o dito RNA de fita dupla é microRNA processado a partir de um pré-microRNA
30 compreendendo a sequência de nucleotídeos de SEQ ID Nº 15, SEQ ID Nº 16, SEQ ID Nº 20 ou SEQ ID Nº 21.

42. Método de acordo com qualquer uma das reivindicações 30

- a 41, em que a dita fibra é uma fibra em plumas.
43. Método de acordo com qualquer uma das reivindicações 30 a 41, em que a dita fibra é uma fibra crespa.
44. Método de acordo com qualquer uma das reivindicações 30 a 43, compreendendo ainda
- (a) cultivar plantas obtidas de acordo com o dito método; e
- (b) isolar as fibras das ditas plantas que produzem fibras.
45. Método para diminuir o comprimento de uma fibra de uma planta que produz fibra, compreendendo a etapa de introduzir um gene quimérico em uma célula da dita planta que produz fibra, em que o dito gene quimérico compreende os seguintes fragmentos de DNA operativamente ligados:
- (d) um promotor que pode ser expresso em planta como definido na reivindicação 1b, 5, 6 ou 7;
- (e) uma região de DNA codificando uma proteína β -1,3-glicanase; e
- (f) uma região terminal 3' compreendendo sinais de término da transcrição e poliadenilação que funcionam em células da dita planta.
46. Método de acordo com a reivindicação 17, em que a dita região de DNA que codifica a dita proteína β -1,3-glicanase codifica uma sequência de aminoácidos que tem pelo menos 95% de identidade de sequência com a sequência de aminoácidos de SEQ ID N° 11 ou SEQ ID N° 12.
47. Gene quimérico como definido em qualquer uma das reivindicações 30 a 46.
48. Célula de uma planta que produz fibra compreendendo um gene quimérico como definido na reivindicação 47.
49. Planta que produz fibra, compreendendo um gene quimérico como definido na reivindicação 47.
50. Planta que produz fibra de acordo com a reivindicação 49, em que as fibras da dita planta têm o comprimento aumentado comparado com plantas de controle não-transformadas.
51. Planta que produz fibra de acordo com qualquer uma das

reivindicações 49 ou 50, em que a dita planta é algodão.

52. Planta de algodão de acordo com a reivindicação 51, que é *Gossypium hirsutum*.

53. Semente da planta que produz fibra como definido em qualquer uma das reivindicações 49 a 52, a dita semente compreendendo um gene quimérico como definido na reivindicação 47.

54. Fibras produzidas como definidas nos métodos de qualquer uma das reivindicações 30 a 46.

```

1  -----ttaaattataaatttaaaagtaaaattacatttttaacctaataaaaa-gagaaaaagtatatcta
1  gtcttgcacctaatggaaaaattgttatgttgattaaattataaaattataattttgacctataaaaaatgaaaaaatatatcta

66  attttcttcgaaaaatggaaagaaaaattataaaatt-----tatggcattttctaaaaaaattctgaaatttgatgaaattataataaaaaaaa
101  gtttcttcgaaaaatggaaagaaaaataaaattgatcacattataaaatttatggcattttctaaaaaaattctgaaatttgatgaaattataataaaaaaaa

142  a-----aattata-----aaatccgaagcattaccagaagatggatccacaaatcacaca
201  agtttaaaaacatatagatttcaagaatagtgggaaaattatatgtgaacaacactgaagaaatccaaagcattagcagaaaaatggatccacaaatcacaca

190  aacaatcaatgaaaaagtaataattgaaagtgaagcatttaattttgatagccataatacttcctgctgtaatttataggttctcattaaatgcaatt
301  aacaatcagtgaaaaagtaataattgaaagtgaagcatttaattttgatagccataatacttcctgctgtaatttataggttctcattaaatgcaatt

290  aaattatatcgcacaccttttgaaatgaaataaaatgacacaagaggaagcggttcattctattttt--ctttcaatcgcccatcaaaataccaaaaat
401  aaattatatgtgcacctttttgaaatgaaata-aatgacaca-----gttcattctatttttttcttctcaatcgcccatcaaaataccgaaaaat

388  gtaactacatgcaaaaa-atcaaatatgaaaaatatctcataatttttgataatttttaataatatgtgtgtgttcaaaaacgtaaaatgtattgaaaaatttatgatgg
488  gtaactacattaaaaaagatc-----gaaaaatatctcataatttttgataatttttaataagattgtgtgttcaaggcgta-atgtactaaaaaaatttatgatgg

487  tgttgtgtgtgtatgtccataaaaattcaatgtactcacatttatcaaatgtatacttttgagagaagttattttgataaatactcaagtttttttttatagat
581  tgttgtgcgtgtatgtccataaaaattcaatgtatttcgcatgtatcaaatgtaaaatttgacacaaagttattctataataataatcaagttattttttatacat

587  gggaaaaatttttaaaattatttttgatgtgaaatgtatatataaaattttaatttcgatacatataataatataatgtaaaaatttttaaaatttta
681  gagatacatctcaaaattattttt-----atataatccgaaaa-----

687  ataataacaaattaagaaaaataattttataaaatttttccgattaaaaataaattctggaaagaagaaatgtcaaacacttttttcaataaaatagtagat
718  -----atcataacgta-----cgatcaaa-----ctagaaaagaggaagtggtcaaaaacctatttcattatatgtcaaaatagat

787  gggacacgatacccttcattgcatttgatatctcaggtgtgccaataaaactcggaatcctttttgaaaaaaaac--ttccagagagagtataaaaatccagcag
784  gggacacgatacccttcattgcatttgatatctcatttatgtccaataaaactcagaatcctttttgaaaaaaaatccagagagagtgataaaaatccagcag

885  taggcacaagaaaaacgagcaccagttattgactttcctttgtaaaaaaaagagtgctgagatcaagaaatataagtgaaatatgggtccaagattttctg
884  tgtgcacaagaaaaacgagcaccagttattgacattcctttgtaaaaaaaagagtgatcaagaaatataagtgaaatatgggtccaacattttctg

984  ggtttttaaatctaagca
984  ggtttttaaatctcagca

