



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

(11) PI 0808008-9 B1

(22) Data do Depósito: 15/02/2008

(45) Data de Concessão: 09/08/2016



* B R P I 0 8 0 8 0 0 8 B 1 *

(54) Título: MOLÉCULA DE ÁCIDO NUCLEICO ISOLADA, CASSETE DE EXPRESSÃO, VETOR DE EXPRESSÃO, MÉTODOS PARA EXCISÃO DE SEQUÊNCIAS ALVO DE UMA PLANTA, PARA PRODUZIR UMA PLANTA COM RENDIMENTO AUMENTADO, E/OU TOLERÂNCIA A ESTRESSE AUMENTADA, E/OU QUALIDADE NUTRITIONAL AUMENTADA, E/OU TEOR DE ÓLEO AUMENTADO OU MODIFICADO DE UMA SEMENTE OU BROTO PARA A PLANTA, E, USO DA MOLÉCULA DE ÁCIDO NUCLEICO

(51) Int.Cl.: C12N 15/82; A01H 5/00

(30) Prioridade Unionista: 16/02/2007 US 60/890217

(73) Titular(es): BASF PLANT SCIENCE GMBH

(72) Inventor(es): HEE-SOOK SONG, CHRISTIAN DAMMANN

“MOLÉCULA DE ÁCIDO NUCLEICO ISOLADA, CASSETE DE EXPRESSÃO, VETOR DE EXPRESSÃO, MÉTODOS PARA EXCISÃO DE SEQUÊNCIAS ALVO DE UMA PLANTA, PARA PRODUZIR UMA PLANTA COM RENDIMENTO AUMENTADO, E/OU TOLERÂNCIA A ESTRESSE AUMENTADA, E/OU QUALIDADE NUTRITIONAL AUMENTADA, E/OU TEOR DE ÓLEO AUMENTADO OU MODIFICADO DE UMA SEMENTE OU BROTO PARA A PLANTA, E, USO DA MOLÉCULA DE ÁCIDO NUCLEICO”

CAMPO DA INVENÇÃO

10 A presente invenção refere-se ao campo de biotecnologia agrícola. São descritas aqui as construções de expressão com especificidade de células para a germinação de embriões, plantas transgênicas compreendendo estas construções de expressão, e métodos para fazer e usar estas construções de DNA e plantas transgênicas.

15 FUNDAMENTOS DA INVENÇÃO

Em culturas de grãos de importância agronômica, a formação de semente é o objetivo final do desenvolvimento da planta. As sementes são colhidas para uso em alimentos, rações e produtos industriais. A utilidade e o valor destas sementes são determinados pela quantidade e qualidade de proteína, óleo, e amido contidos nas mesmas. Por sua vez, a qualidade e a quantidade de semente produzida podem ser afetadas por condições ambientais em qualquer ponto antes da fertilização até o amadurecimento da semente. Em particular, estresse em ou em torno do tempo de fertilização pode ter um impacto substancial sobre o desenvolvimento da semente. Os membros da família das gramíneas (Poaceae) que incluem os grãos de cereais, produzem frutos de uma semente, secos. Este tipo de fruta é, falando estritamente, uma cariopse mas é comumente chamado um cerne ou grão. A cariopse de um revestimento de fruta ou pericarpo circunda a semente e adere

firmemente a um revestimento de semente. A semente consiste de um embrião ou germe e um endosperma encerrado pela epiderme nucelar e um revestimento de semente. Conseqüentemente, o grão compreende a semente e seu revestimento ou pericarpo. A semente compreende o embrião e o endosperma.

Uma planta de milho fértil contém ambos os tecidos reprodutivos macho e fêmea, comumente conhecidos como o cabelo e a espiga, respectivamente. Os tecidos do cabelo formam os grãos de pólen haplóide com dois núcleos em cada grão que, quando abrigados como antese, contatam os fios da espiga fêmea. A espiga pode estar na mesma planta como a que abriga o pólen, ou em uma planta diferente. A célula de pólen desenvolve uma estrutura conhecida como tubo de pólen, que se estende através de um fio do cabelo fêmea individual para o óvulo. Os dois núcleos machos andam através deste tubo para alcançar o ovo fêmea haplóide na base do fio de cabelo. Um destes núcleos machos funde com e fertiliza os núcleos de ovos haplóides fêmeas para formar o zigoto, que é diplóide em número de cromossomos e irá se tornar o embrião dentro do cerne. O núcleo macho restante funde com e fertiliza um segundo núcleo fêmea para formar o núcleo do endosperma primário, que é triplóide em número e irá se tornar o endosperma do cerne, ou semente, da planta de milho. Os óvulos não fertilizados não produzem cernes e os tecidos não fertilizados eventualmente degeneram.

O cerne consiste de várias partes, algumas derivadas de tecido maternal e outras do processo de fertilização. Maternalmente, o cerne herda vários tecidos, incluindo um pericarpo circundante, protetor, e um pedicelo. O pedicelo é um tecido tipo talo curto, que fixa o cerne ao sabugo e provê transferência de nutrientes do tecido maternal para o cerne. O cerne contém tecidos resultantes das atividades de fertilização, incluindo o novo embrião assim como o endosperma. O embrião é um progenitor em miniatura da

próxima geração, contendo células para o crescimento de raiz e broto de uma nova planta jovem de milho. Ele é também um tecido em que óleos e proteínas são armazenadas no cerne. O endosperma funciona mais como um tecido nutritivo e provê a energia na forma de amido armazenado, proteínas e óleo, necessários para a germinação e crescimento inicial do embrião.

Considerando a regulação complexa que ocorre durante o desenvolvimento do embrião e do cerne em plantas superiores, e considerando que este é comumente grão que é a fonte primária de nutrição para humanos e animais, ferramentas chave necessárias para melhorar esta fonte nutricional incluem promotores genéticos que podem dirigir a expressão de genes melhoradores da nutrição. Por outro lado, o embrião é altamente sensível para estresses. Estresses nas plantas podem ser causados por tanto agentes bióticos como abióticos. Por exemplo, as causas bióticas de estresse incluem infecção com um patógeno, alimentação de inseto, e parasitismo por outra planta, como visco, e pasto por animais ruminantes. Os estresses abióticos incluem, por exemplo, água disponível excessiva ou insuficiente, luz insuficiente, extremos de temperatura, produtos químicos sintéticos como herbicidas, vento excessivo, extremos de pH do solo, disponibilidade limitada de nutrientes, e poluição do ar. Ainda, as plantas sobrevivem e com frequência florescem, mesmo sob condições desfavoráveis, usando vários mecanismos internos e externos para evitar ou tolerar estresse. As respostas fisiológicas das plantas ao estresse refletem mudanças na expressão de genes.

Enquanto a manipulação de genes induzidos por estresse pode desempenhar um papel importante na melhora da tolerância da planta aos estresses, foi mostrado que a expressão constitutiva de genes indutíveis por estresses tem um impacto negativo severo sobre o crescimento e desenvolvimento da planta quando o estresse não está presente (Kasuga 1999). Assim, existe a necessidade na técnica para promotores dirigindo a expressão que é diferenciada de modo temporal e/ou espacial, para

proporcionar um meio para controlar e dirigir a expressão do gene em células ou tecidos específicos em tempos críticos, especialmente para prover tolerância ou evitação ao estresse. Particularmente, estresses de seca e/ou densidade de milho com frequência resultam em rendimento reduzido. Para estabilizar o desenvolvimento de plantas e rendimento de grão sob meios desfavoráveis, manipulação de hormônios e suprimento nutricional para embrião (eixo e escutelo) durante a germinação de sementes é de interesse. Assim, existe uma necessidade para seqüências de regulação da transcrição que dirigem a expressão do gene em embrião ou escutelo sob condições de estresse abiótico.

Um outro problema bem conhecido na técnica de biotecnologia de plantas é a deleção de marcador. O marcador selecionável é utilizável durante o processo de transformação para selecionar, e identificar, organismos transformados, mas tipicamente não provê uma função utilizável uma vez que o organismo transformado foi identificado e contribui substancialmente para a falta de aceitação destes produtos de "alimento genético" dentre os consumidores (Kuiper 2001), e poucos marcadores são disponíveis que não são baseados nestes mecanismos (Hare 2002). Assim, existem tentativas múltiplas para desenvolver técnicas por meio das quais o DNA marcador possa ser excisado do genoma da planta (Ow 1995; Gleave 1999). O versado na técnica está familiarizado com vários sistemas para a remoção sítio-dirigida de seqüências de ácido nucleico recombinantemente introduzidas. Eles são principalmente baseados no uso de recombinases específicas da seqüência. Vários sistemas de recombinação específicos de seqüência são descritos, como o sistema Cre/lox de bacteriófago P1 (Dale 1991; Russell 1992; Osborne 1995), o sistema FLP/FRT de levedura (Kilby 1995; Lyznik1996), uma recombinase Gin de fago Mu, uma recombinase Pin de *E. coli*, o sistema R/RS do plasmídeo pSR1 (Onouchi1995; Sugita2000), o sistema attP/bacteriófago Lambda (Zubko 2000). É uma

desvantagem conhecida destes métodos bem conhecidos na técnica anterior que a excisão não é homogênea através das plantas completas, assim levando a padrões de excisão de tipo mosaico, que requerem ciclos adicionais laboriosos de seleção e regeneração.

5 Os promotores que conferem expressão melhorada durante o amadurecimento da semente ou grão são descritos (como os promotores de hordeína de cevada, ver e pedido de patente US 20040088754). Os promotores que dirigem a expressão específica de embrião ou específica de semente em dicotiledôneas (por exemplo, promotor conglucina de soja;
10 Chen 1988; o promotor de napina, Kridl 1991) não são em geral capazes de dirigir uma expressão similar em monocotiledôneas. Infelizmente, relativamente poucos promotores especificamente dirigindo este aspecto da fisiologia foram identificados (ver por exemplo US20040163144).

Porque os promotores específicos de semente ou grão, que são
15 descritos, incluem os associados com genes codificando proteínas de armazenamento de semente de planta, como genes codificando: hordeínas de cevada, glutelinas de arroz, orizinas, prolaminas, ou globulinas; gliadinas ou gluteninas de trigo; zeínas ou glutelinas de milho; glutelinas de aveia; kafirinas de sorgo, penisetinas de milho miúdo, ou secalinas de centeio. No
20 entanto, por um lado, a expressão destes promotores é com frequência danificada ou de nível de expressão baixo. Além disso, foi notado que a melhora das plantas de cultura com transgenes múltiplos ("empilhamento" é de interesse crescente. Por exemplo, um único híbrido de milho pode compreender construções de RNA recombinante conferindo não somente
25 resistência a insetos, mas também resistência a um herbicida específico. Importante, as sequências regulatórias apropriadas são necessárias para dirigir a expressão desejada de cada um destes ou outros transgenes de interesse. Além disso, é importante que os elementos regulatórios sejam distintos um de cada outro. Preocupações associadas com a utilização de sequências

regulatórias similares para dirigir a expressão de genes múltiplos incluem, mas não são limitadas a: (a) formação de pares ao longo de regiões homólogas, troca genética e perda da região interveniente ou dentro de um plasmídeo antes da integração, ou dentro do genoma da planta, pós-integração, (b) laços de grampo de cabelo causados por duas cópias da sequência em orientação oposta um adjacente ao outro, novamente com possibilidades de excisão e perda destas regiões regulatórias; (c) competição dentre cópias diferentes da mesma região de promotor para ligação de fatores de transcrição específicos de promotor ou outras proteínas de ligação a DNA regulatório.

Assim, existe uma grande necessidade na técnica para a identificação de novas sequências que possam ser usadas para expressão de transgenes selecionados em plantas economicamente importantes, especialmente em plantas monocotiledôneas. Também, existe a necessidade na técnica sequências regulando a transcrição que permitam a expressão em embrião ou escutelo durante a semente germinando antes. Assim, é um objetivo da presente invenção prover cassetes de expressão novos e alternativos para a expressão preferencial ou específica de embrião. O objetivo é resolvido pela presente invenção.

SUMÁRIO DA INVENÇÃO

A presente invenção refere-se a uma molécula de ácido nucleico isolada compreendendo uma sequência reguladora da transcrição de plantas, em que a sequência reguladora da transcrição compreende

i) uma primeira sequência de ácido nucleico compreendendo a sequência de promotor de um gene regulado para seca, respondente ao frio e/ou ABA, por exemplo um gene de planta gene rd29 respondente à dessecação ou proteína 78 induzida por baixa temperatura como definido na Figura 4 ou um equivalente funcional ou um homólogo do mesmo, (a seguir “promotor cor78”),

e operativamente ligada à mesma

ii) uma segunda sequência de ácido nucleico compreendendo o primeiro intron de um gene de planta codificando a Metalotioneína 1 (a seguir “MET1”) como definido na Fig. 5 ou um equivalente funcional ou um homólogo do mesmo (a seguir “intron MET1”).

A sequência de promotor de um gene regulado para seca, respondente ao frio e/ou ABA, compreende promotores de uma classe bem conhecida de proteínas. Preferivelmente, ela é uma sequência de promotor de um gene regulado para seca, respondente ao frio e/ou ABA. Exemplos de genes controlados por gene regulado para seca, respondente ao frio e/ou ABA são rd29A, rd29B, lti, ou cor78.

A sequência da Figura 4 sequência representa RD29A e proteína 78 induzida por baixa temperatura. O promotor cor78 é também conhecido como sequência de promotor rd29. BLAST com base de dados Genbank refere-se por exemplo ao número de acesso Genbank AB019226, que compreende outros genes. Nesta sequência do genoma, o promotor cor78 reside bp11743 a bp12328 em direção 5’-3’. O gene que reside a jusante desta região do promotor é proteína 78 induzida por baixa temperatura (por exemplo bp 12650 a 12698, 12784 a 12966, 13063 a 13536, e 13621 a 15047).

O gene de planta RD29 respondente à dessecação é por exemplo publicado sob Acesso NM_001036984 com a 2131 bp mRNA, datado PLN 09-JUN-2006, VERSÃO NM_001036984.1 GI:79330663 e é descrito como um gene que codifica uma proteína que é induzida em expressão em resposta a falta d'água, como frio, elevado teor de sal e dessecação. A resposta parece ser via ácido abscísico. A região de promotor contém dois elementos respondentes a ABA (ABREs) que são requeridos para a expressão respondente a desidratação de rd29B como elementos cis-atuantes. Proteína é um membro de uma família de genes com outros membros encontrados em plantas, animais e fungos. Ela é descrita como

similar a proteína respondente a relacionada com estresse [Arabidopsis thaliana] (TAIR:At4g25580,1). RD29 é também publicado como cor78 sob número de Acesso GB:AAA32776.1.

Assim, em uma forma de realização, o promotor controlando a transcrição de um mRNA codificando uma das referidas proteínas em uma planta é usado, por exemplo o promotor de um gene codificando uma proteína respondente a estresse, por exemplo de Arabidopsis thaliana (TAIR:At4g25580,1), por exemplo rd29A, rd29B, lti ou cor78 (GB:AAA32776.1). Preferivelmente, o promotor tem a sequência como mostrada em Figura 4 ou de um homólogo da mesma, por exemplo de um ortólogo.

Preferivelmente, o promotor cor78 é derivado de uma planta dicotiledônea. Em uma forma de realização o homólogo é um homólogo previsto como ortólogo encontrado em Viridiplantae; preferivelmente encontrado em Streptophyta, mais preferido em Embryophyta; Tracheophyta, ainda mais preferido encontrado em Spermatophyta. Mais preferido o homólogo é identificado de Magnoliophyta, mais preferido é a forma eudicotilédones, ainda mais preferido o homólogo é de eudicotilédones de núcleo. Assim, ele é preferivelmente de rosídeos; mais preferido de eurosídeos II; ainda mais preferido de Brassicales, ainda mais preferido o homólogo é por exemplo de Brassicaceae, preferivelmente de Arabidopsis.

Em uma forma de realização o homólogo é de um cultivar de Arabidopsis thaliana, por exemplo de C24, ou de um ecotipo como "Columbia"

Uma outra indicação de que dois polipeptídeos são substancialmente similares um com relação ao outro, além de terem substancialmente a mesma função, é que um agente, por exemplo, um anticorpo, que especificamente liga a um dos polipeptídeos, também especificamente liga ao outro. Assim, em uma forma de realização, o

homólogo de RD29 é uma proteína que pode ser identificada em um teste Western Blot por ligação a um anticorpo monoclonal gerado para a ligação específica a uma proteína tendo o aminoácido como mostrado em Fig. 4, respectivamente.

5 Em uma forma de realização, a percentagem do gene cujo promotor é usado é um ácido nucleico ou uma proteína com uma sequência de aminoácido ou identidade de ácido nucleico de pelo menos 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, por exemplo, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%,
10 87%, 88%, 89%, e ainda 90% ou mais, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo menos 99% para a sequência de aminoácidos ou a sequência de ácido nucleico como mostrado em Fig. 4.

 Preferivelmente o gene, cujo promotor é usado, é um gene mostrando o padrão de expressão de uma proteína 78 induzida por baixa
15 temperatura ou um gene RD29. Preferivelmente a sequência de promotor conectada ao referido gene é 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, por exemplo, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, e ainda 90% ou mais, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo
20 menos 99% para a sequência de ácido nucleico como descrito por SEQ ID NO.: 1. Referidos promotores podem ser obtidos usando pelo menos 10, 15, 20, 30 ou mais nucleotídeos consecutivos de SEQ ID NO.: 1, 5 ou 6 ou de uma proteína 78 induzida por baixa temperatura ou um gene RD29, por exemplo como mostrado em Figura s 4a a 4c usando técnicas padrões para a
25 identificação ou clonagem de sequências de ácido nucleico como, mas não excluindo outros, buscas em bases de dados, hibridização ou técnicas com base em PCR.

 O gene Met1 é publicado em Sasaki, T., et al., natureza 420 (6913), 312-316 (2002) e sob Acesso número AP002540 com um DNA de

249 bp, linear, PLN 19-OCT-2004, VERSÃO AP002540.2 GI:13872872.

A sequência foi apresentada em 21-JUN-2000 e em 27 de abril, 2001, esta versão de sequência substituiu gi:8698578. Genes foram previstos a partir dos resultados integrados dos seguintes GENSCAN, 5 FGENESH, GeneMark.hmm, GlimmerM, RiceHMM, SplicePredictor, sim4, gap2, BLASTN e BLASTX. A sequência genômica foi pesquisada contra a base de dados NCBI NonRedundant Protein database, nr e a base de dados de sequência de cDNA em RGP ou DDBJ. Homologias de proteína das regiões de codificação foram pesquisadas contra bases de dados NCBI NonRedundant 10 Protein database com BLASTP. ESTs representa as sequências de cDNA identificadas usando BLASTN com o correspondente número de acesso DDBJ e RGP clone ID. Os cDNAs de comprimento completo representam as sequências de cDNA identificadas usando BLASTN com o número de acesso correspondente DDBJ. Um gene com identidade ou homologia significativa 15 para uma proteína é classificado com base no nome da proteína para indicar o nível de homologia, como mesmo nome, "proteína putativa" e "proteína tipo-". Um gene sem homologia significativa para qualquer proteína mas com homologia EST ou de cDNA de comprimento completo (cobrindo quase todo o comprimento da sequência parcial) é classificado como uma proteína 20 'desconhecida'. Um gene previsto por dois ou mais programas de previsão de genes é classificado como uma proteína "hipotética" de acordo com o padrão IRGSP. Um gene previsto por um programa de previsão de gene único é também classificado como uma proteína "hipotética" e é incluído como um aspecto miscelâneo da sequência. A orientação da sequência é de T7 a SP6 do 25 clone PAC. Esta sequência de clone P0434B04 tem uma sobreposição com P0416D03(DDBJ: AP002872) clone a extremidade 5' e com P0009G03(DDBJ: clone AP002522) em extremidade 3'. A informação detalhada sobre sobreposição e qualidade de montagem junto com a anotação de sua entrada é disponível em GenomaSeq.

Preferivelmente, um intron do gene MET1 como mostrado em Fig. 5 ou um homólogo do mesmo é usado, por exemplo de um ortólogo.

Preferivelmente, referido primeiro intron é derivado de um gene MET1 de uma planta monocotiledônea. O homólogo é preferivelmente um ortólogo, por exemplo de Viridiplantae; por exemplo de Streptophyta, mais preferido de Embryophyta, ainda mais preferido de Tracheophyta. O homólogo é por exemplo de Spermatophyta, preferivelmente de Magnoliophyta, mais preferido de Liliopsida, ainda mais preferido Poales, ainda mais preferido de Poaceae. Em uma forma de realização, o homólogo é de BEP, clade, por exemplo de Ehrhartoideae, mais preferido de Oryzeae, por exemplo de Oryza.

Em uma forma de realização, um ortólogo é usado, que é derivado de *Oryza sativa*, por exemplo de um cultivar como o grupo de cultivar japonica.

Preferivelmente o homólogo é um ortólogo e mostra substancialmente o mesmo fenótipo em um teste de depleção como o tipo selvagem.

Uma outra indicação de que dois polipeptídeos são substancialmente similares um ao outro, além de terem substancialmente a mesma função, é que um agente, por exemplo, um anticorpo, que especificamente liga a um dos polipeptídeos, também especificamente liga ao outro. Assim, em uma forma de realização, o homólogo de MET1 é uma proteína que pode ser identificada em um teste Western Blot por ligação a um anticorpo monoclonal gerado para ligação específica a uma proteína tendo o aminoácido como mostrado em Fig. 5.

Em uma forma de realização, o homólogo percentual de um gene cujo intron é usado, por exemplo o primeiro intron, está codificando a proteína com uma sequência de aminoácido ou identidade de ácido nucleico de pelo menos 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%,

70%, por exemplo, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, e ainda 90% ou mais, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo menos 99% para a sequência de aminoácidos como mostrado em Fig. 5.

5 Preferivelmente, o homólogo de um gene cujo intron é usado está codificando a proteína mostrando o padrão de expressão do gene MET1. Preferivelmente a sequência de ácido nucleico do intron de referido gene é 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, por exemplo, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%,
10 84%, 85%, 86%, 87%, 88%, 89%, e ainda 90% ou mais, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo menos 99% para a sequência de ácido nucleico como descrito por SEQ ID NO.: 3. Referidos promotores podem ser obtidos usando pelo menos 10, 15, 20, 30 ou mais nucleotídeos consecutivos de SEQ ID NO.: 3, 7 ou 8 ou de um gene MET1, por exemplo
15 como mostrado em Figura 5 em técnicas padrões para identificação ou clonagem de sequências de ácido nucleico como, mas não excluindo outros, buscas em bases de dados, hibridização, ou técnicas com base em PCR.

As comparações de sequência podem ser realizadas usando um algoritmo de alinhamento de sequência Smith-Waterman (ver por exemplo,
20 Waterman 1995). O programa localS, versão 1.16, é preferivelmente usado com seguintes parâmetros: correspondência: 1, penalidade de falta de correspondência: 0.33, penalidade de espaço aberto: 2, penalidade de espaço estendido: 2.

A distância entre o promotor cor78 e o intron, por exemplo o
25 primeiro intron de MET1 é preferivelmente curta. Em uma forma de realização, a molécula de ácido nucleico compreende por exemplo uma sequência de ligador que está localizada entre o promotor cor78 e o primeiro nucleotídeo de referido intron com por exemplo um comprimento de 0 bp a 100bp, por exemplo 10bp a 90 bp, ou 20 a 80bp, por exemplo 20, 30, 40, 50,

60, 70, 80 ou 90 bp.

Em uma outra forma de realização da presente invenção, referido intron, por exemplo o primeiro intron de MET1, está localizado na sequência de um outro intron de uma sequência nucleotídica transcrita sob o controle de sequência nucleotídica reguladora da transcrição.

Além disso, em uma forma de realização, referida molécula de ácido nucleico compreende por exemplo um 5'UTR que está localizado entre a sequência cor78 de promotor e o primeiro nucleotídeo do intron MET1.

Um possível arranjo é representado por pares de base 8199 a 9656 de SEQ ID NO: 2, compreendendo o promotor cor79 e o primeiro intron do gene MET1, ou por pares de base 8134 a 9656, compreendendo o promotor cor79, o primeiro intron do gene MET1 e o 5'UTR.

O intron da sequência nucleotídica transcrita sob o controle de sequência nucleotídica reguladora da transcrição está por exemplo localizado em 5'UTR ou em outro local próximo de uma sequência cor78 de promotor, por exemplo ele está localizado dentro dos primeiros 100bp da extremidade 5' de uma região de codificação ou dentro dos primeiros 100bp após o códon de início de transcrição.

Conseqüentemente, em uma forma de realização, a invenção refere-se a uma molécula de ácido nucleico isolada compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição de plantas compreendendo

i) um primeiro ácido nucleico selecionado dentre o grupo consistindo de

- a) um polinucleotídeo como definido em SEQ ID NO:1;
- b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:1;
- c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente

200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e

ii) um segundo ácido nucleico selecionado dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:3;

b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:3;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:3; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento da mesma,

em que referida primeira e referida segunda sequências de ácido nucleico são funcionalmente ligadas e heterólogas a cada outra.

Preferivelmente, a molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição é capaz de mediar a transcrição de uma sequência de ácido nucleico operativamente ligada em uma planta ou célula de planta. Mais preferivelmente, a molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição é capaz de mediar transcrição de uma sequência de ácido nucleico em uma planta monocotiledônea ou célula de planta monocotiledônea. Prefere-se que a

molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição seja específica de tecido. Mais preferivelmente, a molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição é preferivelmente expressada em embrião ou escutélion. Prefere-se também que a molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição seja preferivelmente expressada sob condições de estresse abiótico ou biótico. As condições de estresse biótico são selecionadas dentre o grupo consistindo de fungos, nematódeos, inseto, vírus, e bactérias e combinações dos mesmo. As condições de estresse abiótico são selecionadas dentre o grupo consistindo de seca, frio, calor, sal, salinidade, densidade de população de planta elevada, nitrogênio, luz UV, e combinações dos mesmos. O mais preferivelmente, a molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição é preferivelmente expressada sob condição de seca.

Outra forma de realização da presente invenção refere-se a uma molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição como descrito acima, em que o ácido nucleico compreende pelo menos dois, preferivelmente mais, por exemplo três, quatro, cinco, seis, sete, oito, nove, dez ou mais, por exemplo até todos, motivos de promotor de núcleo selecionados dentre o grupo consistindo de sequências como definidos na SEQ ID NOs: 10, 12, 14, 17, 20, 22, 25, 27, 29, 31, 33, 35, 37, 39, 42, 44, 46, 48, 50, 53, 55, 57, 59, 63, 66, 68, 70, 72, 74, 77, 79, 82, 84, 86, 88, 90, 94, 96, 98, 102, 104, 108, 112, 114, 117, 121, 123, 125, 127, 129, 131, 137, e 138. Preferivelmente referidos motivos de promotor são dispostos sobre o filamento negativo ou positivo de uma sequência reguladora da transcrição como descrito em tabela 8 e 9. Os referidos motivos de promotor ainda mais preferidos são posicionados na mesma ou similar posição na sequência reguladora da transcrição como

descrito em tabela 8 e 9.

Ainda outra forma de realização da presente invenção refere-se a uma molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição como descrito acima, em que o ácido nucleico compreende pelo menos dois, preferivelmente mais, por exemplo três, quatro, cinco, seis, sete, oito, nove, dez ou mais, por exemplo até todos os motivos de promotor selecionados dentre o grupo consistindo das sequências como definidas em SEQ ID NOs: 9, 11, 13, 15, 16, 18, 19, 21, 23, 24, 26, 28, 30, 32, 34, 36, 38, 40, 41, 43, 45, 47, 49, 51, 52, 54, 56, 58, 60, 61, 62, 64, 65, 67, 69, 71, 73, 75, 76, 78, 80, 81, 83, 85, 87, 89, 91, 92, 93, 95, 97, 99, 100, 101, 103, 105, 106, 107, 109, 110, 111, 113, 115, 116, 118, 119, 120, 122, 124, 126, 128, 130, 132, 133, 134, 135, e 136. Preferivelmente referidos motivos de promotor são dispostos sobre o filamento negativo ou positivo de uma sequência reguladora da transcrição como descrito em tabela 8 e 9. Ainda mais preferido referidos motivos de promotor estão posicionados na mesma ou similar posição em uma sequência reguladora da transcrição como descrito em tabelas 8 e 9.

Outra forma de realização da presente invenção refere-se a uma construção de expressão compreendendo uma molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição funcionalmente ligada à mesma uma sequência de ácido nucleico. Preferivelmente a sequência de ácido nucleico funcionalmente ligada confere a uma planta um traço ou propriedade selecionado dentre o grupo consistindo de rendimento aumentado, resistência aumentada sob condições de estresse, qualidade nutricional e/ou teor de óleo aumentados de uma semente ou um broto, e excisão do marcador de seleção. A qualidade nutricional e/ou teor de óleo aumentados podem compreender um teor aumentado de pelo menos um composto selecionado dentre o grupo consistindo de vitaminas, carotinóides, antioxidantes, ácidos graxos

insaturados, ácidos graxos poli-insaturados, ou proteínas com teor alterado de aminoácido. Prefere-se também que a transcrição da sequência de ácido nucleico funcionalmente ligada na construção de expressão resulta em expressão da proteína ou expressão de uma sequência ribonucleotídica funcional capaz de conferir função de pelo menos um gene na planta alvo. O RNA funcional compreende pelo menos um dentre o grupo consistindo de : RNA anti-sentido, RNA sentido, dsRNA, microRNA, ta-siRNA, snRNA, RNAi, ou combinações dos mesmos.

A forma de realização da presente invenção provê uma planta ou uma semente produzida por uma planta transgênica transformada com uma molécula de ácido nucleico compreendendo um polinucleotídeo codificando a sequência reguladora da transcrição funcionalmente ligada ao ácido nucleico. Em uma outra forma de realização preferida, a semente produzida pela planta transgênica expressa a sequência de proteína ou uma RNA funcional capaz de conferir função de pelo menos um gene na planta alvo, em que a semente ou planta tem resistência aumentada sob condições de estresse e/ou rendimento aumentado, e/ou qualidade nutricional e/ou teor de óleo aumentados de uma semente ou de um broto. Mais preferivelmente, a semente ou planta é uma monocotiledônea. Preferivelmente, a semente ou planta é selecionada dentre o grupo consistindo de milho, trigo, arroz centeio, aveia, centeio, sorgo, erva-castelhana ou coix. Mais preferivelmente, a semente ou planta é uma planta de cereal selecionada dentre o grupo consistindo de milho, trigo, centeio, arroz, aveia, centeio, e sorgo, ainda mais preferivelmente de milho, trigo, e arroz, o mais preferivelmente a semente ou planta é milho. Outras formas de realização da invenção referem-se a sementes, partes e células da planta monocotiledônea da invenção. Preferivelmente, as partes de plantas selecionadas dentre o grupo consistindo de: células, protoplastos, culturas de tecido de células, calos, grupos de células, embriões, pólen, óvulos, sementes, flores, grãos, espigas, sabugos, folhas, cascas, caules, raízes, pontas de raízes,

anteros, e fios de cabelo.

Outra forma de realização da invenção refere-se a um método para o rendimento aumentado e/ou tolerância ao estresse aumentada em uma planta, em que o método compreende as etapas de

5 A) introduzir em uma planta uma construção de expressão compreendendo

i) uma primeira sequência de ácido nucleico selecionado dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:1;

10 b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:1;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

15 d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo,
20 e operativamente ligado ao mesmo a

ii) um segundo ácido nucleico selecionado dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:3;

25 b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:3;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:3; e

d) um polinucleotídeo hibridizando sob condições equivalentes

a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

5 e operativamente ligado a pelo menos um ácido nucleico que é heterólogo em relação à primeira ou segunda sequência de ácido nucleico e é capaz de conferir a uma planta um rendimento aumentado e/ou tolerância ao estresse aumentada, e

10 B) selecionar plantas transgênicas, em que as plantas tem rendimento aumentado e/ou tolerância ao estresse aumentada sob condições de estresse como comparado com as plantas de tipo selvagem ou segregante nulo.

Vários ácidos nucleicos são conhecidos do versado na técnica para obter rendimento e/ou resistência a estresse. Os ácidos nucleicos podem 15 incluir, mas não são limitados a, polinucleotídeos codificando um polipeptídeo envolvido em biossíntese de fitohormônios, regulação de fitohormônios, regulação de ciclo celular, ou metabolismo de carboidratos.

Outra forma de realização da invenção refere-se a um método para conferir qualidade nutricional e/ou teor de óleo aumentados de uma 20 semente ou um broto a uma planta, em que o método compreende as etapas de

A) introduzir em uma planta uma construção de expressão compreendendo

i) um primeiro ácido nucleico selecionado dentre o grupo consistindo de

25 a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:1;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente

200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo,

e operativamente ligado a

ii) um segundo ácido nucleico selecionado dentre o grupo consistindo de

10 a) um polinucleotídeo como definido em SEQ ID NO:3;

b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:3;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente

15 200 bases consecutivas do polinucleotídeo de SEQ ID NO:3; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

e operativamente ligado a pelo menos um ácido nucleico que é heterólogo em relação à referida primeira ou referida segunda sequência de ácido nucleico e é apropriado para conferir a uma planta uma qualidade nutricional e/ou teor de óleo aumentados de uma semente ou um broto, e

25 B) selecionar plantas transgênicas, em que as plantas têm qualidade nutricional e/ou teor de óleo aumentados de uma semente ou um broto como comparado com as plantas de tipo selvagem ou segregante nulo.

A qualidade nutricional e/ou teor de óleo são definidos como acima. Os exemplos mais específicos são dados aqui abaixo. A sequência

reguladora da transcrição é descrita acima, mais preferivelmente a sequência reguladora da transcrição é específica do embrião.

Ainda outra forma de realização da invenção refere-se a um método para excisão de uma sequência alvo, por exemplo uma sequência de marcador de uma planta, referido método compreendendo as etapas de

A) construção de um cassete de expressão por operativamente ligando uma sequência nucleotídica reguladora da transcrição compreendendo

i) um primeiro ácido nucleico selecionado dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:1;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e

ii) um segundo ácido nucleico selecionado dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:3;

b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:3;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:3; e

d) um polinucleotídeo hibridizando sob condições equivalentes

a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

5 e operativamente ligados a pelo menos um ácido nucleico que é heterólogo em relação à referida primeira ou referida segunda sequência de ácido nucleico e é apropriado para induzir excisão de sequência alvo, por exemplo uma sequência de marcador de uma planta, e

10 B) inserir referido cassete de expressão dentro de uma planta compreendendo pelo menos uma sequência alvo, por exemplo uma sequência de marcador para prover uma planta transgênica, em que referida planta expressa referido sequência de ácido nucleico heteróloga, e

C) selecionar plantas transgênicas, que demonstram excisão de referida sequência alvo, por exemplo da sequência de marcador.

15 Preferivelmente a sequência alvo é uma sequência de marcador, por exemplo um gene de resistência a antibiótico ou herbicida.

Em uma forma de realização preferida da invenção, a sequência nucleotídica expressada a partir da sequência reguladora da transcrição quimérica da invenção não está codificando uma beta-glucuronidase (GUS), ou não é um método para a expressão de um gene GUS
20 com o fim de alcançar a coloração mediando GUS.

BREVE DESCRIÇÃO DOS DESENHOS

Fig. 1. Mapa de plasmídeo pBPSMM368:At.Cor78::BPS1.1::GUS::NOS

25 Fig. 2. Coloração GUS histoquímica de construções de promotor At.Cor78. São mostrados exemplos representativos para os padrões de coloração típicos das construções. A: At.Cor78::BPS1.1::GUS::NOS (pBPSMM368); B: At.Cor78::GUS::NOS (pBPSMM250); C: At.Cor78::Ubi-intron1::GUS::NOS (pBPSMM346)

Fig. 3. Expressão induzida por estresse de seca ou estável controlada por construção de promotor pBPSMM346 em milho. Plantas transgênicas em estágio de cinco folhas foram submetidas a estresse de seca retirando a água. As amostras foram retiradas das folhas nos pontos de tempo indicados. RNA foi isolado das amostras de folha e analisado com RT-PCR quantitativo. A expressão GUS foi normalizada contra um gene de controle interno em cada amostra. Resultados são mostrados como aumento de vezes de níveis de expressão comparados com o ponto de tempo 0 que é fixado como 1.

Fig. 4a a 4c: Sequência nucleotídica e sequência aminoácido do gene RD29A e o gene de proteína 78 induzida por baixa temperatura.

Fig. 5: Sequência nucleotídica do gene MET1.

DEFINIÇÕES

Salvo especificado em contrário, os termos técnicos são usados de acordo com o uso convencional. As definições de termos comuns em biologia molecular podem ser encontradas em Lewin, Genes V publicado por Oxford University Press, 1994 (ISBN 0-19-854187-9); Kendrew et al (eds.), The Encyclopedia of Molecular Biology, publicado por Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); e Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, publicado por VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

Deve-se entender que esta invenção não é limitada à metodologia particular, protocolos, linhagens de células, espécies ou gêneros de plantas, construções, e reagentes descritos como tal. Também se deve entender que a terminologia usada aqui é para fins de descrever formas de realização particulares apenas, e não destinada a limitar o escopo da presente invenção, que será limitado somente pelas reivindicações anexas. Deve-se notar que como usado aqui e nas reivindicações anexas, as formas singulares "o" e "um" incluem formas de referência no plural, salvo especificado em

contrário no contexto. Assim, por exemplo, referência a "um vetor" é uma referência a um ou mais vetores e inclui equivalentes dos mesmos conhecidos do versado na técnica, e assim em diante.

O termo "cerca de" é usado aqui para significar aproximadamente, grosseiramente, em torno de, ou na região de. Quando o termo "cerca de" é usado em conjunto com uma faixa numérica, ele modifica esta faixa por extensão dos limites acima e abaixo dos valores numéricos especificados. Em geral, o termo "cerca de" é usado aqui para modificar um valor numérico acima e abaixo do valor dado por uma variância de 20 por cento, preferivelmente 10 por cento até ou menos (maior ou menor).

Como usado aqui, a palavra "ou" significa que um membro de uma lista particular e também inclui qualquer combinação de membros desta lista.

O termo "gene" é usado amplamente para fazer referência a qualquer segmento de ácido nucleico associado com uma função biológica. Assim, genes incluem sequências de codificação e/ou as sequências regulatórias requeridas para sua expressão. Por exemplo, gene refere-se a um fragmento de ácido nucleico que expressa mRNA ou RNA funcional, ou codifica uma proteína específica, e que inclui sequências regulatórias. Genes podem incluir as sequência de DNA não de codificação e/ou as sequências regulatórias, em que as regiões de não codificação podem ser transcritas em tal microRNA. Os genes também incluem segmentos não expressados que, por exemplo, formam sequências de reconhecimento para outras proteínas e/ou RNA. Genes podem ser obtidos de várias fontes, incluindo clonagem de uma fonte de interesse ou sintetizando de uma informação de sequência conhecida ou prevista via abordagens químicas ou de biologia molecular, e podem incluir sequências projetadas para ter os parâmetros desejados.

O termo gene "nativo" ou tipo selvagem" refere-se a um gene que está presente no genoma da célula não transformada, isto é, uma célula

não tendo uma mutação conhecida.

Um "gene marcador ou "sequência de marcador" codifica um traço selecionável ou triável.

O termo "gene quimérico" refere-se a qualquer gene que contém:

1) sequências de DNA, incluindo sequências regulatórias e de codificação, que não são encontradas juntas na natureza,

2) sequências codificando partes de sequências de codificação, por exemplo codificando para proteínas, não naturalmente juntas, ou

3) partes de sequências regulatórias, por exemplo promotores, que não naturalmente juntos.

Assim, um gene quimérico pode compreender sequências regulatórias e sequências de codificação que são derivadas de diferente fontes, ou compreendem sequências regulatórias e sequências de codificação derivadas da mesma fonte, mas dispostas em um modo diferente ao encontrado em natureza.

Um "transgene" refere-se a um gene que foi introduzido no genoma por transformação e é estavelmente mantido. Transgenes podem incluir, por exemplo, genes que são ou heterólogos ou homólogos para os genes de uma planta particular a ser transformada. Além disso, transgenes podem compreender genes nativos inseridos em um organismo não nativo, ou genes quiméricos. O termo "gene endógeno" refere-se a um gene nativo em seu local natural no genoma de um organismo. Um gene "estranho" gene refere-se a um gene não normalmente encontrado no organismo hospedeiro mas que é introduzido por transferência de genes.

Um "oligonucleotídeo" correspondente a uma sequência nucleotídica da invenção, por exemplo, para uso em reações de sondagem ou amplificação, pode ter cerca de 30 ou menos nucleotídeos em comprimento (por exemplo, 9, 12, 15, 18, 20, 21 ou 24, ou qualquer número entre 9 e 30).

Geralmente iniciadores específica são nucleotídeos de comprimento ascendente. Para uma especificidade ótima e eficácia de custo, os iniciadores de 16 a 24 nucleotídeos em comprimento podem ser preferidos. Os versados na técnica são conhecem bem o projeto de iniciadores para processo de uso como PCR. Se necessário, a sondagem pode ser feita com fragmentos de restrição completos do gene aqui descrito pode ter 100's ou ainda 1,000's de comprimento de nucleotídeos.

Os termos "polipeptídeo", "peptídeo", "oligopeptídeo", "polipeptídeo", "produto de gene", "produto de expressão" e "proteína" são usado de modo interpermutável aqui para fazer referência a um polímero ou oligômero de resíduos de aminoácidos consecutivos. Como usado aqui, o termo " sequência de aminoácido" ou uma "sequência de polipeptídeo " refere-se a uma lista de abreviaturas, letras, caracteres ou palavras representando os resíduos de aminoácidos. Os aminoácidos podem ser referidos aqui ou por seus símbolos de três letras comumente conhecido ou por símbolos de uma letra como recomendado por IUPAC-IUB Biochemical Nomenclature Commission. As abreviaturas usadas aqui são códigos de uma letra convencionais para os aminoácidos: A, alanina; B, asparagina ou ácido aspártico; C, cisteína; D ácido aspártico; E, glutamato, ácido glutâmico; F, fenilalanina; G, glicina; H histidina; I isoleucina; K, lisina; L, leucina; M, metionina; N, asparagina; P, prolina; Q, glutamina; R, arginina; S, serina; T, treonina; V, valina; W, triptofano; Y, tirosina; Z, glutamina ou ácido glutâmico (ver L. Stryer, Biochemistry, 1988, W. H. Freeman and Company, New York. A letra "x" como usada aqui dentro de uma sequência de aminoácido pode fazer referência a qualquer resíduo de aminoácido.

"Sequência de codificação" refere-se a uma sequência de DNA ou RNA que codifica para uma sequência de aminoácido específica e exclui as sequências de não codificação. Ela pode constituir uma " sequência de codificação não interrompida", isto é, faltando um intron, como em um cDNA

ou pode incluir um ou mais introns ligados por junções de emendas apropriadas. Um "intron" é uma sequência de RNA que está contida no transcrito primário mas que é removida através da clivagem e re-ligação do RNA dentro da célula para criar o mRNA maduro que pode ser traduzido em uma proteína.

Os termos matriz de leitura aberta e "ORF" referem-se para a sequência de aminoácidos codificada entre os códons de iniciação e de terminação de tradução de uma sequência de codificação. Os termos " códon de iniciação " e " códon de terminação " referem-s a uma unidade de três nucleotídeos adjacentes (‘códon’) em uma sequência de codificação que especifica a iniciação e terminação de cadeia, respectivamente, de síntese de proteína (tradução de mRNA).

Um "RNA funcional" refere-se a um RNA anti-sentido, ribozima, ou outro RNA que não é traduzido.

O termo " transcrito de RNA " refere-se ao produto resultante de transcrição catalisada por RNA polimerase de uma sequência de DNA. Quando o RNA transcrito é uma cópia complementar preferida da sequência de DNA, ele é referido como o transcrito primário ou pode ser uma sequência de RNA derivada do processamento pós-transcripcional do transcrito primário e é referido como o RNA maduro. " RNA mensageiro " (mRNA) refere-se ao RNA que não tem introns e que pode ser traduzido em proteína pela célula. "cDNA" refere-se a um DNA de filamento único ou duplo que é complementar a e derivado de mRNA.

“Sequência nucleotídica reguladora da transcrição“, "sequências regulatórias", e " sequências regulatórias apropriadas", referem-se cada a sequências nucleotídicas influenciando a transcrição, processamento de RNA ou estabilidade, ou tradução da sequência nucleotídica associada (ou funcionalmente ligada) a ser transcrita. A sequência nucleotídica reguladora da transcrição pode ter várias localizações com relação às sequências

nucleotídicas a serem transcritas. A sequência nucleotídica reguladora da transcrição pode estar localizada a montante (5' sequências não de codificação), dentro de ou a jusante (sequências 3' não de codificação) da sequência a ser transcrita (por exemplo, a sequência de codificação). As

5 sequências nucleotídicas reguladoras da transcrição podem ser selecionadas dentre o grupo compreendendo melhoradores, promotores, sequências líderes de tradução, introns, sequências 5'-não traduzidas, sequências 3'-não traduzidas, e sequências de sinal de poliadenilação. Elas incluem sequências naturais e sintéticas, assim como sequências, que podem ser uma combinação

10 de sequências sintéticas e naturais. Como notado acima, o termo "sequência nucleotídica reguladora da transcrição" é não limitado a promotores. No entanto, preferivelmente a sequência nucleotídica reguladora da transcrição da invenção compreende pelo menos uma sequência de promotor (por exemplo, uma sequência localizada a montante de um início de transcrição de um gene

15 capaz de induzir a transcrição das sequências a jusante). Em uma forma de realização preferida, a sequência nucleotídica reguladora da transcrição da invenção compreende a sequência de promotor do gene correspondente e – opcionalmente e preferivelmente – a região não traduzida 5' nativa do referido gene. Além disso, a região não traduzida 3' e/ou a região de poliadenilação de

20 referido gene também pode ser empregada.