```

Fig. 1

```

1 atgctgtttttaactcaactcctctcttaacagatggcgtgataatgggtgttttgctatgggtttgaacgggcaacaaattccatctccaggagatgttta
1 atggtgtttttaactcaactcctctctcttaacagatggcgtgataatgggtgttttgctatgggtttgaacgggcaacaaattccatctccaggagatgttta

101 ttaatcttttcaaaaactagtggcataaaacaaatatcagggtctaccagccttaccctgaagtgtctgaagcagcaagggtgatacctctctcgat
101 ttaatctttacaaaaactagtggcataaaacaaatatcagggtctaccagccttaccctgaagtgtctgaagcagcaagggtgatacctctctcgat

201 gagtacgacaaaacgagacatacaaaagcctcgaaacggatcaaaagtgcagccgatgcatgggttaacaccaacatcgtcccttataagggaagatgttcaa
201 gggccgagaaaacgagacatacaaaagcctcgaaagatcaaaagtgcagccgatgcatgggttaacaccaacatcgtcccttataagggaagatgttccag

301 ttcagggttcacatcatattgggaatgaagccattccaggacagtcaggctcttacattcctggtgccatgaacaacataaatgaaactcgtcggcctcatttg
301 ttcaagtgtgatacactattgggaatgaagccatttcaggacaaatcaagctcttacattcctgatagccatgaacaacataaatgaaactcgtcggcctcatttg

401 ggctaggcaacgaagggttacgaccgtgtgtcccgatgaatgccctaagtacacctccttcagacggcgcttttggaagcgatataaacatcgat
401 ggttaggcacgacgaagggttacgaccgtgtgtcccgatgaatgccctaagtacacctccttcagacggcgcttttggaagcgatataaacatcgat

501 catgactagtatcatggccattctgggttcgacaggattcggccctcctctgatcaaatgtgtacccttattttggcctatggcctcagaccccaactcataatttc
501 catgactagtatcatggccattctgggttcgacaggattcggccctcctctgatcaaatgtgtacccttattttggcctatggcctcagaccccaactcataatttc

601 ctcaactacgccttgttcacctcgaccgcacccggtgggtggtcgaccaaggcttggaatactacaacctctttgacggcgatgggtcgatgctttcaatggccg
601 ctcgattacgccttgttcacctcgaccgcacccggtgggtggtcgaccaaggcttggaatactacaacctctttgacggcgatgggtcgatgctttcaatggccg

701 ccctagataagatcggcttcggccaaattactctcattgttagccgaaactggatggccgaccgccggtaacgagccttacacgagtgctcgcgaacgctca
701 ccctagataagatcggcttcggccaaattactctcattgttagccgaaactggatggccgaccgccggtaacgagccttacacgagtgctcgcgaacgctca

801 aacttataacaagaacttgttgaaatcatgtgacgcagaaagggtactccgaaaaagacctgaaatataaatgccgacggtttttcttcgagatggttcaacgag
801 aacttataacaagaacttgttgaaatcatgtgacgcagaaagggtactccgaaaaagacctgaaatataaatgccgacggtttttcttcgagatggttcaacgag

901 aacttgaagcaacccacagttgagcagaaatttcggattcttcttccccaaatatgaacctgttttatccatttttgggtga
901 gatttgaagcaacccacagttgagcagaaatttcggattcttcttccccaaatatgaacctgttttatccatttttgggtga