"Promotor" refere-se a uma sequência nucleotídica, geralmente a montante (5') para sua sequência de codificação, que controla a expressão da sequência de codificação ao prover o reconhecimento para RNA polimerase e outros fatores (por exemplo, fatores de transcrição trans-atuantes) requeridos

25 para a transcrição apropriada. "Promotor" inclui um promotor mínimo que é uma sequência de DNA curta composta de uma caixa TATA e outras sequências que servem para especificar a iniciação do sítio de transcrição, ao que elementos regulatórios (por exemplo, *cis*-elementos) são adicionados para controle de expressão. "Promotor" também refere-se a uma sequência

nucleotídica que inclui um promotor mínimo mais elementos regulatórios que é capaz de controlar a expressão de uma sequência de codificação ou RNA funcional. Este tipo de sequência de promotor consiste de elementos proximal e mais distal a montante, os últimos elementos são com frequência referidos como melhoradores. Assim, um "melhorador" é uma sequência de DNA que pode estimular a atividade do promotor e pode ser um elemento inato do promotor ou um elemento heterólogo inserido para melhorar o nível ou especificidade de tecido de um promotor. Ele é capaz de operar em ambas as orientações (normal ou *flipped*), e é capaz de funcionar ainda quando movimentado ou a montante ou a jusante a partir do promotor. Tanto os melhoradores com outros elementos de promotor a montante ligam proteína de ligação a DNA específicas de sequência que mediam seus efeitos. Promotores pode ser derivados em sua totalidade de um gene nativo, ou ser compostos de diferentes elementos, derivados de diferente promotores encontrados em natureza, ou ainda ser compostos de segmentos sintéticos de DNA. Um promotor também pode conter sequências de DNA que estão envolvidas na ligação de fatores de proteína que controlam a eficácia da iniciação de transcrição em resposta a condições fisiológicas ou de desenvolvimento. Como usado aqui, o termo "cis-elemento" refere-se a um elemento regulatório transcripcional *cis*-atuante que confere um aspecto do controle global da expressão de genes. Um *cis*-elemento pode funcionar para ligar fatores de transcrição, fatores de proteína *trans*-atuantes que regulam a transcrição. Alguns *cis*-elementos ligam mais do que um fator de transcrição, e fatores de transcrição podem interagir com diferentes afinidades com mais do que um *cis*-elemento. Os promotores da presente invenção desejavelmente contém *cis*-elementos que podem conferir ou modular a expressão do gene. Os Cis- elementos podem ser identificados por várias técnicas, incluindo análise da deleção, isto é, deletando um ou mais nucleotídeos da extremidade 5' ou interna para um promotor; a análise de proteína de ligação de DNA

usando *DNase I footprinting*, interferência de metilação, testes de mobilidade - deslocamento de eletroforese, *footprinting* genômico in vivo, por PCR mediado por ligação, e outros testes convencionais, ou por similaridade de sequência de DNA com motivos de cis-elementos conhecidos por métodos de comparação de sequência de DNA convencionais. A estrutura fina de um *cis* - elemento pode ser ainda estudada por mutagenese (ou substituição) de um ou mais nucleotídeos ou por outros métodos convencionais. Os *Cis*- elementos podem ser obtidos por síntese química ou por isolamento de promotores que incluem estes elementos, e eles podem ser sintetizados com nucleotídeos de flanco adicionais que contém sítios de enzima de restrição utilizáveis para facilitar a manipulação subsequente.

O "sítio de iniciação" é a posição circundando o primeiro nucleotídeo que faz parte da sequência transcrita, que é também definido como posição + 1. Com relação a este sítio, todas as outras seqüências do gene e suas regiões de controle são numeradas. As seqüências a jusante (isto é seqüências codificando outras proteínas na direção 3') são denominadas positivas, enquanto as seqüências a montante (principalmente as regiões de controle na direção 5') são denominadas negativas.

Os elementos de promotor, particularmente um elemento TATA, que são inativos ou que tem atividade de promotor muito reduzida na ausência de ativação a montante são referidos como "promotores mínimos ou de núcleo". Na presença de um fator de transcrição apropriado, o promotor mínimo funciona para permitir a transcrição. Um "promotor mínimo ou de núcleo" assim consiste somente de todos os elementos basais necessários para a iniciação da transcrição, por exemplo uma caixa TATA e/ou um iniciador.

O termo "intron" refere-se a seções de DNA (sequências intervenientes) dentro de um gene que não codifica parte da proteína que o gene produz, e que é emendado fora do mRNA que é transcrito do gene antes de ser exportado do núcleo celular. A sequência de intron refere-se à

sequência de ácido nucleico de um intron. Assim, introns são as regiões de
sequências de DNA que são transcritas junto com a sequência de codificação
(exons) mas são removidos durante a formação de mRNA maduro. Introns
podem ser posicionados dentro da região de codificação real ou nos líderes
5 não traduzidos 5' ou 3' do pré-mRNA (mRNA não emendado). Introns no
transcrito primário são excisados e as sequências de codificação são
simultaneamente e precisamente ligadas para formar o mRNA maduro. As
junções de introns e exons formam o sítio de emenda. A sequência de um
intron começa com GU e termina com AG. Além disso, em plantas, dois
10 exemplos de introns AuU-AC foram descritos: intron 14 do gene de proteína
tipo RecA e intron 7 do gene G5 de *Arabidopsis thaliana* são introns At-AC.
Pre-mRNAs contendo introns tem três sequências curtas que são - além de
outras sequências- essenciais para o íntron ser precisamente emendado. Estas
sequências no sítio de emenda 5', o sítio de emenda 3', e o ponto de
15 ramificação. A emenda de mRNA é a remoção de seqüências intervenientes
(introns) presentes em transcritos de mRNA primários e união ou ligação de
sequências de exon. Isto é também conhecido como *cis*- emenda que une dois
exons no mesmo RNA com a remoção da sequência interveniente (intron). Os
elementos funcionais de um intron compreendendo seqüências que são
20 reconhecidas e ligadas pelos componentes de proteína específicos do
espliceosoma (por exemplo, seqüências de consenso de emenda nas
extremidades dos introns). A interação dos elementos funcionais com o
espliceosoma resulta na remoção da sequência de intron do mRNA prematuro
e a re-união das seqüências de exon. Introns tem três seqüências curtas que
25 são essenciais - apesar de não suficientes - para o intron ser precisamente
emendado. Estas seqüências são o sítio de emenda 5', o sítio de emenda 3' e o
ponto de ramificação. A sequência de ponto de ramificação é importante na
emenda e na seleção do sítio de emenda em plantas. A sequência de ponto de
ramificação está geralmente localizada 10-60 nucleotídeos a montante do sítio

de emenda 3'. As sequências de planta demonstram desvios de sequência no ponto de ramificação, as sequências de consenso sendo CURAY ou YURAY.

"Expressão constitutiva" refere-se à expressão usando um promotor constitutivo ou regulado. "Condicional" e "expressão regulada" referem-se à expressão controlada por um promotor regulado.

"Promotor constitutivo" refere-se a um promotor que é capaz de expressar a matriz de leitura aberta (ORF) que ele controle em todos ou quase todos os tecidos de planta durante todos ou quase todos os estágios de desenvolvimento da planta. Cada um dos elementos ativadores de transcrição não demonstra uma especificidade de tecido absoluta, mas mediam a ativação transcripcional na maioria das partes de planta a um nível de pelo menos 1% do nível alcançado na parte da planta em que a transcrição é mais ativa.

"Promotor regulado" refere-se aos promotores que dirigem a expressão de gene não constitutivamente, mas em um modo regulado temporariamente e/ou espacialmente, e inclui tanto promotores específicos de tecido como indutíveis. Isto inclui sequências naturais e sintéticas, assim como sequências que podem ser uma combinação de sequências naturais e sintéticas. Os promotores diferentes podem dirigir a expressão de um gene em tecidos ou tipos de células diferentes, ou em estágios de desenvolvimento diferentes, ou em resposta a diferentes condições ambientais. Novos promotores de vários tipos utilizáveis em células de plantas estão sendo constantemente descobertos, numerosos exemplos podem ser encontrados na compilação por Okamuro et al (1989). Os promotores regulados típicos utilizáveis em plantas incluem, mas não limitados a promotores indutivos por agentes de segurança, promotores derivados do sistema indutível por tetraciclina, promotores derivados de sistemas indutíveis por salicilato, promotores derivados de sistemas indutíveis por álcool, promotores derivados de sistema indutível por glucocorticóides, promotores derivados de sistemas indutíveis por patógenos, e promotores derivados de sistemas indutíveis por

ecdisona.

"Promotor específico de tecido" refere-se a promotores regulados que não são expressados em todas as células de plantas mas somente em um ou mais tipos de células em órgão específicos (como folhas
5 ou sementes), tecidos específicos (como embrião ou cotilédone), ou tipos de células específicos (como parênquima de folhas ou células de armazenamento em sementes). Estes também incluem promotores que são temporalmente regulados, como em embriogenese prematura ou tardia, durante o amadurecimento das frutas nas sementes ou frutos em desenvolvimento, em
10 folha completamente diferenciada, ou no início da senescência.

"Promotor indutível" refere-se aos promotores regulados que podem ser ligados em um ou mais tipos de células por um estímulo externo, como químico, luz, hormônio, estresse ou um patógeno.

"Ligado operativamente" ou "ligado funcionalmente" refere-se, preferivelmente, à associação de seqüências de ácido nucleico em
15 fragmento de ácido nucleico único, de modo que a função de um é afetada pelo outro. Mais especificamente, refere-se a uma primeira seqüência (s) estando posicionada de modo suficientemente próximo a uma segunda seqüência (s) de modo que a primeira seqüência (s) pode exercer influência
20 sobre a segunda seqüência (s) ou uma região sob controle desta segunda seqüência. Por exemplo, uma seqüência de DNA regulatória é dita como sendo "operativamente ligada a" ou "associada com" uma seqüência de DNA que codifica para um RNA ou um polipeptídeo se as duas seqüências estiverem situadas de modo que a seqüência de DNA regulatória afeta a
25 expressão da seqüência de DNA de codificação (isto é, que a seqüência de codificação ou RNA funcional está sob o controle transcripcional do promotor).

A maior parte das seqüências ativantes, por exemplo, um intron, estão dentro de cerca de 300 a 400 pares de base do promotor que é

melhorado.

Em uma forma de realização, o local do intron, por exemplo um intron INME, é em 5' UTR do gene controlado. (Rose et al. RNA 2002, 8:1444-1453; Callis et al., 1987, Genes & Dev 1:1183-1200; PF56400).

5 O ligador entre a região do promotor e o intron é, por exemplo abaixo de 400 pares de base, por exemplo 200, 100, 50 ou menos. Preferivelmente, a primeira sequência de uma molécula de ácido nucleico é ligada à segunda sequência por um ligador em torno de 100 nucleotídeos ou menos.

10 Em uma outra forma de realização da invenção, mais do que uma sequência de ativação é empregada e as sequências de ativação estão em cerca de 16 a cerca de 80 pares de base uma de cada outra. A sequência de codificação pode ser operativamente ligada a sequências regulatórias em uma orientação sentido ou anti-sentido, por exemplo sendo transcrita em um
15 mRNA codificando um polipeptídeo ou um fragmento de uma proteína ou sendo transcrito em uma molécula de RNA regulatória ou enzimática, por exemplo um RNA anti-sentido, um RNAi, uma Ribozima, um miRNA, um ta-siRNA, um dsRNA, snRNA ou outros, como descrito aqui ou na literatura relevante, ou combinações dos mesmos.

20 "Expressão" refere-se a uma transcrição e/ou tradução de um gene endógeno, ORF ou porção do mesmo, ou um transgene em plantas. Por exemplo, no caso de construções anti-sentido, a expressão pode fazer referência a uma transcrição do DNA anti-sentido apenas. Além disso, a expressão refere-se a uma transcrição e acúmulo estágio de RNA sentido
25 (mRNA) ou RNA funcional. Expressão também pode fazer referência à produção de proteína.

"Expressão específica" é a expressão de produto de genes que é limitada a um ou poucos tecidos de plantas (limitação espacial) e/ou a um ou poucos estágios de desenvolvimento de plantas (limitação temporal). Sabe-

se que dificilmente uma especificidade existe: promotores parecerem ser preferivelmente ligados em alguns tecidos, enquanto em outros tecidos ele podem não ter ou apenas uma pequena atividade. Este fenômeno é conhecido como uma expressão de vazamento. No entanto, com expressão específica desta invenção significa-se uma expressão preferível em um ou em poucos tecidos de planta.

O "padrão de expressão" de um promotor (como ou sem melhorador) é o nível do padrão de expressão que mostra onde, na planta, e em qual estágio de desenvolvimento, a transcrição é iniciada pelo referido promotor. Os padrões de expressão de um conjunto de promotores são ditos como sendo complementares quando o padrão de expressão de um promotor mostra pouca sobreposição com o padrão de expressão do outro promotor. O nível de expressão de um promotor pode ser determinado por medida da concentração em "estado estável" de um mRNA repórter transcrito padrão. Esta medida é indireta porque a concentração do mRNA repórter é dependente não apenas de sua taxa de síntese, mas também da taxa com a qual o mRNA é degradado. Assim, o nível de estado estável é o produto das taxas de síntese e das taxas de degradação. A taxa de degradação pode ser considerada, no entanto, como prosseguindo em uma taxa fixa, quando as seqüências transcritas são idênticas e, assim, este valor pode servir como uma medida das taxas de síntese. Quando os promotores são comparados deste modo, as técnicas disponíveis para os versados na técnica são hibridização, análise S1-RNase, northern blots e RT-PCR competitivo. Esta lista de técnicas não representa, de nenhum modo, todas as técnicas disponíveis, mas ao contrário, descreve procedimentos comumente usados para analisar a atividade de transcrição e os níveis de expressão de mRNA. A análise dos pontos de partida de transcrição em praticamente todos os promotores revelou que não se tem, geralmente, base única em que começa a transcrição, mas ao contrário mais ou um conjunto mesmo agrupado de sítios de iniciação, cada

um contando por alguns pontos de partida do mRNA. Porque esta distribuição varia de promotor a promotor, as sequências do mRNA repórter em cada uma das populações irão diferir uma de cada outra. Porque cada espécie de mRNA é mais ou menos propensa à degradação, não se pode esperar uma taxa de

5 degradação única para diferentes mRNAs repórteres. Foi mostrado para várias sequências de promotor eucariótico que a sequência circundando o sítio de iniciação (iniciador) desempenha um papel importante na determinação do nível de expressão de RNA dirigida por este promotor específico. Isto inclui também parte das sequências transcritas. A fusão direta de sequências de

10 promotor para repórter deveria, assim, levar a níveis sub-ótimas de transcrição. Um procedimento comumente usado para analisar padrões de expressão e níveis é através da determinação do nível de "estado estável" de acúmulo de proteína em uma célula. Os candidatos comumente usados para o gene repórter, bem conhecidos do versado na técnica, são beta-glucuronidase

15 (GUS), cloranfenicol acetil transferase (CAT) e proteínas com propriedades fluorescentes, como proteína fluorescente verde (GFP) de *Aequora victoria*. Em princípio, no entanto, muitas mais proteínas são apropriadas para este fim, desde que a proteína não interfira com as funções essenciais da planta. Para quantificação e determinação de localização, várias ferramentas são

20 apropriadas. Os sistemas de detecção podem prontamente ser criados ou disponíveis que são baseados em, por exemplo, detecção imunoquímica, enzimática, fluorescente e quantificação. Os níveis de proteína podem ser determinados em extratos de tecido de plantas ou em tecido intacto usando análise in situ de expressão de proteína. Geralmente, as linhagens

25 transformadas individuais com uma construção de repórter de promotor quimérico irá variar em seus níveis de expressão do gene repórter. Também é observado com frequência o fenômeno que estas transformantes não expressam qualquer produto detectável (RNA ou proteína). A variabilidade em expressão é comumente dada a "efeitos de posição", apesar dos

mecanismos moleculares subjacentes a esta inatividade não serem geralmente claros.

"Super-expressão" refere-se ao nível de expressão em células transgênicas ou organismos que excedem os níveis de expressão em células normais ou não transformadas (não transgênicas) ou organismos.

"Seqüência 5' não de codificação" ou "sequência 5' não traduzida" ou "-região" refere-se a uma sequência nucleotídica localizada 5' (a montante) à sequência de codificação. Ela está presente no mRNA completamente processado a montante da códon de iniciação e pode afetar o processamento do transcrito primário para mRNA, estabilidade do mRNA ou eficiência de tradução (Turner 1995).

"Seqüência 3' não de codificação" ou "sequência 3' não traduzida" ou "-região" refere-se a uma sequência nucleotídica localizada 3' (a jusante) à sequência de codificação e inclui seqüências de sinal de poliadenilação e outras seqüências codificando sinais regulatórios capazes de afetar o processamento de mRNA ou a expressão do gene. O sinal de poliadenilação é geralmente caracterizado por afetar a adição de traços de ácido poliadenílico à extremidade 3' do precursor de mRNA. O uso de diferentes seqüências 3' de não codificação é exemplificado por Ingelbrecht et al, 1989.

O termo "sequência líder de tradução" refere-se à porção de sequência de DNA de um gene entre o promotor e a sequência de codificação que é transcrita em RNA e está presente no mRNA completamente processado (5') do códon de partida de tradução. A sequência líder de tradução pode afetar o processamento do transcrito primário para mRNA, estabilidade de mRNA ou eficiência de tradução.

"Peptídeo de sinal" refere-se à extensão amino terminal de um polipeptídeo, que é traduzido em conjunto com o polipeptídeo formando um peptídeo precursor e que é requerido para sua entrada na via secretória. O

termo "sequência de sinal" refere-se a uma sequência nucleotídica que codifica o peptídeo de sinal. O termo "peptídeo de trânsito" como usado aqui refere-se a uma parte de um polipeptídeo expressado (preferivelmente da extensão amino terminal de um polipeptídeo), que é traduzido em conjunto com o polipeptídeo formando um peptídeo precursor e que é requerido para sua entrada em um organelo de célula (como os plastídeos, (por exemplo, cloroplastos) ou mitocôndria). O termo "sequência de trânsito" refere-se a uma sequência de nucleotídeo que codifica o peptídeo de trânsito.

"Inibição anti-sentido" refere-se à produção de transcritos de RNA anti-sentido capazes de suprimir a expressão de proteína de um gene endógeno ou um transgene.

"Silenciamento de genes" refere-se a supressão dependente de homologia de genes virais, transgenes, ou genes nucleares endógenos. O silenciamento de genes pode ser transcripcional, quando a supressão é devido a diminuída transcrição dos genes afetados, ou pós-transcripcional, quando a supressão é devido a um *turnover* aumentado (degradação) de espécies de RNA homólogas aos genes afetados (English 1996). O silenciamento de genes inclui silenciamento de gene induzido por vírus (Ruiz *et al.* 1998).

Os termos " sequência de DNA heteróloga", "segmento de DNA exógeno" ou " ácido nucleico heterólogo," como usados aqui, referem-se, cada, a uma sequência que se origina de uma fonte estranha à célula hospedeira particular ou, se da mesma fonte, é modificada desta forma original. Assim, um gene heterólogo em uma célula hospedeira inclui um gene que é endógeno para a célula hospedeira particular mas que foi modificada através de, por exemplo, o uso de embaralhamento de DNA. Os termos também incluem cópias múltiplas não de ocorrência natural de sequência de DNA de ocorrência natural. Assim, os termos referem-se a um segmento de DNA que é estrangeiro ou heterólogo para a célula, ou homólogo para a células mas em uma posição dentro do ácido nucleico da

célula hospedeira em que o elemento não é comumente encontrado. Os segmentos de DNA exógenos são expressados para dar polipeptídeos exógenos. Uma sequência de DNA "homóloga" é uma sequência de DNA que é naturalmente associada com uma célula hospedeira em que ela é introduzida.

"Homólogo a" no contexto da identidade de sequência nucleotídica refere-se à similaridade entre a sequência nucleotídica de duas moléculas de ácido nucleico ou entre as sequências de aminoácido de duas moléculas de proteína. As estimativas desta homologia são providas por hibridização ou DNA-DNA ou DNA-RNA sob condições de estringência como é bem conhecido pelo versado na técnica (como descrito por Haines e Higgins (eds.), *Nucleic Acid Hybridizations*, IRL Press, Oxford, U.K.), ou pela comparação de similaridade de sequência entre dois ácidos nucleicos ou proteínas.

O termo "substancialmente similar" refere-se a sequências de nucleotídeo e de aminoácidos que representam equivalentes funcional e/ou estrutural ou ortólogos de sequências descritos aqui, em particular de sequências de *Oryza sativa*, *Arabidopsis thaliana* ou *Brassica napus* descritas aqui.

Em seu sentido mais amplo, o termo "substancialmente similar" quando usado aqui com relação a uma sequência nucleotídica significa que a sequência nucleotídica é parte de um gene que codifica um polipeptídeo tendo substancialmente a mesma estrutura e função que um polipeptídeo codificado por um gene para a sequência nucleotídica de referência, por exemplo, a sequência nucleotídica compreende um promotor de um gene que é o ortólogo do gene correspondente à sequência nucleotídica de referência, assim como sequências de promotor que são estruturalmente relacionadas a uma sequência de promotores particularmente exemplificada aqui, isto é, as sequências substancialmente similares de promotor hibridizam

para o complemento de uma sequência de promotores exemplificada aqui sob condições de estringência elevadas ou muito elevadas. Por exemplo, sequências nucleotídicas alteradas que simplesmente refletem a degenerescência do código genético mas, mesmo assim, codificam as

5 sequências de aminoácidos que são idênticas a uma sequência de aminoácidos particular são substancialmente similares às sequências particulares. O termo "substancialmente similar" também inclui sequências nucleotídicas em que a sequência foi modificada, por exemplo, para otimizar a expressão em células

10 particulares, assim como sequências nucleotídicas codificando um polipeptídeo variante tendo uma ou mais substituições de aminoácidos com relação ao polipeptídeo (não modificado) codificado pela sequência de referência, cuja(s) substituição(ões) não altera(m) a atividade do polipeptídeo variante com relação ao polipeptídeo não modificado.

Em seu sentido mais amplo, o termo "substancialmente

15 similar" quando usado aqui com relação ao polipeptídeo significa que o polipeptídeo tem substancialmente a mesma estrutura e função que o polipeptídeo de referência. Preferivelmente, o homólogo é um ortólogo. Se um gene com uma sequência substancialmente similar é um ortólogo de gene, por exemplo ou cor78 diferente de gene Met1, pode ser testado em um teste

20 simples. O gene endógeno, por exemplo cor78 ou Met é primeiro deletado e então o ortólogo potencial é introduzido na célula ou planta depletada. Um ortólogo deve re-estabelecer um fenótipo similar ou substancialmente similar como o tipo selvagem, preferivelmente o fenótipo é idêntico ao tipo selvagem. O termo pode, em uma forma de realização, significar que as

25 sequências de aminoácidos que são substancialmente similares a uma sequência particular são as que a identidade de aminoácido global é pelo menos 60% ou maior para as presentes sequências. Modificações que resultam em sequências de nucleotídeos ou aminoácidos equivalentes são bem conhecidas do versado na técnica. Em uma forma de realização, a

porcentagem de identidade de sequência de aminoácido entre o polipeptídeo substancialmente similar e o de referência é pelo menos 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, por exemplo, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, e ainda 90% ou mais, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo menos 99%. Uma outra indicação de que os dois polipeptídeos são substancialmente similares entre si, além de terem substancialmente a mesma função, é que um agente, por exemplo, um anticorpo, que especificamente liga a um dos polipeptídeos, também especificamente liga ao outro.

As comparações de sequência podem ser realizadas usando um algoritmo de alinhamento de sequência Smith-Waterman (ver por exemplo, Waterman 1995). O programa localS, versão 1.16, é preferivelmente usado com seguintes parâmetros: correspondência: 1, penalidade de falta de correspondência: 0,33, penalidade de espaço aberto: 2, penalidade de espaço estendido: 2.

Além disso, a sequência nucleotídica que é "substancialmente similar" a uma sequência nucleotídica de referência é referida como sendo "equivalente" à sequência nucleotídica de referência. O versado na técnica reconhece que as sequências nucleotídicas equivalente englobadas por esta invenção também podem ser definidas por sua capacidade de hibridizar, sob condições baixas, moderadas e/ou estridentes (por exemplo, 0,1 X SSC, 0,1% SDS, 65°C), com as sequências nucleotídicas que estão dentro do escopo literal das presentes reivindicações.

O que se significa por "substancialmente a mesma atividade" quando usado em referência a um fragmento de polinucleotídeo ou polipeptídeo é que o fragmento tem pelo menos 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, por exemplo, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%,

89%, e ainda 90% ou mais, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo menos 99% da atividade do polinucleotídeo de comprimento completo ou polipeptídeo de comprimento completo.

"Gene alvo" refere-se a um gene por exemplo a um gene sobre o replicon, que expressa a sequência de codificação alvo desejada, RNA funcional, ou proteína. Em uma forma de realização, o gene alvo não é essencial para a replicação do replicon. Adicionalmente, os genes alvos podem compreender genes não virais nativos inseridos em um organismo não nativo ou genes quiméricos e estará sob o controle de sequências regulatórias apropriadas. Assim, as sequências regulatórias no gene alvo podem vir de qualquer fonte, incluindo o vírus. Os genes alvos pode incluir sequências de codificação que são ou heterólogas ou homólogas para os genes de uma planta particular a ser transformada. Em uma forma de realização, no entanto, os genes alvos podem ser genes que não incluem genes virais nativos. Os genes alvos típicos incluem, mas não são limitados a genes codificando uma proteína estrutural, uma proteína de armazenamento de semente, uma proteína que transporta resistência a herbicida, e uma proteína que transporta resistência a insetos. As proteínas codificadas por genes alvos são conhecidas como "proteínas estrangeiras". A expressão de um gene alvo em uma planta irá tipicamente produzir um traço de planta alterado.

O termo "traço de planta alterado" significa qualquer mudança fenotípica ou genotípica em uma planta transgênica para o hospedeiro de planta tipo selvagem ou não transgênico.

O termo "transformação" refere-se à transferência de um fragmento de ácido nucleico no genoma de uma célula hospedeira, resultando em uma herança geneticamente estável. As células hospedeiras contendo os fragmentos de ácido nucleico transformados são referidas como células "transgênicas" e organismos compreendendo células transgênicas são referidos como "organismos transgênicos". Os exemplos de métodos de

transformação de plantas e células de plantas incluem transformação mediada por *Agrobacterium* (De Blaere 1987) e tecnologia de bombardeio de partículas (US 4 945 050). As plantas completas podem ser regeneradas de células transgênicas por métodos bem conhecidos do versado na técnica (ver, 5 por exemplo, Fromm 1990).

"Transformado", "transgênico" e "recombinante" referem-se a um organismo hospedeiro como uma bactéria ou uma planta em que uma molécula de ácido nucleico heteróloga foi introduzida. A molécula de ácido nucleico pode ser estavelmente integrada no genoma. Por exemplo, plantas ou 10 calos "transformados", "transformantes" e "transgênicos" passaram pelo processo de transformação e contêm um gene estrangeiro integrado em seu cromossomo. O termo "não transformado" refere-se a plantas normais que não passaram pelo processo de transformação.

"Transientemente transformado" refere-se a células em que 15 transgenes e DNA estrangeiro foram introduzidos (por exemplo, por métodos como transformação mediada por *Agrobacterium* ou bombardeio biolístico), mas não são selecionadas para manutenção estável.

"Estavelmente transformadas" refere-se a células que foram selecionadas e regeneradas em um meio de seleção após transformação.

20 "Cromossomicamente integrado" refere-se à integração de um gene estrangeiro ou construção de DNA no DNA hospedeiro por ligações covalentes. No caso dos genes não serem "cromossomicamente integrados", eles podem ser "expressados transientemente". A expressão transiente de um gene refere-se à expressão de um gene que não é integrado no cromossomo 25 hospedeiro mas funciona independentemente, ou como parte de um plasmídeo autonomamente replicante ou cassete de expressão, por exemplo, ou como parte de outro sistema biológico como um vírus.

"Expressão transiente" refere-se à expressão em células em que um vírus ou um transgene é introduzido por infecção viral ou por tais

métodos como transformação mediada por *Agrobacterium*, eletroporação, ou bombardeio biolístico, mas não selecionadas por sua manutenção estável.

"Geneticamente estável" e "herdável" refere-se a elementos genéticos cromossomicamente integrados que são estavelmente mantidos na planta e estavelmente herdados por progênie através de gerações sucessivas.

"Transformante primário" e "geração T0" referem-se a plantas transgênicas que são da mesma geração genética que o tecido, que foi inicialmente transformado (isto é, não tendo passado através de meiose e fertilização desde a transformação).

"Transformantes secundários" e as "gerações T1, T2, T3, etc..) referem-se a plantas transgênicas derivadas de transformantes primários através de um ou mais ciclos de fertilização meiótica. Eles podem ser derivados por auto-fertilização de transformantes primários ou secundários ou cruzamentos de transformantes primários ou secundários com outras plantas transformadas ou não transformadas.

"Tipo selvagem" refere-se a um vírus ou organismo encontrado na natureza sem qualquer mutação conhecida.

Um segregante nulo é progênie (ou linhagens derivadas da progênie) de uma planta transgênica que não contém o transgene devido à segregação mendeliana.

Os termos "genoma" ou "DNA genômico" faz referência à informação genética herdável de um organismo hospedeiro. O referido DNA genômico compreende o DNA do núcleo (também referido como DNA cromossômico) mas também o DNA dos plastídeos (por exemplo cloroplastos) e outros organelos celulares (por exemplo mitocôndrias). Preferivelmente, os termos genoma ou DNA genômico faz referência ao DNA cromossômico do núcleo.

O termo "DNA cromossômico" ou "sequência de DNA cromossômico" deve ser entendido como o DNA genômico do núcleo celular

independente do estado do ciclo celular. O DNA cromossômico pode, assim, ser organizado em cromossomas ou cromatídeos, eles podem ser condensados ou não enrolados. Uma inserção no DNA cromossômico pode ser demonstrada e analisada por vários métodos bem conhecidos como por exemplo análise de reação de cadeia polimerase (PCR), análise Southern blot, hibridização *in situ* com fluorescência (FISH), e PCR *in situ*.

O termo "ácido nucleico" refere-se a desoxirribonucleotídeos ou ribonucleotídeos e polímeros dos mesmos em forma de filamento único ou duplo, compostos por monômeros (nucleotídeos) contendo um açúcar, fosfato, e uma base, que é uma purina ou pirimidina. Salvo especificamente limitado, o termo engloba ácidos nucleicos contendo análogos conhecidos de nucleotídeos naturais, que tem propriedades de ligação similares como o ácido nucleico de referência são metabolizados em um modo similar aos nucleotídeos de ocorrência natural. Salvo especificado em contrário, uma sequência particular de ácido nucleico também implicitamente engloba variantes conservativamente modificados dos mesmos (por exemplo, substituições de códons degenerados) e sequências complementares, assim como a sequência explicitamente indicada. Especificamente, substituições de códons degenerados podem ser obtidas por gerações de sequências em que a terceira posição de um ou mais códons selecionados (ou todos) é substituído com base mista e/ou resíduos de deoxiinosina (Batzner 1991; Ohtsuka 1985; Rossolini 1994). Um "fragmento de ácido nucleico" é uma fração de uma dada molécula de ácido nucleico. Em plantas superiores, ácido desoxirribonucleico (DNA) é o material genético enquanto o ácido ribonucleico (RNA) está envolvido na transferência de informação contida dentro do DNA nas proteínas. O termo "sequência nucleotídica" refere-se a um polímero de DNA ou RNA que pode ser de filamento único ou duplo, opcionalmente contendo bases de nucleotídeos sintéticas, não naturais ou alteradas, capazes de incorporação nos polímeros de DNA ou RNA. Os

termos "ácido nucleico" ou "sequência de ácido nucleico" também podem ser usados de modo interpermutável com gene, cDNA, DNA e RNA codificados por um gene.

A invenção engloba composições de ácido nucleico ou proteína isoladas ou substancialmente purificadas. No contexto da presente invenção, uma molécula de DNA "isolada" ou "purificada" ou um polipeptídeo "isolado" ou "purificado" é uma molécula de DNA ou polipeptídeo que, pela mão do homem, existe além de seu meio ambiente nativo e é, assim, não um produto da natureza. Uma molécula de DNA isolada ou polipeptídeo pode existir em uma forma purificada ou pode existir em um meio não nativo como, por exemplo, uma célula hospedeira transgênica. Por exemplo, uma molécula de ácido nucleico ou proteína "isolada" ou "purificada", ou porção biologicamente ativa da mesma, é substancialmente isenta de outro material celular, ou meio de cultura quando produzida por técnicas recombinantes, ou substancialmente livre de precursores químicos ou outros produtos químicos quando quimicamente sintetizada. Preferivelmente, um ácido nucleico "isolado" é livre de sequências (preferivelmente sequências codificando proteína) que naturalmente flanqueiam o ácido nucleico (*isto é*, sequências localizadas nas extremidades 5' e 3' do ácido nucleico) no DNA genômico do organismo do qual o ácido nucleico é derivado. Por exemplo, em várias formas de realização, a molécula de ácido nucleico isolada pode conter menos do que cerca de 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb, ou 0,1 kb de sequências nucleotídicas que naturalmente flanqueiam a molécula de ácido nucleico em DNA genômico da célula da qual o ácido nucleico é derivado. A proteína que é substancialmente livre de material celular inclui preparações de proteína ou polipeptídeo tendo menos do que cerca de 30%, 20%, 10%, 5%, (por peso seco) de proteína contaminante. Quando a proteína da invenção, ou porção biologicamente ativa da mesma) é recombinantemente produzida, preferivelmente o meio de cultura representa menos do que cerca de 30%,

20%, 10%, ou 5% (por peso seco) de precursores químicos ou produtos químicos não proteína de interesse. As sequências nucleotídicas da invenção incluem tanto sequências de ocorrência natural assim como formas mutantes (variantes). Estas variantes irão continuar a possuir a atividade desejada, isto é, ou a atividade do promotor ou a atividade do produto codificado pela matriz de leitura aberta da sequência nucleotídica não variante.

O termo "variante" com relação a uma sequência (por exemplo, uma sequência de polipeptídeo ou de ácido nucleico como – por exemplo – a sequência nucleotídica reguladora da transcrição da invenção) se destina a significar sequências substancialmente similares. Para sequências nucleotídicas compreendendo uma matriz de leitura aberta, as variantes incluem as sequências que, devido à degenerescência do código genético, codificam a sequência de aminoácido idêntica da proteína nativa. As variantes alélicas de ocorrência natural, como estas, podem ser identificadas com o uso de técnicas de biologia molecular bem conhecidas como, por exemplo, técnicas com reação de cadeia polimerase (PCR) e hibridização. As sequências nucleotídicas variantes também incluem sequências nucleotídicas sinteticamente derivadas, como as geradas, por exemplo, usando mutagenese sítio-dirigida e para matriz de leitura aberta, codificam a proteína nativa, assim como as que codificam um polipeptídeo tendo substituições de aminoácido com relação à proteína nativa. Geralmente, variantes de sequência nucleotídica da invenção terão, pelo menos, 40, 50, 60, to 70%, por exemplo, preferivelmente 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, geralmente pelo menos 80%, por exemplo, 81%-84%, pelo menos 85%, por exemplo, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, a 98% e 99% identidade de sequência nucleotídica para a sequência nucleotídica nativa (tipo selvagem ou endógena).

"Variações conservativamente modificadas" de uma sequência de ácido nucleico particular refere-se às sequências de ácido nucleico que

codificam sequências de aminoácidos idênticas ou essencialmente idênticas, ou onde as sequências de ácido nucleico não codificam uma sequência de aminoácido, para sequências essencialmente idênticas. Devido à degenerescência do código genético, um grande número de ácidos nucleicos funcionalmente idênticos codificam qualquer dado polipeptídeo. Por exemplo, os códons CGT, CGC, CGA, CGG, AGA, e AGG codificam, todos, o aminoácido arginina. Assim, em cada posição onde uma arginina é especificada por um códon, o códon pode ser alterado para qualquer códon correspondente descrito sem alterar a proteína codificada. Estas variações de ácido nucleico são "variações silentes" que são uma espécie de "variações conservativamente modificadas." Cada sequência de ácido nucleico descrita aqui que codifica um polipeptídeo também descreve cada possível variação silente, exceto onde especificado ao contrário. Um versado na técnica irá reconhecer que cada códon em um ácido nucleico (exceto ATG, que é comumente o único códon para metionina) pode ser modificado para dar uma molécula funcionalmente idêntica por técnicas padrões. Assim, cada "variação silente" de um ácido nucleico que codifica um polipeptídeo é implícita em cada sequência descrita.

As moléculas de ácido nucleico da invenção podem ser "otimizadas" para expressão melhorada em plantas de interesse (ver, por exemplo, WO 91/16432; Perlak 1991; Murray 1989). Deste modo, as matrizes de leitura aberta em genes ou fragmentos de gene podem ser sintetizadas usando códons preferidos das plantas (ver, por exemplo, Campbell & Gowri, 1990 para uma discussão de uso de códon preferido do hospedeiro). Assim, as sequências nucleotídicas podem ser otimizadas para expressão em qualquer planta. Reconhece-se que todas ou qualquer parte de uma sequência de genes pode ser otimizada ou sintética. Isto é, sequências sintéticas ou parcialmente otimizadas também podem ser usadas. As sequências nucleotídicas variantes e proteínas também englobam sequências e proteínas derivadas de um

procedimento mutagênico e recombinogênico como o embaralhamento de DNA. Com tal procedimento, uma ou mais diferentes sequências de codificação podem ser manipuladas para criar um novo polipeptídeo possuindo as propriedades desejadas. Deste modo, as bibliotecas de polinucleotídeos recombinante são geradas a partir de uma população de sequências de polinucleotídeos relacionadas compreendendo regiões de sequência que tem uma identidade de sequência substancial e podem ser homologamente recombinadas *in vitro* ou *in vivo*. Estratégias para este embaralhamento de DNA são bem conhecidas na técnica (ver, por exemplo, Stemmer 1994; Stemmer 1994; Cramer 1997; Moore 1997; Zhang 1997; Cramer 1998; e US 5.605.797, 9, 11, 13, 15, e 17.837.458).

Por polipeptídeo "variante" pretende-se um polipeptídeo derivado de proteína nativa por deleção (assim chamada truncação) ou adição de um ou mais aminoácidos à extremidade N-terminal e/ou C-terminal da proteína nativa; deleção ou adição de um ou mais aminoácidos em um ou mais sítios na proteína nativa; ou substituição de um ou mais aminoácidos em um ou mais sítios na proteína nativa. Estas variantes irão resultar de, por exemplo, polimorfismo genético ou de manipulação humana. Métodos para tais manipulações são geralmente bem conhecidos na técnica.

Assim, os polipeptídeos podem ser alterados em vários modos incluindo substituições de aminoácidos, deleções, truncações, e inserções. Os métodos para estas manipulações são geralmente bem conhecidos. Por exemplo, as variantes de sequência de aminoácido dos polipeptídeos podem ser preparadas por mutações no DNA. Métodos para mutagenese e alterações de sequência nucleotídica são bem conhecidas na técnica (ver, por exemplo, Kunkel 1985; Kunkel 1987; US 4.873.192; Walker & Gaastra, 1983 e as referências citadas). O guia para as substituições de aminoácidos apropriadas que não afetam a atividade biológica da proteína de interesse pode ser encontrado no modelo de Dayhoff *et al.* (1978). As substituições

conservativas, como troca de um aminoácido por outro tendo propriedades similares, são preferidas. As substituições individuais, deleções ou adições que alteram, adicionam ou deletam um único aminoácido ou uma porcentagem pequena de aminoácidos (tipicamente menos do que 5%, mais tipicamente menos do que 1%) em uma sequência codificada são "variações conservativamente modificadas", onde as alterações resultam na substituição de um aminoácido com um aminoácido quimicamente similar. As tabelas de substituições conservativas provendo aminoácidos funcionalmente similares são bem conhecidas na técnica. Os seguintes cinco grupos contém, cada, aminoácidos que são substituições conservativas um para o outro: alifático: Glicina (G), Alanina (A), Valina (V), Leucina (L), Isoleucina (I); Aromático: Fenilalanina (F), Tirosina (Y), Triptofano (W); contendo enxofre: Metionina (M), Cisteína (C); Básico: Arginina (R), Lisina (K), Histidina (H); ácido: Ácido aspártico (D), Ácido glutâmico (E), Asparagina (N), Glutamina (Q). Ver também, Creighton, 1984. Além disso, as substituições individuais, deleções ou adições que alteram, adicionam ou deletam um único aminoácido ou uma porcentagem pequena de aminoácidos em uma sequência codificada são também ' variações conservativamente modificadas'.

"Cassete de expressão" ou "construção de expressão" como usado aqui significa uma sequência de DNA capaz de dirigida a expressão de uma sequência nucleotídica particular em uma célula hospedeira apropriada, compreendendo um promotor operativamente ligada a uma sequência nucleotídica de interesse, que é – opcionalmente - operativamente ligada a sinais de terminação e/ou outros elementos regulatórios. Um cassete de expressão também pode compreender sequências requeridas para a apropriada tradução de sequência nucleotídica. A região de codificação geralmente codifica para uma proteína de interesse mas também pode codificar para um RNA funcional de interesse, por exemplo RNA anti-sentido ou RNA não traduzido, na direção sentido ou anti-sentido. O cassete de expressão

compreendendo a sequência nucleotídica de interesse pode ser quimérico, significando que pelo menos um de seus componentes é heterólogo com relação a pelo menos um de seus outros componentes. O cassete de expressão também pode ser um, que é de ocorrência natural mas que foi obtido em uma
5 forma recombinante utilizável para a expressão heteróloga. Um cassete de expressão pode ser montado completamente extracelularmente (por exemplo, por técnicas de clonagem recombinante). No entanto, um cassete de expressão também pode ser montado usando em parte componentes endógenos. Por exemplo, um cassete de expressão pode ser obtido colocando (ou inserindo)
10 uma sequência de promotor a montante de uma sequência endógena, que assim se torna funcionalmente ligada e controlada pelas referidas sequências de promotor. Do mesmo modo, a sequência de ácido nucleico a ser expressada pode ser colocada (ou inserida) a jusante de uma sequência de promotor endógena, assim formando um cassete de expressão. Em geral, a
15 expressão de sequência nucleotídica no cassete de expressão pode estar sob o controle de um promotor constitutivo ou de um promotor indutível que inicia a transcrição somente quando a célula hospedeira é exposta a algum estímulo externo particular. Os cassetes de expressão da invenção, a sequência reguladora da transcrição ou o promotor também podem ser específicos para
20 um tecido em particular ou órgão ou estágio de desenvolvimento, em particular para o embrião ou o escutélulo. Em uma forma de realização preferida, estes cassetes de expressão irão compreender a região de iniciação de transcrição da invenção ligada a uma sequência nucleotídica de interesse. Este cassete de expressão é preferivelmente provido com uma pluralidade de
25 sítios de restrição para inserção do gene de interesse a estar sob a regulação transcripcional das regiões regulatórias. O cassete de expressão pode conter adicionalmente genes marcadores selecionáveis. O cassete irá incluir na direção 5'-3' de transcrição, uma região de iniciação de transcrição e de tradução, uma sequência de DNA de interesse, e uma região de terminação de

transcrição e de tradução funcional em plantas. A região de terminação pode ser nativa com a região de iniciação de transcrição, pode ser nativa com a sequência de DNA de interesse, ou pode ser derivada de outra fonte. As regiões de terminação convenientes são disponíveis de Ti- plasmídeo de *A. tumefaciens*, como as regiões de terminação de octopina sintase e nopalina sintase, e outras descritas abaixo (ver também, Guerineau 1991; Proudfoot 1991; Sanfacon 1991; Mogen 1990; Munroe 1990; Ballas 1989; Joshi 1987).

"Vetor" é definido para incluir, inter alia, qualquer plasmídeo, cosmídeo, fato ou vetor binário de *Agrobacterium*, em forma circular ou linear de filamento duplo ou único, da qual ele pode ser ou não auto-transmissível ou mobilizável, e que pode transformar hospedeiro procariótico ou eucariótico por integração no genoma celular ou existir extra cromossomicamente (por exemplo, plasmídeo de replicação autônoma com uma origem de replicação).