```

Fig. 2

```
1 gtaaaacaaacttctctacagtgatcttacagtaaaatatggctttgaaaaatatacaacaaaaacatttatcttccaatccatttttaattactgatctacta
1 gtaaaacaaacttctctacagtgatcttacagtgatcttacggtaagtgatggctttgaaaaatatacaacaaaaacattta-----tactgatctacca
101 tatatggttgacag
83 tatatggttgacag
```

Fig. 3

1 mlf1tqlsltdgrdigvcyglngnnlpspgdvlnlftsginnirlyqpypevleaargsgislsmttnediqlatdqsaadawvntninvpykedvq
1 mvfl1tqlsltdgrdigvcyglngnnlpspgdvlnlftsginnirlyqpypevleaargsgislsmgprnediqlakdqsaadawvntninvpykddvq

101 frfiiigneaipegqsssyipgamnnimnslasfglgttkvttvvpnnalstsyppsdgafgsdi tsimtsimailvrqdspllinvpyfayasdpthis
101 fklitigneaisgqsssyipdamnnimnslalfglgttkvttvvpnnalstsyppsdgafgsdi tsimtsimailavqdspllinvpyfayasdpthis

201 lnyalftstapvvvdqgleyylnlfdgmvdafnaaldki fgqitli vaetgwptagnepytsvanaqtynknllnhvtqkgtpkpkrpeyimpptffemfne
201 ldyalftstapvvvdqgleyylnlfdgmvdafnaaldki fgqitli vaetgwptagnepytsvanaqtynknllnhvtqkgtpkpkrpeyimpptffemfne

301 nlkqptveqnf gfffpnmnpvypfw
301 dlkqptveqnf gfffpnmnpvypfw

Fig. 4

GaGluc1#-#Garb#leaves#gDNA CCGATCAAAGT
GhGluc1SGA#FiberMax966#Ghir#leaves#gDNA CCGATCAAAGT
GhGluc1SGD#FiberMax966#Ghir#leaves#gDNA AAGATCAAAGT
GrGluc1#-#Grai#leaves#gDNA AAGATCAAAGT