São especificamente aqui incluídos os vetores shuttle o que se significa um veículo de DNA capaz, naturalmente ou por projeto, de replicação em dois organismos hospedeiros diferentes, que podem ser selecionados, por exemplo de actinomicetes e espécies relacionadas, como *Sacharomyces cerevisiae*, bactérias e eucarióticos (por exemplo células de planta superior, mamíferos, leveduras ou fungos)

Preferivelmente o ácido nucleico no vetor está sob o controle de, e operativamente ligado a um promotor apropriado, ou outros elementos regulatórios para a transcrição em uma célula hospedeira, como uma célula microbiana, por exemplo bacteriana, ou de planta. O vetor pode ser um vetor de expressão bi-funcional que funciona em hospedeiros múltiplos. No caso de DNA genômico, este pode conter seu próprio promotor ou outros elementos regulatórios e no caso de cDNA, este pode estar sob o controle de um promotor apropriado ou outros elementos regulatórios para expressão na célula hospedeira.

"Vetores de clonagem" tipicamente contém um ou um número pequeno de sítios de reconhecimento de endonuclease de restrição em que sequências de DNA estrangeiro podem ser inseridas em um modo determinável sem perda de função biológica essencial do vetor, assim como um gene marcador que é apropriado para uso na identificação e seleção de células transformadas com o vetor de clonagem. Os genes marcadores tipicamente incluem genes que provêm resistência a tetraciclina, resistência a higromicina ou resistência a ampicilina.

Uma "planta transgênica" é uma planta tendo uma ou mais células de planta que contém um vetor de expressão ou um cassete de expressão recombinante, como os cassetes de expressão da invenção.

"Tecido de planta" inclui tecidos ou plantas diferenciadas ou não diferenciadas, incluindo mas não limitado a raízes, caules, brotos, folhas, pólen, sementes, tecido de tumor e várias formas de células e cultura, como células únicas, protoplasto, embriões, e tecido de calo. O tecido de planta pode estar em plantas ou em cultura de órgão, tecido ou células.

Os seguintes termos são usados para descrever as relações de sequência entre dois ou mais ácidos nucleicos ou polinucleotídeos: (a) "sequência de referência", (b) "janela de comparação", (c) "identidade de sequência", (d) porcentagem de identidade de sequência, e (e) "identidade substancial".

(a) Como usado aqui, " sequência de referência " é uma sequência definida usada como uma base para a comparação de sequências. Uma sequência de referência pode ser um subconjunto ou a totalidade de uma sequência especificada; por exemplo, como um segmento de uma sequência de cDNA de comprimento completo ou de gene, ou a sequência de gene ou cDNA completo.

(b) Como usado aqui, "janela de comparação" faz referência a um segmento contíguo e especificado de uma sequência de polinucleotídeo

em que a sequência de polinucleotídeo na janela de comparação pode compreender adições ou deleções (isto é, espaços) comparados com a sequência de referência (que não compreende adições ou deleções) para um alinhamento ótimo das duas sequências. Geralmente, a janela de comparação tem pelo menos 20 nucleotídeos contíguos de comprimento, e opcionalmente pode ter 30, 40, 50, 100 ou mais. O versado na técnica entende que para evitar uma similaridade elevada a uma sequência de referência devido à inclusão de espaços na sequência de polinucleotídeos, uma penalidade de espaço é tipicamente introduzida e é subtraída do número de correspondências.

Os métodos de alinhamento de sequências para comparação são bem conhecidos na técnica. Assim, a determinação da identidade percentual entre quaisquer duas sequências pode ser obtida usando um algoritmo matemático. Os exemplos não limitativos preferidos destes algoritmos matemáticos são o algoritmo de Myers e Miller, 1988; o algoritmo de homologia local de Smith et al, 1981, o algoritmo de alinhamento de homologia de Needleman e Wunsch 1970; o método de busca para similaridade de Pearson e Lipman 1988; o algoritmo de Karlin e Altschul, 1990, modificado como em Karlin e Altschul, 1993.

As implementações de computador destes algoritmos matemáticos podem ser usadas para comparação de sequências para determinar a identidade de sequência. Estas implementações incluem, mas não são limitados a: CLUSTAL em programa PC/Gene (disponível de Intelligenetics, Mountain View, Calif.); o programa ALIGN (Versão 2.0) e GAP, BESTFIT, BLAST, FASTA, e TFASTA em Wisconsin Genetics Software Package, Versão 8 (disponível de Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alinhamentos usando estes programas podem ser realizados usando parâmetros default. O programa CLUSTAL é bem descrito (Higgins 1988, 1989; Corpet 1988; Huang 1992; Pearson 1994). O programa ALIGN é baseado no algoritmo de Myers e

Miller, *supra*. Os programas BLAST de Altschul *et al.*, 1990, são baseados no algoritmo de Karlin e Altschul, *supra*. Os alinhamentos múltiplos, (isto é, de mais do que 2 sequências) são preferivelmente realizados usando o algoritmo Clustal W (Thompson 1994; por exemplo, em software VectorNTI™, versão 9; Invitrogen Inc.) com a matriz de classificação BLOSUM62MT2 com as configurações de default (penalidade de abertura de espaço 15/19, penalidade de extensão de espaço 6.66/0.05; penalidade de separação de espaço, faixa 8; % identidade para retardo de alinhamento 40; usando espaços específicos de resíduo e espaços de resíduo hidrofílico).

Software para realizar análises BLAST é publicamente disponível através do National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo envolve primeiro identificar pares de sequência de classificação elevada (HSPs) por identificação de palavras curtas de comprimento W na sequência de busca, que ou correspondem ou atendem a algum escore de limiar de valor positivo T , quando alinhadas com uma palavra do mesmo comprimento em uma sequência de base de dados. T faz referência ao limiar do escore de palavra vizinha (Altschul 1990). Estas dicas de palavra vizinha inicial atuam como sementes para iniciar as buscas para encontrar HSPs mais longo contendo as mesmas. As dicas de palavras são então estendidas em ambas as direções ao longo de cada sequência de modo que o escore de alinhamento cumulativo pode ser aumentado. Os escores cumulativos são calculados usando, para sequências nucleotídicas, os parâmetros M (escore de retrocesso para um par de resíduos correspondentes, sempre > 0) e N (escore de penalidade para resíduos não correspondentes; sempre < 0). Para sequências de aminoácidos, uma matriz de classificação é usada para calcular o escore cumulativo. A extensão das dicas de palavra em cada direção é parada quando o escore de alinhamento cumulativo cai pela quantidade X de seu valor obtido máximo, o escore cumulativo vai a zero ou abaixo devido ao acúmulo de um ou mais

alinhamentos de resíduo de escore negativo, ou o final de cada sequência é alcançado.

Além de calcular a identidade de sequência percentual, o algoritmo BLAST também realiza uma análise estatística da similaridade entre duas sequências (ver, por exemplo, Karlin & Altschul (1993). Uma medida de similaridade provida pelo algoritmo BLAST é a probabilidade de soma menor ($P(N)$), que provê uma indicação da probabilidade pela qual uma correspondência entre duas sequências de nucleotídeo ou de aminoácidos poderia ocorrer por acaso. Por exemplo, uma sequência de ácido nucleico de teste é considerada similar a uma sequência de referência se a probabilidade de soma menor em uma comparação de sequência de ácido nucleico de teste para a sequência de ácido nucleico de referência for menor do que cerca de 0,1, mais preferivelmente menor do que cerca de 0,01, e o mais preferivelmente menor do que cerca de 0,001.

Para obter alinhamentos com espaço para fins de comparação, Gapped BLAST (em BLAST 2.0) pode ser usado como descrito em Altschul *et al.* 1997. Alternativamente, PSI-BLAST (em BLAST 2.0) pode ser usado para realizar uma busca iterada que detecta relações distantes entre as moléculas. Ver Altschul *et al.*, *supra*. Quando usando BLAST, Gapped BLAST, PSI-BLAST, os parâmetros de default dos programas respectivos (por exemplo BLASTN para sequências nucleotídicas, BLASTX para proteínas) pode ser usado. O programa BLASTN (para sequências nucleotídicas) usa, como defaults, um comprimento de palavra (W) de 11, uma expectativa (E) de 10, um corte de 100, $M=5$, $N=-4$, e uma comparação de ambos os filamentos. Para sequências de aminoácidos, programa BLASTP usa, como defaults, um comprimento de palavra (W) de 3, uma expectativa (E) de 10, e a matriz de escore BLOSUM62 (ver Henikoff & Henikoff, 1989). Ver <http://www.ncbi.nlm.nih.gov>. O alinhamento também pode ser realizado manualmente por inspeção.

Para fins da presente invenção, comparação de sequências nucleotídicas para determinação de identidade percentual de sequência para a sequência de promotores aqui descrita, é preferivelmente feita usando o programa BlastN (versão 1.4.7 ou posterior) com seus parâmetros de default ou qualquer programa equivalentes. Por " programa equivalente " destina-se qualquer programa de comparação de sequência que, para as duas sequências em questão, gera um alinhamento tendo correspondências de resíduo de nucleotídeo ou de aminoácido idênticas e uma identidade de sequência percentual idêntica quando comparadas com o alinhamento correspondente gerado pelo programa preferido.

(c) Como usado aqui, "identidade de sequência" ou "identidade" no contexto de duas sequências de ácido nucleico ou polipeptídeo faz referência aos resíduos em que as duas sequências que são iguais quando alinhadas para correspondência máxima sobre uma janela de comparação específica. Quando a porcentagem da identidade de sequência é usada na referência a proteínas, reconhece-se que as posições de resíduo que não são idênticas com frequência diferem por substituições de aminoácidos conservativas, onde os resíduos de aminoácidos são substituídos por outros resíduos de aminoácidos com propriedades químicas similares (por exemplo, carga ou hidrofobicidade) e, assim, não mudam as propriedades funcionais da molécula. Quando sequências diferem em substituições conservativas, a identidade de sequência percentual pode ser ajustada para cima para corrigir a natureza conservativa da substituição. As sequências que diferem por estas substituições conservativas são ditas como tendo "similaridade de sequência" ou "similaridade". Meios para fazer este ajuste são bem conhecidos do versado na técnica. Tipicamente, isto envolve classificar uma substituição conservativa como uma falta de correspondência parcial em vez de uma completa, assim aumentando a identidade de sequência percentual. Assim, por exemplo, onde um aminoácido idêntico recebe um escore de 1, e uma

substituição não conservativa recebe um escore de zero, uma substituição conservativa recebe um escore entre zero e 1. O escore de substituições conservativas é calculado, por exemplo, como implementado no programa PC/GENE (Intelligenetics, Mountain View, Calif.).

5 (d) Como usado aqui, "percentagem de identidade de sequência" significa o valor determinado por comparação de suas sequências otimamente alinhadas sobre uma janela de comparação, em que a promotor da sequência polinucleotídica na janela de comparação pode compreender adições ou deleções (isto é, espaços) comparados com a sequência de
10 referência (que não compreende adições ou deleções) para um alinhamento ótimo das duas sequências. A porcentagem é calculada por determinação do número de posições em que base de ácido nucleico idêntica ou resíduo de aminoácido ocorre em ambas as sequências para dar o número de posições correspondidas, dividindo o número de posições correspondidas pelo número
15 total de posições na janela de comparação e multiplicando o resultado por 100 para dar a porcentagem de identidade de sequência.

 (e) (i) O termo "identidade substancial" de sequências polinucleotídicas significa que um polinucleotídeo compreende uma sequência que tem pelo menos 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%,
20 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, ou 79%, preferivelmente pelo menos 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, ou 89%, mais preferivelmente pelo menos 90%, 91%, 92%, 93%, ou 94%, e o mais preferivelmente pelo menos 95%, 96%, 97%, 98%, ou 99% de identidade de sequência, comparado com a sequência de referência usando
25 um dos programas de alinhamento descrito usando parâmetros padrões. Um versado na técnica irá reconhecer que estes valores podem ser ajustados de modo apropriado para determinar a identidade de proteínas correspondente codificadas por duas sequências nucleotídicas ao levar em conta a degenerescência do códon, similaridade de aminoácidos, posicionamento na

matriz de leitura e outros. A identidade substancial de sequências de aminoácidos para estes fins normalmente significa identidade de sequência de pelo menos 60% ou 70%, mais preferivelmente pelo menos 80%, 90%, e o mais preferivelmente pelo menos 95%.

5 Outra indicação que as sequências nucleotídicas são substancialmente idênticas é se as duas moléculas hibridizam uma em cada outra sob condições estringentes (ver abaixo). Geralmente, condições estringentes são selecionadas para ser cerca de 5°C menores do que o ponto de fusão térmico (T_m) para a sequência específica em uma potência iônica e
10 em pH definidos. No entanto, condições estringentes englobam temperaturas na faixa de cerca de 1°C a cerca de 20°C, dependendo do grau de estringência desejada, como de outra forma qualidade aqui. Os ácidos nucleicos que não hibridizam um no outro sob condições estringentes são ainda substancialmente idênticos se os polipeptídeos que eles codificam são
15 substancialmente idênticos. Isto pode ocorrer, por exemplo, quando uma cópia de um ácido nucleico é criada usando a degenerescência de códon máxima permitida pelo código genético. Uma indicação de que duas sequências de ácido nucleico são substancialmente idênticas é quando o polipeptídeo codificado pelo primeiro ácido nucleico é imunologicamente
20 reativo cruzado com o polipeptídeo codificado pelo segundo ácido nucleico.

(ii) O termo "identidade substancial" no contexto de um peptídeo indica que um peptídeo compreende a sequência com pelo menos 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, ou 79%, preferivelmente 80%, 81%, 82%,
25 83%, 84%, 85%, 86%, 87%, 88%, ou 89%, mais preferivelmente pelo menos 90%, 91%, 92%, 93%, ou 94%, ou ainda mais preferivelmente, 95%, 96%, 97%, 98% ou 99%, identidade de sequência para a sequência de referência sobre a janela de comparação especificada. Preferivelmente, o alinhamento ótimo é conduzido usando o algoritmo de alinhamento de homologia de

Needleman e Wunsch (1970). Uma indicação de que duas sequências de peptídeos são substancialmente idênticas é que uma em que um peptídeo é imunologicamente reativo com anticorpos criados contra o segundo peptídeo. Assim, um peptídeo é substancialmente idêntico a um segundo peptídeo, por exemplo, onde os dois peptídeos diferem somente por uma substituição conservativa.

Para comparação de sequência, tipicamente uma sequência atua como uma sequência de referência em que as sequências de teste são comparadas. Quando usando um algoritmo de comparação de sequência, as sequências de teste e referência são entradas em um computador, as coordenadas da subsequência são designadas, se necessário, e os parâmetros do programa de algoritmo de sequência são designados. O algoritmo de comparação de sequência então calcula a identidade de sequência percentual para a (s) sequência (s) de teste com relação à sequência de referência, com base nos parâmetros de programa designados.

Como notado acima, outra indicação de que duas sequências de ácido nucleico são substancialmente idênticas é que as duas moléculas hibridizam uma com a outra sob condições estridentes. A frase "hibridizando especificamente para" refere-se à ligação, duplação ou hibridização de uma molécula somente em uma sequência nucleotídica particular sob condições estridentes quando esta sequência está presente em uma mistura complexa (por exemplo, celular total) DNA ou RNA. "Liga (m) substancialmente" refere-se a hibridização complementar entre ácido nucleico de sonda e um ácido nucleico alvo e engloba faltas de correspondências pequenas que podem ser acomodadas por redução da estridência do meio de hibridização de modo a obter a detecção desejada da sequência de ácido nucleico alvo.

"Condições de hibridização estridentes" e "condições de lavagem de hibridização estridentes" no contexto das experiências de hibridização de ácido nucleico, como hibridização Southern e Northern são

dependentes de sequência, e são diferentes sob diferentes parâmetros ambientais⁹s. T_m é a temperatura (sob potência iônica e pH definidos, em que 50% da sequência alvo hibridizam para uma sonda perfeitamente correspondida. A especificidade é tipicamente a função de lavagens de pós-hibridização, os fatores críticos sendo a potência iônica e a temperatura da solução de lavagem final. Para híbridos de DNA-DNA, o T_m pode ser aproximado da equação de Meinkoth e Wahl, 1984:

$$T_m = 81,5^{\circ}\text{C} + 16,6 (\log_{10} M) + 0,41 (\%GC) - 0,61 (\% \text{ forma}) - 500 / L$$

Onde M é a molaridade de cátions monovalentes, % GC é a porcentagem de nucleotídeos guanina e citosina no DNA. % forma é a porcentagem de formamida na solução de hibridização, e L é o comprimento do híbrido nos pares de base. T_m é reduzido por cerca de 1 °C para cada 1% de falta de correspondência, assim hibridização T_m e/ou condições de lavagem podem ser ajustado para hibridizar em sequências da identidade desejada. Por exemplo, se sequências com >90% identidade são procuradas, o T_m pode ser diminuído em 10°C. Geralmente, condições estridentes são selecionadas para ser cerca de 5°C menor do que o ponto de fusão térmico I para a sequência específica e seu complemento em uma potência iônica e pH definidos. No entanto, condições severamente estridentes podem usar uma hibridização e/ou lavagem a 1, 2, 3, ou 4°C menores do que o ponto de fusão térmico I; condições moderadamente estridentes podem usar uma hibridização e/ou lavagem a 6, 7, 8, 9, ou 10°C menor do que ponto de fusão térmico I; condições de baixa estringência podem usar uma hibridização e/ou lavagem a 11, 12, 13, 14, 15, ou 20°C menor do que o ponto de fusão térmico I. Usando a equação, as composições de hibridização e lavagem, e T desejado, o versado na técnica irá compreender que variações na estringência de soluções de hibridização e/ou lavagem são inerentemente descritas. Se o grau desejado de falta de correspondência resultar em um T de menos do que 45°C (solução aquosa) ou 32°C (solução de formamida), prefere-se aumentar a

concentração de SSC de modo que uma maior temperatura pode ser usada. Um guia extensivo para a hibridização de ácidos nucleicos é encontrado em Tijssen, 1993. Geralmente, condições de hibridização e de lavagem são selecionadas para serem cerca de 5°C, menores do que o ponto de fusão térmica T_m para a sequência específica na potência iônica e pH definidos.

Um exemplo de condições de lavagem altamente estridentes é 0,15 M NaCl a 72°C durante cerca de 15 minutos. Um exemplo de estridentes condições de lavagem é uma lavagem 0,2 X SSC a 65°C durante 15 minutos (ver, Sambrook, *infra*, para uma descrição de tampão SSC). Com frequência, a lavagem de alta estridência é precedida por uma lavagem de baixa estridência para remover o sinal de sonda de fundo. Um exemplo de lavagem de estridência média para um duplex de, por exemplo, mais do que 100 nucleotídeos, é 1 X SSC a 45°C durante 15 minutos. Um exemplo de lavagem de baixa estridência para um duplex de, por exemplo, mais do que 100 nucleotídeos, é 4 a 6 X SSC a 40°C durante 15 minutos. Para sondas curtas (por exemplo, cerca de 10 a 50 nucleotídeos), condições estridentes tipicamente envolvem concentrações de sal de menos do que cerca de 1,5 M, mais preferivelmente cerca de 0,01 a 1,0 M, concentração íon Na (ou outros sais) a pH 7,0 a 8,23, e a temperatura é tipicamente pelo menos cerca de 30°C e pelo menos cerca de 60°C para sondas longas (por exemplo, >50 nucleotídeos). Condições estridentes também podem ser obtidas com a adição de agentes desestabilizantes, como formamida. Em geral, uma relação de sinal para ruído de 2 X (ou maior) do que o observado para uma sonda não relacionada no teste particular de hibridização indica a detecção da hibridização específica. Os ácidos nucleicos que não hibridizam um em cada outro sob condições estridentes são ainda substancialmente idênticos se as proteínas que eles codificam forem substancialmente idênticas. Isto ocorre, por exemplo, quando uma cópia de um ácido nucleico é criada usando a degenerescência de códon máxima permitida pelo código genético.

As condições muito estringentes são selecionadas para serem iguais ao T_m para uma sonda particular. Um exemplo de condições estringentes para hibridização de ácidos nucleicos complementares que tem mais do que 100 resíduos complementares em um filtro em um Southern ou Northern blot é 50% formamida, por exemplo, hibridização em 50% formamida, 1 M NaCl, 1% SDS a 37°C, e uma lavagem em 0,1 x SSC a 60 to 65°C. As condições exemplares de baixa estringência incluem hibridização com uma solução tampão de 30 a 35% formamida, 1 M NaCl, 1% SDS (dodecil sulfato de sódio) a 37°C, e uma lavagem em 1 X a 2 X SSC (20 X SSC=3,0 M NaCl/0.3 M citrato trissódico) a 50 a 55°C. As condições de estringência moderadas exemplares incluem hibridização em 40 a 45% formamida, 1,0 M NaCl, 1% SDS a 37°C, e uma lavagem em 0,5 X a 1 X SSC a 55 a 60°C.

Os seguintes são exemplos de conjuntos de condições de hibridização / lavagem que podem ser usados para clonar sequências nucleotídicas ortólogas que são substancialmente idênticas a sequências nucleotídicas de referência da presente invenção: uma sequência nucleotídica de referência preferivelmente hibridiza para a sequência nucleotídica de referência em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C.

"Embaralhamento de DNA" é um método para introduzir

mutações ou rearranjos, preferivelmente aleatoriamente, em uma molécula de DNA ou para gerar trocas de sequências de DNA entre duas ou mais moléculas de DNA, preferivelmente aleatoriamente. A molécula de DNA resultante do embaralhamento de DNA é uma molécula de DNA embaralhada que é uma molécula de DNA não de ocorrência natural derivada de pelo menos uma molécula de DNA gabarito. O DNA embaralhado preferivelmente codifica um polipeptídeo variante modificado com relação ao polipeptídeo codificado pelo DNA gabarito, e pode ter uma atividade biológica alterada com relação ao polipeptídeo codificado pelo DNA gabarito.

"Molécula de DNA recombinante" é uma combinação de sequências de DNA que são unidas juntas usando tecnologia de DNA recombinante e procedimentos usados para unir juntas as sequências de DNA como descrito, por exemplo, em Sambrook *et al.*, 1989.

A presente invenção é especialmente utilizável para aplicações em plantas monocotiledôneas. O termo "planta monocotiledônea" inclui plantas de uma variedade de níveis de plóides, incluindo aneuplóide, poliplóide, diplóide, haplóide e hemizigoto. São ainda incluídas as plantas maduras, sementes, brotos e mudas, e partes, material de propagação (por exemplo sementes e frutas) e culturas, por exemplo culturas de células, derivadas das mesmas. As plantas monocotiledôneas anuais e perenes são organismos hospedeiros preferidos para a geração de plantas transgênicas.

Além disso, a presente invenção inclui plantas maduras, sementes, brotos e mudas, e partes, material de propagação e culturas derivadas dos mesmos, por exemplo, culturas de células. As plantas maduras se refere às plantas em qualquer estágio de desenvolvimento além do de muda. O termo muda refere-se a uma planta imatura jovem em um estágio de desenvolvimento inicial em que ela é ainda dependente dos assimilados armazenados dentro da semente (por exemplo, no endosperma, perisperma ou cotilédones. São incluídos todos os gêneros das subfamílias Bambusoideae

(por exemplo, o gênero bambu), eropogonoideae (por exemplo, o gênero *Saccharum*, *Sorghum*, ou *Zea*), Arundineae (por exemplo, o gênero *Phragmites*), Oryzoideae (por exemplo, o gênero *Oryza*), Panicoideae (por exemplo, o gênero *Panicum*, *Pennisetum*, e *Setaria*), Pooideae (Festuciadeae)

5 (por exemplo, o gênero *Poa*, *Festuca*, *Lolium*, *Trisetum*, *Agrostis*, *Phleum*, *Dactylis*, *Alopecurus*, *Avena*, *Triticum*, *Secale*, e *Hordeum*). São preferidos *Avena sativa* (aveia), *Bambusa* sp. e *Bambusa bambos* (bambu), *Saccharum officinarum* (cana de açúcar), *Triticum dicoccum* (trigo Emmer), *Triticum monococcum* (trigo Einkorn), *Triticum spelta* (trigo espelta), *Triticum durum*

10 (trigo), *Triticum turgidum*, *Triticum aestivum* (trigo), *Zea mays* (milho/milho branco), *Panicum miliaceum* (milhete comum), *Pennisetum thipoides* (milheto peniseto), *Hordeum vulgare* ou *H. sativum* (cevada), *Oryza sativa* (arroz), *Zizania aquatica* (arroz selvagem), *Secale cereale* (centeio), *Sorghum bicolor* (*S. vulgare*) (sorgo). Mais preferidos são trigo (*Triticum* spp.), arroz

15 (*Oryza* spp.), centeio (*Hordeum* spp.), aveias (*Avena* spp.), centeio (*Secale* spp.), milho (*Zea mays*), sorgo e milheto (*Pennisetum* spp). São preferidas todas as espécies de trigo, especialmente da família *Triticum* (incluindo tanto trigo do inverno como da primavera), mais especialmente *Triticum* spp.: comum (*T. aestivum*), duro (*T. durum*), espelta (*T. spelta*), *Triticum dicoccum*

20 (trigo Emmer), *Triticum turgidum*, e *Triticum monococcum* (trigo Einkorn), com *T. aestivum* sendo particularmente preferido. O método da invenção pode ser usado para produzir plantas transgênicas a partir de trigos da primavera, como, por exemplo, Bobwhite, Marshall, PIVOT1, UC702, e Panewawa assim como de trigos do inverno, como, por exemplo, HY368, Neeley,

25 FL302, RH91, R332, R1269 e R585. Outros genótipos de trigo apropriados incluem, mas não são limitados a Yecora Rojo, Karl e Anza. No entanto, deve ser apontado que a invenção não é limitada a algumas variedades mas é altamente independente do genótipo.

A palavra "planta" refere-se a qualquer planta, particularmente

a plantas agronomicamente utilizáveis (por exemplo plantas de semente) e "célula de planta" é uma unidade estrutural e fisiológica da planta que compreende uma parede de células mas pode também fazer referência a um protoplasto. A célula de planta pode estar na forma de uma célula única isolada ou uma célula cultivada, ou como uma parte de unidade organizada superior como, por exemplo, um tecido de planta, ou um órgão de planta diferenciado em uma estrutura que está presente em qualquer estágio de um desenvolvimento de planta. Estas estruturas incluem um ou mais órgãos de planta incluindo, mas não limitados a fruto, broto, caule, folha, pétala da flor, etc. Preferivelmente, o termo "planta" inclui plantas completas, órgãos / estruturas vegetativos do broto, (por exemplo folhas, caules e tubérculos), raízes, flores, e órgãos / estruturas florais (por exemplo brácteas, sépalas, pétalas, estames, carpelos, anteros e óvulos⁰, sementes (incluindo embrião, endosperma, e revestimento da semente) e frutas (o ovário maduro), tecidos de plantas (por exemplo, tecido vascular, tecido triturado, e outros) e células (por exemplo células de guarda, células de ovos, tricomas e outros) e a progênie dos mesmos.

"Aumento significativo" é um aumento que é maior do que a margem de erro inerente na técnica de medição, preferivelmente um aumento em cerca de 2 vezes ou maior.

"Significativamente menor" significa que a diminuição é maior do que a margem de erro inerente na técnica de medição, preferivelmente uma diminuição em cerca de 2 vezes ou maior.

DESCRIÇÃO DETALHADA DA INVENÇÃO

Uma primeira forma de realização da invenção refere-se a uma planta ou célula de planta compreendendo um cassete de expressão, referido cassete de expressão compreendendo

a) uma primeira molécula de ácido nucleico selecionada dentre o grupo consistindo de

i) um polinucleotídeo como definido em SEQ ID NO:1;

ii) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70%, ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:1; e

iii) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e operativamente ligada à mesma

b) uma segunda molécula de ácido nucleico selecionada dentre o grupo consistindo de

i) um polinucleotídeo como definido em SEQ ID NO:3;

ii) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70%, ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:3; e

iii) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais

preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

e operativamente ligado a pelo menos um ácido nucleico que é heterólogo em relação à referida primeira ou referida segunda sequência de ácido nucleico de uma sequência reguladora da transcrição e é capaz de modo apropriado de conferir a uma planta um rendimento aumentado, e/ou tolerância ao estresse aumentada, e/ou qualidade nutricional e/ou teor de óleo aumentados de uma semente ou um broto.

Preferivelmente, a sequência nucleotídica reguladora da transcrição quimérica leva o referido DNA heterólogo a ser predominantemente expressado no embrião germinando. Este perfil de expressão é especialmente utilizável para as seguintes aplicações:

i) melhorada resistência contra fatores de estresse: como descrito acima na seção da técnica anterior, o embrião é muito sensível contra todos os tipos de fatores de estresse biótico e abiótico (seca, frio, doenças, etc). Estes fatores de estresse tem um efeito imediato sobre o rendimento e qualidade da cultura. A maior parte dos promotores bem conhecidos não tem ou tem uma capacidade de expressão baixa durante este estágio. A especificidade de regulação da transcrição descrita aqui é especialmente utilizável para expressar genes de resistência ao estresse "em demanda", isto é, no tempo correto em níveis elevados. Além disso, devido à especificidade

no endosperma amidoso, é possível procurar novos modos de resistência ao esperma. Devido ao endosperma amidoso ser o tecido, que nutre o embrião, pode-se aumentar a resistência ao estresse via suplementação aumentada no embrião com nutrientes.

5 ii) qualidade nutricional aumentada de uma semente ou um broto: O perfil de expressão das sequências reguladoras da transcrição de ácido nucleico quiméricas permite a conversão de ingredientes da semente (cerne) ou para mudar a distribuição dos ingredientes na semente. Por exemplo, pode-se converter carboidratos (amido) em óleo ou outros
10 ingredientes de alto valor (por exemplo vitaminas) ou pode-se deslocar a localização dos ingredientes do endosperma para o embrião, assim provendo brotos com melhorado valor nutricional.

 iii) rendimento aumentado: Rendimento aumentado está parcialmente relacionado com a resistência ao estresse (ver acima sob i)). No
15 entanto, o perfil de expressão das sequências reguladoras da transcrição de ácido nucleico quiméricas permite mesmo sem fatores de estresse o aumento do rendimento por otimização do crescimento do embrião, que irá diretamente afetar o crescimento da muda. Pode-se também obter uma germinação prematura sob condições de campo e outros traços, que irão levar a um maior
20 ou rendimento prematuro.

 iv) excisão de sequência marcada. Como descrito acima a excisão homogênea de sequências, especialmente sequências de marcador, é um aspecto ainda não resolvido no campo de biotecnologia. A maior parte das plantas demonstra padrões de excisão de tipo mosaico, com áreas de excisão
25 com sucesso e áreas sem excisão. Para obter uma excisão homogênea ou substancialmente homogênea, o mecanismo de excisão precisa ser ativado preferivelmente em um estágio prematuro de desenvolvimento, quando o organismo não consiste de muitas plantas. Além disso, a ativação (isto é, expressão de enzima mediando a excisão) precisa ser forte. Ambas as

exigências são atendidas pelo perfil de expressão das sequências de ácido nucleico reguladoras da transcrição quiméricas descritas aqui. A atividade de transcrição forte em germinação de embriões prematura permite uma excisão do marcador eficiente neste estágio, a partir do qual uma planta com
 5 sequência alvo livre (por exemplo livre de marcador) é gerada.

"Transcrição específica de embrião germinando" no contexto desta invenção significa que a transcrição de uma sequência de ácido nucleico por um elemento de regulação da transcrição em um modo que a transcrição de referida sequência de ácido nucleico na planta germinante, preferivelmente
 10 o embrião germinando contribui para mais do que 90%, preferivelmente mais do que 95%, mais preferivelmente mais do que 99% da quantidade completa do RNA transcrito a partir da referida sequência de ácido nucleico na planta completa, semente ou broto durante o estágio de desenvolvimento específico.

1. A sequência de ácido nucleico reguladora da transcrição quimérica

15 Em sua forma a mais geral, a molécula de ácido nucleico compreendendo um polinucleotídeo codificando a sequência nucleotídica reguladora da transcrição quimérica compreende

i) uma primeira molécula de ácido nucleico selecionada dentre o grupo consistindo de

20 a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70%, ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:1; e

25 c) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais

preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e

ii) uma segunda molécula de ácido nucleico selecionada dentre

o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:3;

b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70%, ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:3; e

c) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

em que referida primeira e referida segunda sequências de

ácido nucleico são funcionalmente ligadas e heterólogas a cada outra.

O termo "derivado" quando usado no contexto de regiões de DNA como promotores, sequências de ácido nucleico reguladoras da transcrição, ou sequências ativantes a montante, refere-se a situações onde a região do DNA que é "derivada" é obtida a partir de ou com base na região de DNA de ocorrência natural ou outra região de DNA de fonte. A região de DNA que é "derivada" pode diferir, geralmente através de mutação deliberada, a partir da região de DNA de ocorrência natural ou outra região de DNA de fonte.

1.1 Derivados e variantes da sequência nucleotídica reguladora da transcrição quimérica da invenção e seus elementos funcionais

A invenção descrita aqui contempla que, além das específicas sequências nucleotídicas reguladoras da transcrição quiméricas e seus elementos específicos descritos aqui, derivados e variantes das referidas sequências podem ser empregados.

Derivados de sequências nucleotídicas reguladoras da transcrição quiméricas e seus elementos específicos podem incluir, mas não são limitados a, deleções de sequência, mutações de ponto único ou múltiplo, alterações em um sítio de enzima de restrição particular, sítio de enzima, adição de elementos funcionais, ou outros meios de modificação molecular. Esta modificação pode ou não melhorar, ou de outra forma alterar a atividade de regulação da transcrição de referidas sequências.

Por exemplo, o versado na técnica pode delimitar os elementos funcionais dentro das sequências e deletar quaisquer elementos não essenciais. Os elementos funcionais podem ser modificados ou combinados para aumentar a utilidade ou expressão das sequências da invenção para qualquer aplicação em particular. Os fragmentos funcionalmente equivalentes de uma sequência nucleotídica reguladora da transcrição da invenção também podem ser obtidos por remoção ou deleção de sequências não essenciais sem

deletar as essenciais. A limitação da sequência nucleotídica reguladora da transcrição a seus elementos de mediação da transcrição essenciais pode ser realizada *in vitro* por mutações de deleção de tentativa-e-erro, ou *in silico* usando rotinas de pesquisa de elemento de promotor. As regiões essenciais para a atividade do promotor com frequência demonstram agrupamentos de alguns elementos promotores conhecidos. Esta análise pode ser realizada usando algoritmos de computador como PLACE ("Plant Cis-atuantes Regulatory DNA Elements"; Higo 1999), a base de dados BIOBASE "Transfac" (Biologische Datenbanken GmbH, Braunschweig; Wingender 2001) ou a base de dados PlantCARE (Lescot 2002). Especialmente preferidos são os fragmentos equivalentes de sequências nucleotídicas reguladoras da transcrição, que são obtidos por deleção da região codificando a região 5'- não traduzida do mRNA, assim somente provendo a região de promotor (não transcrita). A região 5'- não traduzida pode ser facilmente determinada por métodos bem conhecidos na técnica (como análise 5'- RACE). Assim, algumas das sequências nucleotídicas reguladoras da transcrição invenção são fragmentos equivalentes de outras sequências.

Como indicado acima, mutantes de deleção do promotor da invenção podem ser aleatoriamente preparados e então testados. Com esta estratégia, uma série de construções são preparadas, cada contendo uma porção diferente do clone (um subclone), e estas construções são então triadas para atividade. Um meio apropriado para triar a atividade consiste em fixar uma construção de promotor deletada, que contém um segmento deletado para um marcador selecionável ou triável, e isolar somente as células expressando o gene marcador. Deste modo, várias construções diferentes, de promotor deletado, são identificadas que ainda retém a atividade desejada, ou mesmo melhorada. O menor segmento, que é requerido para atividade, é assim identificado através de comparação das construções selecionadas. Este segmento pode então ser usado para a construção de vetores para a expressão

de genes exógenos.

Os meios para mutageneizar ou criar deleções em um segmento de DNA codificando qualquer sequência de promotor são bem conhecidos do versado na técnica e descritos por exemplo na patente US 6.583.338, incorporada aqui por referência em sua totalidade. Algumas sequências nucleotídicas variantes da presente invenção retêm a atividade biológica (isto é, regulam a transcrição com um perfil como definido acima). Um exemplo de uma variante de sequência regulatória é um promotor formado por uma ou mais deleções a partir de um promotor maior. A porção 5' de um promotor até a caixa TATA próximo de um sítio de início de transcrição pode ser às vezes deletado sem abolir a atividade do promotor, como descrito por Zhu *et al.*, (1995) The Plant Cell 7:1681-1689. Um modo de rotina para remover parte de uma sequência de DNA consiste em usar uma exonuclease em combinação com uma amplificação de DNA para produzir deleções abrigadas unidirecionais de clones de DNA de filamento duplo. Um kit comercial para este fim é vendido sob o nome comercial Exo-Size.TM. (New England Biolabs, Beverly, Mass.). As variantes biologicamente ativas também incluem, por exemplo, as sequências de promotor nativas da invenção tendo uma ou mais substituições, deleções ou inserções de nucleotídeos.

Derivados e variantes também incluem homólogos, parálogos e ortólogos de outras espécies, como mas não limitadas to, bactérias, fungos, e plantas. "Homólogo" é um termo genético usado na técnica para indicar uma sequência de polinucleotídeo ou polipeptídeo possuindo um grau elevado de parentesco de sequência para uma sequência de referência. Este parentesco pode ser quantificado por determinação do grau de identidade e/ou similaridade entre as duas sequências, como acima definido. Dentro deste termo genérico, estão os termos "ortólogo", e "parálogo". "Parálogo" refere-se a um polinucleotídeo ou polipeptídeo dentro da mesma espécie que é

funcionalmente similar. "Ortólogo" refere-se a um polinucleotídeo ou polipeptídeo que é o equivalente funcional do polinucleotídeo ou polipeptídeo em outra espécie. Um gene ortólogo significa preferivelmente um gene, que está codificando uma proteína ortóloga. Mais especificamente, o termo

5 "ortólogo" denota um polipeptídeo ou proteína obtido de uma espécie que é a contra-parte funcional de um polipeptídeo ou proteína de uma espécie diferente. As diferenças de sequência dentre os ortólogos são o resultado de especialização.

Preferivelmente, a atividade de regulação da transcrição de

10 uma variante ou derivado de sequências nucleotídicas reguladoras da transcrição quiméricas é substancialmente igual (ou equivalente) que para as sequências nucleotídicas reguladoras da transcrição líquido especificamente descritas aqui, isto é, que a expressão é regulada no modo específico do embrião germinando como descrito acima. Além disso, a atividade de

15 regulação da transcrição de um derivado ou variante pode variar da atividade de sua sequência parental, especialmente com relação ao nível de expressão. O nível de expressão pode ser maior ou menor do que o nível de expressão da sequência parental. Ambas as derivações podem ser vantajosas dependendo da sequência de ácido nucleico de interesse a ser expressada. As preferidas

20 são as sequências funcionais equivalentes, que – em comparação com sua sequência parental – não deriva do nível de expressão de referida sequência parental por mais do que 50%, preferivelmente 25%, mais preferivelmente 10% (como a ser preferivelmente julgado ou por expressão de mRNA ou proteína (por exemplo, expressão de gene repórter)). Ainda mais preferidas

25 são as sequências equivalentes que demonstram uma expressão aumentada em comparação com sua sequência parental, preferivelmente um aumento em pelo menos 50%, mais preferivelmente em pelo menos 100%, o mais preferivelmente em pelo menos 500%. Este perfil de expressão é preferivelmente demonstrado usando genes repórter operativamente ligados à

referida sequência nucleotídica reguladora da transcrição. Os genes repórter preferidos (Schenborn 1999) neste contexto são proteína de fluorescência verde (GFP) (Chui 1996; Leffel 1997), cloranfenicol transferase, luciferase (Millar 1992), β -glucuronidase ou β -galactosidase. Especialmente preferida é a β -glucuronidase (Jefferson 1987). Outros métodos para testar a regulação são bem conhecidos na técnica e incluem Northern blots, e RT-PCR (ver, por exemplo, Sambrook et al., supra, aqui incorporados por referência).

Em uma forma de realização preferida a sequência nucleotídica reguladora da transcrição é descrita por uma molécula de ácido nucleico isolada compreendendo um polinucleotídeo codificando a sequência selecionada dentre o grupo consistindo de

i) a sequência descrita por SEQ ID NO:1;

ii) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas da sequência descrita por SEQ ID NOs:1,

iii) a sequência nucleotídica tendo a sequência identidade de pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% para a sequência descrita por SEQ ID NO:1,

iv) a sequência nucleotídica capaz de hibridizar (preferivelmente sob condições de baixa estringência, mais preferivelmente sob condições de média estringência, o mais preferivelmente sob condições de alta estringência como definido acima na seção DEFINIÇÃO; por exemplo sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a

50°C, preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C) to a sequência descrita por SEQ ID NO: 1, ou o complemento do mesmo;

v) a sequência nucleotídica capaz de hibridizar (preferivelmente sob condições de baixa estringência, mais preferivelmente sob condições de média estringência, o mais preferivelmente sob condições de alta estringência como definido acima na seção DEFINIÇÃO; por exemplo sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C) para um ácido nucleico compreendendo 50 a 200 ou mais nucleotídeos consecutivos (como 50 ou 100, preferivelmente 150 ou 200, mais preferivelmente 250 ou 400 nucleotídeos consecutivos, o mais preferivelmente a sequência completa) de uma sequência descrita por SEQ ID NO: 1, ou o complemento do mesmo;

vi) a sequência nucleotídica que é o complemento ou complemento reverso de qualquer uma das previamente mencionadas sequências nucleotídicas sob i) a v).

Em outra forma de realização preferida a sequência ativante a montante é descrita por uma molécula de ácido nucleico compreendendo um polinucleotídeo codificando a sequência selecionada dentre o grupo

consistindo de

i) a sequência descrita por SEQ ID NO: 3,

ii) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas da sequência descrita por SEQ ID NO:3,

iii) a sequência nucleotídica tendo a sequência identidade de pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% para a sequência descrita por SEQ ID NO: 3,

iv) a sequência nucleotídica capaz de hibridizar (preferivelmente sob condições de baixa estringência, mais preferivelmente sob condições de média estringência, o mais preferivelmente sob condições de alta estringência como definido acima na seção DEFINIÇÃO; por exemplo sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C) to a sequência descrita por SEQ ID NO: 3, ou o complemento do mesmo;

v) a sequência nucleotídica capaz de hibridizar (preferivelmente sob condições de baixa estringência, mais preferivelmente sob condições de média estringência, o mais preferivelmente sob condições de alta estringência como definido acima na seção DEFINIÇÃO; por exemplo sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio

(SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C) para um ácido nucleico compreendendo 50 a 200 ou mais nucleotídeos consecutivos (como 50 ou 100, preferivelmente 150 ou 200, mais preferivelmente 250 ou 400 nucleotídeos consecutivos, o mais preferivelmente a sequência completa) de uma sequência descrita por SEQ ID NO: 3, ou o complemento do mesmo;

vi) a sequência nucleotídica que é o complemento ou complemento reverso de qualquer uma das previamente mencionadas sequências nucleotídicas sob i) a v).

As sequências especificadas sob ii), iii), iv) v) e vi) de qualquer uma das c sequências reguladoras da transcrição quiméricas especificadas definidas acima são preferivelmente capazes de modificar a transcrição em uma célula ou organismo de planta monocotiledônea, mais preferivelmente elas são capazes de induzir a expressão específica de embrião. Preferivelmente, as sequências especificadas sob iv) ou v) estão hibridizando sob condições estridentes com a sequência alvo especificada.

Preferivelmente, a identidade das sequências nucleotídicas é determinada usando o programa BlastN (versão 1.4.7 ou posterior) com seus parâmetros de default (comprimento da palavra (W) de 11, uma expectativa (E) de 10, um corte de 100, M=5, N=-4, e uma comparação de ambos os filamentos) ou qualquer programa equivalente.

Em técnicas de hibridização, toda ou parte de uma sequência

de nucleotídeo conhecida é usada como uma sonda que seletivamente hibridiza em outras sequências nucleotídicas correspondentes presentes em uma população de fragmentos de DNA genômico clonado ou fragmentos de cDNA (isto é, bibliotecas de genômico ou cDNA) de um organismo selecionado. As sondas de hibridização podem ser fragmentos de DNA genômico, fragmentos de cDNA, fragmentos de RNA ou outros oligonucleotídeos, e podem ser rotuladas com um grupo detectável como ^{32}P , ou qualquer outro marcador detectável. Assim, por exemplo, sondas para a hibridização podem ser feitas por rotulação de oligonucleotídeos sintéticos com base na sequência da invenção. Os métodos para a preparação de sondas para hibridização e para a construção de bibliotecas de cDNA e genômico são geralmente bem conhecidos na técnica e descritos em Sambrook *et al.* (1989). Em geral, sequências que hibridizam para as sequências descritas aqui terão pelo menos cerca de 60% a 70% e ainda cerca de 80% 85%, 90%, 95% a 98% ou mais identidade com as sequências descritas. Isto é, a similaridade de sequência de sequências pode estar na faixa, compartilhando pelo menos cerca de 60% a 70%, e ainda cerca de 80%, 85%, 90%, 95% a 98% de similaridade de sequência.

Um motivo típico de promotor de núcleo da presente invenção, como mostrado em tabelas 8 e 9, compreende quatro nucleotídeos bem definidos flanqueados por um cordão de quaisquer nucleotídeos em ambas as extremidades 5' e 3' do motivo do promotor. Por exemplo, motivo de promotor GAPB/GAP.01 compreende 15 nucleotídeos “aaaaATGAatagaaa” como definido na SEQ ID NO:26, em que “nnnnATGAnnnnnnn” é motivo de promotor de núcleo GAPB/GAP.01 como descrito em SEQ ID NO:27, em que n é selecionado dentre o grupo consistindo de a, c, t e g. O número de n ou na ext 5' ou na extremidade 3' do motivo de promotor de núcleo corresponde a um número de nucleotídeos flanqueando o motivo de promotor de núcleo no motivo de promotor correspondente na extremidade 5' ou 3',

respectivamente. A porcentagem de identidade do motivo de promotor de núcleo para o motivo de promotor correspondente é calculada, por exemplo, como ilustrado abaixo. Quando dois “n” no motivo de promotor de núcleo GAPB/GAP.01 de SEQ ID NO:26 são idênticos aos nucleotídeos nas posições correspondentes do motivo de promotor GAPB/GAP.01 de SEQ ID NO:27, por exemplo, $n(2) = a$ e $n(4) = a$, a identidade percentual é determinada por divisão do número total dos nucleotídeos idênticos (incluindo os quatro nucleotídeos de núcleo) pelo comprimento completo do motivo do promotor, $6/15 = 40\%$. Quando cinco “n” em motivo de promotor de núcleo GAPB/GAP.01 de SEQ ID NO:26 são idênticos aos nucleotídeos nas posições correspondentes do motivo de promotor GAPB/GAP.01 de SEQ ID NO:27, por exemplo, $n(1) = a$, $n(3) = a$, $n(10) = t$, $n(12) = g$, e $n(14) = a$, a identidade percentual é $9/15 = 60\%$.

Os motivos de promotor e motivos de promotor de núcleo identificado em Ar.cor78 são como mostrados em tabelas 8 e 9. Mais especificamente, SEQ ID NOs: 10, 12, 14, 17, 20, 22, 25, 27, 29, 31, 33, 35, 37, 39, 42, 44, 46, 48, 50, 53, 55, 57, 59, 63, 66, 68, 70, 72, 74, 77, 79, 82, 84, 86, 88, 90, 94, 96, 98, 102, 104, 108, 112, 114, 117, 121, 123, 125, 127, 129, 131, 137, e 138 são motivos de promotor de núcleo identificados em promotor At. cor78. SEQ ID NOs: 9, 11, 13, 15, 16, 18, 19, 21, 23, 24, 26, 28, 30, 32, 34, 36, 38, 40, 41, 43, 45, 47, 49, 51, 52, 54, 56, 58, 60, 61, 62, 64, 65, 67, 69, 71, 73, 75, 76, 78, 80, 81, 83, 85, 87, 89, 91, 92, 93, 95, 97, 99, 100, 101, 103, 105, 106, 107, 109, 110, 111, 113, 115, 116, 118, 119, 120, 122, 124, 126, 128, 130, 132, 133, 134, 135, e 136 são motivos de promotor identificados em promotor At.sor78. Promotor de Arabidopsis cor78 (SEQ ID NO:1) é descrito por exemplo em (Horvath 1993) e (Gilmour *et al.* 1991; Nordin *et al.* 1991). Em alguns casos, um motivo de promotor de núcleo corresponde a motivos de promotor localizados em posições diferentes de At.cor78 com diferentes 5' e 3' nucleotídeos de flanco, como mostrado em tabela 9.

A presente invenção ainda refere-se a uma molécula de ácido nucleico compreendendo um polinucleotídeo codificando uma sequência reguladora da transcrição compreendendo pelo menos dois motivos de promotor de núcleo selecionados dentre o grupo consistindo de sequências

5 como definidas em SEQ ID NOs: 10, 12, 14, 17, 20, 22, 25, 27, 29, 31, 33, 35, 37, 39, 42, 44, 46, 48, 50, 53, 55, 57, 59, 63, 66, 68, 70, 72, 74, 77, 79, 82, 84, 86, 88, 90, 94, 96, 98, 102, 104, 108, 112, 114, 117, 121, 123, 125, 127, 129, 131, 137, e 138. Preferivelmente os motivos de promotor de núcleo de SEQ ID NOs: 10, 12, 14, 17, 20, 22, 25, 27, 29, 31, 33, 35, 37, 39, 42, 44, 46,

10 48, 50, 53, 55, 57, 59, 63, 66, 68, 70, 72, 74, 77, 79, 82, 84, 86, 88, 90, 94, 96, 98, 102, 104, 108, 112, 114, 117, 121, 123, 125, 127, 129, 131, 137, e 138 são 30-40%, preferivelmente 40-50%, ou mais preferivelmente 50-60% ou mais idênticos a SEQ ID NOs: 9, (11,18), 13, (15, 16), 19, (21, 23, 61, 92), (24, 75), (26, 40), 28, 30, (32, 51, 80, 105, 109), 34, 36, 38, 41, (43, 64, 132), 45,

15 47, 49, 52, (54, 99), (56, 119), (58, 60), 62, 65, 67, 69, 71, 73, 76, 78, 81, 83, 85, (87,100,106, 110), (89, 91), 93, 95, 97, 101, 103, 107, 111, (113, 115), (116,118), 120, 122, 124, 126, 128, (130, 134, 135), 136, e 133, respectivamente. Mais preferivelmente os motivos de promotor de núcleo de SEQ ID NOs: 10, 12, 14, 16, 19, 22, 24, 27, 29, 31, 33, 35, 37, 39, 41, 44, 46,

20 48, 50, 52, 55, 57, 59, 61, 65, 68, 70, 72, 74, 76, 79, 81, 84, 86, 88, 90, 92, 96, 98, 100, 104, 106, 110, 114, 116, 119, 123, 125, 127, 129, 131, 133, 139, e 140 são 60-70%, preferivelmente 70-80%, ou mais preferivelmente 80-90% ou mais idênticos a SEQ ID NOs:9, (11,18), 13, (15, 16), 19, (21, 23, 61, 92), (24, 75), (26, 40), 28, 30, (32, 51, 80, 105, 109), 34, 36, 38, 41, (43, 64, 132),

25 45, 47, 49, 52, (54, 99), (56, 119), (58, 60), 62, 65, 67, 69, 71, 73, 76, 78, 81, 83, 85, (87,100,106, 110), (89, 91), 93, 95, 97, 101, 103, 107, 111, (113, 115), (116,118), 120, 122, 124, 126, 128, (130, 134, 135), 136, e 133, respectivamente. O mais preferivelmente os motivos de promotor de núcleo de SEQ ID NOs: 10, 12, 14, 16, 19, 22, 24, 27, 29, 31, 33, 35, 37, 39, 41, 44,

46, 48, 50, 52, 55, 57, 59, 61, 65, 68, 70, 72, 74, 76, 79, 81, 84, 86, 88, 90, 92, 96, 98, 100, 104, 106, 110, 114, 116, 119, 123, 125, 127, 129, 131, 133, 139, e 140 são 90-95% ou preferivelmente 95-99% idênticos a SEQ ID NOs:9, (11,18), 13, (15, 16), 19, (21, 23, 61, 92), (24, 75), (26, 40), 28, 30, (32, 51, 80, 105, 109), 34, 36, 38, 41, (43, 64, 132), 45, 47, 49, 52, (54, 99), (56, 119), (58, 60), 62, 65, 67, 69, 71, 73, 76, 78, 81, 83, 85, (87,100,106, 110), (89, 91), 93, 95, 97, 101, 103, 107, 111, (113, 115), (116,118), 120, 122, 124, 126, 128, (130, 134, 135), 136, e 133, respectivamente.

1.2. Elementos regulatórios e funcionais adicionais para o cassete de expressão e vetores da invenção

Um cassete de expressão da invenção pode compreender outros elementos regulatórios. O termo neste contexto deve ser entendido em um significado amplo compreendendo todas as sequências que podem influenciar a construção ou função do cassete de expressão. Os elementos regulatórios podem por exemplo modificar a transcrição e/ou tradução em organismo procariótico ou eucariótico. Em uma forma de realização preferida, o cassete de expressão da invenção compreendia a sequência de regulação da transcrição e - opcionalmente elementos regulatórios adicionais - cada ligado operativamente à sequência de ácido nucleico a ser expressada (ou a sequência nucleotídica reguladora da transcrição).

Uma variedade de sequências regulatórias transcricionais 5' e 3' são disponíveis para uso na presente invenção. Os terminados de transcrição são responsáveis pela terminação de transcrição e poliadenilação de mRNA correta. A sequência de DNA regulatória não traduzida 3' preferivelmente inclui de cerca de 50 a cerca de 1,000, mais preferivelmente cerca de 100 a cerca de 1,000, pares de base de nucleotídeo e contém sequências de terminação de transcrição e tradução de plantas. Os terminadores transcricionais apropriados e os que são conhecidos como funcionando em plantas incluem o terminador CaMV 35S, o terminador tmL,

o terminador de nopalina sintase, o terminador de ervilha rbcS E9, o terminador para o transcrito T7 do gene octopina sintase de *Agrobacterium tumefaciens*, e a extremidade 3' dos genes de inibidor de protease I ou II de batata e tomate, apesar de outros elementos 3' bem conhecidos do versado na técnica também poderem ser empregados. Alternativamente, pode-se usar um terminador de gama coixina, oleosina 3 ou outro do gênero Coix.

Como a sequência de DNA entre o sítio de iniciação de transcrição e o início da sequência de codificação, isto é, a sequência líder não traduzida, podem influenciar a expressão dos genes, pode-se também desejar empregar uma sequência de líder particular. As sequências de líder particulares são contempladas para incluir as que incluem sequências, previstas para dirigir a expressão ótima do gene fixado, isto é, para incluir uma sequência de líder de consenso preferida, que pode aumentar ou manter a estabilidade do mRNA e evitar a iniciação inapropriada de tradução. A escolha destas sequências será bem conhecida do versado na técnica à luz da presente descrição. As sequências de que são derivadas dos genes que são altamente expressados em plantas serão as mais preferidas.

Os elementos regulatórios preferidos também incluem a região 5' não traduzida, íntrons e a região 3' não traduzida de genes.

Estas sequências que foram verificadas como melhorando a expressão de genes em plantas transgênicas incluem sequências de íntrons (ver abaixo para detalhes) e sequências de líder viral (por exemplo, de TMV, MCMV e AMV; Gallie 1987). por exemplo, várias sequências de líder não traduzidas derivadas de vírus são conhecidas como melhorando a expressão. Especificamente, as sequências de líder de vírus do mosaico do tabaco (TMV), vírus da mancha clorótica do milho (MCMV), e vírus do mosaico da alfafa (AMV) foram demonstradas como sendo efetivas para melhorar a expressão (por exemplo, Gallie 1987; Skuzeski 1990). Outros líderes bem conhecidos na técnica incluem mas não são limitados a: líderes de

Picornavírus, por exemplo, líder EMCV (Encephalomyocarditis 5' região não de codificação) (Elroy-Stein 1989); líderes de Potyvírus, por exemplo, líder TEV (vírus de ataque ao tabaco); líder de MDMV (Vírus do mosaico do ananismo de milho); líder da proteína de ligação de cadeia pesada de imunoglobulina humana H (BiP), (Macejak 1991); líder não traduzido do mRNA de proteína de revestimento do vírus do mosaico da alfafa (AMV RNA 4), (Jobling 1987; líder do vírus do mosaico do tabaco (TMV), (Gallie 1989; e líder do vírus da mancha clorótica do milho (MCMV) (Lommel 1991. Ver também, Della-Cioppa 1987. Os elementos regulatórios como o elemento omega de TMV (Gallie 1989), podem ainda ser incluídos onde desejado. Os exemplos adicionais de melhoradores incluem elementos do promotor de CaMV 35S, genes de octopina sintase (Ellis *et al.*, 1987), o gene actina I de arroz, o gene de desidrogenase de álcool de milho (Callis 1987), o gene shrunken I de milho (Vasil 1989), elemento Omega de TMV (Gallie 1989) e promotores de eucarióticos não plantas (por exemplo levedura; Ma 1988). Vetores para uso de acordo com a presente invenção podem ser construídos para incluir o elemento melhorador ocs. Este elemento foi primeiro identificado como um melhorador palindrômico de 16 bp a partir do gene de octopina sintase (ocs) de ulysses (Ellis 1987), e está presente em pelo menos 10 outros promotores (Bouchez 1989). O uso de um elemento melhorador, como os elementos ocs e particularmente cópias múltiplas do elemento, irá atuar para aumentar o nível de transcrição de promotores adjacentes quando aplicados no contexto da transformação de plantas.

Os elementos regulatórios preferidos adicionais são sequências de melhorador ou sequências de poliadenilação. As sequências de poliadenilação preferidas são as de genes de plantas ou genes de T-DNA de *Agrobacterium* (como por exemplo as sequências de terminador dos genes OCS (octopina sintase) ou NOS (nopalina sintase)).

Um cassete de expressão da invenção (ou um vetor derivado

do mesmo, isto é, um vetor compreendendo um cassete de expressão da invenção) pode compreender elementos funcionais adicionais, que devem ser entendidos no sentido amplo como todos os elementos que influenciam a construção, propagação ou função de um cassete de expressão ou um vetor ou um organismo transgênico compreendendo os mesmos. Estes elementos funcionais podem incluir origem de replicação (para permitir a replicação em bactérias; para o ORI de pBR322 ou o P15A ori; Sambrook 1989), ou elementos requeridos para a transferência de T-DNA de *Agrobacterium* (como por exemplo as bordas à esquerda e/ou à direita do T-DNA.

Além disso, os cassetes de expressão podem ser construídos e empregados na marcação intracelular de um produto de gene específico dentro das células de uma planta transgênica ou para dirigir a proteína para o meio extracelular. Isto será geralmente obtido unindo a sequência de DNA codificando uma sequência de peptídeo de transito ou sinal para a sequência de codificação de um gene particular. O peptídeo de transito ou sinal resultante irá transportar a proteína para um destino intracelular ou extracelular particular, respectivamente, e será então pós-translacionalmente removido. Os peptídeos de transito ou sinal atuam por facilitação do transporte das proteínas através das membranas intracelulares, por exemplo membranas do vacúolo, vesícula, plastídeo e mitocondriais, enquanto os peptídeos de sinal dirigem proteínas através da membrana extracelular. Ao facilitar o transporte da proteína em compartimentos dentro e fora da célula, estas seqüências podem aumentar o acúmulo do produto de gene protegendo as mesmas da degradação proteolítica. Estas seqüências também permitem que seqüências de mRNA adicionais de genes altamente expressados sejam fixadas à sequência de codificação dos genes. Porque mRNA sendo traduzido por ribossomas é mais estável do que mRNA nu, a presença de mRNA traduzível na frente do gene pode aumentar a estabilidade global do transcrito e mRNA do gene e, assim, aumentar a síntese do produto de gene. Porque as

seqüências de transito e sinal são geralmente removidas pós-translacionalmente do produto de tradução inicial, o uso destas seqüências permite a adição de seqüências extras traduzidas que podem não aparecer no polipeptídeo final. A marcação de algumas proteínas pode ser desejável a fim de melhorar a estabilidade da proteína (US 5 545 818).

1.3 Montagem da seqüência de ácido nucleico reguladora da transcrição quimérica, cassetes de expressão, e vetores da invenção

Uma ligação operável em relação a qualquer seqüência de ácido nucleico reguladora da transcrição quimérica, cassete de expressão ou vetor da invenção pode ser realizada usando técnicas de recombinação e clonagem padrões bem conhecidas na técnica (ver por exemplo, Maniatis 1989; Silhavy 1984; Ausubel 1987). Muitas abordagens ou métodos foram desenvolvidos e usados para a clonagem de genes. Exemplos destes são a clonagem por digestão da enzima de restrição e ligação de extremidades compatíveis, clonagem de T-A diretamente a partir do produto de PCR, clonagem unidirecional fixada a TOPO, e clonagem com base em recombinação. A clonagem com base em recombinação é um dos métodos de clonagem mais versáteis disponíveis devido à sua elevada eficiência de clonagem e sua aplicação ampla para a clonagem de vários genes sem levar em conta os sítios disponíveis de enzima de restrição. A clonagem de recombinação usa o sistema de recombinação lambda para clonar genes em vetores que contém seqüências de recombinação para o maquinário de lambda recombinase. A clonagem de recombinação usa as recombinases específicas do sítio, que junto com proteínas associadas em alguns casos, reconhecem seqüências específicas de bases em uma molécula de ácido nucleico e trocam os segmentos de ácido nucleico flanqueando estas seqüências. As recombinases e proteínas associadas são coletivamente referidas como "proteínas de recombinação". As recombinases específicas de sítio são proteínas que estão presentes em muitos organismos (por exemplo vírus e

bactérias) e foram caracterizadas como tendo tanto propriedades de endonuclease como ligase. Muitas das recombinases específicas de sítio conhecidas pertencem à família integrase de recombinases incluindo o sistema integrase / att de bacteriófago lambda. Um exemplo de uma aplicação do sistema integrase / att do bacteriófago lambda é a reação de clonagem LR como descrito em US 5.888.732 e US 6.277.608 e pedido de patente publicado U.S. 2002/0007051 A1 e pedido internacional WO 02/081711 A1, todos sendo incorporados aqui por referência. A reação de clonagem LR é comercialmente disponível como a tecnologia GATEWAYTM (disponível de Invitrogen Corporação, Carlsbad, California). A reação de clonagem LR é catalisada pela mistura LR Clonase Enzyme que compreende proteínas de recombinação lambda Int, Xis, e IHF de proteína codificada por E. coli.

Um cassete de expressão também pode ser montado por inserção da sequência de ácido nucleico reguladora da transcrição quimérica da invenção no genoma da planta. Esta inserção irá resultar em uma ligação operável para uma sequência de ácido nucleico de interesse, que como tal já existia no genoma. Pela inserção, o ácido nucleico de interesse é expressado em um modo específico do embrião germinando devido às propriedades de regulação da transcrição da sequência nucleotídica reguladora da transcrição quimérica. A inserção pode ser dirigida ou ocorrer por acaso. Preferivelmente a inserção é dirigida e realizada por exemplo recombinação homóloga. Por este procedimento, um promotor natural pode ser trocado contra a sequência nucleotídica reguladora da transcrição quimérica da invenção, assim modificando o perfil de expressão de um gene endógeno. A sequência nucleotídica reguladora da transcrição também pode ser inserida em um modo, que o mRNA anti-sentido de um gene endógeno seja expressado, assim induzindo o silenciamento do gene.

Uma ligação operável pode - por exemplo - compreender um arranjo sequencial da sequência nucleotídica reguladora da transcrição

quimérica da invenção (por exemplo o super-promotor) com a sequência de ácido nucleico a ser expressada e – opcionalmente – elementos regulatórios adicionais, como por exemplo, elementos de poliadenilação ou de terminação da transcrição, melhoradores, introns *etc*, em um modo em que a sequência nucleotídica reguladora da transcrição pode preencher sua função no processo de expressão da sequência de ácido nucleico de interesse sob as condições apropriadas. O termo “condições apropriadas” significa preferivelmente a presença do cassete de expressão em uma célula de planta. São preferidos os arranjos, em que a sequência de ácido nucleico de interesse a ser expressada é colocada a jusante (isto é, em direção 3’-) da sequência nucleotídica reguladora da transcrição quimérica da invenção em um modo em que ambas as sequências são covalentemente ligadas. Opcionalmente, as sequências adicionais podem ser inseridas entre as duas sequências. Estas sequências podem ser por exemplo sítios de ligador ou clonagem múltipla. Além disso, as sequências podem ser inseridas codificando para partes das proteínas de fusão (no caso uma proteína de fusão da proteína codificada pelo ácido nucleico de interesse se destina a ser expressada). Preferivelmente, a distância entre a sequência de ácido nucleico de interesse a ser expressada e a sequência nucleotídica reguladora da transcrição da invenção não é maior do que 200 pares de base, preferivelmente não é maior do que 100 pares de base, mais preferivelmente não é maior do que 50 pares de base.

Virtualmente qualquer composição de DNA pode ser usada para a distribuição às plantas monocotiledôneas recipientes ou células de planta para, por fim, produzir plantas transgênicas férteis de acordo com a presente invenção. Por exemplo, segmentos ou fragmentos de DNA na forma de vetores e plasmídeos, ou segmentos ou fragmentos de DNA linear, em alguns casos contendo somente o elemento de DNA a ser expressado na planta, e outros, podem ser empregados. A construção de vetores, que podem ser empregados em conjunto com a presente invenção, serão bem conhecidos

do versado na técnica à luz da presente descrição (ver, por exemplo, Sambrook 1989; Gelvin 1990).

A presente invenção ainda provê um vetor recombinante ou outra construção de DNA apropriado para transformação de plantas (incluindo mas não limitado a cosmídeos, YACs (cromossomos artificiais de levedura), BACs (cromossomos artificiais de bactérias) e cromossomos artificiais de plantas) contendo o cassete de expressão da invenção, células hospedeiras de monocotiledôneas compreendendo o cassete de expressão ou vetor, por exemplo, compreendendo um plasmídeo. O cassete de expressão ou vetor pode(preferivelmente) aumentar o genoma de uma planta monocotiledônea transformada ou pode ser mantido extra cromossomicamente. O cassete de expressão ou vetor da invenção pode estar presente no núcleo, cloroplasto, mitocôndria e/ou plastídeo das células da planta. Preferivelmente, o cassete de expressão ou vetor da invenção está compreendido no DNA cromossômico do núcleo da planta. Em algumas formas de realização, contempla-se que pode-se desejar empregar vetores virais competentes na replicação em transformação de monocótilas. Estes vetores incluem, por exemplo, vírus do ananismo do tribo (WDV), vetores "shuttle", como pW1-11 e PW1-GUS (Ugaki 1991). Estes vetores são capazes de replicação autônoma em células de milho, assim como em *E. coli*, e como tal, podem prover uma sensibilidade aumentada para a detecção de DNA fornecido às células transgênicas. Um vetor de replicação também pode ser utilizável para a distribuição de genes flanqueados por sequências de DNA de elementos transposáveis, como Ac, Ds, ou Mu.

A construção de DNA de acordo com a invenção e quaisquer vetores daí derivados podem compreender outros elementos funcionais. O termo "outros elementos funcionais" deve ser entendido em sentido amplo. Ele refere-se preferivelmente a todos os elementos que afetam a geração, multiplicação, função, uso ou valor de referida construção de DNA ou vetores

compreendendo referida construção de DNA, ou células ou organismos compreendendo os acima mencionados. Estes outros elementos funcionais podem incluir, mas não são limitados a:

i) origens de replicação que asseguram a replicação dos cassetes de expressão ou vetores de acordo com a invenção em, por exemplo, *E. coli*. Exemplos que pode ser mencionados são ORI (origem de replicação de f DNA), o pBR322 ori ou P15A ori (Sambrook *et al.* 1989).

ii) sítios de clonagem múltiplos (MCS) para permitir e facilitar a inserção de uma ou mais sequências de ácido nucleico.

iii) sequências que tornam possível a recombinação homóloga ou inserção no genoma de um organismo hospedeiro.

iv) elementos, por exemplo sequências de borda, que tornam possível a transferência mediada por *Agrobacterium* em células de planta para a transferência e integração no genoma da planta, como, por exemplo, a borda à direita ou esquerda do T-DNA ou a região vir.

A molécula de DNA recombinante introduzida usada para a transformação aqui pode ser circular ou linear, de filamento duplo ou filamento único. Geralmente, o DNA está na forma de DNA quimérico, com DNA plasmídeo, que também pode conter regiões de codificação flanqueadas por sequências regulatórias, que promovem a expressão do DNA recombinante presente na planta resultante. Geralmente, a molécula de DNA recombinante introduzida será relativamente pequena, isto é, menor do que cerca de 30 kb para minimizar qualquer susceptibilidade a degradação física, química ou enzimática que é conhecida como aumentando à medida que aumenta o tamanho da molécula de nucleotídeo. Como notado acima, o número de proteínas, transcritos de RNA ou suas misturas, que é introduzido no genoma na planta, é preferivelmente pré-selecionado e definido, por exemplo de um a cerca de 5-10 como os produtos do DNA introduzido podem ser formados.

A presente invenção também provê uma planta monocotiledônea (preferivelmente uma planta transgênica), semente e partes desta planta, e plantas de progênie desta planta, incluindo linhagens endogâmicas e híbridos.

5 A invenção também provê um método de cruzamento de plantas, por exemplo, para preparar uma planta transgênica cruzada fértil. O método compreende cruzar uma planta transgênica fértil compreendendo um cassete de expressão particular da invenção com ela mesma ou com uma segunda planta, por exemplo, uma faltando o cassete de expressão particular,
10 para preparar uma semente de uma planta transgênica cruzada fértil compreendendo o cassete de expressão particular. A semente é então plantada para obter uma planta transgênica cruzada fértil. A planta pode ser preferivelmente uma monocotiledônea (preferivelmente como definido acima). A planta transgênica cruzada fértil pode ter o cassete de expressão
15 particular herdado através de um parental feminino ou através de um parental masculino. A segunda planta pode ser uma planta endogâmica. A planta transgênica cruzada fértil pode ser um híbrido. Também estão incluídas na presente invenção sementes de qualquer uma destas plantas transgênicas cruzadas férteis.

20 2. Traços ou propriedades vantajosos a serem expressados pelo cassete de expressão da invenção

As sequências nucleotídicas reguladoras da transcrição quiméricas da invenção são utilizáveis para modificar o fenótipo de uma planta. Várias mudanças no fenótipo de uma planta transgênica são desejáveis
25 e podem ser obtidos usando o perfil de expressão vantajoso (isto é, expressão específica de embrião germinando) de sequências nucleotídicas reguladoras da transcrição descritas aqui. Estes resultados podem ser obtidos provendo a expressão de produtos heterólogos ou a expressão aumentada de produtos endógenos nas plantas. Alternativamente, os resultados podem ser obtidos

provendo uma redução da expressão de um ou mais produtos endógenos, particularmente enzimas ou co-fatores na planta. Geralmente, as sequências nucleotídicas reguladoras da transcrição quiméricas podem ser empregadas para expressar um segmento de ácido nucleico que é operativamente ligado ao referido promotor como, por exemplo, uma matriz de leitura aberta, ou uma porção da mesma, uma sequência anti-sentido, uma sequência codificando para a sequência de RNA sentido ou filamento duplo, uma sequência codificando para uma sequência de RNA micro, ou um transgene em plantas. Estas mudanças resultam em uma alteração no fenótipo da planta transformada.

A escolha de um DNA heterólogo para expressão em uma célula hospedeira de planta monocotiledônea de acordo com a invenção irá depender do fim da transformação. Um dos fins principais de transformação de plantas de cultura consiste em adicionar traços comercialmente desejáveis, agronomicamente importantes ou de produto final à planta.

Apesar de numerosas seqüências de ácido nucleico serem apropriadas para serem expressadas pela sequência de ácido nucleico reguladora da transcrição quimérica da invenção, o mais preferivelmente o ácido nucleico está conferindo, quando da expressão, à planta monocotiledônea um traço ou propriedade selecionada dentre o grupo consistindo de

i) melhorada resistência contra pelo menos um fator de estresse,

ii) qualidade nutricional aumentada de uma semente ou um broto,

iii) rendimento aumentado, e

iv) excisão de uma sequência alvo, por exemplo excisão de uma sequência de marcador de seleção.

2.1. Princípios básicos

Dois métodos principais para o controle d expressão são conhecidos, super-expressão e sub-expressão. A super-expressão pode ser obtida por inserção de uma ou mais de uma cópia extra do gene selecionado. No entanto, isto não é conhecido para plantas ou sua progênie, originalmente transformada com uma ou mais do que uma cópia extra de uma sequência de nucleotídeo, para demonstrar os efeitos da sub-expressão, assim como da super-expressão. Para sub-expressão, existem dois métodos principais, que são comumente referidos na técnica como "regulação negativa anti-sentido" e "regulação negativa sentido" (regulação negativa sentido é também referida "co-supressão"). Genericamente, estes processos são referidos como "silenciamento de genes". Ambos os métodos levam a uma inibição de expressão de um gene alvo.

Assim, expressão da sequência de ácido nucleico sob a sequência reguladora da transcrição quimérica pode resultar em transcrição de um mRNA e expressão de uma proteína, ou expressão de um RNA anti-sentido, RNA sentido, dsRNA, microRNA, ta-siRNA, snRNA, RNAi, ou qualquer combinação dos mesmos.

O mRNA ou o RNA regulatório expressado/transcrito no método da invenção pode ser por exemplo

a) a sequência de RNA de filamento duplo de ácido nucleico (dsRNA) como descrito acima;

b) uma sequência anti-sentido de ácido nucleico. Englobados são os métodos em que a sequência anti-sentido de ácido nucleico é dirigida contra um gene (isto é sequências de DNA genômico incluindo a sequência de promotor) ou um gene transcrito (isto é sequências de RNA) incluindo as regiões não traduzidas 5' and 3'. Também englobadas são sequências de ácido nucleico α -anoméricas;

c) uma sequência anti-sentido de ácido nucleico em combinação com a ribozima;

d) uma sequência sentido de ácido nucleico para induzir co-supressão;

e) a sequência de ácido nucleico codificando o fator negativo dominante;

5 f) um fator de ligação a DNA-, RNA- ou proteína ou anticorpos contra os genes, RNA's ou proteínas;

g) sequências virais de ácido nucleico que ocasionam a degradação de RNA;

10 h) um microRNA ou micro-RNA (miRNA) que foi projetado para marcar o gene de interesse a fim de induzir uma ruptura do mRNA ou inibição da tradução do gene de interesse e assim silenciar a expressão do gene;

i) um ta-siRNA que foi projetado para marcar o gene de interesse a fim de induzir uma ruptura do mRNA (ou talvez a inibição da
15 tradução) do gene de interesse e assim silenciar a expressão do gene, e/ou.

O que se segue é uma descrição breve dos métodos preferidos individuais. Em geral, aqui, o significado do termo "expressão" deve incluir o significado dos termos "transcrição" e/ou "tradução" onde apropriado.

20 a) RNA regulatório expressado no método da invenção pode ser por exemplo a sequência de RNA de filamento duplo de ácido nucleico (dsRNA) por exemplo para a redução ou deleção de atividade de uma molécula de ácido nucleico ou polipeptídeo cuja atividade deve ser reduzida no processo da invenção.

25 Como uma alternativa polinucleotídeos anti-sentido e polinucleotídeos sentido, RNA de filamento duplo (dsRNA) pode ser usado para reduzir a expressão de um gene. O termo dsRNA, como usado aqui, refere-se a híbridos de RNA compreendendo dois filamentos de RNA. O dsRNA pode ser linear ou circular em estrutura. Os RNAs hibridizando podem ser substancialmente ou completamente complementares. Por

“substancialmente complementares”, significa-se que quando os dois RNAs hibridizando são otimamente alinhados usando o algoritmo de Needleman e Wunsch como descrito acima, as porções hibridizando são pelo menos 95% complementares. Preferivelmente, o dsRNA terá pelo menos 100 pares em comprimento. Métodos para fazer e usar o dsRNA são bem conhecidos a técnica. Um método compreende a transcrição simultânea de dois filamentos de DNA complementares in vivo (ver, por exemplo, Patente U.S. No. 5.795.715). DsRNA pode ser introduzido em uma planta ou célula de planta diretamente por processamentos de transformação padrões. Alternativamente, dsRNA pode ser expressado em uma planta por transcrição de dois RNAs complementares.

O método de regular os genes por meio de RNA de filamento duplo (“RNA de filamento duplo interferência”; dsRNAi) foi descrito extensivamente para organismos animais, levedura, fungos e plantas, por exemplo para *Neurospora*, *Zebrafish*, *Drosophila*, camundongos, planária, humanos, *Trypanosoma*, *petúnia* ou *Arabidopsis* (por exemplo Matzke MA et al. (2000) *Plant Mol. Biol.* 43: 401-415; Fire A. et al. (1998) *Nature* 391: 806-811; WO 99/32619; WO 99/53050; WO 00/68374; WO 00/44914; WO 00/44895; WO 00/49035; WO 00/63364). Além disso, RNAi é também documentado como uma ferramenta vantajosa para a repressão de genes em bactérias, como *E. coli*, por exemplo por Tchurikov et al. [*J. Biol. Chem.*, 2000, 275 (34): 26523-26529]. Fire et al. Chamaram o fenômeno de RNAi para RNA interferência. As técnicas e métodos descritos nas referências acima são expressamente referidos aqui. A supressão de genes eficiente também pode ser observada no caso de expressão transiente ou após a transformação transiente, por exemplo, como a consequência de uma transformação biolística (Schweizer P et al. (2000) *Plant J* 2000 24: 895-903). Os métodos dsRNAi são baseados no fenômeno de que a introdução simultânea de filamento complementar e contra-filamento de um gene

transcrito ocasiona cerca de uma supressão altamente efetiva da expressão do gene em questão. O fenótipo resultante é muito similar ao de um mutante de nocaute análogo (Waterhouse PM et al. (1998) Proc. Natl. Acad. Sci. USA 95: 13959-64).

5 Tuschl et al., Gens Dev., 1999, 13 (24): 3191–3197, foram capazes de demonstrar que a eficiência do método RNAi é uma função do comprimento do duplex, o comprimento dos *overhangs* da extremidade 3', e da sequência nestes *overhangs*.

10 Com base no trabalho de Tuschl et al. e considerando os princípios subjacentes são conservados entre diferentes espécies, as seguintes diretrizes podem ser dadas ao versado na técnica. Assim, a molécula de dsRNA da invenção ou usada no processo da invenção atende preferivelmente a pelo menos um dos seguintes princípios:

15 - para alcançar bons resultados, as regiões não traduzidas 5' e 3' da sequência de ácido nucleico em uso e regiões próximas ao códon de partida devem ser, em geral, evitadas como estas regiões são mais ricas em sítios de ligação de proteína regulatória e interações entre as sequências de RNAi e estas proteínas regulatórias podem levar a interações indesejadas.

20 - em plantas, as regiões não traduzidas 5' e 3' da sequência de ácido nucleico em uso e regiões próximas ao códon de partida, preferivelmente 50 a 100 nt a montante do códon de partida, dão bons resultados e assim não devem ser evitadas.

25 - preferivelmente um região do mRNA usado é selecionada, que está 50 a 100 nt (= nucleotídeos ou bases) a jusante do códon de partida de AUG;

- somente sequências de dsRNA (= RNA de filamento duplo) de exons são utilizáveis para o método, as sequências de introns não tem efeito;

- o teor de G/C nesta região deve ser maior do que 30% e

menor do que 70%, idealmente em torno de 50%;

- uma estrutura secundária possível do mRNA alvo é menos importante para o efeito do método RNAi.

O método dsRNAi pode ser particularmente efetivo e vantajoso para a redução da expressão de uma molécula de ácido nucleico cuja atividade deve ser reduzida no processo da invenção. Como descrito inter alia em WO 99/32619, as abordagens dsRNAi são claramente superiores às abordagens tradicionais anti-sentido.

Assim, a invenção pode ser usada para a expressão de moléculas de RNA de filamento duplo (moléculas de dsRNA) que, quando introduzidas em um organismo, com vantagem dentro de uma planta (ou uma célula, tecido, órgão ou semente derivadas da mesma) ocasiona uma atividade metabólica alterada pela redução da expressão de uma molécula de ácido nucleico cuja atividade deve ser reduzida no processo da invenção.

Em uma molécula de RNA de filamento duplo, por exemplo um dsRNA para reduzir a expressão de uma proteína codificada por uma molécula de ácido nucleico cuja atividade deve ser reduzida no processo da invenção, um dos dois filamentos de RNA é essencialmente idêntico a pelo menos parte de uma sequência de ácido nucleico, e o outro filamento de RNA respectivo é essencialmente idêntico a pelo menos parte do filamento complementar de uma sequência de ácido nucleico.

O termo “essencialmente idêntico” refere-se ao fato de que a sequência de dsRNA também pode incluir inserções, deleções e mutações de pontos individuais em comparação com a sequência alvo enquanto ainda ocasionando cerca de uma redução efetiva na expressão. Preferivelmente, a homologia como definida acima chega a pelo menos 30%, preferivelmente pelo menos 40%, 50%, 60%, 70% ou 80%, muito especialmente preferivelmente pelo menos 90%, o mais preferivelmente 100%, entre o filamento "sentido" de um dsRNA inibitório e parte de segmento de uma

sequência de ácido nucleico da invenção incluindo em uma forma de realização preferida da invenção suas regiões endógenas 5'- e 3' não traduzidas ou entre o filamento "anti-sentido" e o filamento complementar de uma sequência de ácido nucleico, respectivamente. A parte de segmento
5 chega a pelo menos 10 bases, preferivelmente pelo menos 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 ou 30 bases, especialmente preferivelmente pelo menos 40, 50, 60, 70, 80 ou 90 bases, muito especialmente preferivelmente pelo menos 100, 200, 300 ou 400 bases, o mais preferivelmente pelo menos 500, 600, 700, 800, 900 ou mais bases ou pelo menos 1000 ou 2000 bases ou
10 mais em comprimento. Em outra forma de realização preferida da invenção, a parte do segmento chega a 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 ou 27 bases, preferivelmente a 20, 21, 22, 23, 24 ou 25 bases. Estas sequências curtas são preferidas em animais e plantas. As sequências mais longas preferivelmente entre 200 e 800 bases são preferidas em animais não mamíferos,
15 preferivelmente em invertebrados, em levedura, fungos ou bactérias, mas elas também são utilizáveis em plantas. RNAs de filamento duplo longos são processados em organismos em muitos siRNAs (=RNAs de interferência pequena- curta) por exemplo pela proteína Dicer, que é uma enzima Rnase III específica para *ds*. Como uma alternativa, um dsRNA "essencialmente
20 idêntico" também pode ser definido como uma sequência de ácido nucleico, que é capaz de hibridizar com parte de um gene transcrito (por exemplo em 400 mM NaCl, 40 mM PIPES pH 6,4, 1 mM EDTA a 50°C ou 70°C durante 12 a 16 h).

O dsRNA pode consistir de um ou mais filamentos de
25 ribonucleotídeos polimerizados. A modificação de tanto a estrutura dorsal de açúcar -fosfato e de nucleosídeos pode além disso estar presente. Por exemplo, as ligações fosfodiéster do RNA natural podem ser modificadas em tal modo que elas englobam pelo menos um nitrogênio ou heteroátomo de enxofre. Bases podem sofrer modificação em tal modo que a atividade de, por

exemplo, adenosina deaminase, é limitada. Estas e outras modificações são descritas aqui abaixo em métodos para estabilizar RNA anti-sentido.

O dsRNA pode ser preparado enzimaticamente; ele também pode ser sintetizado quimicamente, ou no todo ou em parte. O dsRNA curto de até 30 bp, que efetivamente media RNA interferência, pode ser por exemplo, eficientemente gerado por digestão parcial de gabaritos longos de dsRNA usando ribonuclease III de *E. coli* (RNase III). (Yang, D., et al. (2002) Proc. Natl. Acad. Sci. USA 99, 9942.)

A estrutura de filamento duplo pode ser formada partindo de um filamento auto-complementar, único, ou partindo de dois filamentos complementares. Em um filamento único, auto-complementar, sequência “sentido” e “anti-sentido” pode ser ligada por uma sequência de ligando (“ligador”) e formar, por exemplo, uma estrutura de grampo de cabelo. Preferivelmente, a sequência de ligando pode tomar a forma de um intron, que é emendado após a síntese de dsRNA. A sequência de ácido nucleico codificando um dsRNA pode conter outros elementos como, por exemplo, sinais de terminação de transcrição ou sinais de poliadenilação. Se dois filamentos do dsRNA devem ser combinados em uma célula ou organismo com vantagem em uma planta, isto pode ser ocasionado em cerca de vários modos:

a) transformação da célula ou do organismo, com vantagem de uma planta, com um vetor englobando os dois cassetes de expressão;

b) cotransformação da célula ou do organismo, com vantagem de uma planta, com dois vetores, um dos quais engloba os cassetes de expressão com o filamento “sentido” enquanto o outro engloba os cassetes de expressão com o filamento “anti-sentido”;

c) supertransformação da célula ou do organismo, com vantagem de uma planta, com um vetor englobando os cassetes de expressão com o filamento “sentido”, após a célula ou o organismo ter sido

transformada com um vetor englobando os cassetes de expressão com o filamento “anti-sentido” ou vice versa;

d) hibridização por exemplo cruzamento de dois organismos, com vantagem de plantas, cada tendo sido transformado com um vetor, um dos quais engloba o cassete de expressão com o filamento “sentido” enquanto o outro engloba o cassete de expressão com o filamento “anti-sentido”;

e) introdução de uma construção compreendendo dois promotores que levam à transcrição da sequência desejada de ambas as direções, e/ou

f) infecção da célula ou do organismo, com vantagem de uma planta, com um vírus engenheirado, que é capaz de produzir a molécula de dsRNA desejada.

A formação do RNA duplex pode ser iniciada ou de fora da célula ou de dentro da célula. E o dsRNA for sintetizado fora da célula alvo ou organismo, ele pode ser introduzido no organismo ou na célula do organismo por injeção, micro-injeção, eletroporação, partículas de alta velocidade, por feixe de laser, ou mediado por compostos químicos (DEAE-dextrano, fosfato de cálcio, lipossomas).

Assim, em uma forma de realização, a presente invenção refere-se a um dsRNA pelo que o filamento sentido de referida molécula de RNA de filamento duplo de ácido nucleico tem um homologia de pelo menos 30%, 35%, 40%, 45%, 50%, 55% ou 60%, preferivelmente 65%, 70%, 75% ou 80%, mais preferivelmente 85%, 90%, 95%, 96%, 97%, 98% ou 99% para a sequência de ácido nucleico.

Outra forma de realização da invenção é uma molécula de dsRNA, compreendendo um fragmento de pelo menos 10 pares de base (= bases, nt, nucleotídeos), preferivelmente pelo menos 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45 ou 50, especialmente preferivelmente pelo menos 55, 60, 70, 80 ou 90 pares de base, muito especialmente

preferivelmente pelo menos 100, 200, 300 ou 400 pares de base, o mais preferivelmente pelo menos 500, 600, 700, 800, 900 ou mais pares de base ou pelo menos 1000 ou 2000 pares de base de uma molécula de ácido nucleico com um homologia de pelo menos 50%, 60%, 70%, 80% ou 90%,
5 preferivelmente 100% para uma molécula de ácido nucleico.

Em outra forma de realização preferida da invenção a sequência codificada ou seu segmento de parte da molécula de dsRNA chega a 17, 18, 19, 20, 21, 22, 23, 24, 25, 26 ou 27 bases, preferivelmente a 20, 21, 22, 23, 24 ou 25 bases, pelo que a homologia da sequência é essencialmente
10 95%, 96%, 97%, 98%, ou preferido 99% ou 100%.

Em uma forma de realização preferida da invenção, o filamento sentido e anti-sentido do RNA de filamento duplo são covalentemente ligados ou são ligados por outro, por exemplo ligações químicas fracas como ligações de hidrogênio a cada outro e o filamento anti-sentido é essencialmente o complemento do filamento de RNA sentido.
15

Como mostrado em WO 99/53050, o dsRNA também pode englobar uma estrutura de grampo de cabelo, por ligação dos filamentos “sentido” e “anti-sentido” por um “ligador” (por exemplo um intron). As estruturas de dsRNA auto-complementar são preferidas porque elas apenas
20 requerem a expressão de uma construção e sempre englobam os filamentos complementares em uma relação equimolar.

Os cassetes de expressão codificando o filamento “anti-sentido” ou “sentido” do dsRNA ou o filamento auto-complementar do dsRNA são preferivelmente inseridos em um vetor e estavelmente inseridos
25 no genoma de uma planta, usando os métodos descritos aqui abaixo (por exemplo usando marcadores de seleção), a fim de assegurar a permanente expressão do dsRNA. A expressão transiente com vetores bacterianos ou virais é similarmente utilizável.

O dsRNA pode ser introduzido usando uma quantidade que

torna possível pelo menos uma cópia por célula. Uma quantidade maior (por exemplo pelo menos 5, 10, 100, 500 ou 1 000 cópias por célula) pode ocasionar cerca de uma redução mais eficiente.

5 Como já foi descrito, 100 % sequência identidade entre dsRNA e um gene transcrito de uma molécula de ácido nucleico a ser reduzida, não é necessariamente requerido a fim de ocasionar uma redução efetiva na expressão. A vantagem é, assim, que o método é tolerante com relação dos desvios de sequência que podem estar presentes como uma consequência de mutações genéticas, polimorfismos ou divergências
10 evolucionárias. Assim, por exemplo, usando o dsRNA, que foi gerado partindo de uma molécula de ácido nucleico a ser reduzida de acordo com o processo da invenção.