Fig. 5

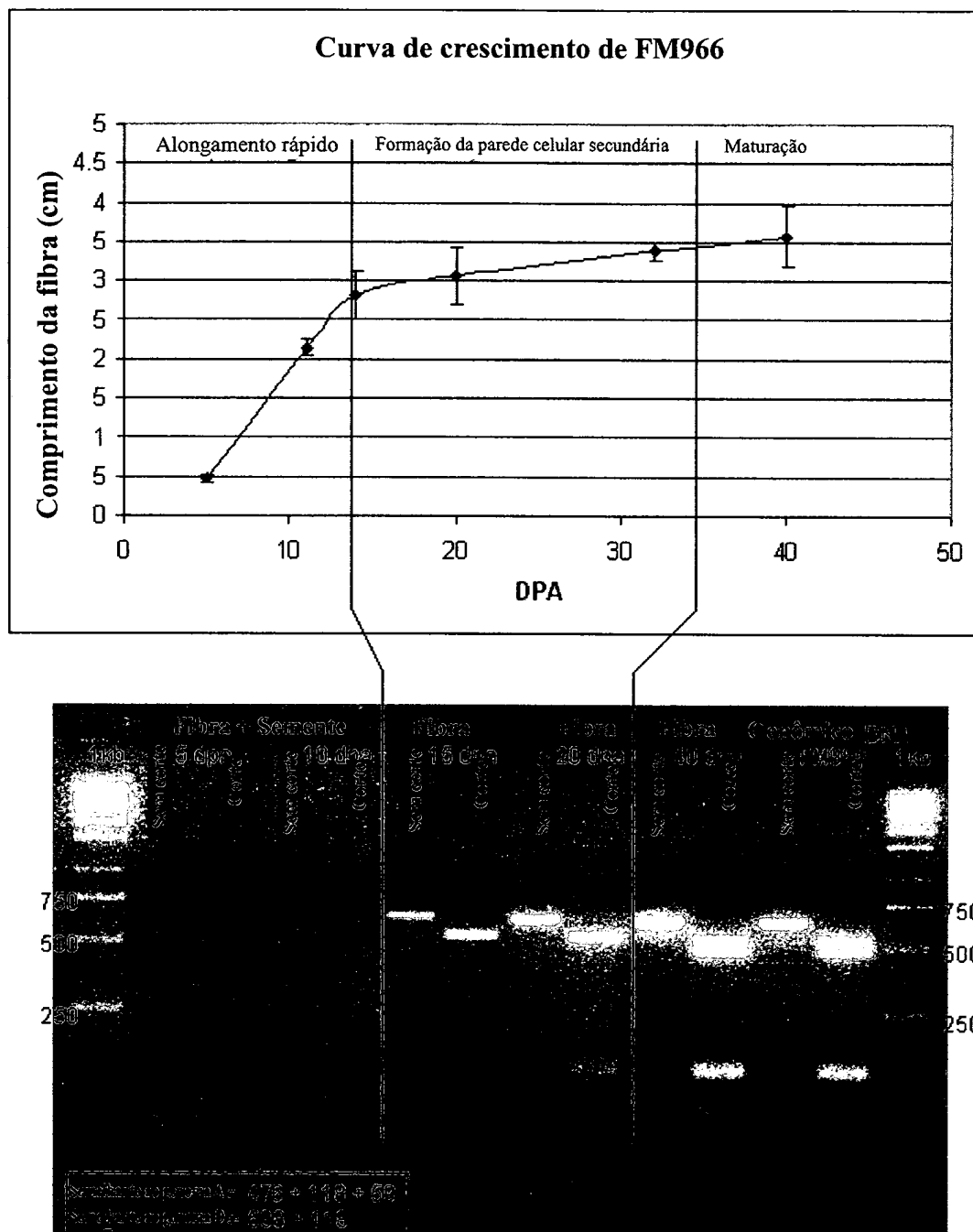


Fig. 6

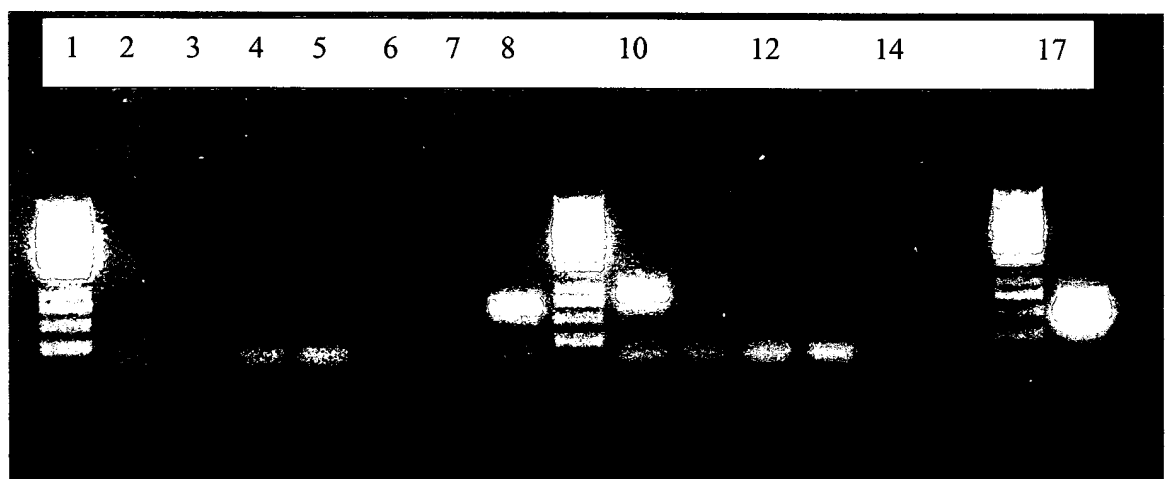


Fig. 7

RESUMO

Patente de invenção: **"EXPRESSÃO DIFERENCIAL DE ALELOS SUBGÊNOMICOS ESPECÍFICOS NO ALGODÃO E USOS DOS MESMOS"**.

- 5 A presente invenção refere-se a métodos e meios para alterar as propriedades da fibra tais como o comprimento da fibra em plantas que produzem fibra tal como o algodão. Adicionalmente, são providos promotores para plantas com um perfil de expressão preferencial para fibra ou seletiva para fibra.