O dsRNA pode ser sintetizado ou in vivo ou in vitro. Para esta finalidade, uma sequência de DNA codificando um dsRNA pode ser
15 introduzida em um cassete de expressão sob o controle de pelo menos um elemento de controle genético (como, por exemplo, promotor, melhorador, silenciador, doador de emenda ou aceitador de emenda ou sinal de poliadenilação). As construções apropriadas vantajosas são descritas aqui abaixo. A poliadenilação não é requerida, nem os elementos para a iniciação
20 da tradução precisam estar presentes.

Um dsRNA pode ser sintetizado quimicamente ou enzimaticamente. RNA polimerases celulares ou bacteriofago RNA polimerases (como, por exemplo T3, T7 ou SP6 RNA polimerase) podem ser usados para este fim. Apropriados métodos para a expressão in-vitro de RNA
25 são descritos (WO 97/32016; US 5.593.874; US 5.698.425, US 5.712.135, US 5.789.214, US 5.804.693). Antes da introdução em uma célula, tecido ou organismo, um dsRNA que foi sintetizado in vitro ou quimicamente ou enzimaticamente pode ser isolado em um grau maior ou menor da mistura de reação, por exemplo, por extração, precipitação, eletroforese, cromatografia

combinações destes métodos. O dsRNA pode ser introduzido diretamente na célula ou também pode ser aplicados extracelularmente (por exemplo no espaço intersticial). Em uma forma de realização da invenção, o método RNAi leva a apenas uma perda parcial da função do gene e assim permite ao
5 versado estudar o efeito da dose de gene no organismo desejado e sintonizar de acordo com o processo da invenção. Em outra forma de realização preferida isto leva a uma perda total de função e assim aumento da produção do produto químico fino. Além disso, permite ao versado na técnica estudar as funções múltiplas de um gene.

10 A transformação estável de uma planta com uma construção de expressão, que ocasiona cerca da expressão do dsRNA é preferida, no entanto. Os métodos apropriados são descritos aqui abaixo.

b) O RNA regulatório expressada em método da invenção pode ser por exemplo uma sequência de ácido nucleico anti-sentido, por
15 exemplo para a redução ou deleção de uma molécula de ácido nucleico ou polipeptídeo cuja atividade deve ser reduzida no processo da invenção.

Uma sequência de DNA exógena pode ser projetada para regular de modo negativo uma sequência de ácido nucleico. Isto é por exemplo alcançado operativamente ligando com a sequência de ácido
20 nucleico reguladora da transcrição quimérica da invenção, um DNA exógeno em uma orientação anti-sentido ou um DNA projetado de modo que uma molécula de RNA formadora de grampo de cabelo é gerada quando da transcrição. A supressão de genes pode ser efetiva contra um gene nativo de planta associado com um traço, por exemplo para prover plantas com níveis
25 reduzidos de uma proteína codificada pelo gene nativo ou com níveis melhorados ou reduzidos de um metabolito afetado. Por exemplo, a sequência e ácido nucleico reguladora da transcrição quimérica da invenção pode ser operativamente ligada a um DNA heterólogo projetado de modo que o RNA de forma de grampo de cabelo é formado para supressão de um gene nativo

em embriões de milho. Diferentes tipos de arranjos de DNA exógenos resultante em supressão de genes são conhecidos do versado na técnica e incluem mas não são limitados ao seguintes. Publicação internacional WO 94/01550 descreve construções de DNA onde o RNA anti-sentido foi

5 estabilizado com um segmento auto-complementar 3'. Outros elementos de filamento duplo formadores de grampo de cabelo em RNA transcrito são descritos em WO 98/05770 onde o RNA anti- sentido é estabilizado por repetição de formação de grampo de cabelo de poli(CG) nucleotídeos e Publicação de pedido de patente No. 2002/0048814 A1 descreve RNA sentido

10 ou anti- sentido estabilizado por uma cauda poli(T)-poli(A). Publicação de pedido de patente U.S. No. 2003/0018993 A1 descreve RNA sentido ou anti-sentido que é estabilizado por uma repetição invertida de uma subsequência de região 3' não traduzida do gene NOS. Publicação de pedido de patente U.S. No. 2003/0036197 A1 descreve um RNA estabilizado por duas regiões

15 complementares de RNA tendo homologia para uma sequência alvo.

Métodos para suprimir uma proteína específica por prevenção do acúmulo de seu mRNA por meio de tecnologia “anti-sentido” podem ser usados amplamente e foram descritos extensivamente, incluindo para plantas; Sheehy et al. (1988) Proc. Natl. Acad. Sci. USA 85: 8805-8809;

20 US 4.801.34100; Mol JN et al. (1990) FEBS Lett 268(2): 427-430. A molécula anti-sentido de ácido nucleico hibridiza com, ou se liga ao mRNA celular e/ou o DNA genômico codificando a proteína alvo a ser suprimida. Este processo suprime a transcrição e/ou tradução da proteína alvo. Hibridização pode ser ocasionada em de modo convencional via a formação

25 de um duplex estável, ou no caso de DNA genômico, pela molécula anti-sentido de ácido nucleico ligando ao duplex do DNA genômico pela interação específica na ranhura larga da hélice do DNA.

Em uma forma de realização, uma molécula de ácido nucleico “anti-sentido” compreende uma sequência nucleotídica, que é pelo menos em

parte complementar a uma molécula de ácido nucleico "sentido" codificando uma proteína, por exemplo, complementar ao filamento de codificação de uma molécula de cDNA de filamento duplo ou complemento a uma sequência de mRNA codificando. Assim, uma molécula de ácido nucleico anti-sentido
5 pode ligar via ligações hidrogênio a uma molécula de ácido nucleico sentido. A molécula de ácido nucleico anti-sentido pode ser complementar a um filamento de codificação completo de uma molécula de ácido nucleico conferindo a expressão do polipeptídeo a ser reduzida no processo da invenção ou compreendendo a molécula de ácido nucleico cuja atividade deve
10 ser reduzida no processo da invenção ou para apenas uma porção do mesmo. Assim, uma molécula de ácido nucleico anti-sentido pode ser anti-sentido para uma "região de codificação" do filamento de codificação de uma sequência nucleotídica de uma molécula de ácido nucleico da presente invenção.

15 O termo "região de codificação" refere-se à região da sequência nucleotídica compreendendo códons, que são traduzidos em resíduos de aminoácido.

Em outra forma de realização, a molécula anti-sentido de ácido nucleico é anti-sentido para uma "região não de codificação" do mRNA
20 flanqueando a região de codificação de uma sequência nucleotídica. O termo "região não de codificação" refere-se a sequências 5' e 3' que flanqueiam a região de codificação que não são traduzidas em um polipeptídeo, isto é, também referido como as regiões não traduzidas 5' e 3' (5'-UTR ou 3'-UTR). Com vantagem, a região não de codificação está na área de 50 bp, 100 bp,
25 200bp ou 300 bp, preferivelmente 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp ou 1000 bp a montante e/ou a jusante da região de codificação.

Dadas as sequências de filamento de codificação codificando a molécula de polipeptídeo ou de ácido nucleico a ser reduzida no processo da invenção, por exemplo tendo a atividade acima mencionada, por exemplo a

atividade de um polipeptídeo com a *atividade* da proteína cuja atividade deve ser reduzida no processo da invenção como descrito aqui, moléculas de ácido nucleico anti-sentido podem ser projetadas de acordo com as regras de formação de pares de base de Watson e Crick.

5 Em uma outra forma de realização, a molécula de ácido nucleico anti-sentido pode ser um ácido nucleico α -anomérico. Estas moléculas α -anoméricas de ácido nucleico formam híbridos específicos de filamento duplo com RNA complementar em que - em oposição aos β -ácidos nucleicos convencionais - os dois filamentos correm em paralelo um com o
10 outro (Gautier C et al. (1987) Nucleic Acids Res. 15: 6625-6641). Além disso, a molécula de ácido nucleico anti-sentido também pode compreender e 2'-O-metilribonucleotídeos (Inoue et al. (1987) Nucleic Acids Res. 15: 6131-6148), ou análogos de RNA-DNA quiméricos (Inoue et al. (1987) FEBS Lett 215: 327-330).

15 As moléculas de ácido nucleico anti-sentido da invenção são tipicamente administradas a uma célula ou geradas *in situ* de modo que elas hibridizam com ou ligam ao mRNA celular e/ou DNA genômico codificando um polipeptídeo tendo a atividade de proteína cuja atividade deve ser
20 reduzida no processo da invenção ou codificando uma molécula de ácido nucleico tendo a atividade de uma molécula de ácido nucleico cuja atividade deve ser reduzida no processo da invenção e assim inibir a expressão da proteína, por exemplo, por inibição da transcrição e/ou tradução e levando à atividade crescente de química fina acima mencionada.

25 A molécula anti-sentido da presente invenção compreende também uma molécula de ácido nucleico compreendendo sequências nucleotídicas complementares para a região regulatória de uma sequência nucleotídica codificando o polipeptídeo de ocorrência natural da invenção, por exemplo as sequências de polipeptídeo como mostrado na listagem de sequência, ou identificados de acordo com os métodos descritos aqui, por

exemplo, seus promotores e/ou melhoradores, por exemplo para formar estruturas de hélice tripla que evitam a transcrição do gene em células alvo. Ver geralmente, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; e Maher, L.J. (1992) *Bioassays* 14(12):807-15.

c) O RNA regulatório expressado no método da invenção pode ser por exemplo uma sequência anti-sentido de ácido nucleico combinada com uma ribozima, por exemplo para a redução ou a deleção da atividade da molécula de ácido nucleico ou polipeptídeo cuja atividade deve ser reduzida no processo da invenção.

É vantajoso combinar a estratégia anti-sentido acima mencionada com um método ribozima. As moléculas catalíticas de RENA ribozimas podem ser adaptadas a qualquer RNA alvo e clivar a estrutura dorsal de fosfodiéster em posições específicas, assim funcionalmente desativando o RNA alvo (Tanner NK (1999) *FEMS Microbiol. Rev.* 23(3): 257-275). A ribozima per se não é assim modificada, mas é capaz de clivar mais moléculas de RNA alvo em um modo análogo, assim adquirindo as propriedades de uma enzima. A incorporação de sequências de ribozima em RNAs “anti-sentido” confere esta propriedade de clivagem de RNA de tipo enzima para precisamente estes RNAs “anti-sentido” e assim aumentar sua eficiência quando inativando o RNA alvo. A preparação e o uso de moléculas apropriadas de RNA “anti-sentido” de ribozima é descrito, por exemplo, por Haseloff et al. (1988) *Nature* 33410: 585-591.

Além disso, a molécula anti-sentido de ácido nucleico da invenção pode ser também uma ribozima. Ribozimas são moléculas de RNA catalíticas com atividade de ribonuclease, que são capazes de clivar um ácido nucleico de filamento único, como mRNA, de modo que elas tem, uma região complementar. Deste modo, ribozimas (por exemplo ribozimas tipo "cabeça de martelo"; Haselhoff e Gerlach (1988) *Nature* 33410: 585-591) podem ser

usadas para clivar cataliticamente o mRNA de uma enzima a ser suprimida e para evitar a tradução. A tecnologia de ribozima pode aumentar a eficiência da estratégia anti-sentido. Métodos para expressar ribozimas para reduzir as proteínas específicas são descritos em (EP 0 291 533, EP 0 321 201, EP 0 360 257). A expressão de ribozima também foi descrita para células de planta (Steinecke P et al. (1992) EMBO J 11(4): 1525-1530; de Feyter R et al. (1996) Mol. Gen. Genet. 250(3): 329-338). As apropriadas sequências alvo e ribozimas podem ser identificadas por exemplo como descrito por Steinecke P, Ribozimas, Methods in Cell Biology 50, Galbraith et al. eds, Academic Press, Inc. (1995), pp. 449-460 calculando-se as estruturas secundárias de RNA e RNA alvo de ribozima e por sua interação [Bayley CC et al. (1992) Plant Mol. Biol. 18(2): 353-361; Lloyd AM e Davis RW et al. (1994) Mol. Gen. Genet. 242(6): 653-657]. Por exemplo, derivados de tetrahimena L-19 IVS RNA, que têm regiões complementares para o mRNA da proteína a ser suprimido podem ser construídos (ver também US 4.987.071 e US 5.116.742). Com uma alternativa, estas ribozimas também podem ser identificadas com uma bibliotecas para uma variedade de ribozimas via um processo de seleção (Bartel D e Szostak JW (1993) Science 261: 1411-1418).

d) O RNA regulatório expressada em método da invenção pode ser por exemplo uma sequência (sentido) de ácido nucleico para induzir a co-supressão, por exemplo para a redução ou deleção da atividade de uma molécula de ácido nucleico ou polipeptídeo cuja atividade deve ser reduzida no processo da invenção.

Como usado aqui "supressão de genes" significa qualquer um dos métodos bem conhecidos para a supressão de um RNA transcrito ou produção de proteína traduzida de um RNA transcrito, incluindo supressão pós-transcricional de genes e supressão transcricional. A supressão pós-transcricional de genes é mediada pelo RNA de filamento duplo tendo homologia para um gene marcado para supressão. A supressão de gene por

RNA transcrito de uma construção exógena de DNA compreendendo uma repetição invertida de pelo menos parte de uma unidade de transcrição é um aspecto comum de métodos de supressão de genes conhecidos como supressão anti-sentido, co-supressão e RNA interferência. Mais particularmente, a supressão pós-transcricional de genes por inserção de uma construção de DNA exógena com DNA orientado anti-sentido para regular a expressão de genes em células de planta é descrita em US 5.107.065 e US 5.759.829. A supressão transcricional pode ser mediada por RNA transcrita de filamento duplo tendo homologia para a sequência DNA do promotor para efetuar o que é chamado trans-supressão de promotor. A supressão de genes pós-transcricional por inserção de uma construção de DNA exógeno com DNA orientado sentido para regular a expressão de genes em plantas é descrita em US 5.283.184 e US 5.231.020.

A expressão de uma sequência de ácido nucleico em orientação sentido pode levar à co-supressão dos correspondentes genes endógenos homólogos. A expressão de RNA sentido com homologia para um gene endógeno pode reduzir ou de fato eliminar a expressão do gene endógeno, em um modo similar ao que foi descrito para as seguintes abordagens anti-sentido: Jorgensen et al. (1996) *Plant Mol. Biol.* 31(5): 957-973, Goring et al. (1991) *Proc. Natl. Acad. Sci. USA* 88: 1770-1774], Smith et al. (1990) *Mol. Gen. Genet.* 224: 447-481, Napoli et al. (1990) *Plant Cell* 2: 279-289 ou Van der Krol et al. (1990) *Plant Cell* 2: 291-99. Neste contexto, a construção introduzida pode representar o gene homólogo a ser reduzido ou no total ou somente em parte. A aplicação desta técnica às plantas foi descrita por exemplo por Napoli et al. (1990) *The Plant Cell* 2: 279-289 e em US 5.034.103.23. Além disso, a estratégia de co-supressão acima descrita pode combinada com vantagem com o método RNAi como descrito por Brummell et al., 2003, *Plant J.* 33, pp793-800. Pelo menos em plantas, é vantajoso usar promotores fortes ou muito fortes em abordagens de co-

supressão Trabalho recente, por exemplo por Schubert et al., (Plant Journal 2004, 16, 2561-2572) tem indicado que efeitos de co-supressão são dependentes em um nível de limiar de gene específico, acima do qual a co-supressão ocorre.

5 e) O mRNA expressado no método da invenção pode ser por exemplo sequências de ácido nucleico codificando um proteína negativa dominante, por exemplo para a redução ou deleção da atividade do polipeptídeo cuja atividade deve ser reduzida no processo da invenção

A função ou atividade de uma proteína também pode ser
 10 eficientemente reduzida por expressão de uma variante negativa dominante de referida proteína. O versado é familiar com métodos para reduzir a função ou atividade da proteína por meio de co- expressão de sua forma negativa dominante [Lagna G e Hemmati-Brivanlou A (1998) Current Topics in developmental Biology 36: 75-98; Perlmutter RM e Alberola-Ila J (1996)
 15 Current Opinion in Immunology 8(2): 285-90; Sheppard D (1994) American Journal of Respiratory Cell & Molecular Biology 11(1): 1-6; Herskowitz I (1987) Nature 329 (6136): 219-22].

Uma variante negativa dominante pode ser realizada por exemplo por mudança de um aminoácido de um polipeptídeo.

20 Esta mudança pode ser determinada por exemplo por comparação auxiliada por computador ("alinhamento"). Estas mutações para obter uma variante negativa dominante são preferivelmente realizadas ao nível das sequências de ácido nucleico. Uma correspondente mutação pode ser realizada por exemplo por mutagenese in vitro mediada por PCR usando
 25 iniciadores apropriados de oligonucleotídeos por meio do que a mutação desejada é introduzida. Para esta finalidade, métodos são usados com que o versado é familiar. Por exemplo, o kit "LA PCR in vitro Mutagenese Kit" (Takara Shuzo, Kyoto) pode ser usado para este fim. Também é possível e conhecido dos versados na técnica que a deleção ou mudança de domínios

funcionais, por exemplo TF ou outros componentes de sinalização que podem ligar mas não ativar, podem alcançar a red da atividade da proteína.

f) Ainda outra forma de realização da invenção o mRNA controlado codifica um fator de ligação de DNA ou de proteína.

5 Estes fatores fixam à sequência genômica do gene alvo endógeno, preferivelmente nas regiões regulatórias, e ocasionar a repressão do gene endógeno. O uso de tal método torna possível a redução na expressão de um gene endógeno sem ser necessário manipular de modo recombinante a sequência do último. Estes métodos para a preparação de fatores relevantes
10 são descritos em Dreier B et al. (2001) J. Biol. Chem. 276(31): 29466-78 e (2000) J. Mol. Biol. 303(4): 489-502, Beerli RR et al. (1998) Proc. Natl. Acad. Sci. USA 95(25): 14628-14633; (2000) Proc. Natl. Acad. Sci. USA 97(4): 1495-1500 e (2000) J. Biol. Chem. 275(42): 32617-32627), Segal DJ e Barbas CF, 3rd (2000) Curr. Opin. Chem. Biol. 4(1): 3410-39, Kang JS e Kim JS (2000) J. Biol. Chem. 275(12): 8742-8748, Kim JS et al. (1997) Proc. Natl. Acad. Sci. USA 94(8): 3616-3620, Klug A (1999) J. Mol. Biol. 293(2): 215-218, Tsai SY et al. (1998) Adv. Drug Deliv. Rev. 30(1-3): 23-31, Mapp AK et al. (2000) Proc. Natl. Acad. Sci. USA 97(8): 3930-3935, Sharrocks AD et al. (1997) Int. J. Biochem. Cell Biol. 29(12): 1371-1387 e Zhang L et al. (2000)
20 J. Biol. Chem. 275(43): 33850-33860. Exemplos para a aplicação desta tecnologia em plantas foram descritos em WO 01/52620, Ordiz MI et al., (Proc. Natl. Acad. Sci. USA, Vol. 99, Issue 20, 13290 – 13295, 2002) ou Guan et al., (Proc. Natl. Acad. Sci. USA, Vol. 99, Issue 20, 13296 – 13301, 2002)

25 Estes fatores pode ser selecionados usando qualquer porção de um gene. Este segmento é preferivelmente localizado em região do promotor. Para os fins de supressão de genes, no entanto, ele também pode estar localizado na região dos exons ou introns de codificação. O versado pode obter os relevantes segmentos de Genbank por busca nas bases de dado ou

partindo de um cDNA cujo gene não está presente no Genbank por triagem de uma biblioteca genômica para clones genômicos correspondentes.

Também é possível primeiro identificar as sequências em uma cultura alvo, que engloba uma molécula de ácido nucleico ou que codifica o polipeptídeo cuja atividade deve ser reduzida no processo da invenção, então encontrar o promotor e reduzir a expressão pelo uso dos fatores acima mencionados.

O versado conhece os métodos requeridos para isto.

Além disso, fatores que são introduzidos em uma célula também podem ser os que inibem, eles mesmos, a proteína alvo. Os fatores de ligação a proteína, podem, por exemplo, ser aptâmeros (Famulok M e Mayer G (1999) Curr. Top Microbiol. Immunol. 243: 123-36) ou anticorpos ou fragmentos de anticorpos ou anticorpos de cadeia única. A obtenção destes fatores foi descrita, e o versado é familiar com os mesmos. Por exemplo, um anticorpo scFv citoplásmico foi empregado para a modulação da atividade da proteína de fitocromo A em plantas de tabaco geneticamente modificadas (Owen M et al. (1992) Biotechnology (NY) 10(7): 790-794; Franken E et al. (1997) Curr. Opin. Biotechnol. 8(4): 411-416; Whitelam (1996) Trend Plant Sci. 1: 286-272).

A expressão de genes também pode ser suprimida por compostos sintéticos de baixo peso molecular feitos por encomenda, por exemplo para o tipo poliamida, Dervan PB e Bürli RW (1999) Current Opinion in Chemical Biology 3: 688-693; Gottesfeld JM et al. (2000) Gene Expr. 9(1-2): 77-91. Estes oligômeros consistem de unidades de 3-(dimetilamino)propilamina, N-metil-3-hidroxipirrol, N-metilimidazol e N-metilpirróis; eles podem ser adaptados a cada porção do DNA de filamento duplo em tal modo que eles ligam a sequência-especificamente para a ranhura grande e bloqueiam a expressão das sequências de gene localizadas nesta posição. Os métodos apropriados foram descritos em Bremer RE et al. (2001)

Bioorg. Med. Chem. 9(8): 2093-103], Ansari AZ et al. [(2001) Chem. Biol. 8(6): 583-92, Gottesfeld JM et al. (2001) J. Mol. Biol. 309(3): 615-29, Wurtz NR et al. (2001) Org. Lett 3(8): 1201-3, Wang CC et al. (2001) Bioorg. Med. Chem. 9(3): 653-7], Urbach AR e Dervan PB (2001) Proc. Natl. Acad. Sci. USA 98(8): 434103-8 e Chiang SY et al. (2000) J. Biol. Chem. 275(32): 24246-54.

g) O RNA regulatório expressado no método da invenção pode ser por exemplo uma sequência viral de ácido nucleico que provoca a degradação de RNA, por exemplo para a redução ou deleção da atividade de uma molécula de ácido nucleico ou polipeptídeo cuja atividade deve ser reduzida no processo da invenção.

A inativação ou regulação negativa também pode ser ocasionada de modo eficiente por indução da degradação de DNA específica pelo organismo, com vantagem na planta, com a ajuda do sistema de expressão viral(Amplikon) (Angell, SM et al. (1999) Plant J. 20(3): 357-362). As sequências de ácido nucleico com homologia para os transcritos a serem suprimidos são introduzidas em uma planta por estes sistemas - também referidos como “VIGS” (silenciamento de genes induzido viral), com a ajuda dos vetores virais. Então, a transcrição é desligada, provavelmente mediada pelos mecanismos de defesa da planta contra os vírus. As técnicas e métodos apropriados são descritos em Ratcliff F et al. (2001) Plant J. 25(2): 237-45, Fagard M e Vaucheret H (2000) Plant Mol. Biol. 43(2-3): 285-93, Anandalakshmi R et al. (1998) Proc. Natl. Acad. Sci. USA 95(22): 13079-84 e Ruiz MT (1998) Plant Cell 10(6): 937-46.

g) O RNA regulatório expressado no método da invenção pode ser por exemplo um microRNA (ou micro-RNA) que foi projetado para marcar o gene de interesse a fim de induzir uma ruptura ou inibição translacional do mRNA do gene de interesse e, assim, silenciar a expressão do gene ou de um cassete de expressão assegurando a expressão do primeiro, por

exemplo para a redução ou deleção da atividade de uma molécula de ácido nucleico ou polipeptídeo cuja atividade deve ser reduzida no processo da invenção.

A regulação de expressão de gene também pode ser obtida por modificação da expressão de microRNAs. Evidência crescente demonstra que os miRNAs da planta tem uma ampla faixa de funções regulatórias na identidade do meristema, divisão celular, separação e órgãos, polaridade dos órgãos e resposta ao estresse abiótico (Bartel D 2004). Muitos miRNAs de plantas tem um padrão de expressão espacial ou temporal único. A super-expressão ou expressão ectópica de miRNA de planta pode mudar a morfologia e fisiologia de uma planta. O precursor de miRNA de planta também pode ser engenheirado para marcar um gene repórter em Arabidopsis (Parizotto E A et al., 2004).

MicroRNAs (miRNAs) tem emergidos como reguladores à base de RNA evolucionariamente conservados, de expressão de genes em plantas e animais. MiRNAs (~ 21 a 25 nt) surgem dos precursores maiores em uma estrutura de laço do tronco que são transcritos em genes não de codificação de proteína. miRNA marca um mRNA específico para suprimir a expressão do gene pós-transcricional (*isto é* degrada mRNA) ou níveis de tradução (*isto é* inibe a síntese de proteína) (Bartel D 2004, Cell 116, 281-297). MiRNAs podem ser eficientemente projetados para marcar especificamente e regular de modo negativo os genes selecionados. Os determinantes de seleção de marcação de miRNAs de plantas naturais foram analisados por Schwab e colaboradores (Schwab et al. 2005,. 2005 Dev. Cell 8, 517-527). Este trabalho foi estendido ao projeto e uso de miRNAs artificiais (amiRNAs) para eficientemente regular de modo negativo os genes de marcação, resultando em conceitos e regras para o projeto de amiRNAs efetivos para silenciamento dirigido de genes (Highly Specific Gene Silencing por Artificial microRNAs in Arabidopsis, Schwab et al., Plant Cell 2006 18

(4)) e uma ferramenta na rede para o projeto eficiente de amiRNA (<http://wmd.weigelworld.org>)).

i) O RNA regulatório expressada no método da invenção pode ser por exemplo um RNA interferência transatuate pequeno (ta-siRNA) ou de um cassete de expressão assegurando a expressão do primeiro, por exemplo para a redução ou deleção da atividade de uma molécula de ácido nucleico ou talvez de um polipeptídeo cuja atividade deve ser reduzida no processo da invenção.

siRNA trans-atuantes (ta-siRNA) é um siRNA endógeno recentemente identificado (Peragine et al. 2004; Vazquez et al. 2004). A geração de ta-siRNA requer a clivagem de mRNA mediada por miRNA e a subsequente síntese de dsRNA por uma RNA polimerase dependente de RNA (RdRP), RDR6 (Allen et al. 2005). ta-siRNA regula a expressão dos genes em um modo similar ao de miRNA.

Um RNA interferência pequeno transatuate (ta-siRNA) pode ser projetado para marcar o gene de interesse a fim de induzir uma ruptura do mRNA do gene de interesse e assim silenciar a expressão do gene.

Métodos empregando ta-siRNAs utilizável para a repressão ou inativação de um produto de gene de acordo com o processo da presente invenção são descritos em US 60/672976 e 60/718645.

As sequências de ácido nucleico como descritas em itens acima são expressadas na célula ou organismo por transformação/ transfecção da célula ou organismo ou são introduzidas na célula ou organismo por método conhecidos, por exemplo como descrito em item A).

2.2 Traços agronomicamente relevantes

As sequências nucleotídicas reguladoras da transcrição quiméricas podem ser preferivelmente empregadas para conferir à planta monocotiledônea transformada um traço agronomicamente relevante. Estes traços incluem, mas não são limitados a, resistência a herbicida, tolerância a

herbicida, resistência a inseto, tolerância a inseto, resistência a doença, tolerância a doença (viral, bacteriana, fungal, nematódeos), tolerância a estresse, resistência a estresse, como exemplificado por resistência ou tolerância a seca, calor, frio, congelamento, umidade excessiva, estresse a sal e estresse oxidativo, rendimento aumentado, teor e valor dos alimentos, teor e valor de ração aumentados, aparência física, esterilidade masculina, esterilidade feminina, secagem, capacidade de ficar em pé, prolifacácia, quantidade e qualidade de amido, quantidade e qualidade de óleo, quantidade e qualidade de proteína, composição de aminoácido, e outros. Apesar de numerosas sequências de ácido nucleico serem apropriadas para serem expressadas pela sequência reguladora de ácido nucleico da transcrição quimérica da invenção (por exemplo, em combinação com promotores preferidos, por exemplo com o super-promotor) o mais preferivelmente o ácido nucleico está conferindo, quando da expressão, a uma planta monocotiledônea um traço agronomicamente relevante selecionada dentre o grupo consistindo de

- i) aumentada resistência ou tolerância contra pelo menos um fator de estresse,
- ii) qualidade nutricional aumentada de uma semente ou um broto,
- iii) rendimento aumentado.

Um dos traços mais economicamente relevantes é o rendimento. O rendimento é fortemente afetado por dano em qualquer tipo ao embrião e muda jovem. Assim, qualquer tipo de traço que proteja a muda jovem e embrião ou melhore seu desempenho é vantajoso com relação ao rendimento. Assim, um traço resultando em resistência a estresse (ver abaixo) também pode resultar em rendimento aumentado. Assim, outra forma de realização da invenção refere-se a um método para conferir rendimento aumentado e/ou tolerância ao estresse aumentada a uma planta, referido

método compreendendo as etapas de

A) introduzir em uma planta uma construção de expressão compreendendo um polinucleotídeo codificando a sequência reguladora da transcrição de plantas, em que o polinucleotídeo codificando a sequência reguladora da transcrição compreende

i) uma primeira molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:1;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C, para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e operativamente ligada à mesma

ii) uma segunda molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:3;

b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:3;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:3; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C, para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

e operativamente ligado a pelo menos um ácido nucleico que é heterólogo em relação á referida primeira ou referida segunda sequência de ácido nucleico a sequência reguladora da transcrição e é capaz de conferir a uma planta um rendimento aumentado e/ou tolerância ao estresse aumentada, e

B) selecionar plantas transgênicas, em que as plantas tem

rendimento aumentado e/ou tolerância ao estresse aumentada como comparado com as plantas de tipo selvagem ou segregante nulo.

O rendimento aumentado e/ou tolerância ao estresse aumentada e a sequência de ácido nucleico heteróloga correspondente a ser expressada são definidos como acima. Exemplos mais específicos são dados aqui abaixo. A sequência nucleotídica reguladora da transcrição quimérica preferida é descrita acima.

2.2.1 Resistência a estresse ou tolerância aumentada

As sequências nucleotídicas reguladoras da transcrição podem ser preferivelmente empregadas para conferir às planta monocotiledôneas transformadas uma resistência a estresse aumentada (ou melhorada) (preferivelmente para obter uma planta resistente ao estresse ou tolerante ao estresse). Por exemplo "resistente" significa uma planta que não demonstra substancialmente mudanças fenotípicas como uma consequência de administração de agente, infecção com um patógeno, ou exposição a estresse. Por "tolerante", significa-s uma planta que, apesar de poder demonstrar algumas mudanças fenotípicas como uma consequência de infecção, não tem uma capacidade reprodutiva substancialmente diminuída ou metabolismo substancialmente alterado.

Varias sequências de ácidos nucleicos são conhecidas dos versados na técnica para obter esta resistência ao estresse. Referidas sequências podem incluir mas não são limitadas a polinucleotídeos codificando um polipeptídeo envolvido em biossíntese de fitohormônios, regulação de fitohormônios, regulação de ciclo celular, ou metabolismo de carboidratos. O fator de estresse é preferivelmente definido como acima. A sequência de ácido nucleico heteróloga a ser expressada (por exemplo, tanto como um RNA de sentido, anti-sentido como de filamento duplo) pode codificar um polipeptídeo relacionado ao estresse como descrito acima (ou uma parte do mesmo; preferivelmente uma parte de pelo menos 5, mais

preferivelmente pelo menos 10, e mais preferivelmente pelo menos 30 aminoácidos consecutivos). A sequência nucleotídica reguladora da transcrição quimérica é descrita acima.

5 O fator de estresse e a sequência de ácido nucleico heteróloga a ser expressada são preferivelmente definidos como acima. A sequência nucleotídica reguladora da transcrição quimérica preferida é descrita acima.

A resistência a estresse, que pode ser obtida com vantagem, é preferivelmente contra um fator de estresse abiótico ou biótico. O fator de estresse biótico pode ser selecionado dentre o grupo consistindo de resistência
10 a fungos, resistência a nematódeos, resistência a insetos, resistência a vírus, e resistência a bactérias. Preferivelmente, o fator de estresse biótico é uma doença de origem em semente (principalmente doenças fúngicas, por exemplo, alfaroba comum (*Tilletia tritici*) principalmente em trigo; listra de folhas (*Pyrenophora graminea*), e fuligem solta (*Ustilago nuda*)
15 principalmente em cevada).

O fator de estresse abiótico pode ser selecionado dentre o grupo consistindo de estresse por seca, umidade excessiva, calor, resfriamento, congelamento, frio, sal, nitrogênio, densidade da população de plantas, luz UV e oxidativo. Preferivelmente, a resistência ao estresse é obtida
20 induzindo vigor prematuro

Várias sequências de ácidos nucleicos são conhecidas dos versados na técnica para obter tal resistência ao estresse. Referidas sequências podem incluir mas não estão limitadas a polinucleotídeos codificando um polipeptídeo envolvido em biossíntese de fitohormônios, regulação de
25 fitohormônios, regulação de ciclo celular, ou metabolismo de carboidratos. Exemplos mais específicos são dados abaixo.

A invenção é aplicável a todas as plantas monocotiledôneas como milho, trigo, arroz, cevada, aveia, cevada, sorgo, milhete, tritcale, erva-castelhana ou coix, mas é preferivelmente aplicável a plantas de cereal

produzindo grãos, da família Pooideae como milho, trigo, arroz, cevada, aveia, cevada, sorgo, milhete, ou triticale, preferivelmente a milho, cevada e trigo, mais preferivelmente a milho.

5 Outras formas de realização da invenção referem-se a sementes, partes e células de uma planta monocotiledônea da invenção. Preferivelmente, as partes da planta são selecionada dentre o grupo consistindo de: células, protoplastos, culturas de tecido de células, calos, grupos de células, embriões, pólen, óvulos, sementes, flores, grãos, espigas, sabugos, folhas, cascas, caules, raízes, pontas de raízes, anteros, e fios de
10 cabelo.

Indiretamente, a tolerância ao estresse aumentada pode dar origem a um ou mais traços que promovem aspectos de características agronômicas de grão melhoradas, enchimento de grão, aborto de grãos diminuído, transporte aumentado de nutrientes e semelhantes.

15 2.2.1.1 Tolerância e Resistência a Insetos

Um aspecto importante da presente invenção refere-se à introdução de genes conferindo resistência a insetos em plantas. Genes de resistência a inseto potenciais, que podem ser introduzidos, incluem genes de toxina de cristal *Bacillus thuringiensis* ou genes Bt (Watrud 1985). Genes
20 podem prover resistência a pragas de lepidópteros ou coleópteros como broca de milho européia (ECB) e verme de raiz de milho (CRW). Genes de toxina Bt preferidos para uso nestas formas de realização incluem os genes CryIA (b) e CryIA (c). Genes de endotoxina de outras espécies de *B. thuringiensis*, que afetam o desenvolvimento ou crescimento de insetos, também podem ser
25 empregados a este respeito. Inibidores de protease também podem prover resistência a inseto (Johnson 1989), e assim terão utilidade na transformação de plantas. O uso de um gene II inibidor de protease, pinII, de tomate ou batata é considerado ser particularmente utilizável. Ainda mais vantajoso é o uso de um gene pinII em combinação com um gene de toxina Bt, o efeito

combinado dos quais foi descoberto pelos presentes inventores para produzir atividade inseticida sinérgica. Outros genes, que codificam inibidores do sistema digestivo de insetos, ou os que codificam enzimas ou cofatores que facilitam a produção de inibidores, também podem ser utilizáveis. Inibidores de cistatina e amilase, como os de trigo e cevada, podem exemplificar este grupo.

Também, genes codificando lecitinas podem conferir propriedades alternativas ou adicionais. Lecitinas (originalmente denominadas fitohemaglutininas) são proteínas de ligação a carboidratos multivalentes, que têm a capacidade de aglutinar células do sangue vermelho de uma faixa de espécies. As lecitinas foram recentemente identificadas como agentes inseticidas com atividade contra gorgulhos. ECB e verme de raiz (Murdock 1990; Czapla & Lang, 1990). Genes de lecitina contemplados para ser utilizáveis incluem, por exemplo, aglutinina de cevada e germe de trigo (WGA) e lecitinas de arroz (Gatehouse 1984), com WGA sendo preferido.

Genes controlando a produção de polipeptídeos grandes ou pequenos contra insetos quando introduzidos em pragas de insetos, como, por exemplo, peptídeos líticos, hormônios de peptídeos e toxinas e venenos, formam um outro aspecto da invenção. Por exemplo, é contemplado que a expressão do hormônio da juventude esterase, dirigido para pragas de insetos específicas, também podem resultar em atividade inseticida, ou talvez causarem a cessão de metamorfose (Hammock 1990).

Plantas transgênicas expressando genes, que codificam enzimas que afetam a integridade cuticular de inseto, formam um outro aspecto da invenção. Estes genes incluem os codificando, por exemplo, quitinase, proteases, lipases e também genes para a produção de nicomicina, um composto que inibe a síntese de quitina, a introdução de qualquer um dos quais é contemplada produzir plantas de milho resistentes a insetos. Genes que codificam atividades que afetam a muda de insetos, como os que afetam a

produção de ecdisteróide UDP-glucosil transferase, também estão dentro do escopo dos transgenes utilizáveis da presente invenção.

Genes que codificam enzimas que facilitam a produção de compostos que reduzem a qualidade nutricional da planta hospedeira em pragas de insetos também são englobados pela presente invenção. Pode ser possível, por exemplo, conferir atividade inseticida em uma planta alterando sua composição de esteróis. Esteróis são obtidos pelos insetos de sua dieta e são usados para síntese hormonal e estabilidade de membrana. Consequentemente, alterações na composição de esteróis na planta pela expressão de novos genes, por exemplo, os que diretamente promovem a produção de esteróis indesejáveis ou os que convertem esteróis desejáveis em formas indesejáveis, podem ter um efeito negativo sobre o crescimento e/ou desenvolvimento de inseto e assim dotar a planta com atividade inseticida. Lipoxigenases são enzimas de plantas ocorrendo naturalmente que mostram 15
exibir efeitos antinutricionais sobre insetos e reduzir a qualidade nutricional de sua dieta. Consequentemente, outras formas de realização da invenção referem-se a plantas transgênicas com atividade de lipoxigenase melhorada, que podem ser resistentes à alimentação dos insetos.

A presente invenção também provê métodos e composições pelos quais são alcançadas mudanças qualitativas ou quantitativas em metabólitos secundários de plantas. Um exemplo refere-se a transformar plantas para produzir DIMBOA, que é contemplado, conferirá resistência a broca de milho européia, verme de raiz e várias outras pragas de insetos do milho. Genes candidatos que são particularmente considerados para uso a este respeito incluem os genes no locus bx que estão envolvidos, como se sabe, na via de DIMBOA sintética (Dunn 1981). A introdução de genes que podem regular a produção de maisina, e genes envolvidos na produção de durina em sorgo, também é contemplada ser de uso para facilitar resistência a verme de espiga e verme de raiz, respectivamente.

Tripsacum dactyloides é uma espécie de grama que é resistente a certos insetos, incluindo verme de raiz de milho. É antecipado que genes codificando proteínas que são tóxicas para insetos ou estão envolvidas na biossíntese de compostos tóxicos para insetos serão isolados de *Tripsacum* e que estes novos genes serão utilizáveis para conferir resistência a insetos. Sabe-se que a base da resistência a insetos em *Tripsacum* é genética, porque referida resistência foi transferida para *Zea mays* via cruzamentos sexuais (Branson & Guss, 1972).

Outros genes codificando proteínas caracterizadas como tendo atividade inseticida potencial também podem ser usadas como transgenes de acordo com o mesmo. Estes genes incluem, por exemplo, o inibidor de cowpea trypsin (CpTI; Hilder 1987) que pode ser usado como um repressor de verme de raiz; genes codificando avermectina (Campbell 1989; Ikeda 1987) que podem ser provados como sendo particularmente utilizáveis como um repressor de verme de raiz de milho; genes de proteínas inativando ribossoma; e até genes que regulam estruturas de plantas. Milho transgênico incluindo genes de anticorpos antiinsetos e genes que codificam enzimas que podem converter um inseticida em não tóxico (pró-inseticida) aplicado no exterior da planta em um inseticida no interior da planta também são contemplados.

2.2.1.2 Tolerância e resistência a estresse ou ambiental

A melhora de uma capacidade de planta para tolerar vários estresses ambientais como, mas não limitados a estresses de seca, umidade em excesso, nitrogênio, resfriamento, congelamento, temperatura alta, sal e oxidativo, também pode ser efetuada através da expressão de genes heterólogos ou super expressão de genes homólogos. Os benefícios podem ser realizados em termos de resistência aumentada a temperatura de congelamento através da introdução de uma proteína “anti congelamento” como a de Flounder do inverno (Cutler 1989) ou derivados de genes

5 sintéticos da mesma. Tolerância a resfriamento melhorada também pode ser conferida através da expressão aumentada de glicerol-3-fosfato acetiltransferase em cloroplastos (Murata 1992; Wolter 1992). Resistência a estresse oxidativo (frequentemente exacerbada por condições como
10 temperaturas de resfriamento em combinação com intensidade alta de luz) pode ser conferida pela expressão de superóxido (Gupta 1993), e pode ser melhorada por glutathione redutase (Bowler 1992). Estas estratégias podem permitir tolerância a congelamento em campos recentemente surgidos bem como estender variedades de rendimento mais alta com maturidade posterior a zonas de maturidade relativas mais prematuras.

A expressão de novos genes que afetam favoravelmente o teor de água na planta, potencial de água total, potencial osmótico e turgescência pode conferir a capacidade da planta tolerar seca. Como usado, os termos “resistência à seca” e “tolerância à seca” são usados para referir-se a uma
15 tolerância ou resistência aumentada da planta a estresse induzido por uma redução na disponibilidade de água, quando comparado a circunstâncias normais, e à capacidade da planta funcionar e sobreviver em ambientes com baixo teor de água, e funcionar de um modo relativamente superior. Neste aspecto da invenção é proposto, por exemplo, que a expressão de um gene
20 codificando a biossíntese de solutos osmoticamente ativos pode conferir proteção contra seca. Dentro desta classe de genes estão DNAs codificando manitol desidrogenase (Lee e Saier, 1982) e trealose-6-fosfato sintase (Kaasen 1992). Através da ação subsequente de fosfatases nativas na célula ou pela introdução e co-expressão de uma fosfatase específica, estes genes
25 introduzidos resultarão no acúmulo tanto de manitol como de trealose, respectivamente, ambas as quais são bem documentadas como compostos protetores capazes de mitigar os efeitos do estresse. O acúmulo de manitol em tabaco transgênico foi verificado e resultados preliminares indicam que plantas expressando altos níveis deste metabólito são capazes de tolerar um

estresse osmótico aplicado (Tarczynski 1992).

5 Similarmente, a eficácia de outros metabólitos na proteção tanto da função enzimática (por exemplo, alanopina ou ácido propiônico) ou integridade da membrana (por exemplo, alanopina) foi documentada (Loomis
10 1989), e, conseqüentemente, a expressão do gene codificando a biossíntese destes compostos pode conferir resistência a seca de um modo similar a ou complementar a manitol. Outros exemplos de metabólitos ocorrendo naturalmente que são osmoticamente ativos e/ou provêm algum efeito protetor direto durante seca e/ou dissecação incluem derivados de açúcar
15 como frutose, eritritol (Coxson 1992), sorbitol, dulcitol (Karsten 1992), glucosilglicerol (Reed 1984; Erdmann 1992), sacarose, estaquiose (Koster & Leopold 1988; Blackman 1992), ononitol e pinitol (Vernon & Bohnert 1992), e rafinose (Bernal-Lugo & Leopold 1992). Outros solutos osmoticamente ativos, que não são açúcares, incluem, mas não estão limitados a prolina e
20 glicina-betaína (Wyn-Jones e Storey, 1981). O crescimento de canópia continuado e arranjo reprodutivo aumentado durante épocas de estresse podem ser aumentados pela introdução e expressão de genes como os que controlam os compostos osmoticamente ativos discutidos acima e outros destes compostos, como representado em uma forma de realização exemplar
25 pela enzima mioinositol 0-metiltansferase.

 É contemplado que a expressão de proteínas específicas também pode aumentar a tolerância à seca. Três classes de proteínas embriogênicas posteriores são designadas com base em similaridades estruturais (ver Dure 1989). Todas as três classes destas proteínas são
25 demonstradas em sementes em maturação (isto é, dessecantes). Nestes 3 tipos de proteínas, o tipo II (tipo desidrina) está geralmente implicado em tolerância à seca e/ou dessecação em partes de plantas de vegetais (por exemplo Mundy e Chua, 1988; Piatkowski 1990; Yamaguchi-Shinozaki 1992). Recentemente, verificou-se que a expressão de um LEA tipo III

(HVA-1) em tabaco influencia a altura da planta, tolerância à seca e amadurecimento) (Fitzpatrick, 1993). A expressão de genes estruturais de todos os três grupos pode, conseqüentemente, conferir tolerância à seca. Outros tipos de proteínas induzidas durante o estresse de água incluem tiol

5 proteases, aldolases e transportadores de transmembrana (Guerrero 1990), que podem conferir várias funções protetoras e/ou do tipo reparo durante estresse de seca. A expressão de um gene que efetua biossíntese de lipídeo e, conseqüentemente, composição de membrana também pode ser utilizável para conferir resistência à seca na planta.

10 Muitos genes que melhoram a resistência à seca têm modos complementares de ação. Assim, combinações destes genes podem ter efeitos aditivos e/ou sinérgicos na melhora da resistência à seca em milho. Muitos destes genes também melhoram a tolerância ao congelamento (ou resistência); os estresses físicos incorridos durante congelamento e seca são similares na

15 natureza e podem ser mitigados de modo similar. Benefício pode ser conferido via expressão constitutiva ou específica de tecido destes genes, mas o meio preferido de expressar estes novos genes pode ser através do uso de um promotor induzido por turgescência (como os promotores para os genes induzidos por turgescência descritos em Guerrero et al. 1990 e Shagan 1993).

20 Padrões de expressão especiais e temporais destes genes podem capacitar o milho com uma melhor resistência a estresse.

A expressão de genes que estão envolvidos com traços morfológicos específicos que permitem extrações de água aumentadas de solo seco pode ser de benefício. Por exemplo, a introdução e expressão de genes

25 que afetam as características da raiz podem melhorar a absorção de água. A expressão de genes que melhoram as condições reprodutivas durante épocas de estresse pode ser de valor significativo. Por exemplo, a expressão de DNAs que melhoram a sincronia de abrigo de pólen e a receptividade das flores de plantas do sexo feminino, isto é, fios de cabelo, pode ser de benefício. Além

disso, a expressão de genes que minimizam o aborto dos cernes de cereais durante épocas de estresse pode aumentar a quantidade de grão a ser coletado e conseqüentemente ser de valor. A regulação de níveis de citoquinina em monocotiledôneas, como milho, por introdução e expressão de um gene isopentenila transferase com sequências regulatórias apropriadas pode melhorar o rendimento e resistência a estresse de monocotiledôneas (Gan 1995).

Dado o papel global da água na determinação do rendimento, é contemplado que possibilitar que as plantas utilizem água mais eficientemente, através da introdução e expressão de novos genes, melhorará o desempenho global mesmo quando a disponibilidade de água no solo não é limitante. Introduzindo genes que melhoram a capacidade das plantas maximizar o uso de água através de uma faixa total de estresses com relação à disponibilidade de água, a estabilidade de rendimento ou consistência de desempenho de rendimento pode ser realizada.

Proteção melhorada da planta em fatores de estresse abióticos como seca, calor ou resfriamento também pode ser obtida – por exemplo – super expressando polipeptídeos anticongelamento de *Myoxocephalus Scorpius* (WO 00/00512), *Myoxocephalus octodecemspinosus*, o ativador CBF1 de transcrição de *Arabidopsis thaliana*, glutamato desidrogenases (WO 97/12983, WO 98/11240), genes de proteínas dependentes de cálcio (WO 98/26045), calcineurinas (WO 99/05902), caseína quinase de levedura (WO 02/052012), farnesiltransferases (WO 99/06580; Pei ZM *et al.* 1998), ferritina (Deak M *et al.* 1999), oxalato oxidase (WO 99/04013; Dunwell JM 1998), fator DREB1A (“elemento B 1A de resposta a desidratação”; Kasuga M *et al.* 1999), genes de manitol ou síntese de trealose-fosfato fosfatase (WO 97/42326) ou inibindo genes como trehalase (WO 97/50561).

Um uso para sequências quiméricas é proteger o embrião de dano por frio durante a germinação. Um fator importante é dano oxidativo. O

super-promotor pode dirigir, isto é, catalase, ascorbato peroxidase, superóxido dismutase e semelhantes. O frio afeta a enzima COX também através de uma membrana rígida. Para expressão de estresse por seca de glutamina sintase e glicina betaína sintase pode ser benéfica.

5 2.2.1.3 Tolerância e Resistência a Doenças

Propõe-se que resistência aumentada a doenças pode ser realizada através da introdução de genes em plantas. É possível produzir resistência a doenças causadas por vírus, bactérias, fungos, patógenos de raiz, insetos e nematódeos. Também é contemplado que o controle de organismos produzindo micotoxina pode ser realizado através da expressão de genes introduzidos.

Resistência a vírus pode ser produzida através da expressão de novos genes. Por exemplo, é demonstrado que a expressão de uma proteína de revestimento viral em uma planta transgênica pode conferir resistência a infecção da planta por esse vírus e talvez outros vírus intimamente relacionados (Cuozzo 1988, Hemenway 1988, Abel 1986). É contemplado que a expressão de genes alvos anti-sentido em funções virais essenciais pode conferir resistência a referido vírus. Por exemplo, um gene alvo anti-sentido no gene responsável por replicação de ácido nucleico pode inibir referida replicação e levar à resistência ao vírus. Acredita-se que interferência com outras funções virais através do uso de genes anti-sentido também pode aumentar resistência a vírus. Além disso é proposto que pode ser possível obter resistência a vírus através de outras abordagens, incluindo, mas não limitadas ao uso de vírus de satélite.

É proposto que resistência aumentada a doenças causadas por bactérias e fungos pode ser realizada através da introdução de novos genes. É contemplado que genes codificando os assim chamados “antibióticos de peptídeos,” proteínas relacionadas a patogênese (PR), resistência a toxina, e proteínas afetando interações hospedeiro-patógeno como características

morfológicas serão utilizáveis. Antibióticos de peptídeos são sequências de polipeptídeo, que são inibitórias para o crescimento de bactérias e outros microorganismos. Por exemplo, as classes de peptídeos referidas como cecropinas e magaininas inibem o crescimento de muitas espécies de bactérias e fungos. É proposto que a expressão de proteínas PR em plantas pode ser utilizável para conferir resistência a doença bacteriana. Estes genes são induzidos após ataque por patógenos em uma planta hospedeira e são divididos em pelo menos cinco classes de proteínas (Bol 1990). Incluídas dentre as proteínas PR estão beta-1,3-glucanases, quitinases, e osmotina e outra proteínas que acredita-se funcionem na resistência da planta a organismos doentes. Outros genes são identificados que têm propriedades antifúngicas, por exemplo, UDA (lecitina de espinho de urtiga) e heveína (Broakgert 1989; Barkai-Golan 1978). Sabe-se que certas doenças de plantas são causadas pela produção de fitotoxinas. A resistência a estas doenças pode ser obtida através da expressão de um novo gene que codifica uma enzima capaz de degradar ou de outro modo inativar a fitotoxina. A expressão de novos genes que alteram as interações entre a planta hospedeira e patógeno pode ser utilizável na redução da capacidade do organismo doente invadir os tecidos da planta hospedeira, isto é, um aumento na cerosidade da cutícula foliar ou outras características morfológicas.

Nematódeos de plantas parasíticas são uma causa de doença em muitas plantas. É proposto que pode ser possível tornar uma planta resistente a estes organismos através da expressão de novos genes. É antecipado que o controle de infestações de nematódeos pode ser realizado alterando a capacidade do nematódeo reconhecer ou atacar uma planta hospedeira e/ou capacitar a planta a produzir compostos nematicidas, incluindo mas não limitadas a proteínas.

Além disso, uma resistência a fungos, insetos, nematódeos e doenças pode ser obtida pelo acúmulo marcado de certos metabólitos ou

proteínas. Estas proteínas incluem mas não estão limitadas a glucosinolatos (defesa contra herbívoros), quitinases ou glucanases e outras enzimas que destroem a parede celular de parasitas, proteínas inativando ribossomas (RIPs) e outras proteínas de resistência de planta e reação a estresse como são

5 induzidas quando são feridas ou atacadas por micróbios, ou quimicamente, por exemplo, por ácido salicílico, ácido jasmônico ou etileno, ou lisozimas de fontes não plantas, como, por exemplo, T4-lisozima ou lisozima de uma variedade de mamíferos, proteínas inseticidas como a endotoxina de *Bacillus thuringiensis*, inibidor de alfa-amilase ou inibidores de protease (inibidor de

10 tripsina de feijão de corda), lecitinas como aglutinina de germe de trigo, RNAses ou ribozimas. Outros exemplos são ácidos nucleicos que codificam a endoquitinase *Trichoderma harzianum* chit42 (GenBank Acc. No.: S78423) ou a proteína de citocromo N-hidroxilante, multi-funcional, P-450 (CYP79) de sorgo bicolor (GenBank Acc. No.: U32624), ou equivalentes funcionais

15 destes. O acúmulo de glucosinolatos como proteção de pragas (Rask L *et al.* 2000; Menard R *et al.* 1999), a expressão de endotoxinas de *Bacillus thuringiensis* (Vaeck *et al.* 1987) ou a proteção contra ataque por fungos, pela expressão de quitinases, por exemplo de feijões (Broglie e al. 1991), são vantajosos. Resistência a pragas como, por exemplo, a praga de arroz

20 *Nilaparvata lugens* em plantas de arroz pode ser obtida expressando a aglutinina de lecitina de fura-neve (*Galanassim nivalis*) (Rao *et al.* 1998). A expressão de genes cryIA(b) e cryIA(c) sintéticos, que codificam D-endotoxinas de *Bacillus thuringiensis* específicas de lepidópteros pode dar origem à resistência a pragas de insetos em várias plantas (Goyal RK *et al.*

25 2000). Outros genes que são apropriados para defesa contra patógenos compreendem a “proteína de inibição de poligalacturonase” (PGIP), taumatina, invertase e peptídeos antimicrobianos como lactoferrina (Lee TJ *et al.* 2002). Outras sequências de ácido nucleico que podem ser vantajosamente usadas aqui incluem traços para controle de insetos (U.S. Nos. 6.063.597;

6.063.756; 6.093.695; 5.942.664; e 6.110.464), resistência a doenças fúngicas (U.S. Pat. Nos. 5.516.671; 5.773.696; 6.121.436; 6.316.407; e 6.506.962), resistência a vírus (U.S. Pat. Nos. 5.304.730 e 6.013.864), resistência a nematódeos (US 6.228.992), e resistência a doença bacteriana (US 5.516.671).

A sequência de ácido nucleico heteróloga a ser expressada pode codificar um polipeptídeo relacionado a estresse (ou uma parte de pelo menos 5, mais preferivelmente pelo menos 10, mais preferivelmente pelo menos 30 aminoácidos consecutivos). Sequência de nucleotídeo regulando transcrição quimérica preferida é descrita acima.

2.2.2 Qualidade nutricional aumentada de uma semente ou um broto

As sequências reguladoras da transcrição nucleotídicas quiméricas podem ser preferivelmente empregadas para conferir à planta monocotiledônea transformada qualidade nutricional aumentada (ou melhorada) de uma semente ou um broto. Assim, outra forma de realização da invenção refere-se a um método para conferir qualidade nutricional e/ou teor de óleo aumentados de uma semente ou um broto a uma planta, referido método compreendendo as etapas de

A) introduzir em uma planta uma construção de expressão compreendendo um polinucleotídeo codificando a sequência reguladora da transcrição de plantas, em que o polinucleotídeo codificando a sequência reguladora da transcrição compreende

i) uma primeira molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:1; e

- c) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e operativamente ligadas to
- 15 ii) uma segunda molécula de ácido nucleico selecionada dentre o grupo consistindo de
- a) um polinucleotídeo como definido em SEQ ID NO:3;
- b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 20 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:3; e
- c) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais 25 preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM

EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

e operativamente ligado a pelo menos um ácido nucleico que é heterólogo em relação à referida primeira ou referida segunda sequência de ácido nucleico e é apropriado para conferir a uma planta uma qualidade nutricional e/ou teor de óleo aumentados de uma semente ou um broto, e

B) selecionar plantas transgênicas, em que as plantas tem qualidade nutricional e/ou teor de óleo aumentados de uma semente ou um broto como comparado com as plantas de tipo selvagem ou segregante nulo.

A qualidade nutricional e/ou teor de óleo podem compreender um teor aumentado de pelo menos um composto selecionada dentre o grupo consistindo de vitaminas, carotinóides, antioxidantes, ácidos graxos insaturados, e ácidos graxos poli-insaturados. A sequência de ácido nucleico heteróloga a ser expressada (por exemplo, ou como um RNA sentido, anti-sentido ou de filamento duplo) pode codificar um polipeptídeo relacionado com o traço (ou uma parte do mesmo, preferivelmente uma parte de pelo menos 5, mais preferivelmente pelo menos 10, o mais preferivelmente pelo menos 30 aminoácidos consecutivos) como descrito abaixo, ou um equivalente funcional do mesmo, que é capaz de ocasionar o mesmo fenótipo do que qualquer um do referido polipeptídeo. A sequência nucleotídica reguladora da transcrição quimérica é descrita acima.

A qualidade nutricional e a sequência de ácido nucleico heteróloga correspondente a ser expressada são definidas aqui abaixo. A sequência nucleotídica reguladora da transcrição quimérica é descrita acima, o mais preferido é o super-promotor. A planta monocotiledônea à qual os métodos desta invenção são preferivelmente aplicados pode ser selecionada

dentre o grupo consistindo de milho, trigo, arroz, centeio, aveia, centeio, sorgo, erva-castelhana ou coix. Preferivelmente a planta é uma planta de cereal selecionada dentre o grupo consistindo de milho, trigo, centeio, arroz, aveia, centeio, e sorgo, ainda mais preferivelmente de milho, trigo, e arroz, o
5 mais preferivelmente a planta é uma planta de milho.

Uma qualidade nutricional aumentada pode – por exemplo - resultar em uma ou mais das seguintes propriedades: modificar a composição de ácido graxo em uma planta, alterar o teor de aminoácido de uma planta, ou aumentar a concentração de um metabólito de planta.

10 Genes podem ser introduzidos em plantas monocotiledôneas, particularmente cereais comercialmente importantes como milho, trigo ou arroz, para melhorar o grão para o qual o cereal é primariamente cultivado. Uma ampla faixa de novas plantas transgênicas produzidas deste modo pode ser visada dependendo do uso final particular do grão.

15 Por exemplo, o maior uso de grãos de milho é para ração ou alimentos. A introdução dos genes que alteram a composição do grão pode melhorar muito o valor da ração ou alimento. Os componentes primários do grão de milho são amido, proteína, e óleo. Cada um destes componentes primários de grão de milho pode ser melhorado por alteração de seu nível ou
20 composição. Vários exemplos podem ser mencionados para fins ilustrativas mas de nenhum modo fornecendo uma lista exaustiva das possibilidades.

A proteína de muitos grãos de cereal é subótima para fins de ração e alimento, especialmente quando alimentando suínos aves e humanos. A proteína é deficiente em vários aminoácidos que são essenciais na dieta
25 destas espécies, requerendo a adição de suplementos ao grão. Os aminoácidos essenciais limitativos podem incluir lisina, metionina, triptofano, treonina, valina, arginina, e histidina. Alguns aminoácidos se tornam limitativos somente após o grão ser suplementado com outros adicionais para as formulações da ração. Por exemplo, quando o grão é suplementado com

farinha de soja para atender às exigências de lisina, metionina se torna limitada. Os níveis destes aminoácidos essenciais em sementes e grãos podem ser elevado por mecanismos, incluem, mas não são limitados a, introdução de genes para aumentar a biossíntese dos aminoácidos, diminuir a degradação dos aminoácidos, aumentar o armazenamento de aminoácidos em proteínas, ou aumentar o transporte dos aminoácidos para as sementes ou grãos.

Um mecanismo para aumentar a biossíntese dos aminoácidos é introduzir genes que desregulam as vias de biossíntese de aminoácido de modo que a planta não pode mais controlar de modo apropriado os níveis que são produzidos. Isto pode ser feito por etapas de desregulação ou de desvio com as vias de biossíntese de aminoácidos que são normalmente reguladas por níveis do produto final de aminoácido da via. Exemplos incluem a introdução de genes que codificam versões desreguladas das enzimas aspartoquinase ou ácido di-hidrodipicolínico (DHDP)-sintase para aumentar produção de lisina e treonina, e antranilato sintase para aumentar produção de triptofano. Redução do catabolismo dos aminoácidos pode ser obtida por introdução de sequências de DNA que reduzem ou eliminam a expressão de genes codificando enzimas que catalisam etapas nas vias catabólicas como a enzima lisina-cetoglutarato reductase.

A composição de proteína do grão pode ser alterada para melhorar o equilíbrio de aminoácidos em vários modos incluindo elevar a expressão de proteínas nativas, diminuir a expressão das com uma composição fraca, mudar a composição de proteínas nativas, ou introduzir genes codificando proteínas completamente novas, possuindo uma composição superior. DNA pode ser introduzido que diminui a expressão de membros da família zeína de proteínas de armazenamento. Este DNA pode codificar ribozimas ou sequências anti-sentido dirigidas para afetar a expressão de proteínas zeína ou expressão de reguladores de expressão de zeína como o produto de gene opaque-2. Ca composição de proteína do grão

pode ser modificada através de fenômeno de co-supressão, *isto é*, inibição de expressão de um gene endógeno através de expressão de um gene estrutural idêntico ou do fragmento de gene introduzido através de transformação (Goring 1991). Além disso, o DNA introduzido pode codificar enzimas, que degradam zeína. A diminuição em expressão de zeína que é obtida pode ser acompanhada por aumentos em proteínas com composição de aminoácido mais desejável ou aumentos em outros componentes principais da semente como amido. Alternativamente, um gene quimérico pode ser introduzido que compreende a sequência de codificação para uma proteína nativa de composição de aminoácido adequada como para uma das proteínas globulina ou zeína 10 kD de milho e um promotor outra sequência regulatória projetada para elevar a expressão de referida proteína. A sequência de codificação de referido gene pode incluir códons adicionais ou de substituição para os aminoácidos essenciais. Além disso, a sequência de codificação obtida de outra espécie, ou, por sequência parcialmente ou completamente sintética codificando completamente uma sequência de peptídeo única projetada para melhorar a composição de aminoácido de uma semente pode ser empregada.

A introdução de genes que alteram o teor de óleo do grão pode ser de valor. Aumentos em teor de óleo podem resultar em aumentos em teor de energia metabolizável e densidade das sementes para uso em ração e alimentos. Os genes introduzidos podem codificar enzimas que removem ou reduzem limitações de taxa ou etapas reguladas em biossíntese de ácido graxo ou lipídeos. Estes genes podem incluir, mas não são limitados aos que codificam acetil-CoA carboxilase, ACP-aciltransferase, beta-cetoacil-ACP sintase, mais outras atividades biossintéticas de ácidos graxos conhecidas. Outras possibilidades são genes que codificam proteínas que não possuem atividade enzimática como proteína veículo acila. Os exemplos adicionais incluem 2-acetiltransferase, complexo oleosina piruvato desidrogenase, acetil CoA sintetase, ATP citrato liase, ADP-glucose pirofosforilase e genes dos

shuttles de carnitina-CoA-acetil-CoA. Antecipa-se que a expressão de genes relacionados com a biossíntese de óleo serão marcados para o plastídeo, usando uma sequência de peptídeo de plastídeo de trânsito e preferivelmente expressada em um embrião de semente. Genes pode ser introduzidos que alteram o equilíbrio de ácidos graxos presentes no óleo provendo alimentos mais saudáveis ou nutritivos. O DNA introduzido pode também codificar sequências que bloqueiam a expressão de enzimas envolvidas em biossíntese de ácido graxo, alterando as proporções de ácidos graxos presentes no grão como descrito abaixo.

10 Genes podem ser introduzidos que melhoram o valor de nutrição do componente de amido do grão, por exemplo por aumento do grau de ramificação, resultando em uma melhorada utilização do amido em vacas por retardar seu metabolismo.

15 Além de afetar a maior parte dos componentes do grão, os genes podem ser introduzidos que afetam vários outros aspectos nutrientes, de processamento ou de qualidade do grão, como usado para ração ou alimento. Por exemplo, a pigmentação do grão pode ser aumentada ou diminuída. A melhora e a estabilidade da pigmentação amarela é desejável em algumas rações de animais e pode ser obtida por introdução de genes que resulta em 20 melhorada produção de xantofilas e caroteno por eliminação das etapas de limitação de taxa em sua produção. Estes genes podem codificar formas alteradas de enzimas, fitoeno sintase, fitoeno desaturase, ou licopeno sintase. Alternativamente, o milho branco não pigmento é desejável para a produção de muitos produtos alimentícios e pode ser produzido pela introdução de 25 DNA, que bloqueia ou elimina a etapa nas vias de produção de pigmentos.

A ração ou alimento compreendendo grãos de cereal possui quantidades insuficientes de vitaminas e deve ser suplementada para prover adequado valor nutricional. A introdução de genes que melhoram a biossíntese de vitaminas em sementes pode ser visada incluindo, por exemplo,

vitaminas A, E, B₁₂, colina, e outros. Por exemplo, o grão de milho também não possui um teor de mineral suficiente para um valor nutricional ótimo. Genes que afetam o acúmulo ou disponibilidade de compostos contendo fósforo, enxofre, cálcio, manganês, zinco e ferro, dentre outros, devem ser valiosos. Um exemplo pode ser a introdução de um gene que reduz a produção de ácido fítico ou codifica a enzima fitase, que melhora a ruptura do ácido fítico. Estes genes devem aumentar os níveis de fosfato disponível na dieta, reduzindo a necessidade para suplementação com mineral fosfato.

Numerosos outros exemplos de melhora de cereais para fins de ração e alimentos podem ser descritos. As melhoras podem ainda não necessariamente envolver o grão mas podem, por exemplo melhorar o valor do grão para silagem. Introdução of DNA para obter isto inclui as sequências que alteram a produção de lignina como as que resultam em fenótipo "da nervura central das folhas marrom" associado com um valor superior da ração para gado.

Além de dirigir as melhoras no valor da ração ou alimento, os genes também podem ser introduzidos que melhoram o processamento do grão e melhoram o valor dos produtos resultantes do processamento. O método primário de processamento de alguns grãos como milho é via moagem a úmido. O milho pode ser melhorado através da expressão de novos genes que aumentam a eficácia e eficiência e reduzem o custo do processamento como por diminuição do tempo de infusão.

A melhora do valor dos produtos moídos pode incluir a alteração da quantidade ou da qualidade do amido, óleo, farinha de glúten de milho ou os componentes da ração com glúten de milho. A elevação do teor de amido pode ser obtida através da identificação e eliminação das etapas de limitação da taxa em biossíntese de amido ou por diminuição dos níveis dos outros componentes do grão resultando em aumentos proporcionais em amido. Um exemplo do primeiro pode ser a introdução de genes codificando

enzimas ADP-glucose pirofosforilase com atividade regulatória alterada ou que são expressadas em maiores níveis. Exemplos do último podem incluir inibidores seletivos de, por exemplo, biossíntese de proteína ou óleo expressada durante os estágios posteriores do desenvolvimento do cerne.

5 O óleo é outro produto da moagem a úmido de milho e outros grãos, cujo valor pode ser melhorado por introdução e expressão de genes. A quantidade de óleo que pode ser extraída por moagem a umido pode ser elevada por abordagens com descrito para ração e alimentos acima. As propriedades do óleo também podem ser alteradas para melhorar seu

10 desempenho na produção e uso de óleo de cozinha, molhos, lubrificantes ou outros produtos derivados de óleo ou e melhora de seus atributos para a saúde, quando usado em aplicações relacionadas com alimentos. Novos ácidos graxos também podem ser sintetizados que quando da extração podem servir como matérias primas para síntese química. As mudanças nas propriedades do

15 óleo podem ser obtidas, por alteração do tipo, nível, ou arranjo de lipídeos dos ácidos graxos presentes no óleo. Isto por sua vez pode ser obtido por adição de genes que codificam enzimas que catalisam a síntese de novos ácidos graxos e lipídeos possuindo os mesmos ou por aumento dos níveis de ácidos graxos nativos enquanto possivelmente reduzindo os níveis de precursores.

20 Alternativamente sequências de DNA podem ser introduzidas que retardam ou bloqueiam as etapas em biossíntese de ácido graxo resultando em aumento no precursor de intermediários de ácido graxo. Genes que podem ser adicionados incluem desaturases, epoxidases, hidratases, desidratases, e

25 outras enzimas que catalisam as reações envolvendo os intermediários de ácido graxo. Os exemplos representativos de etapas catalíticas que podem ser bloqueadas incluem as desaturações de ácido esteárico para oleico e ácido oleico para linolênico resultantes em acúmulos respectivos de ácidos esteárico e oleico.

As melhoras em outros produtos de moagem a úmido de

cereais principais, farinha de glúten e ração de glúten também podem ser obtidas por introdução de genes para obter novas plantas. As possibilidades representativas incluem mas não são limitadas às descritas acima para melhora do valor da ração e dos alimentos.

5 Além disso, pode ser ainda considerado que a planta, a ser usada para a produção ou fabricação de compostos biológicos utilizáveis, pode não ter sido produzidas de todo, ou ter sido produzida no mesmo nível da planta previamente. As novas plantas produzindo estes compostos são
10 tornadas possíveis para a introdução e expressão de genes por métodos de transformação. As possibilidades incluem, mas não são limitados a, qualquer composto biológico que é atualmente produzido por qualquer organismo como proteínas, ácidos nucleicos, metabólitos primários e intermediários, polímeros de carboidratos, *etc.* Os compostos podem ser produzidos pela planta, extraídos quando da colheita e/ou processamento e usados para
15 qualquer fim utilizável atualmente reconhecido como fármacos, fragrâncias, enzimas industriais, para apenas nomear alguns.

Outras possibilidades para exemplificar a faixa de traços ou propriedades de grãos potencialmente codificados por genes introduzidos em plantas transgênicas incluem grão com menor susceptibilidade à ruptura para
20 fins de exportação ou com maior tamanho da partícula fina, quando processados por moagem a seco através da introdução de genes que melhoram a síntese de gama-zeína, pipoca com melhorado estouro, qualidade e volume de expansão através de genes que aumentam a espessura do pericarpo, milho com grão mais branco para uso em alimentos através da introdução de genes
25 que efetivamente bloqueiam a expressão of enzimas envolvidas em vias de produção de pigmentos, e bebidas alcoólicas de melhorada qualidade ou milho doce através da introdução de genes que afetam o sabor como o gene *shrunk* (codificando sacarose sintase) para milho doce.

As sequências de ácido nucleico utilizáveis que podem ser

combinadas com o sequência de ácido nucleico de promotor da presente invenção e proporcionar traços de produto final melhorados incluem, sem limitação, as proteínas codificando armazenamento de semente, enzimas de via de ácido graxo, enzimas biossintéticas de tocoferol, enzimas biossintéticas de aminoácido, e enzimas de ramificação de amido. Uma discussão de DNA heterólogos exemplares utilizáveis para a modificação de fenótipos de planta pode ser encontrada em, por exemplo, U.S. Nos. 6.194.636; 6.207.879; 6.232.526; 6.426.446; 6.429.357; 6.433.252; 6.437.217; 6.515.201; e 6.583.338 e WO 02/057471, cada sendo especificamente incorporado aqui por referência em sua totalidade. Estes traços incluem mas não são limitados a:

- Expressão de enzimas metabólicas para uso no setor de alimentos e ração, por exemplo de fitases e celulasas. Especialmente preferidos são ácidos nucleicos como o cDNA artificial que codifica uma fitase microbiana (GenBank Acc. No. A19451) ou equivalentes funcionais dos mesmos.

- Expressão de genes que ocasionam um acúmulo de produtos químicos finos como de tocoferóis, tocotrienóis ou carotenóides. Um exemplo que pode ser mencionada é fitoeno desaturase. Preferidos são ácidos nucleicos que codificam fitoeno desaturase de *Narcissus pseudonarcissus* (GenBank Acc. No.X78815) ou equivalentes funcionais dos mesmos. Preferidas enzimas biossintéticas de tocoferol incluem tyrA, slr1736, ATPT2, dxs, dxr, GGPPS, HPPD, GMT, MT1, tMT2, AANT1, slr 1737, e uma construção anti-sentido para ácido homogentísico dioxigenase (Kridl *et al.* (1991); Keegstra (1989); Nawrath *et al.* (1994); Xia *et al.* (1992); Lois *et al.* (1998); Takahashi *et al.* (1998); Norris *et al.* (1998); Bartley e Scolnik (1994); Smith *et al.* (1997); WO 00/32757; WO 00/10380; Saint Guily *et al.* (1992); Sato *et al.* (2000), todos sendo incorporados aqui por referência.

- produção de amido (U.S. Nos. 5.750,876 e 6.476.295), produção de alto teor de proteína (US 6,380,466), amadurecimento de frutas

(US 5,512,466), nutrição animal e humana melhorada (U.S. Nos. 5.985.605 e 6.171.640), biopolímeros (US 5,958,745 e Publicação Patente U.S. No. 2003/0028917), resistência a estresse ambiental (US 6,072,103), peptídeos farmacêuticos (US 6,080,560), traços de processamento melhorados (US 6.476.295), melhorada digestibilidade (US 6.531.648), baixo teor de rafinose (US 6.166.292), produção industrial de enzima (US 5,543,576), aroma melhorado (US 6.011.199), fixação de nitrogênio (US 5.229.114), produção de semente híbrida (US 5.689.041), e produção biocombustível (US 5.998.700), os elementos genéticos e transgenes descritos nas patentes listados acima são aqui incorporados por referência. Enzimas de ramificação de amido preferidas (para modificação de propriedades de amido) incluem as especificadas em U.S. Nos. 6.232.122 e 6.147.279; e WO 97/22703, todas sendo incorporadas aqui por referência.

- Produção de óleos modificados (US 6.444.876), produção com alto teor de óleo (U.S. Nos. 5.608.149 e 6.476.295), ou teor de ácido graxo modificado (US 6.537.750). As enzimas de via de ácido graxo preferida incluem tioesterases (U.S. Nos. 5.512.482; 5.530.186; 5.945.585; 5.639.790; 5.807.893; 5.955.650; 5.955.329; 5.759.829; 5.147.792; 5.304.481; 5.298.421; 5.344.771; e 5.760.206), diacilglicerol aciltransferases (Publicações de Patentes U.S. 20030115632A1 e 20030028923A1), e desaturases (U.S. Nos. 5.689.050; 5.663,068; 5.614.393; 5.856.157; 6.117.677; 6.043.411; 6.194.167; 5.705.391; 5.663,068; 5.552.306; 6.075.183; 6.051.754; 5.689.050; 5.789.220; 5.057.419; 5.654.402; 5.659.645; 6.100.091; 5.760.206; 6.172.106; 5.952.544; 5.866.789; 5.443.974; e 5.093.249) todas sendo incorporadas aqui por referência.

- As enzimas biossintéticas aminoácido preferidas incluem antranilato sintase (US 5.965.727 e WO 97/26366, WO 99/11800, WO 99/49058), triptofano decarboxilase (WO 99/06581), treonina decarboxilase (U.S. Nos. 5,534.421 e 5.942.660; WO 95/19442), treonina deaminase (WO

99/02656 e WO 98/55601), ácido di-hidrodipicolínico sintase (US 5,258,300), e aspartato quinase (U.S. Nos. 5.367.110; 5.858.749; e 6.040.160) todas sendo incorporadas aqui por referência.

- Produção de nutracêuticos como, por exemplo, ácidos poligraxos insaturados (por exemplo ácido araquidônico, ácido eicosapentaenoico, ou ácido docosahexaenóico) por expressão de elongases e/ou desaturases de ácido graxo, ou produção de proteínas com melhorado valor nutricional como, por exemplo, com um teor elevado de aminoácidos essenciais (por exemplo o gene de albumina com elevado teor de metionina 2S da castanha do Pará). Preferidos são os ácidos nucleicos que codificam a albumina com elevado teor de metionina 2S de *Bertholletia excelsa* (GenBank Acc. No. AB044391), a Δ^6 -acil-lipídeo desaturase de *Physcomitrella patens* (GenBank Acc. No. AJ222980; Girke *et al.* 1998), a Δ^6 -desaturase de *Mortierella alpina* (Sakuradani *et al.* 1999), a Δ^5 -desaturase de *Caenorhabditis elegans* (Michaelson *et al.* 1998), a Δ^5 - ácido graxo desaturase de *Caenorhabditis elegans* (des-5) (GenBank Acc. No. AF078796), a Δ^5 -desaturase de *Mortierella alpina* (Michaelson *et al.* JBC 273:19055-19059), a Δ^6 -elongase de *Caenorhabditis elegans* (Beaudoin *et al.* 2000), a Δ^6 -elongase de *Physcomitrella patens* (Zank *et al.* 2000), ou equivalentes funcionais dos mesmos.

- Produção de proteínas e enzimas de elevada qualidade para fins industriais (por exemplo enzimas, como lipases) ou fármacos (como, por exemplo, anticorpos, fatores de coagulação do sangue, interferons, linfocinas, fator de estimulação de colônia, ativadores de plasminogênio, hormônios ou vacinas, como descrito por Hood EE e Jilka JM 1999). Por exemplo, foi possível produzir avidina recombinante a partir de albúmen de galinha e β -glucuronidase bacteriano (GUS) em uma larga escala em plantas de milho transgênico (Hood *et al.* 1999).

- Obter uma capacidade de armazenamento aumentado em

células que normalmente compreendem menos proteínas de armazenamento ou lipídeos de armazenamento, com o fim de aumentar o rendimento destas substâncias, por exemplo por expressão de acetil-CoA carboxilase. Preferidos ácidos nucleicos são os que codificam acetil-CoA carboxilase de *Medicago* sativa (ACCase) (GenBank Acc. No. L25042), ou equivalentes funcionais dos mesmos. Alternativamente, em alguns cenários, um teor de proteína de armazenamento aumentado pode ser vantajoso para a produção de produto com elevado teor de proteína. As proteínas de armazenamento de semente preferidas incluem zeínas (U.S. Nos. 4.886.878; 4.885.357; 5.215.912; 5.589.616; 5.508.468; 5.939.599; 5.633.436; e 5.990.384; WO 90/01869, WO 91/13993, WO 92/14822, WO 93/08682, WO 94/20628, WO 97/28247, WO 98/26064, e WO 99/40209), proteínas 7S (U.S. 5.003,045 e 5.576.203), proteína da castanha do Pará (US 5,850,024), proteínas livres de fenilalanina (WO 96/17064), albumina (WO 97/35023), beta-conglicinina (WO 00/19839), 11S (US 6,107,051), alfa-hordotionina (U.S. 5.885.802 e 5.885.801), proteínas de armazenamento de semente de arcelina (US 5,270,200), lectinas (US 6,110,891), e glutenina (U.S. 5,990,389 e 5,914,450) todos sendo incorporados aqui por referência.

– Reduzir os níveis de atividade de enzima fosforilase de tubérculo de α -glucano de tipo L (GLTP) fosforilase de tubérculo de α -glucano de tipo G (GHTP) preferivelmente dentro do tubérculo de batata (ver US 5.998.701). A conversão de amidos em açúcares em tubérculos de batata, particularmente quando armazenados em temperaturas abaixo de 7°C., é reduzida em tubérculos demonstrando atividade de enzima GLTP ou GHTP reduzida. A redução do adoçar no frio das batatas permite o armazenamento de batatas em temperaturas mais frias, resultando em dormência prolongada, reduzida incidência de doença, e aumentada vida de armazenamento. A redução da atividade de GLTP ou GHTP dentro do tubérculo da batata pode ser obtida por tais técnicas como supressão de expressão de gene usando RNA

anti-sentido ou RNA de filamento duplo homólogo, o uso de sequências de silenciamento regulatória, de co-supressão. A planta de batata tendo características de armazenamento no frio melhoradas, compreendendo uma planta de batata transformada com um cassete de expressão tendo uma

5 sequência TPT de promotor operativamente ligada a uma sequência de DNA compreendendo pelo menos 20 nucleotídeos de um gene codificando uma α -glucano fosforilase selecionada dentre o grupo consistindo fosforilase de tubérculo de α -glucano de tipo L (GLTP) fosforilase de tubérculo de α -glucano de tipo G (GHTP).

10 Outros exemplos de genes vantajosos são mencionados por exemplo em Dunwell JM, Transgenic approaches to crop improvement, J Exp Bot. 2000; 51 Spec No; pags. 487-96. Uma discussão de DNA heterólogos exemplares utilizáveis para a modificação de fenótipos de plantas pode ser encontrada em, por exemplo, U.S. 6.194.636;

15 Outro aspecto da invenção provê uma construção de DNA em que o promotor com expressão específica ou preferencial para amido-endosperma e/ou de embrião germinando que dirige elementos de DNA de supressão de genes, por exemplo para suprimir uma enzima catabolizando aminoácido.

20 A maturação da semente ou desenvolvimento do grão refere-se ao período iniciando com fertilização em que reservas de alimentos metabolizáveis (por exemplo, proteínas, lipídeos, amido, *etc.*) são depositadas na semente em desenvolvimento, particularmente em órgãos de armazenamento da semente, incluindo o endosperma, testa, camada de

25 aleurona, embrião, epitélio escutelar, resultando em aumento e enchimento de uma semente e terminando com a dessecação da semente.

Os promotores específicos de embrião desta invenção podem ser utilizáveis para minimizar o retardo da produção e outros efeitos fisiológicos adversos em potencial sobre o crescimento do milho e o

desenvolvimento que poderiam ser encontrados por uma expressão constitutiva de alta nível, não indutível, de uma proteína transgênica ou outra molécula em uma planta. Quando cada transgene é fundido a um promotor da invenção, o risco de homologia de sequência de DNA dependente da inativação do transgene (co-supressão) pode ser minimizado.

Pode ser utilizável marcar o próprio DNA dentro de uma célula. Por exemplo, pode ser utilizável marcar o DNA introduzido no núcleo como isto pode aumentar a frequência de transformação. Dentro do próprio núcleo seria utilizável marcar um gene a fim de obter uma integração sítio-específica. Por exemplo, seria utilizável ter um gene introduzido através de transformação substituir um gene existente na célula. Outros elementos incluem os que podem ser regulados por agentes endógenos ou exógenos, por exemplo, por proteínas de dedo de zinco, incluindo proteínas de dedo de zinco de ocorrência natural ou proteínas de dedo de zinco quiméricas. (ver, por exemplo, US 5.789.538, WO 99/48909; WO 99/45132; WO 98/53060; WO 98/53057; WO 98/53058; WO 00/23464; WO 95/19431; e WO 98/54311) ou fatores de transcrição de tipo myb. Por exemplo, uma proteínas de dedo de zinco quimérica pode incluir sequências de aminoácidos, que ligam a uma sequência de DNA específica (o dedo de zinco) e sequências de aminoácidos que ativam (por exemplo, sequências GAL 4) ou reprimem transcrição de sequências ligadas à sequência de DNA específica.

As categorias gerais de genes de interesse para os fins da presente invenção incluem, por exemplo, os genes envolvidos em informação, como dedos de zinco, os envolvidos em comunicação como quinases, e os envolvidos em *housekeeping*, como proteínas de choque térmico. As categorias mais específicas de transgenes incluem genes codificando traços importantes para a qualidade agronômica, resistência a inseto, resistência a doença, resistência herbicida, e características do grão. Ainda outras categorias de transgenes incluem genes para induzir a expressão de produtos

exógenos como enzimas, cofatores, e hormônios de plantas e outros organismos eucarióticos, assim como procarióticos. Reconhece-se que qualquer gene de interesse pode ser operativamente ligado a um promotor da invenção e expressado sob estresse.

5 Em uma forma de realização mais preferida, o promotor da presente invenção modula os genes codificando proteínas que atuam como reguladores do ciclo celular ou que controlam o metabolismo de carboidratos ou níveis de fito-hormônios, como foi mostrado em tabaco e canola com outros promotores de tecido. (Ma, Q. H. *et al.*, 1998; Roeckel, P. *et al.*, 1997)

10 Por exemplo, genes codificando isopentenil transferase ou IAA-M podem ser utilizáveis na modulação do desenvolvimento dos fósculos fêmeas. Outros genes importantes codificam fatores de crescimento e fatores de transcrição. A expressão de nucleotídeos endógenos ou heterólogos selecionados sob a direção de um promotor podem resultar em desenvolvimento continuado ou

15 melhorado dos fósculos fêmeas sob condições adversas.

 A produção de semente pode ser melhorada por alteração da expressão de genes que afetam a resposta de crescimento e desenvolvimento da semente durante o estresse ambiental (Cheikh-N *et al.* 1994) e genes controlando metabolismo de carboidratos para reduzir o aborto de sementes

20 em milho (Zinselmeier *et al.* 1995).

2.3 Excisão de sequência marcada

 A especificidade das sequências de ácido nucleico reguladoras da transcrição quiméricas da invenção em plantas monocotiledôneas torna a mesma especialmente utilizável para a excisão ou deleção marcada de

25 sequências (como sequências de marcador) a partir do genoma de referida planta monocotiledônea. Conhece-se a desvantagem dos métodos conhecidos na técnica anterior em que a excisão não é homogênea através de plantas completas assim levando a padrões de excisão de tipo mosaico, que requerem ciclos adicionais laboriosos de seleção e regeneração. A especificidade dos

promotores da invenção em embriões prematuros permite uma excisão homogênea através de todo o embrião, que irá então prover uma planta isenta da sequência alvo homogênea da planta (por exemplo, marcador).

Outra forma de realização da invenção refere-se a um método para excisão de sequências alvo (por exemplo, sequências de marcador) a partir de uma planta, preferivelmente uma planta monocotiledônea, referido método compreendendo as etapas de

A) construção de um cassete de expressão por operativamente ligando polinucleotídeo codificando a sequência nucleotídica reguladora da transcrição quimérica compreendendo

i) uma primeira molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:1;

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais

preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e

5 operativamente ligado a

ii) uma segunda molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:3;

b) um polinucleotídeo tendo pelo menos 50% preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:3;

10

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

15

i) d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

20

25

a pelo menos uma sequência de ácido nucleico que é

heteróloga em relação à referido sequência nucleotídica reguladora da transcrição quimérica e é apropriada para induzir a excisão de uma sequência alvo de uma planta monocotiledônea, e

B) inserir referido cassete de expressão em uma planta monocotiledônea compreendendo pelo menos uma sequência alvo para prover a planta transgênica, em que referida planta expressa referida sequência de ácido nucleico heteróloga, e

C) selecionar plantas transgênicas, que demonstram excisão de referido marcador.

A excisão é realizada por vários meios, incluindo mas não limitados a:

- indução de deleção de sequência por recombinação sítio-específica usando recombinases sítio-específicas, em que referida recombinase sítio-específica é expressada pela sequência nucleotídica reguladora da transcrição quimérica da invenção,

- indução de deleção de sequência por recombinação homóloga induzida, em que as sequências a serem deletadas são flanqueadas por sequências, referidas sequências tendo uma orientação, um comprimento suficiente e uma homologia para cada outra de modo a permitir a recombinação homóloga entre as mesmas, em que a recombinação homóloga é induzida por uma ruptura do filamento duplo sítio-específica feita pela endonuclease sítio-específica (preferivelmente uma endonuclease *homing*, mais preferivelmente a endonuclease *homing* I-SceI), em que referida endonuclease sítio-específica é expressada pela sequência nucleotídica reguladora da transcrição quimérica da invenção.

Outra forma de realização da invenção refere-se a uma planta, preferivelmente um planta monocotiledônea ou célula de planta compreendendo

A) pelo menos uma sequência alvo, que é estavelmente

inserida no genoma da planta, em que referida sequência alvo é flanqueada por sequências de excisão que são capazes de mediar por interação com uma excisão de enzima mediando a excisão específica de sequência de referida sequência alvo a partir do genoma da planta, e

5 B) um cassete de expressão compreendendo pelo menos uma sequência de ácido nucleico codificando uma enzima mediando a excisão, que é capaz de interagir com referidas sequências de excisão de i), operativamente ligadas a uma sequência nucleotídica reguladora da transcrição quimérica compreendendo

10 i) uma primeira molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 15 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:1; a

c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e

20 d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais 25 preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM

EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e operativamente ligado a

- 5 ii) uma segunda molécula de ácido nucleico selecionada dentre o grupo consistindo de
 - a) um polinucleotídeo como definido em SEQ ID NO:3;
 - b) um polinucleotídeo tendo pelo menos 50%, preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 85% ou 10 90%, o mais preferivelmente pelo menos 95%, 98% ou 99% identidade de sequência para o polinucleotídeo de SEQ ID NO:3;
 - c) um fragmento de pelo menos 50 bases consecutivas, preferivelmente pelo menos 100 bases consecutivas, mais preferivelmente 200 bases consecutivas do polinucleotídeo de SEQ ID NO:1; e
 - 15 d) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais 20 preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, e o mais 25 preferivelmente em 7% dodecil sulfato de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo.

As preferidas moléculas de ácido nucleico compreendendo um polinucleotídeo codificando a sequência nucleotídica reguladora da

transcrição quimérica são descritas acima. A sequência de ácido nucleico heteróloga preferida a ser expressada para obter a excisão de sequência (por exemplo, codificando para uma recombinase ou endonuclease sítio-específica) é descrita aqui abaixo.

5 A planta monocotiledônea à qual os métodos desta invenção são preferivelmente aplicados a pode ser selecionada dentre o grupo consistindo de milho, trigo, arroz, centeio, centeio, milhete, sorgo, erva-castelhana ou coix, tritcale ou aveias e cana de açúcar. Preferivelmente a planta é uma planta de cereal selecionada dentre o grupo consistindo de
10 milho, trigo, centeio, arroz, aveia, centeio, e sorgo, ainda mais preferivelmente de milho, trigo, e arroz, o mais preferivelmente a planta é uma planta de milho.

 A sequência nucleotídica reguladora da transcrição quimérica é preferivelmente definida como acima. A sequência alvo na planta
15 monocotiledônea acima definida ou célula de planta será excisada logo que as sementes de referida planta forem germinadas e o embrião começar a crescer. A partir deste embrião, a planta livre de sequência alvo irá resultar.

 A sequência alvo e o cassete de expressão para a enzima mediando a excisão podem ser combinados em um DNA ou em uma
20 construção diferente. As construções de DNA diferentes podem ser combinadas por outros meios no genoma de uma planta monocotiledônea de célula de planta como – por exemplo - cruzamento de linhagens parentais distintas compreendendo referida sequência alvo e referido cassete de expressão para a enzima mediando a excisão, respectivamente, ou co-
25 transformação ou subsequente transformação.

 Assim, uma outra forma de realização da invenção refere-se a um método para excisar pelo menos uma sequência alvo a partir do genoma de uma planta monocotiledônea ou célula de planta compreendendo as etapas de

A) inserir estavelmente no genoma um construção de ácido nucleico pelo menos uma sequência alvo, que é estavelmente inserida no genoma da planta, em que referida sequência alvo é flanqueada por sequências de excisão, que são capazes de mediar por interação com uma excisão de enzima mediando a excisão específica de sequência, a excisão de referida sequência alvo a partir do genoma da planta, e

B) introduzir em referidas plantas monocotiledôneas ou células de planta um cassete de expressão compreendendo pelo menos uma sequência de ácido nucleico codificando uma enzima mediando a excisão, que é capaz de interagir com referidas sequências de excisão de i), operativamente ligado ao polinucleotídeo codificando a sequência nucleotídica reguladora da transcrição quimérica compreendendo

i) uma primeira molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:1;

b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:1; e

c) um polinucleotídeo hibridizando sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de um polinucleotídeo como definido em SEQ ID NO:1, ou o complemento do mesmo, e operativamente ligado a

ii) uma segunda molécula de ácido nucleico selecionada dentre o grupo consistindo de

a) um polinucleotídeo como definido em SEQ ID NO:3;

b) um polinucleotídeo tendo pelo menos 50% identidade de sequência para o polinucleotídeo de SEQ ID NO:3; e

c) um polinucleotídeo hibridizando sob condições equivalentes

a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5M NaPO₄, 1mM EDTA a 50°C com lavagem em 2xSSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo pelo menos 50 nucleotídeos de uma sequência descrita por SEQ ID NO:3, ou o complemento do mesmo,

5 C) gerar sementes de referida planta monocotiledônea ou células de planta compreendendo tanto referida sequência alvo como referido cassete de expressão, germinando referidas sementes e cultivando plantas a partir das mesmas, e

10 D) selecionar plantas das quais a referida sequência alvo foi excisada.

Em uma forma de realização preferida o método da invenção ainda compreende a etapa de regeneração de uma planta fértil. O método pode ainda incluir a propagação sexual ou assexual ou a prole crescendo ou um descendente da planta regenerada a partir da referida célula de planta.

15 Preferivelmente, excisão (ou deleção) da sequência alvo pode ser realizada por vários meios conhecidos como tal na técnica, incluindo mas não limitados a um ou mais dos seguintes métodos:

20 a) recombinação induzida por uma recombinase específica de sequência, em que referida sequência alvo é flanqueada por sítios de recombinação correspondentes em um modo em que recombinação entre referidos sítios de recombinação de flanco resulta em deleção da sequência alvo no meio do genoma;

25 b) recombinação homóloga entre sequências de homologia A e A' flanqueando referida sequência alvo, preferivelmente induzida por uma ruptura de filamento duplo específico da sequência entre referidas sequências de homologia causado pela endonuclease específica de sequência, em que referidas sequências de homologia A e A' tem um comprimento e homologia suficientes a fim de assegurar uma recombinação homóloga entre A e A', e tendo uma orientação que –quando da recombinação entre A e A' – irá levar à

excisão de referida sequência alvo a partir do genoma de referida planta.

As sequências de excisão e enzimas de excisão preferidas são especificadas abaixo. Em outra forma de realização preferida o mecanismo de deleção /excisão pode ser induzido ou ativado em um modo para evitar deleção /excisão prematura do marcador de função dual. Preferivelmente, assim, expressão e/ou atividade de uma recombinase ou endonuclease específica de sequência preferivelmente empregada pode ser induzida e/ou ativada, preferivelmente pelo método selecionado dentre o grupo consistindo de

a) expressão indutível por ligação operativamente da sequência codificando referida enzima de excisão (por exemplo, recombinase ou endonuclease) em uma sequência reguladora da transcrição quimérica combinada com um promotor indutível ou elemento de promotor,

b) ativação indutível, por emprego de uma enzima de excisão indutível, modificada (por exemplo, uma recombinase ou endonuclease) por exemplo compreendendo um domínio de ligação de ligando, em cuja atividade de referida enzima de excisão modificado pode ser modificada por tratamento do composto tendo atividade de ligação para referido domínio de ligação de ligando.

Preferivelmente, a sequência alvo é um marcador, mais preferivelmente um marcador de seleção (sequências de marcador preferidas são especificadas abaixo). Assim, o método da invenção resulta em uma célula de planta monocotiledônea ou planta, que é isenta de marcador de seleção.

2.3.1 Sequências de excisão e enzimas de excisão preferidas

Para assegurar a deleção / excisão de sequência alvo, a sequência alvo é flanqueada por sequências de excisão, que são capazes de mediar por interação com uma excisão de enzima mediando a excisão específica de sequência de referida sequência alvo a partir do genoma da

planta. Preferivelmente, deleção da sequência alvo pode ser realizada por vários meios conhecidos na técnica, incluindo mas não limitado a um ou mais dos seguintes métodos:

a) recombinação induzida por uma recombinase específica de sequência, em que referida sequência alvo é flanqueada por sítios de recombinação correspondentes em um modo em que recombinação entre referidos sítios de recombinação de flanco resulta em deleção da sequência alvo no meio do genoma,

b) recombinação homóloga entre sequências de homologia A e A' flanqueadas pela referida sequência alvo, preferivelmente induzida por uma ruptura de filamento duplo específico da sequência entre referidas sequências de homologia causada pela endonuclease específica de sequência, em que referidas sequências de homologia A e A' tem um comprimento e homologia suficientes para assegurar a recombinação homóloga entre A e A', e tendo uma orientação que – quando da recombinação entre A e A' – irá levar à excisão de referida sequência alvo a partir do genoma de referida planta.

Assim, para assegurar a deleção / excisão da sequência alvo, a sequência alvo é flanqueada por sequências que permitem a deleção específica de referido cassete de expressão. Referidas sequências pode ter sítios de recombinação para uma recombinase específica de sequência, que são colocadas em um modo que a recombinação induzida entre referidos sítios de recombinação de flanco resulta em deleção de referida sequência alvo do genoma. Estes são vários sítios de recombinação e recombinase específica correspondente de sequências conhecidos na técnica (descrito aqui abaixo), que podem ser empregados para os fins da invenção.

Em outra forma de realização preferida, deleção / excisão de uma sequência alvo é realizada por recombinação intramolecular (preferivelmente intra- cromossômica) homóloga. A recombinação homóloga pode ocorrer de modo espontâneo, mas é preferivelmente induzida por uma

ruptura de filamento duplo específico da sequência (por exemplo, entre as sequências de homologia). Os princípios básicos são descritos em WO 03/004659. Para este fim a sequência alvo é flanqueada por sequências de homologia A e A', em que referidas sequências de homologia tem comprimento e homologia suficientes a fim de assegurar recombinação homóloga entre A e A', e tendo uma orientação que – quando da recombinação entre A e A' – irá leva uma excisão da referida sequência alvo do genoma. Além disso, a sequência flanqueada por referidas sequências de homologia ainda compreende pelo menos uma sequência de reconhecimento de pelo menos 10 pares de base da indução sítio-dirigida de rupturas do filamento duplo do DNA por uma enzima induzindo a ruptura de filamento duplo de DNA específico da sequência, preferivelmente uma DNA-endonuclease específica de sequência, mais preferivelmente uma endonuclease *homing*, o mais preferivelmente uma endonuclease selecionada dentre o grupo consistindo de I-SceI, I-CpaI, I-CpaII, I-CreI e I-ChuI ou quimeras dos mesmos com domínios de ligação de ligando. As apropriadas endonucleases são descritas aqui abaixo.

2.3.1.1 Sítios de recombinação e recombinases

Recombinase específica de sequências e seus sítios de recombinação correspondentes apropriados dentro da presente invenção podem incluir mas não são limitados ao sistema Cre/lox do bacteriófago P1 (Dale EC e Ow DW 1991; Russell SH *et al.* 1992; Osborne BI *et al.* 1995), o sistema FLP/FRT de levedura (Kilby NJ *et al.* 1995; Lyznik LA *et al.* 1996), a Gin recombinase de fago Mu, a Pin recombinase de E. coli ou o sistema R/RS do plasmídeo pSR1 (Onouchi H *et al.* 1995; Sugita Ket *et al.* 2000). A recombinase (por exemplo Cre ou FLP) interage especificamente com suas sequências de recombinação correspondentes (sequência lox de 34 bp e sequência FRT de 47 bp, respectivamente) a fim de deletar ou inverter as sequências interpostas. A deleção de marcador de seleção padrão em plantas

que foi flanqueada por duas sequências lox pelo Cre é descrita (Dale EC e Ow DW 1991). Os sítios de recombinação preferidos para as apropriadas recombinases são descritos em tabela 1 abaixo:

Tabela 1. Recombinase específica de sequências apropriada

Recombinase	Organismo de origem	Sítios de recombinação
CRE	Bacteriofago P1	5'-AACTCTCATCGCTTCGGATAACTTCCTGTTATCCGAAA CATATCACTCACTTTGGTGATTTCACCGTAACTGTCTATGA TTAATG-3'
FLP	Saccharomyces cerevisiae	5'-GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAA AGTATAGGAACTTC-3'
R	pSR1 Plasmídeos	5'-CGAGATCATATCACTGTGGACGTTGATGAAAGAATAC GTTATTCTTTCATCAAATCGT

5 2.3.1.2 As sequências de homologia

Com referência às sequências de homologia (por exemplo, A, A') de "comprimento suficiente" preferivelmente refere-se a sequências com um comprimento de pelo menos 20 pares de base, preferivelmente pelo menos 50 pares de base, especialmente preferivelmente pelo menos 100 pares de base, muito especialmente preferivelmente pelo menos 250 pares de base, o mais preferivelmente pelo menos 500 pares de base.

Com referência às sequências de homologia (por exemplo, A, A'), de "homologia suficiente" preferivelmente refere-se às sequências com pelo menos 70%, preferivelmente 80%, por preferência pelo menos 90%, especialmente preferivelmente pelo menos 95%, muito especialmente preferivelmente pelo menos 99%, o mais preferivelmente 100%, de homologia dentro destas sequências de homologia sobre um comprimento de pelo menos 20 pares de base, preferivelmente pelo menos 50 pares de base, especialmente preferivelmente pelo menos 100 pares de base, muito especialmente preferivelmente pelo menos 250 pares de base, o mais preferivelmente pelo menos 500 pares de base.

As sequências de homologia A e A' são preferivelmente organizadas na forma de uma repetição direta. O termo "repetição direta" significa uma localização subsequente de duas sequências no mesmo filamento de uma molécula de DNA molécula na mesma orientação, em que

estas duas sequências atendem às exigências acima dadas para a recombinação homóloga entre referidas duas sequências.

Na forma de realização preferida, as sequências de homologia podem ser uma duplicação de uma sequência tendo uso adicional dentro da construção de DNA. Por exemplo, as sequências de homologia podem ser duas sequências de terminador de transcrição. Uma destas sequências de terminador pode ser operativamente ligada ao traço agronomicamente valioso ou relevante, enquanto a outra pode ser ligada ao marcador de seleção de função dupla, que está localizada em 3'-direção do gene de traço.

Recombinação entre as duas sequências de terminador irá excisar a sequência alvo (por exemplo, um gene marcador) mas irá reconstituir o terminador do gene de traço. Em outro exemplo, as sequências de homologia podem ser duas sequências de promotor. Uma destas sequências de promotor pode ser operativamente ligada ao traço agronomicamente valioso ou relevante, enquanto a outra pode ser ligada a uma sequência alvo (por exemplo, um marcador de seleção), que está localizado em 5'-direção do gene de traço. Recombinação entre as duas sequências de promotor irá excisar a sequência alvo (por exemplo, um gene marcador) mas irá reconstituir o promotor do gene de traço. O versado na técnica sabe bem que as sequências de homologia não precisam ser limitadas a um único elemento funcional (por exemplo promotor ou terminador), mas podem compreender ou se estender a outras sequências (por exemplo sendo parte da região de codificação do gene de traço e a sequência de terminador respectiva de referido gene de traço).

2.3.1.3. Enzima Induzindo ruptura de filamento duplo

Preferivelmente, deleção / excisão de uma sequência alvo (por exemplo, um gene marcador) é realizada por recombinação homóloga entre as sequências de homologia acima especificadas induzidas por uma ruptura de filamento duplo específico da sequência, preferivelmente entre as sequências de homologia que devem recombinar. Os métodos gerais são descritos por

exemplo em WO 03/004659, incorporado aqui por referência em sua totalidade. Várias enzimas apropriadas para indução de ruptura de filamento duplo específico da sequências (a seguir em conjunto “endonuclease”) são conhecidas na técnica. A endonuclease pode ser por exemplo selecionada
5 dentre o grupo compreendendo:

1. Endonucleases de restrição (tipo II), preferivelmente endonucleases homing como descrito em detalhes abaixo.

2. Transposases, por exemplo uma transposase elemento P (Kaufman PD e Rio DC 1992) ou AcDs (Xiao YL e Peterson T 2000). Em
10 princípio, todas as transposases ou integrases são apropriadas desde que elas tenham especificidade de sequência (Haren L *et al.* 1999; Beall EL, Rio DC 1997).

3. nucleases quiméricas como descrito em detalhes abaixo.

4. Enzimas que induzem rupturas de filamento duplo no
15 sistema imune, como o sistema RAG1/RAG2 (Agrawal A *et al.* 1998).

5. Endonucleases de intron Grupo II. Modificações da sequência de intron permite que os introns de grupo II sejam dirigidos para virtualmente qualquer sequência em um DNA filamento duplo, onde os introns de grupo II podem subsequentemente inserir por meio de um
20 mecanismo de emenda reversa (Mohr *et al.* 2000; Guo *et al.* 2000). Durante este mecanismo de emenda reversa, uma ruptura de filamento duplo é introduzida no DNA alvo, o RNA de íntron excisado clivando o filamento sentido enquanto a porção de proteína da endonuclease de íntron de grupo II hidrolisa o filamento anti-sentido (Guo *et al.* 1997). Se for apenas desejado
25 induzir a ruptura de filamento duplo sem alcançar uma emenda reversa completa, como é o caso na presente invenção, é possível recorrer a, por exemplo, endonucleases de íntron de grupo II que faltam a atividade de transcriptase reversa. Apesar de não evitar a geração de ruptura de filamento duplo, o mecanismo de emenda reversa não pode prosseguir até completar.

As enzimas apropriadas não são apenas as enzimas naturais, mas também enzimas sintéticas. As enzimas preferidas são todas as endonucleases cuja sequência de reconhecimento é conhecida e que pode ou ser obtida na forma de suas proteínas (por exemplo por purificação) ou expressadas usando sua sequência de ácido nucleico.

Em uma forma de realização preferida, endonuclease específica de sequência é empregada para a indução específica de rupturas de filamento duplo e subsequente recombinação homóloga induzida. O termo “endonuclease- DNA específica de sequência ” geralmente refere-se a todas as enzimas, que são capazes de gerar rupturas de filamento duplo em DNA de filamento duplo em um modo específica de sequência em uma ou mais sequências de reconhecimento. A referida clivagem de DNA pode resultar em extremidades cegas, ou em assim chamadas extremidades "aderentes" do DNA (tendo um *overhang* 5'- ou 3'). O sítio de clivagem pode estar localizado dentro ou fora da sequência de reconhecimento. Vários tipos de endonucleases pode ser empregados. Endonucleases podem ser, por exemplo, do tipo de Classe II ou Classe IIs. endonucleases de restrição R-M Classe IIs catalisam a clivagem de DNA em sequências diferentes da sequência de reconhecimento, isto é elas clivagem em uma sequência de DNA em um número particular de nucleotídeos fora da sequência de reconhecimento (Szybalski *et al.* 1991). O seguinte pode ser mencionado a título de exemplo, mas não por limitação:

1. Endonucleases de restrição (por exemplo, tipo II ou IIs), preferivelmente endonucleases *homing* como descrito em detalhes abaixo.

2. Nucleases quiméricas ou sintéticas como descrito em detalhes abaixo.

Diferente das recombinases, as enzimas de restrição tipicamente não ligam DNA, mas somente clivam DNA. As enzimas de restrição são descritas, por exemplo, no catálogo on line do New England

Biolabs (www.neb.com), no catálogo on line Promega (www.promega.com) e Rao *et al.* (2000). Nesta invenção “ligação” das extremidades de DNA resultante da clivagem pela endonuclease é realizada por fusão de recombinação homóloga das sequências de homologia.

5 Preferivelmente, a endonuclease é selecionada em um modo em que suas sequências de reconhecimento correspondentes são raramente, se de todo, encontradas no genoma não modificado do organismo de planta alvo. Idealmente, a única cópia (ou cópias) da sequência de reconhecimento no genoma é (ou são) a (as) introduzida (s) para a construção de DNA da
10 invenção, assim eliminando a possibilidade de outro DNA no genoma ser excisado ou re-arranjado quando uma endonuclease específica de sequência é expressada.

Um critério para selecionar uma endonuclease apropriada é o comprimento de sua sequência de reconhecimento correspondente. Referida
15 sequência de reconhecimento tem um comprimento apropriado para permitir a clivagem rara, mais preferivelmente clivagem somente na(s) sequência(s) de reconhecimento(s) composta na construção de DNA da invenção. Um fator determinando o comprimento mínimo de referida sequência de reconhecimento é – de um ponto de vista estatístico - o tamanho do genoma
20 do organismo hospedeiro. Em uma forma de realização preferida a sequência de reconhecimento tem um comprimento de pelo menos 10 pares de base, preferivelmente pelo menos 14 pares de base, mais preferivelmente pelo menos 16 pares de base, especialmente preferivelmente pelo menos 18 pares de base, o mais preferivelmente pelo menos 20 pares de base.

25 A enzima de restrição que cliva uma sequência de reconhecimento de 10 pares de bases é descrita em Huang B *et al.* 1996.

As enzimas apropriadas não são apenas as enzimas naturais, mas também enzimas sintéticas. As enzimas preferidas são todas as endonucleases específicas de sequência cuja sequência de reconhecimento é

conhecida e que pode ou ser obtida na forma de suas proteínas (por exemplo por purificação) ou expressadas usando sua sequência de ácido nucleico.

São especialmente preferidas as endonucleases de restrição (enzimas de restrição) que não tem ou têm apenas algumas sequências de reconhecimento –além das sequências de reconhecimento presentes na construção de recombinação transgênica - na sequência de DNA cromossômica de um organismo eucariótico particular. Isto evita outras rupturas de filamento duplo em *loci* indesejados no genoma. Isto é porque as endonucleases *homing* são especialmente preferidas (Review: (Belfort M e Roberts RJ 1997; Jasin M 1996; Internet: <http://rebase.neb.com/rebase/rebase.homing.html>). Devido às suas sequências de reconhecimento longas, elas não tem, ou tem somente algumas, outras sequências de reconhecimento no DNA cromossômico de organismos eucarióticos na maior parte dos casos.

As sequências codificando estas endonucleases *homing* podem ser isoladas por exemplo no genoma de cloroplasto de *Chlamydomonas* (Turmel M *et al.* 1993). Elas são pequenas (18 a 26 kD) e suas matrizes de leitura aberta (ORF) tem um "uso de códon" que é apropriado diretamente para expressão nuclear em eucariotos (Monnat RJ Jr *et al.* 1999). As endonucleases *Homing* que são muito especialmente preferivelmente isoladas são as endonucleases *homing* I-SceI (WO96/14408), I-SceII (Sarguiel B *et al.* 1990), I-SceIII (Sarguiel B *et al.* 1991), I-CeuI (Marshall 1991), I-CreI (Wang J *et al.* 1997), I-ChuI (Cote V *et al.* 1993), I-TevI (Chu *et al.* 1990; Bell-Pedersen *et al.* 1990), I-TevII (Bell-Pedersen *et al.* 1990), I-TevIII (Eddy *et al.* 1991), Endo SceI (Kawasaki *et al.* 1991), I-CpaI (Turmel M *et al.* 1995a) e I-CpaII (Turmel M *et al.* 1995b).

Outras endonucleases *homing* são detalhadas nos sites de Internet acima mencionados, e cujos exemplos podem ser mencionados incluindo endonucleases *homing* como F-SceI, F-SceII, F-SuvI, F-TevI, F-

TevII, I-AmaI, I-AniI, I-CeuI, I-CeuAIIP, I-ChuI, I-Cmoel, I-CpaI, I-CpaII, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIVP, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HspNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NclIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PbolIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PpbIP, I-PpoI, I-SPBetaIP, I-ScaI, I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-SexIP, I-SneIP, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiS3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPA1P, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-PspI, PI-Rma43812IP, PI-SPBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII, H-DreI, I-BasI, I-BmoI, I-PogI, I-TwoI, PI-MgaI, PI-PabI, PI-PabII.

São preferidas neste contexto as endonucleases *homing* cujas sequências de gene já são conhecidas, como, por exemplo, F-SceI, I-CeuI, I-ChuI, I-DmoI, I-CpaI, I-CpaII, I-CreI, I-CsmI, F-TevI, F-TevII, I-TevI, I-TevII, I-AniI, I-CvuI, I-DdiI, I-HmuI, I-HmuII, I-LlaI, I-NanI, I-MsoI, I-NitI, I-NjaI, I-PakI, I-PorI, I-PpoI, I-ScaI, I-Ssp6803I, PI-PkoI, PI-PkoII, PI-PspI, PI-TfuI, PI-TliI. São especialmente preferidas as endonucleases *homing* comercialmente disponíveis, como I-CeuI, I-SceI, I-DmoI, I-PpoI, PI-PspI ou PI-SceI. Endonucleases com sequências de reconhecimento particularmente longas, e que somente raramente clivam (se de todo) dentro de um genoma incluem: I-CeuI (sequência de reconhecimento de 26 bp), PI-PspI (sequência de reconhecimento de 30 bp), PI-SceI (sequência de reconhecimento de 39 bp), I-SceI (sequência de reconhecimento de 18 bp) e I-PpoI (sequência de reconhecimento de 15 bp). As enzimas podem ser isoladas de seus organismos de origem do modo que o versado conhece, e/ou sua sequência de codificação de ácido nucleico pode ser clonada. As sequências de várias enzimas são depositadas em GenBank. Especialmente preferidas são as endonucleases *homing* I-SceI, I-CpaI, I-CpaII, I-CreI e I-ChuI. Sequências

codificando referidas nucleases são conhecidas na técnica e – por exemplo – especificadas em WO 03/004659 (por exemplo, as SEQ ID NO: 2, 4, 6, 8, e 10 de WP 03/004659 aqui incorporado por referência).

A sequência de ácido nucleico heteróloga a ser expressada
5 pode preferivelmente codificar um polipeptídeo como descrito por qualquer uma dentre SEQ ID NO: 22, ou um equivalente funcional da mesma, que é capaz de ocasionar o mesmo fenótipo diferente do referido polipeptídeo.

Mais preferivelmente a sequência de ácido nucleico a ser expressada é descrita por SEQ ID NO 21 ou 23.

10 Em uma forma de realização preferida, as sequências codificando referidas endonucleases *homing* podem ser modificadas por inserção de uma sequência de íntron. Isto evita a expressão da enzima funcional em organismos hospedeiros procarióticos, e assim facilita os procedimentos de clonagem e transformação (por exemplo, com base em
15 *E.coli* ou *Agrobacterium*). Em organismos de planta, expressão da funcional enzima é realizada, porque as plantas são capazes de reconhecer e "emendar" introns. Preferivelmente, introns são inseridos nas endonucleases *homing* mencionadas como preferidas acima (por exemplo, em I-SceI ou I-CreI).

Em alguns aspectos da invenção, a evolução molecular pode
20 ser empregada para criar uma endonuclease melhorada Polinucleotídeos codificando uma enzima endonuclease candidata podem, por exemplo, ser modulados com protocolos de embaralhamento de DNA. O embaralhamento de DNA é um processo de recombinação e mutação recursivas, realizadas por fragmentação aleatória do conjunto de genes relacionados, seguido por re-
25 arranjo dos fragmentos pelo processo de tipo de referida de cadeia polimerase. Ver, por exemplo, Stemmer (1994); Stemmer (1994); e US 5.605.793, US 5.837.458, US 5.830,721 e US 5. 811.238.

Outras endonucleases sintéticas que podem ser mencionadas a título de exemplo são nucleases quiméricas que são compostas de um domínio

de nuclease não específico e um domínio de ligação de DNA específico de sequência consistindo de dedos de zinco (Bibikova M *et al.* 2001). Estes domínios de dedo de zinco de ligação de DNA podem ser adaptados para se adequar a qualquer sequência de DNA. Os métodos apropriados para preparar os domínios apropriados de dedo de zinco são descritos e conhecidos do versado (Beerli RR *et al.* 2000; Beerli RR *et al.* 2000; Segal DJ e Barbas CF 2000; Kang JS e Kim JS 2000; Beerli RR *et al.* 1998; Kim JS *et al.* 1997; Klug A 1999; Tsai SY *et al.* 1998; Mapp AK *et al.* 2000; Sharrocks AD *et al.* 1997; Zhang L *et al.* 2000).

A endonuclease é preferivelmente expressada como uma proteína de fusão com uma sequência de localização nuclear (NLS). Esta sequência NLS permite o transporte facilitado no núcleo e aumenta a eficácia do sistema de recombinação. Várias sequências NLS são conhecidas do versado e descritas, inter alia, por Jicks GR e Raikhel NV 1995. Preferida para organismos de planta é, por exemplo, a sequência NLS do antígeno grande SV40. Exemplos são dados em WO 03/060133. No entanto, devido ao tamanho pequeno de muitas enzimas DSBI (como, por exemplo, endonucleases *homing*), uma sequência NLS não é necessariamente requerida. Estas enzimas são capazes de passar através de poros nucleares sem outra ajuda qualquer.

Em uma outra forma de realização preferida, a atividade da endonuclease pode ser induzida. Apropriados métodos foram descritos para a recombinase específica de sequências (Angrand PO *et al.* 1998; Logie C e Stewart AF 1995; Imai T *et al.* 2001). Estes métodos empregam proteínas de fusão da endonuclease e o domínio de ligação de ligando para o receptor de hormônio esteróide (por exemplo o receptor de androgênio humano, ou variantes mudadas de receptor de estrogênio humano (como descrito aqui). A indução pode ser efetuada com ligandos, como, por exemplo, estradiol, dexametasona, 4-hidroxitamoxifeno ou

raloxifeno. Algumas endonucleases são ativas como dímero (homo- ou heterodímeros; I-CreI forma um homodímero; I-SecIV forma um heterodímero) (Wernette CM 1998). Dimerização pode ser projetada como um aspecto indutível, por exemplo por troca dos domínios de dimerização natural para o domínio de ligação do ligando de baixo peso molecular. A adição de ligando dimérico então ocasiona a dimerização da proteína de fusão. Os métodos de dimerização indutível correspondentes e a preparação dos ligandos diméricos foram descritos (Amara JF *et al.* 1997; Muthuswamy SK *et al.* 1999; Schultz LW e Clardy J 1998; Keenan T *et al.* 1998).

As sequências de reconhecimento para DNA endonuclease específica de sequência (por exemplo, endonucleases *homing*) são descritas na técnica. "Sequência de reconhecimento " refere-se uma sequência de DNA que é reconhecida pela DNA endonuclease específica de sequência da invenção. A sequência de reconhecimento tipicamente terá pelo menos 10 pares de base de comprimento, tendo mais geralmente de 10 a 30 pares de base de comprimento e na maior parte das formas de realização, menor do que 50 pares de base de comprimento.

"Sequência de reconhecimento" geralmente refere-se às sequências que, sob as condições em uma célula de planta usada nesta invenção, permite o reconhecimento e clivagem pela DNA-endonuclease específica de sequência. As sequências de reconhecimento para as DNA-endonucleases específicas de sequência respectivas são mencionadas em tabela 2 abaixo a título de exemplo, mas não por limitação.

Tabela 2. Sequências de reconhecimento e organismos de origem para endonucleases (por exemplo, endonucleases *homing*; "^" indica o sítio de clivagem da DNA-endonuclease específica de sequência dentro da sequência de reconhecimento).

Enzima DSBII	Organismo de origem	Sequência de reconhecimento
P-Elemento Trans-Posase	Drosophila	5'-CTAGATGAAATAACATAAGGTGG-3'
I-Anil	Aspergillus nidulans	5'-TTGAGGAGGTT^TCTCTGTAAATAANNNNNNNNNNNNNNNN 3'-AACTCCTCCAAAGAGACATTATNNNNNNNNNNNNNNNN^
I-DdiI	Dictyostelium discoideum AX3	5'-TTTTTTGGTCATCCAGAAGTATAT 3'-AAAAAACCCAG^TAGGTCTTCATATA
I-Cvul	Chlorella vulgaris	5'-CTGGGTTCAAACCGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTGCAG^CACTCTGTCAAACC
I-Csml	Chlamydomonas smithii	5'-GTACTAGCATGGGGTCAAATGTCTTTCTGG
I-Cmoel	Chlamydomonas moewusii	5'-TCGTAGCAGCT^CACGGTT 3'-AGCATCG^TCGAGTGCCAA
I-Crel	Chlamydomonas reinhardtii	5'-CTGGGTTCAAACCGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTGCAG^CACTCTGTCAAACC
I-Chul	Chlamydomonas humicola	5'-GAAGGTTTGGCACCTCG^ATGTCGGGTCATC 3'-CTTCCAAACCGTG^GAGCTACAGCCGAGTAG
I-Cpal	Chlamydomonas pallidostigmatica	5'-CGATCCTAAGGTAGCGAA^ATTCA 3'-GCTAGGATTCCATC^GCTTTAAGT
I-CpaiI	Chlamydomonas pallidostigmatica	5'-CCC GGCTAACTC^TGTGCCAG 3'-GGGCCGAT^TGAGACACGGTC
I-Ceul	Chlamydomonas eugametos	5'-CGTA ACTATAACGGTCTTAA^GGTAGCGAA 3'-GCATTGATATTGCCAG^GATTCCATCGCTT
I-Dmol	Desulfurococcus mobilis	5'-ATGCCTTGCCGGGTAA^GTTCCGGCGCGCAT 3'-TACGGAACGGCC^CATTC AAGGCCGCGCTA
I-SceI	Saccharomyces cerevisiae	5'-AGTTACGCTAGGGATAA^CAGGGTAATATAG 3'-TCAATGCGATCCC^TATTGTCCCATTATATC 5'-TAGGGATAA^CAGGGTAAT 3'-ATCCC^TATTGTCCCATTA ("Core"-Sequência)
I-SceII	S.cerevisiae	5'-TTTTGATTCTTTGGTCACCC^TGAAGTATA 3'-AAA ACTAAGAA ACCAG^TGGGCACTTCATAT
I-SceIII	S.cerevisiae	5'-ATTGGAGGTTTTGGTAAC^TATTTATTACC 3'-TAACCTCCAAAACC^ATTGATAAATAATGG
I-SceIV	S.cerevisiae	5'-TCTTTTCTCTTGATTA^GCCCTAATCTACG 3'-AGAAAAGAGAAC^TAATCGGGATTAGATGC
I-SceV	S.cerevisiae	5'-AATAATTTTCT^TCTTAGTAATGCC 3'-TTATTA AAAAGAAGATCATT A^CGG
I-SceVI	S.cerevisiae	5'-GTTATTTAATG^TTTTAGTAGTTGG 3'-CAATAAATTACAAAATCATCA^ACC
I-SceVII	S.cerevisiae	5'-TGTCACATTGAGGTGCATGATTATTAC
PI-SceI	S.cerevisiae	5'-ATCTATGTCGGGTGC^GGAGAAAGAGGTAAT 3'-TAGATACAGCC^CACGCCTCTTTCTCCATTA
F-SceI	S.cerevisiae	5'-GATGCTGTAGGC^ATAGGCTTGGTT 3'-CTACGACA^TCCGTATCCGAACCAA
F-SceII	S.cerevisiae	5'-CTTTCCGCAACA^GTAA AATT 3'-GAAAGGCG^TTGTCATTTTAA
I-Hmul	Bacillus subtilis bacteriofago SPO1	5'-AGTAATGAGCCTAACGCTCAGCAA 3'-TCATTACTCGGATTGC^GAGTCGTT
I-HmulI	Bacillus subtilis bacteriofago SP82	5'-AGTAATGAGCCTAACGCTCAACAANNNNNNNNNNNNNNNN NN- NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
I-LlaI	Lactococcus lactis	5'-CACATCCATAAC^CATATCATTTTT 3'-GTGTAGGTATTGGTATAGTAA^AAA
I-Msol	espécies Monomastix	5'-CTGGGTTCAAACCGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTGCAG^CACTCTGTCAAACC
I-NanI	Naegleria andersoni	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
I-NitI	Naegleria italica	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
I-Njal	Naegleria jamiesoni	5'-AAGTCTGGTGCCA^GCACCCGC 3'-TTCAGACC^ACGGTCGTGGGCG
I-PakI	Pseudomonas aeruginosa akinetum	5'-CTGGGTTCAAACCGTCGTGA^GACAGTTTGG 3'-GACCCAAGTTTTGCAG^CACTCTGTCAAACC
I-PorI	Pyrobaculum organotrophum	5'-GCGAGCCCGTAAGGGT^GTGTACGGG 3'-CGCTCGGGCATT^CCCACACATGCC

Enzima DSBI	Organismo de origem	Sequência de reconhecimento
I-PpoI	Physarum policephalum	5'-TAACTATGACTCTCTTAA^GGTAGCCAAAT 3'-ATTGATACTGAGAG^AATTCCATCGGTTTA
I-ScaI	Saccharomyces capensis	5'-TGTCACATTGAGGTGCACT^AGTTATTAC 3'-ACAGTGTAACCTCCAC^GTGATCAATAATG
I-Ssp6803I	Espécies Synechocystis	5'-GTCGGGGCT^CATAACCCGAA 3'-CAGCCCGAGTA^TTGGGCTT
PI-PfuI	Pyrococcus furiosus VcI	5'-GAAGATGGGAGGAGGG^ACCGGACTCAACTT 3'-CTTCTACCCTCC^TCCCTGGCCTGAGTTGAA
PI-PfuII	Pyrococcus furiosus VcI	5'-ACGAATCCATGTGGAGA^AGAGCCTCTATA 3'-TGCTTAGGTACAC^CTCTTCTCGGAGATAT
PI-PkoI	Pyrococcus kodakaraensis KOD1	5'-GATTTTAGAT^CCCTGTACC 3'-CTAAAA^TCTAGGGACATGG
PI-PkoII	Pyrococcus kodakaraensis KOD1	5'-CAGTACTACG^GTTAC 3'-GTCATG^ATGCCAATG
PI-PspI	Pyrococcus sp.	5'-AAAATCCTGGCAAACAGCTATTAT^GGGTAT 3'-TTTTAGGACCGTTTGTGCGAT^AATACCCATA
PI-TfuI	Thermococcus fumicolans ST557	5'-TAGATTTTAGGT^CGCTATATCCTTCC 3'-ATCTAAAA^TCCAGCGATATAGGAAGG
PI-TfuII	Thermococcus fumicolans ST557	5'-TAYGCNGAYACN^GACGGYTTYT 3'-ATRCGNCT^RTGNCTGCCRAARA
PI-ThyI	Thermococcus hydrothermalis	5'-TAYGCNGAYACN^GACGGYTTYT 3'-ATRCGNCT^RTGNCTGCCRAARA
PI-TliI	Thermococcus litoralis	5'-TAYGCNGAYACNGACGG^YTTYT 3'-ATRCGNCTRTGNC^TGCCRAARA
PI-TliII	Thermococcus litoralis	5'-AAATTGCTTGCAAACAGCTATTACGGCTAT
I-TevI	Bacteriophago T4	5'-AGTGGTATCAAC^GCTCAGTAGATG 3'-TCACCATAGT^TGCGAGTCATCTAC
I-TevII	Bacteriophago T4	5'-GCTTATGAGTATGAAGTGAACACGT^TATTC 3'-CGAATACTCATACTTCACTTGTG^CAATAAG
F-TevI	Bacteriophago T4	5'-GAAACACAAGA^AATGTTTAGTAAANNNNNNNNNNNNNNNN 3'-CTTTGTGTCTTTACAAATCATTTNNNNNNNNNNNNNNNN^
F-TevII	Bacteriophago T4	5'-TTTAATCCTCGCTTC^AGATATGGCAACTG 3'-AAATTAGGAGCGA^AGTCTATACCGTTGAC
H-DreI	<i>E. coli pl-DreI</i>	5'-CAAAACGTCGTAA^GTTCCGGCGCG 3'-GTTTTGCAG^CATTCAAGGCCGCGC
I-BasI	<i>Bacillus thuringiensis fago Bastille</i>	5' AGTAATGAGCCTAACGCTCAGCAA 3' TCATTACGAGTCGAACCTCGGATTG
I-BmoI	<i>Bacillus mojavensis s87-18</i>	5'-GAGTAAGAGCCCG^TAGTAATGACATGGC 3'-CTCATCTCG^GGCATCATTACTGTACCG
I-PogI	<i>Pyrobaculum oguniense</i>	5'-CTTCAGTAT^GCCCCGAAAC 3'-GAAGT^CATACGGGGCTTTG
I-TwoI	<i>Staphylococcus aureus fago Twort</i>	5'-TCTTGCACCTACACAATCCA 3'-AGAACGTGGATGTGTTAGGT
PI-Mgal	<i>Mycobacterium gastri</i>	5'-CGTAGCTGCCCAGTATGAGTCA 3'-GCATCGACGGGTCATACTCAGT
PI-PabI	<i>Pyrococcus abyssi</i>	5'-GGGGGCAGCCAGTGGTCCCGTT 3'-CCCCCGTCGGTCACCAGGGCAA
PI-PabII	<i>Pyrococcus abyssi</i>	5'-ACCCCTGTGGAGAGGAGCCCTC 3'-TGGGGACACCTCTCCTCGGGGAG

Também são englobados desvios menores (degenerações) da sequência de reconhecimento que ainda permite o reconhecimento e clivagem pela específica de sequência DNA-endonuclease em questão. Estes desvios - também em conexão com diferentes condições de armação como, por exemplo, concentração de cálcio ou magnésio - foram descritos (Argast GM *et al.* 1998). Também são englobadas as sequências de núcleo se estas

sequências de reconhecimentos e desvios menores (degenerações) nas mesmas. Sabe-se que as porções internas das sequências de reconhecimento bastam para uma ruptura de filamento duplo induzida e que as externas não são absolutamente relevantes mas podem co-determinar a eficácia de clivagem. Assim, por exemplo, uma sequência de núcleo de 18bp pode ser definida para I-SceI.

2.3.2 Iniciação de Deleção / Excisão

Existem vários meios apropriados para iniciar a deleção / excisão de uma sequência alvo. Preferivelmente a deleção é somente iniciada após a integração com sucesso da sequência alvo no genoma da planta. Por exemplo em casos, onde a sequência alvo é um marcador de seleção, excisão é preferivelmente iniciada após o marcador ter completado com sucesso sua função resultando em inserção da construção de DNA no genoma da célula de planta ou organismo a ser transformado.

Vários meios são disponíveis para o versado na técnica combinar a enzima de excisão com a sequência alvo flanqueada pelas sequências de excisão. Preferivelmente, um enzima de excisão (por exemplo, uma recombinase ou endonuclease) pode ser expressada ou combinado com sua sequência de excisão correspondente (por exemplo, um sítio de recombinação ou reconhecimento), respectivamente, pelo método selecionado dentre o grupo consistindo de:

a) incorporação do cassete de expressão para a enzima de excisão (por exemplo, a recombinase ou endonuclease específica de sequência) em uma construção de DNA, preferivelmente junto com a sequência alvo (por exemplo, um gene marcador) flanqueada por referido excisão sequências,

b) incorporação do cassete de expressão para a enzima de excisão (por exemplo, a recombinase ou endonuclease específica de sequência) em células de planta ou plantas, que já compreendem a sequência

alvo (por exemplo, um gene marcador) flanqueada por referidas sequências de excisão,

c) incorporação do cassete de expressão para a enzima de excisão (por exemplo, a recombinase ou endonuclease específica de sequência) em células de planta ou plantas, que são subsequentemente usadas para plantas mestre ou células para transformação com construções compreendendo a sequência alvo (por exemplo, um gene marcador) flanqueada por referidas sequências de excisão,

d) incorporação do cassete de expressão para a enzima de excisão (por exemplo, uma recombinase ou endonuclease específica de sequência) em uma construção de DNA separada, que é transformada por meio de co-transformação com uma construção de DNA separada compreendendo a sequência alvo (por exemplo, um gene marcador) flanqueada por referidas sequências de excisão.

Assim a sequência alvo e a enzima de excisão (por exemplo, a recombinase ou endonuclease) pode ser combinada em um organismo de planta, célula, compartimento de célula ou tecido por exemplo como a seguir:

1.) Plantas compreendendo inserida em seu genoma a sequência alvo (por exemplo, um gene marcador) flanqueada por sequências de excisão (preferivelmente no DNA cromossômico) são geradas do modo comum um outro cassete de expressão para a enzima de excisão é então combinado com referidas construções de DNA

a) uma segunda transformação com referido segundo cassete de expressão, o

b) cruzamento das plantas compreendendo a sequência alvo com plantas mestre compreendendo o cassete de expressão para a enzima de excisão.

2.) O cassete de expressão codificando para a enzima de excisão pode ser integrado na construção de DNA que já contém a sequência

alvo. Prefere-se inserir a sequência codificando a enzima de excisão entre as sequências permitindo deleção e assim deletar o mesmo do DNA genômico após ter preenchido sua função. Especialmente preferido, expressão da endonuclease é indutível em tal caso (por exemplo sob o controle de um dos

5 promotores indutíveis descritos abaixo), em um modo dependente do desenvolvimento - usando um promotor dependente do desenvolvimento, ou também enzimas de excisão são empregadas cuja atividade é indutível a fim de evitar a deleção prematura do marcador de função dupla antes de sua inserção no genoma.

10 3.) Baseado na técnica de co-transformação, o cassete de expressão para a enzima de excisão pode ser transformado nas células simultaneamente com a construção de DNA compreendendo a sequência alvo, mas em uma molécula de DNA separada (por exemplo, vetor). Co-transformação pode ser estável ou transiente. Em tal caso, a expressão da

15 enzima de excisão é preferivelmente indutível (por exemplo sob o controle de uma das sequências reguladoras da transcrição quimérica indutíveis como descrito acima), apesar do padrão de expressão dependente de desenvolvimento do super-promotor não modificado já estar prevenindo a excisão prematura.

20 4.) Plantas expressando a enzima de excisão também podem atuar como indivíduos parentais. Na progênie a partir do cruzamento entre plantas expressando a enzima de excisão, por um lado, e plantas contendo a sequência alvo por outro lado, a excisão da sequência alvo desejada (por exemplo, por rupturas de filamento duplo e recombinação entre as sequências

25 de homologia) são observadas.

A forma de realização preferida da invenção está relacionada com as construções de DNA compreendendo tanto a sequência alvo (por exemplo, um cassete de expressão, um marcador de seleção; o primeiro cassete de expressão) e um segundo cassete de expressão para a enzima de

excisão (por exemplo, uma sequência codificando endonuclease ou recombinase ligada a um promotor de planta), preferivelmente em um modo em que referido segundo cassete de expressão é junto com referido primeiro cassete de expressão flanqueado por referidas sequências de excisão, que
5 permitem deleção de alvo específica de sequência.

Em outra forma de realização preferida o mecanismo de deleção /excisão pode ser induzido ou ativado em um modo para evitar deleção /excisão pré-matura do marcador de função dupla. Preferivelmente, assim expressão e/ou atividade da enzima de excisão preferivelmente
10 empregada pode ser induzida, preferivelmente pelo método selecionado dentre o grupo consistindo de

a) expressão indutível por ligando operativamente a sequência codificando a referida enzima de excisão (por exemplo, um recombinase ou endonuclease) para um promotor indutível,

b) ativação indutível, por emprego de uma enzima de excisão modificada (por exemplo, uma recombinase ou endonuclease) compreendendo um domínio de ligação de ligando, cuja atividade de referida enzima de excisão modificada pode ser modificada por tratamento da composto tendo atividade de ligação para o referido domínio de ligação de
15 ligando.
20

Expressão do polinucleotídeo codificando a enzima de excisão é preferivelmente controlada por um promotor de excisão, que permite a expressão em um modo sincronizado, de modo que o marcador de função dupla pode realizar sua função como um marcador de seleção negativo antes
25 de se tornar excisado. Os promotores apropriados são por exemplo descritos no Pedido de patente DE 03028884.9. Estes promotores podem ter, por exemplo, especificidade de expressão para o desenvolvimento tardio em todos os estágios como, por exemplo, tecido reprodutivo. O promotor de excisão pode ser selecionado dentre um dos seguintes grupos de promotores:

2.3.3 A sequência alvo a ser excisada

Apesar de várias sequências serem contempladas aqui, cuja excisão pode ser vantajosa, a sequência alvo a mais preferida a ser excisada é uma sequência de marcador. Várias sequências selecionáveis e triáveis de marcador são incluídas sob o termo geral de sequência de marcador. Assim, os métodos da invenção resultam e uma célula de planta monocotiledônea ou planta, que é livre de marcador. Os termos “livre de marcador- ” ou “livre de marcador de seleção” como usados aqui com relação a uma célula ou organismo se destinam a significar uma célula ou um organismo que não é capaz de expressar uma proteína de marcador funcional. A sequência codificando referida proteína de marcador pode estar ausente em parte ou – preferivelmente – completamente.

2.3.3.1 Genes marcadores

Os genes marcadores (por exemplo, marcadores selecionáveis ou triáveis) são usados com frequência a fim de melhorar a capacidade de identificar os transformantes. "Genes marcadores" são genes que conferem um fenótipo distinto às células expressando o marcador gene e assim permitem que estas células transformadas sejam distintas de células que não tem o marcador. Estes genes podem codificar ou marcadores selecionáveis ou triáveis, dependendo de se o marcador confere um traço que se pode "selecionar " por meios químicos *isto é*, através de uso do agente seletivo (por exemplo, herbicida, antibiótico ou outro) ou se ele é simplesmente um traço que se pode identificar através de observação ou teste, *isto é*, por "triagem" (por exemplo, o traço de R-locus, a proteína fluorescente verde (GFP)). Como evidente, muitos exemplos de apropriados genes marcadores são conhecidos na técnica e podem ser empregados na prática da invenção. São incluídos dentro dos termos genes marcadores selecionáveis ou triáveis, também os genes que codificam um " marcador selecionável" cuja secreção pode ser detectada como um meio de identificar ou selecionar para células

transformadas. Exemplos incluem marcadores, que codificam um antígeno secretável que pode ser identificado por interação de anticorpos, ou ainda enzimas secretáveis, que podem ser detectadas por sua atividade catalítica. As proteínas secretáveis estão dentre várias classes, incluindo proteínas pequenas, difundíveis, detectáveis, por exemplo, por ELISA; enzimas ativas pequenas detectáveis em solução extracelular (por exemplo, alfa-amilase, beta-lactamase, fosfinotricina acetiltransferase); e proteínas que são inseridas ou aprisionadas na parede celular (por exemplo, proteínas que incluem uma sequência líder como a encontrado em unidade de expressão de extensina ou PR-S de tabaco).

Com relação aos marcadores selecionáveis secretáveis, o uso de um gene que codifica a proteína que se torna sequestrada na parede celular, e cuja proteína inclui um epítipo único, é considerado para ser particularmente vantajoso. Este marcador de antígeno secretado deve idealmente empregar uma sequência de epítipo que iria prover um fundo baixo em tecido de planta, uma sequência de promotor- líder que iria conferir uma expressão e marcação eficientes através da membrana de plasma, e iria produzir proteína que é ligada na parede celular e ainda acessível aos anticorpos. Uma proteína de parede normalmente secretada modificada para incluir um epítipo único iria atender a todas estas exigências. Um exemplo da proteína apropriada para modificação deste modo é extensina ou glicoproteína rica em hidroxiprolina (HPRG). Por exemplo, a molécula HPRG de milho (Steifel 1990) é bem caracterizada em termos de biologia molecular, expressão e estrutura da proteína. No entanto, qualquer uma dentre a variedade de proteínas ricas em glicina e/ou ultilano (Keller 1989) pode ser modificado pela adição de um sítio antigênico para criar um marcador triável.

Uma forma de realização exemplar do marcador triável secretável refere-se ao uso da sequência de milho codificando HPRG da proteína de parede, modificado para incluir um epítipo de 15 resíduos da pro-

região de interleucina de murinos, no entanto, virtualmente qualquer epítopo detectável pode ser empregado em tais formas de realização, como selecionado a partir da variedade extremamente ampla de combinações antígeno - anticorpo conhecidas do versado na técnica. O epítopo extracelular
 5 único pode então ser detectado diretamente usando rotulação de anticorpo em conjunto com adjuvantes cromogênicos ou fluorescentes.

Elementos da presente descrição podem ser exemplificados em detalhes através do uso do gene bar e/ou GUSs, e também através de uso de vários outros marcadores. Com evidente à luz desta descrição, numerosos
 10 outros possíveis marcadores de genes selecionáveis e/ou triáveis serão evidentes para o versado na técnica além do especificado aqui abaixo. Assim, será entendido que a seguinte discussão é exemplar em vez de exaustiva, À luz das técnicas descritas aqui e as técnicas recombinantes que são conhecidos na técnica, a presente invenção torna possível a introdução de qualquer gene,
 15 incluindo genes marcadores, em uma célula recipiente para gerar uma planta transformada.

A sequência de marcador pode ser expressada por qualquer sequência reguladora da transcrição ou promotor tendo capacidade de expressão em células de planta (sequências apropriadas de promotor são
 20 descritas abaixo). As mais preferidas são as sequências de marcador, que são empregadas em transformação de plantas, triagem e seleção. Marcadores permitem que células ou organismos transgênicos (por exemplo, plantas ou células de planta) sejam identificados após transformação. Eles podem ser divididos em marcador de seleção positivo (conferindo uma vantagem
 25 seletiva), marcador de seleção negativo (compensando uma desvantagem de seleção), e marcador de contra-seleção (conferindo uma desvantagem de seleção), respectivamente. Estes marcadores podem incluir mas não são limitados a:

2.3.3.1.1 Marcadores de seleção negativos

Os marcadores de seleção negativos conferem uma resistência a um composto biocida, como um inibidor metabólico (por exemplo, 2-deoxiglucose-6-fosfato, WO 98/45456), antibióticos (por exemplo, canamicina, G 418, bleomicina ou higromicina) ou herbicidas (por exemplo, fosfinotricina ou glifosato). Os materiais de planta transformada (por exemplo, células, tecidos ou plântulas), que expressam genes marcadores, são capazes de desenvolver em presença de concentrações do composto de seleção correspondente (por exemplo, antibiótico ou herbicida), que suprime o crescimento de um tecido de tipo selvagem não transformado. Os marcadores negativos especialmente preferidos de seleção são os que conferem resistência a herbicidas. Exemplos, que podem ser mencionados, are:

- Fosfinotricina acetiltransferases (PAT; também chamado resistência Bialophos[®]; bar; de Block 1987; Vasil 1992, 1993; Semanas 1993; Becker 1994; Nehra 1994; Wan & Lemaux 1994; EP 0 333 033; US 4,975,374). Preferidos são o gene bar de *Streptomyces hygroscopicus* ou o gene pat de *Streptomyces viridochromogenes*. PAT inativa o ingrediente ativo no herbicida bialafos, fosfinotricina (PPT). PPT inibe glutamina sintetase, (Murakami 1986; Twell 1989) causando acúmulo rápido de amônio e morte celular.

- 5-enolpiruvilshikimato-3-fosfato sintase alterado (EPSPS) confere resistência a Glifosato[®] (N-(fosfonometil)glicina) (Hinchee 1988; Shah 1986; Della-Cioppa 1987). Onde um gene mutante de EPSP sintase é empregado, benefício adicional pode ser realizado através de incorporação do peptídeo de trânsito de cloroplasto apropriado, CTP (EP-A1 0 218 571).

- Enzimas degradando Glifosato[®] (Glifosato[®] oxidoreductase; gox),

- desalogenases inativantes de Dalapon[®] (deh)

- acetolactato sintases inativando sulfoniluréia e/ou

imidazolinona (*ahas* ou ALS; por exemplo variantes *ahas*/ALS mudadas com, por exemplo, a mutação S4, XI12, XA17, e/ou Hra (EP-A1 154 204)

- nitrilases degradando Bromoxynil® (bxn; Stalker 1988)

- genes resistência a canamicina ou geneticina (G418) (NPTII;

5 NPT ou neo; Potrykus 1985) codificando por exemplo, para neomicina fosfotransferases (Fraley 1983; Nehra 1994)

- 2-Desoxiglucose-6-fosfato fosfatase (DOG^R1-Produto de gene; WO 98/45456; EP 0 807 836) conferindo resistência contra 2-desoxi glucose (Randez-Gil 1995).

10 - hidromicina fosfotransferase (HPT), que media a resistência a hidromicina (Vanden Elzen 1985).

- di-hidrofolato reductase alterada (Eichholtz 1987) conferindo resistência contra metotrexato (Thillet 1988);

- genes de antranilato sintase mudados que conferem
15 resistência a 5-metil triptofano.

Os genes marcadores selecionáveis negativos adicionais de origem bacteriana que conferem resistência a antibióticos incluem o gene *aadA*, que confere resistência ao antibiótico espectinomicina, gentamicina acetil transferase, estreptomicina fosfotransferase (SPT), aminoglicosídeo-3-
20 adenil transferase e o determinante de resistência a bleomicina (Hayford 1988; Jones 1987; Svab 1990; Hille 1986).

Especialmente preferidos são os marcadores de seleção negativos que conferem resistência contra os efeitos tóxicos impostos por D-aminoácidos como por exemplo, D-alanina e D-serina (WO 03/060133;
25 Erikson 2004). Especialmente preferidos como marcadores de seleção negativos neste contexto são o gene *daol* (EC: 1.4. 3.3: GenBank Acc.No. U60066) da levedura *Rhodotorula gracilis* (*Rhodospiridium toruloides*) e o gene *dsdA* de *E. coli* (D-serina desidratase (D-serina deaminase) [EC: 4.3. 1.18; GenBank Acc.No. J01603).

Os materiais de planta transformados (por exemplo, células, embriões, tecidos ou plântulas) que expressam estes genes marcadores são capazes de desenvolvimento na presença de concentrações do composto de seleção correspondente (por exemplo, antibiótico ou herbicida) que suprime o crescimento de um tecido de tipo selvagem não transformado. As plantas resultantes podem ser cruzadas e hibridizadas em modo comum. Duas ou mais gerações devem ser cultivadas a fim de assegurar que a integração genômica é estável e hereditária. Os correspondentes métodos são descritos (Jenes 1993; Potrykus 1991).

Além disso, genes repórteres podem ser empregados para permitir a triagem visual, que pode ou não (dependendo do tipo de gene repórter) requerer suplementação com um substrato como um composto de seleção.

Vários esquemas de tempo podem ser empregados para os vários genes marcadores de seleção negativos. Em caso de genes de resistência (por exemplo, contra herbicidas ou D-aminoácidos) a seleção é preferivelmente aplicada através da fase de indução de calos durante cerca de 4 semanas e além de pelo menos 4 semanas em regeneração. Este esquema de seleção pode ser aplicado para todos os regimes de seleção. Além disso, é possível (apesar de não explicitamente preferido) permanecer na seleção também em todo o esquema de regeneração incluindo a formação de raízes.

Por exemplo, com o gene de resistência a fosfinotricina (bar) como o marcador seletivo, a fosfinotricina em uma concentração de cerca de 1 a 50 mg/L pode ser incluído no meio. Por exemplo, com o gene *dao1* como o marcador seletivo, D-serina ou D-alanina a uma concentração de cerca de 3 a 100 mg/L pode ser incluído no meio. As concentrações típicas para seleção são 20 a 40 mg/L. Por exemplo, com os genes *ahas* mudados como o marcador seletivo, PURSUIT™ a uma concentração de cerca de 3 a 100 mg/L pode ser incluído no meio. As concentrações típicas para seleção são 20 a 40

mg/L.

2.3.3.1.2 Marcador de seleção positivo

Além disso, marcador de seleção positivo pode ser empregado. Os marcadores de seleção positivos são os não resultam em destoxificação do composto biocida, mas conferem uma vantagem por regeneração, crescimento, propagação, multiplicação aumentados ou melhorados como os de uma célula ou organismo compreendendo este tipo de marcador. Exemplos são isopenteniltransferase (uma enzima chave da regeneração facilitando a biossíntese de citoquinina de células de planta transformadas por seleção em meio isento de citoquinina; Ebinuma 2000a; Ebinuma 2000b; por exemplo de cepa:PO22; Genbank Acc.No. AB025109). Os marcadores de seleção positivos adicionais, que conferem uma vantagem de crescimento a células de planta transformadas em comparação com as não transformadas, são descritos por exemplo, em EP-A 0 601 092. Os marcadores de seleção estimuladores do crescimento podem incluir (mas não devem ser limitados a) □-Glucuronidase (em combinação com por exemplo, uma citoquinina glucuronídeo), manose-6-fosfato isomerase (em combinação com manose), UDP-galactose-4-epimerase (em combinação com por exemplo, galactose), em que manose-6-fosfato isomerase em combinação com manose é especialmente preferido.

2.3.3.1.3 Marcador de contra-seleção

A sequência alvo a ser excisada pode não somente compreender um marcador de seleção negativo ou um marcador de seleção positivo (para facilitar a seleção e isolamento de plantas transformadas com sucesso) mas podem também compreender um -marcador de contra- seleção para avaliar a subsequente excisão do marcador com sucesso. Em uma forma de realização preferida tanto o marcador de seleção positivo com o negativo e o marcador de contra seleção são flanqueados pelas sequências de excisão e são ambos deletados/ excisados por adição da enzima de excisão. Os marcadores de contra-seleção são especialmente apropriados para selecionar

organismos com definidas sequências deletadas compreendendo referido marcador (Koprek 1999). Os marcadores de contra-seleção são sequências codificando para enzimas que são capazes de converter um composto não tóxico em um composto tóxico. Em consequência, somente as células irão sobreviver ao tratamento com referido composto não tóxico quando estão faltando o referido marcador de contra-seleção, assim permitindo seleção de células que sofreram uma deleção de sequência com sucesso (por exemplo, de marcador). Os marcadores de contra-seleção conhecidos na técnica são por exemplo

10 a) citosina deaminases (CodA) em combinação com 5-fluorocitosina (5-FC) (WO 93/01281; US 5,358,866; Gleave AP *et al.* 1999; Perera RJ *et al.* 1993; Stougaard J 1993; EP-A1 595 837; Mullen CA *et al.* 1992; Kobayashi T *et al.* 1995; Schlaman HRM & Hooykaas PFF 1997; Xiaohui Wang H *et al.* 2001; Koprek T *et al.* 1999; Gleave AP *et al.* 1999; Gallego ME 1999; Salomon S & Puchta H 1998; Thykjaer T *et al.* 1997; Serino G 1997; Risseuw E 1997; Blanc V *et al.* 1996; Corneille S *et al.* 2001).

20 b) enzimas Cytochrome P-450 em combinação com o pro-herbicida sulfoniluréia R7402 (2-metiletil-2-3-di-hidro-N-[(4,6-dimetoxipirimidina-2-il)aminocarbonil]-1,2-benzoisotiazol-7-sulfonamida-1,1-dióxido) (O'Keefe DP *et al.* 1994; Tissier AF *et al.* 1999; Koprek T *et al.* 1999; O'Keefe DP 1991).

25 c) hidrolases de ácido indol acético como por exemplo, o produto de gene tms2 de *Agrobacterium tumefaciens* em combinação com naftalacetamida (NAM) (Fedoroff NV & Smith DL 1993; Upadhyaya NM *et al.* 2000; Depicker AG *et al.* 1988; Karlin-Neumannn GA *et al.* 1991; Sundaresan V *et al.* 1995; Cecchini E *et al.* 1998; Zubko E *et al.* 2000).

 d) Haloalcano desalogenases (produto de gene dhlA) de *Xanthobacter autotropicus* GJ10 em combinação com 1,2-dichloroetano

(DCE) (Naested H *et al.* 1999; Janssen DB *et al.* 1994; Janssen DB 1989).

e) Timidina quinases (TK), por exemplo, de vírus Herpes Simplex tipo 1 (TK HSV-1), em combinação com aciclovir, ganciclovir ou 1,2-deoxi-2-fluoro- β -D-arabinofuranosil-5-iodouracila (FIAU) (Czako M & Marton L 1994; Wigler M *et al.* 1977; McKnight SL *et al.* 1980; McKnight SL *et al.* 1980; Preston *et al.* 1981; Wagner *et al.* 1981; St. Clair *et al.* 1987).

Vários outros sistemas de contra-seleção são conhecidos na técnica (ver por exemplo pedido internacional WO 04/013333; p.13 a 20 para um sumário; aqui incorporado por referência).

2.3.3.1.4. Marcadores triáveis

Os marcadores triáveis (também chamados genes repórteres ou proteínas; Schenborn E, Groskreutz D. 1999) que podem ser empregadas incluem, mas não são limitados a, beta-glucuronidase (GUS; Jefferson *et al.* 1987) ou gene uidA que codifica uma enzima para os vários substratos cromogênicos são conhecidos; um gene de R-locus, que codifica um produto que regula a produção de pigmento de antocianina (cor vermelha) em tecido de plantas (Dellaporta 1988); um gene beta-lactamase (Sutcliffe 1978), que codifica uma enzima para que vários substratos cromogênicos são conhecidos (por exemplo, PADAC, uma cefalosporina cromogênica); um gene xylE (Zukowsky 1983) que codifica uma catecol dioxigenase que pode converter catecóis cromogênicos; um gene α -amilase (Ikuta 1990); um gene tirosinase (Zukowsky 1983) que codifica uma enzima capaz de oxidar tirosina em DOPA e dopaquinona que por sua vez condensa para formar o composto melanina facilmente detectável ; gene β -galactosidase, que codifica uma enzima que são substratos cromogênicos; um gene luciferase (lux) (Ow 1986; Millar *et al.* 1992), que permite uma detecção bioluminescente; ou ainda um gene aequorina (Prasher 1985), que pode ser empregado em detecção de bioluminescência sensível a cálcio, ou um gene de proteína fluorescente verde (GFP) (Niedz 1995; Chui WL *et al.* 1996; Leffel SM *et al.* 1997; Sheen *et al.*

1995; Haseloff *et al.* 1997; Reichel *et al.* 1996; Tian *et al.* 1997).

Genes do complexo de milho R são contemplados como sendo particularmente utilizáveis como marcadores triáveis. O complexo de gene R em milho codifica uma proteína que atua para regular a produção de pigmentos de antocianina na maior parte das sementes e tecido de planta. Um gene do complexo R foi aplicado a transformação de milho, porque a expressão deste gene em células transformadas não prejudica as células. Assim, um gene R introduzido em tais células irá causar a expressão do pigmento vermelho e, se estavelmente incorporado, pode ser visualmente classificado como um setor vermelho. Se uma linhagem de milho é dominante para genes codificando os intermediários enzimáticos na via de biossíntese antocianina (C2, A1, A2, Bz1 e Bz2), mas transporta um alelo recessivo no locus R, transformação de qualquer célula da qual a linhagem com R irá resultar em formação de pigmento vermelho. As linhagens exemplares incluem Wisconsin 22 que contém o alelo rg-Stadler e TR112, um derivado K55 que é r-g, b, P1. Alternativamente qualquer genótipo de milho pode ser utilizado se os alelos C1 e R forem introduzidos juntos.

É ainda proposto que as regiões regulatórias de gene R podem ser empregadas em construções quiméricas a fim de prover mecanismos para controlar a expressão de genes quiméricos. Uma maior diversidade de expressão fenotípica é conhecida no locus R do que em qualquer outro locus (Coe 1988). É contemplado que regiões regulatórias obtidos de regiões 5' para o gene R gene estrutural devem ser valiosas para dirigir a expressão de genes, por exemplo, resistência a inseto, resistência a seca, tolerância a herbicida ou outras regiões de codificação de proteína. Para os fins da presente invenção, acredita-se que qualquer um dos vários membros da família de gene R pode ser empregada com sucesso (por exemplo, P, S, Lc, *etc.*). No entanto, o mais preferido será geralmente Sn (particularmente Sn:bol3). Sn é um membro dominante do complexo de gene R e é funcionalmente similar aos loci R e B

em que Sn controla a deposição específica de tecido de pigmentos de antocianina em algumas mudas e células de planta, assim seu fenótipo é similar a R.

Um outro marcador triável contemplado para uso na presente invenção é luciferase de vaga-lume, codificado para o gene lux. A presença do gene lux em células transformadas pode ser detectada usando, por exemplo, filme de raio X, contagem de cintilação, espectrofotometria fluorescente, câmaras de vídeo com pouca luz,, câmaras de contagem de fótons, ou luminometria de múltiplas cavidades. Também é visado que este sistema pode ser desenvolvido para triagem populacional para bioluminescência, como em placas de cultura de tecidos ou ainda para triagem de plantas completas. Onde se deseja o uso de um gene marcador triável como lux ou GFP, benefício pode ser obtido por criação de uma fusão de gene entre o gene marcador triável e um gene marcador selecionável, como por exemplo, uma fusão de genes GFP-NPTII. Isto pode permitir, por exemplo, seleção de células transformadas seguido por triagem de plantas ou sementes transgênicas.

2.3.3.1.5. Marcador de função dupla

Em uma forma de realização preferida da invenção a sequência alvo é um marcador de função dupla. O termo marcador de função dupla refere-se a um marcador que combina em uma sequência a oportunidade de ser empregado como um marcador de contra- seleção ou negativo. A escolha, cujo efeito é obtido, depende do substrato empregado no processo de triagem. O mais preferivelmente o marcador de função dupla é uma D-aminoácido oxidase. Esta enzima é capaz de converter D-aminoácidos. Alguns D-aminoácidos são tóxicos para plantas e são destoxificados por ação da enzima. Outros D-aminoácidos são prejudiciais às plantas mas são convertidos em compostos tóxicos pela enzima.

O termo D-aminoácido oxidase (abreviado DAAO, DAMOX,

ou DAO) faz referência à enzima convertendo o D-aminoácido em um ácido 2-oxo, por - preferivelmente - empregando oxigênio (O_2) como um substrato e produzindo peróxido de hidrogênio (H_2O_2) como um co-produto (Dixon M & Kleppe K. *Biochim. Biophys. Acta* 96 (1965) 357-367; Dixon M & Kleppe K *Biochim. Biophys. Acta* 96 (1965) 368-382; Dixon M & Kleppe *Biochim. Biophys. Acta* 96 (1965) 383-389; Massey V *et al. Biochim. Biophys. Acta* 48 (1961) 1-9. Meister A & Wellner D Flavoproteína aminoácido oxidase. Em: Boyer, P.D., Lardy, H. e Myrbäck, K. (Eds.), 1963)

DAAO pode ser descrito pela Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) com EC (Enzyme Commission) número EC 1.4.3.3. Geralmente uma enzima DAAO da classe EC 1.4.3.3. é uma FAD flavoenzima que catalisa a oxidação de D-aminoácidos neutros e básicos em seus ceto-ácidos correspondentes. DAAOs foram caracterizados e sequenciados em fungos e vertebrados onde eles são conhecidos como estando localizados nos peroxisomas. O termo D-amino oxidase ainda compreende D-aspartato oxidases (EC 1.4.3.1) (DASOX) (Negri A *et al.* 1992), que são enzimas estruturalmente relacionadas com DAAO catalisando a mesma reação mas ativos somente para D-aminoácidos dicarboxílicos. Nesta invenção DAAO da classe EC 1.4.3.3. é preferido.

Em DAAO, uma histidina conservada foi mostrada (Miyano M *et al.* 1991) como sendo importante para a atividade catalítica da enzima. Em uma forma de realização preferida da invenção um DAAO faz referência a uma proteína compreendendo o seguinte motivo de consenso:

[LIVM]-[LIVM]-H*-[NHA]-Y-G-x-[GSA]-[GSA]-x-G-x₅-G-x-A

em que resíduos de aminoácidos em parênteses representam resíduos alternativos para a posição respectiva, x representa qualquer resíduo de aminoácido, e números de índice indicam o número respectivo de resíduos de aminoácido consecutivos. A abreviatura para os resíduos de aminoácidos

individuais tem seu significado IUPAC padrão, como definido acima. Outras enzimas DAAO potenciais compreendendo referido motivo são descritas em tabela 3.

Tabela 3. D-aminoácido oxidases apropriadas a partir de vários organismos.

5 Acc.-No. refere-se a sequência de proteína da base de dados SwisProt.

No. acesso	Nome do gene	Descrição	Organismo de fonte	Comprimento
Q19564	F18E3.7	D-aminoácido oxidase putativa (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Caenorhabditis elegans	334
P24552		D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Fusarium solani (subsp. pisi) (Nectria haematococca)	361
P14920	DAO, DAMOX	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Homo sapiens (Humano)	347
P18894	DAO, DAO1	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Mus musculus (camundongo)	346
P00371	DAO	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Sus scrofa (Pig)	347
P22942	DAO	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Oryctolagus cuniculus (Rabbit)	347
O35078	DAO	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Rattus norvegicus (Rat)	346
P80324	DAO1	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Rhodospiridium toruloides (Yeast) (Rhodotorula gracilis)	368
U60066	DAO	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Rhodospiridium toruloides, cepa TCC 26217	368
Q99042	DAO1	D-aminoácido oxidase (EC 1.4.3.3) (DAMOX) (DAO) (DAAO)	Trigonopsis variabilis (Yeast)	356
P31228	DDO	D-aspartato oxidase (EC 1.4.3.1) (DASOX) (DDO)	Bos taurus (Bovino)	341
Q99489	DDO	D-aspartato oxidase (EC 1.4.3.1) (DASOX) (DDO)	Homo sapiens (Humano)	341
Q9C1L2	NCU06558.1	(AF309689) D-aminoácido oxidase putativa G6G8.6 (proteína hipotética)	Neurospora crassa	362
Q7SFW4	NCU03131.1	proteína hipotética	Neurospora crassa	390
Q8N552		Similar a D-aspartato oxidase	Homo sapiens (Humano)	369
Q7Z312	DKFZP686F04272	proteína hipotética DKFZp686F04272	Homo sapiens (Humano)	330
Q9VM80	CG11236	proteína CG11236 (GH12548p)	Drosophila melanogaster (mosca das frutas)	341
O01739	F20H11.5	proteína F20H11.5	Caenorhabditis elegans	383
O45307	C47A10.5	proteína C47A10.5	Caenorhabditis elegans	343

No. acesso	Nome do gene	Descrição	Organismo de fonte	Comprimento
Q8SZN5	CG12338	RE73481p	<i>Drosophila melanogaster</i> (mosca das frutas)	335
Q9V5P1	CG12338	proteína CG12338 (RE49860p)	<i>Drosophila melanogaster</i> (mosca das frutas)	335
Q86JV2		Similar a <i>Bos taurus</i> (Bovino). D-aspartato oxidase (EC 1.4.3.1) (DASOX) (DDO)	<i>Dictyostelium discoideum</i> (mofo de lodo)	599
Q95XG9	Y69A2AR.5	proteína hipotética	<i>Caenorhabditis elegans</i>	322
Q7Q7G4	AGCG53627	AgCP5709 (Fragmento)	<i>Anopheles gambiae</i> str. PEST	344
Q7PWY8	AGCG53442	AgCP12432 (Fragmento)	<i>Anopheles gambiae</i> str. PEST	355
Q7PWX4	AGCG45272	AgCP12797 (Fragmento)	<i>Anopheles gambiae</i> str. PEST	373
Q8PG95	XAC3721	D-aminoácido oxidase	<i>Xanthomonas axonopodis</i> (pv. citri)	404
Q8P4M9	XCC3678	D-aminoácido oxidase	<i>Xanthomonas campestris</i> (pv. campestris)	405
Q9X7P6	SCO6740, SC5F2A.23C	D-aminoácido oxidase putativa	<i>Streptomyces coelicolor</i>	320
Q82MI8	DAO, SAV1672	D-aminoácido oxidase putativa	<i>Streptomyces avermitilis</i>	317
Q8VCW7	DAO1	D-aminoácido oxidase	<i>Mus musculus</i> (camundongo)	345
Q9Z302		D-aminoácido oxidase	<i>Cricetulus griseus</i> (hamster chinês)	346
Q9Z1M5		D-aminoácido oxidase	<i>Cavia porcellus</i> (porquinho-da-índia)	347
Q922Z0		Similar a D-aspartato oxidase	<i>Mus musculus</i> (camundongo)	341
Q8R2R2		proteína hipotética	<i>Mus musculus</i> (camundongo)	341
P31228		D-aspartato oxidase	<i>B.taurus</i>	341

D-aminoácido oxidase (EC-número 1.4.3.3) pode ser isolada de vários organismos, incluindo mas não limitados suínos, humanos, ratos, leveduras, bactérias ou fungos. Organismos de exemplo são *Candida tropicalis*, *Trigonopsis variabilis*, *Neurospora crassa*, *Chlorella vulgaris*, e

5 *Rhodotorula gracilis*. Um polipeptídeo apropriado metabolizando D-aminoácido pode ser uma enzima eucariótica, por exemplo de uma levedura (por exemplo *Rhodotorula gracilis*), fungo, ou animal ou pode ser uma enzima procariótica, por exemplo, de uma bactéria como *Escherichia coli*.

Exemplos de apropriados polipeptídeos que metabolizam D-aminoácidos são como mostrados em tabela 4.

Tabela 4. D-aminoácido oxidases apropriadas a partir de vários organismos.

Acc.-No. refere-se a sequência de proteína de base de dados SwisProt

GenBank Acc.-No,	Organismo de fonte
Q19564	Caenorhabditis elegans. F18E3.7.
P24552	Fusarii solani (subsp. pisi) (Nectria haematococca).
JX0152	Fusarium solani
P14920	Homo sapiens (Humano)
P18894	Mus musculus (camundongo)
P00371	Sus scrofa (suíno)
P22942	Oryctolagus cuniculus (coelho)
O35078	Rattus norvegicus (Rato)
P80324	Rhodospiridium toruloides (levedura) (Rhodotorula gracilis)
Q99042	Trigonopsis variabilis
Q9Y7N4	Schizosaccharomyces pombe (levedura de fissão) SPCC1450
O01739	Caenorhabditis elegans.F20H11.5
Q28382	Sus scrofa (suíno).
O33145	Mycobacterium leprae
Q9X7P6	Streptomyces coelicolor.SCSF2A.23C
Q9JXF8	Neisseria meningitidis (serogrupo B).
Q9Z302	Cricetulus griseus (hamster chinês)
Q921M5	D-AMINO ACID OXIDASE. Cavia parcellus (porquinho da índia)

5

Preferivelmente a D-aminoácido oxidase é selecionada dentre as enzimas codificadas pela sequência de ácido nucleico ou as sequências de aminoácidos correspondentes selecionadas dentre a seguinte tabela 5:

Tabela 5: D-aminoácido oxidases apropriadas de vários organismos. Acc.-No. refere-se à sequência de proteína da Base de dados Genbank.

GenBanc Acc.-No	Organismo
U60066	Rhodospiridium toruloides (Yeast)
Z71657	Rhodotorula gracilis
A56901	Rhodotorula gracilis
AF003339	Rhodospiridium toruloides
AF003340	Rhodospiridium toruloides
U53139	Caenorhabditis elegans
D00809	Nectria haematococca
Z50019.	Trigonopsis variabilis
NC_003421	Schizosaccharomyces pombe (levedura de fissão)
AL939129.	Streptomyces coelicolor A3(2)
AB042032	Candida boidinii

10

DAAO é uma enzima bem caracterizada e sua estrutura de

crystal e seu mecanismo catalítico foram determinados por espectroscopia de raios-X de alta resolução (Umhau S. *et al.* 2000). Ela é uma flavoenzima localizada no peroxisoma e sua função reconhecida em animais é a destoxificação de D-aminoácidos (Pilone MS 2000). Além disso, ela permite
5 que as leveduras usem os D-aminoácidos para crescimento (Yurimoto H *et al.* 2000). Como demonstrado acima, DAAO de várias espécies diferentes foi caracterizada e mostrada como tendo afinidades de substrato levemente diferentes (Gabler M *et al.* 2000), mas em geral elas demonstram uma especificidade de substrato ampla, oxidativamente desaminando todos os D-
10 aminoácidos (exceto D-glutamato e D-aspartato para enzimas EC 1.4.3.3. classe DAAO; Pilone MS 2000).

A atividade de DAAO é encontrada em muitos eucariotes (Pilone MS 2000), mas não se tem registro de atividade de DAAO em plantas. A baixa capacidade para o metabolismo de D-aminoácido em plantas tem
15 como consequência principal o modo como as plantas respondem a D-aminoácidos.

Em uma forma de realização preferida D-aminoácido oxidase expressada da construção de DNA da invenção tem preferivelmente atividade enzimática contra pelo menos um dos aminoácidos selecionados dentre o
20 grupo consistindo de D-alanina, D-serina, D-isoleucina, D-valina, e seus derivados.

As D-aminoácido oxidases apropriadas também incluem fragmentos, mutantes, derivados, variantes e alelos dos polipeptídeos exemplificados acima. Os apropriados fragmentos, mutantes, derivados,
25 variantes e alelos são os que retêm a característica funcional da D-aminoácido oxidase como definido acima. Mudanças em uma sequência, para produzir um mutante, variante ou derivado, pode ser por uma ou mais dentre adição, inserção, deleção ou substituição de um ou mais nucleotídeos no ácido nucleico, levando à adição, inserção deleção ou substituição de um ou mais

aminoácidos no polipeptídeo codificado. Como evidente, mudanças no ácido nucleico que não fazem diferença para a sequência de aminoácido codificada são incluídas.

A D-aminoácido oxidase da invenção pode ser expressada no
 5 citosol, peroxisoma, ou compartimento intracelular da célula de planta. Compartimentalização do polipeptídeo metabolizando D-aminoácido pode ser obtida fusionando a sequência de ácido nucleico codificando o polipeptídeo de DAAO em uma sequência codificando um peptídeo de trânsito para gerar uma proteína de fusão. Os produtos de genes expressados sem estes peptídeos
 10 de trânsito geralmente se acumulam no citosol. A localização de DAAO expressado no peroxisoma produz H_2O_2 que pode ser metabolizado pela enzima degradando H_2O_2 catalase. Maiores níveis de D-aminoácidos podem ser assim requeridos para produzir níveis danificadores de H_2O_2 . Expressão de DAAO em citosol, onde os níveis de catalase atividade são menores, reduz a
 15 quantidade de D-aminoácido requerida para produzir níveis danificadores de H_2O_2 . Expressão de DAAO no citosol pode ser obtido por remoção dos sinais de marcação de peroxisoma ou peptídeos de trânsito da sequência codificando ácido nucleico. Por exemplo, o gene *dao1* (EC: 1.4.3.3: GenBank Acc.No. U60066) da levedura *Rhodotorula gracilis* (*Rhodospiridium toruloides*) foi
 20 clonado como descrito (WO 03/060133). Os últimos nove nucleotídeos codificam o peptídeo de sinal SKL, que guia a proteína para o organelo sub-celular de peroxisoma. Apesar de não terem sido observadas diferenças significantes entre DAAO expressado citossólico e peroxisomal, a construção peroxisomal foi verificada como sendo marginalmente mais efetiva do que a
 25 versão citossólica com relação à inibição da germinação das plantas transgênicas DAAO em 30 mM D-Asn. No entanto, ambas as construções são inibidas de modo significativo mais do que o tipo selvagem e podem assim ser usadas para uma contra-seleção condicional.

As modificações adicionais e uso do marcador de função dupla

são descritas em pedido EP. No. 04006358.8 (SweTree Technologies AB & BASF; IMPROVED CONSTRUCTS FOR MARKER EXCISION BASED ON DUAL-FUNCTION SELECTION MARKER) e pedidos adicionais nacionais e internacionais reivindicando prioridade ao mesmo.

5 2.3.4. Expressão do gene marcador e outras sequências

O gene marcador (ou outras sequências que podem ser expressadas a partir de uma das construções de DNA da invenção) pode ser expressado por qualquer promotor funcional em plantas. Estes promotores incluem, mas não são limitados a, promotores constitutivos, indutíveis, 10 temporariamente regulados, regulados no desenvolvimento, regulados no espaço, regulados quimicamente, respondentes a estresse, específicos de tecido, virais e sintéticos. O promotor pode ser um promotor de gama zeína, um promotor de oleosina ole16, um promotor de globulina, um promotor de actina I, um promotor de actina cl, um promotor de sacarose, um promotor de 15 sintetase, um promotor de INOPS, um promotor de EXM5, um promotor de globulina2, um promotor de β -32, promotor de ADPG-pirofosforilase, um promotor de LtpI, um promotor de Ltp2, um promotor de oleosina ole17, um promotor de oleosina ole18, um promotor de actina 2, um promotor de proteína específica de pólen, um promotor de pectato liase específica de 20 pólen, um promotor de proteína específico de antero, um promotor de gene RTS2 específico de antero, um promotor de gene específico de pólen, um promotor de gene específico de tapetum, um promotor de gene RAB24 específico de tapetum, um promotor de alfa-subunidade de antranilato sintase, um promotor de alfa zeína, um promotor de beta-subunidade de antranilato 25 sintase, um promotor de di-hidrodipicolinato sintase, um promotor de Thil, um promotor de álcool desidrogenase, um promotor de proteína de ligação cab, um promotor de H3C4, um promotor enzima de ramificação de amido de RUBISCO SS, um promotor de ACCase, um promotor de actin3, um promotor de actin7, um promotor de proteína regulatória GF14-12, um

promotor de proteína L9 ribossômica, um promotor de enzima biossintética de celulose, um promotor de S-adenosil-L-homocisteína hidrolase, um promotor de superóxido dismutase, um promotor de receptor C-quinase, um promotor de fosfoglicerato mutase, um promotor de mRNA RCc3 específico de raiz, um promotor de glucose-6 fosfato isomerase, um promotor de pirofosfato-fructose 6-fosfatofosfotransferase, um promotor de ubiquitina, um promotor de beta-cetoacil-ACP sintase, um promotor de 33 kDa photosystem 11, um promotor de proteína envolvendo oxigênio, um promotor de subunidade ATPase vacuolar de 69 kDa, um promotor de proteína tipo Metalotioneína, um promotor de gliceraldeído-3-fosfato desidrogenase, um promotor de proteína de tipo indutível de ABA- e de amadurecimento, um promotor de fenilalanina amônia liase promotor, um promotor de adenosina trifosfatase S-adenosil-L-homocisteína hidrolase, um promotor de α -tubulina, um promotor de cab, um promotor de PEPCase, um promotor de gene R, um promotor de lectina, um promotor de complexo de colheita leve, um promotor de proteína de choque térmico, um promotor de calcona sintase, um promotor de zeína, um promotor de globulina-1, um promotor de ABA, um promotor de ligação a auxina, um promotor de gene UDP glucose flavonóide glicosil-transferase, um promotor de NTI, um promotor de actina, um promotor opaque 2, um promotor de b70, um promotor de oleosina, um promotor de CaMV 35S, um promotor de CaMV 34S, um promotor de CaMV 19S, um promotor de histona, um promotor indutível por turgescência, um promotor de RuBP carboxilase de subunidade pequena de ervilha, um promotor de Ti plasmídeo manopina sintase, um promotor de Ti plasmídeo nopalina sintase, um promotor de petúnia calcona isomerase, um promotor de proteína I rica em glicina de feijão, um promotor de CaMV 35S transcrito, um promotor de patatina de batata, ou um promotor de RuBP carboxilase S-E9 subunidade pequena.

3. Testes de expressão de transgenes

Para confirmar a presença de um DNA exógeno em plantas regeneradas, vários testes podem ser realizados. Estes testes incluem, por exemplo, testes biológicos moleculares, como Southern e Northern blotting e PCR; testes bioquímicos como detecção da presença do produto de proteína, por exemplo, por meios imunológicos (ELISAs e Western blots) ou por função enzimática; testes de partes de plantas como testes de folhas e raiz; e em alguns casos, análise de fenótipos da planta regenerada completa. Testes adicionais utilizáveis para a determinação da eficiência da expressão de transgene e função do promotor também incluem sem limitação hibridização em situ fluorescente (FISH), sequenciamento de DNA direto, análise de eletroforese de gel de campo pulsado, (PFGE), análise de conformação de filamento único (SSCA), teste de proteção, oligonucleotídeo específico do alelo(ASO), análise dot blot, eletroforese de gel gradiente desnaturante, RT-PCR, RT-PCR quantitativo, RFLP e PCR-SSCP. Estes testes são conhecidos do versado na técnica (ver também acima).

4. Plantas da invenção transformadas (transgênicas) e métodos de Preparação

As espécies de plantas monocotiledôneas podem ser transformadas com a construção de DNA da presente invenção por vários métodos conhecidos na técnica. Qualquer tecido de planta capaz de uma propagação clonal subsequente, seja por organogênese ou embriogênese, pode ser transformado. O termo "organogênese," como usado aqui, significa um processo por que brotos e raízes são desenvolvidos sequencialmente a partir de centros meristemático, o termo "embriogênese," como usado aqui, significa um processo por que brotos e raízes se desenvolvem juntos em um modo em combinação (não sequencialmente) juntos a partir de células somáticas ou gametas. O tecido particular selecionado irá variar dependendo dos sistemas de propagação clonal disponíveis para e melhor apropriados para a espécie em particular sendo transformada. Os alvos de tecido exemplares incluem discos de folhas, pólen, embriões, cotilédones, hipocótilos,

megagametófitos, tecidos de calo, tecido meristemático existente (por exemplo, meristemas apicais, brotos axilares, e meristemas de raiz) e tecido de meristema induzido (por exemplo meristema de cotilédone e meristema de utililano).

5 Plantas da presente invenção podem tomar várias formas. As plantas podem ser quimeras de células transformada e células não transformadas; as plantas podem ser transformantes clonais (por exemplo, todas as células transformadas para conter o cassete de expressão); as plantas podem compreender enxertos de tecidos transformados e não transformados
10 (por exemplo, uma carga de raiz transformada enxertada em um enxerto não transformado na espécie de cítricos). As plantas transformadas podem ser propagada por vários meios como por propagação clonal ou por técnicas clássicas de cruzamento. Por exemplo, plantas transformadas de primeira geração (ou T1) podem ser auto-cruzadas para dar plantas transformadas de
15 segunda geração homozigóticas (ou T2) e as plantas T2 ainda propagadas através de técnicas de cruzamento clássicas. Um marcador selecionável dominante (como npt II) pode ser associado com o cassete de expressão para auxiliar no cruzamento.

Assim, a presente invenção provê plantas monocotiledôneas transformadas (transgênicas) e células de planta monocotiledônea, *in planta*
20 ou *ex planta*, incluindo um plastídeo transformado ou outro organelo, por exemplo, núcleo, mitocôndria ou cloroplasto. A presente invenção pode ser usada para transformação de qualquer espécie de planta monocotiledônea, incluindo, mas não limitada a células de uma espécie de planta especificada
25 acima na seção de DEFINIÇÃO. Preferivelmente, plantas transgênicas da presente invenção são plantas de cultura e em particular cereais (por exemplo, milho, alfafa, arroz, centeio, sorgo, trigo, milhete *etc.*), e ainda mais preferivelmente milho, trigo e arroz. Outras formas de realização da invenção são relacionadas com células, culturas de células, tecidos, partes (como

órgãos de plantas, folhas, raízes, etc.) e material de propagação (como sementes) destas plantas monocotiledôneas.

Transformação de plantas monocotiledôneas pode ser realizada com uma molécula de DNA ou única ou molécula de DNA múltipla (isto é, co-transformação) e ambas as técnicas são apropriadas para uso com os cassetes de expressão da presente invenção. Os numerosos vetores de transformação são disponíveis para a transformação de plantas e os cassetes de expressão desta invenção podem ser usados em conjunto com quaisquer um de tais vetores. A seleção de vetor irá depender da técnica de transformação preferida e a espécie alvo para a transformação.

Várias técnicas são disponíveis e conhecidas dos versados na técnica para a introdução de construções dentro de uma célula hospedeira de planta. Estas técnicas geralmente incluem transformação com DNA empregando *A. tumefaciens* ou *A. rhizogenes* como o agente de transformação, lipossomas, precipitação PEG, eletroporação, injeção de DNA, absorção direta de DNA, bombardeio de microprojéteis, aceleração de partículas e outros (ver, por exemplo, EP 295959 e EP 138341). No entanto, células diferentes de células de planta podem ser transformadas com os cassetes de expressão da invenção. As descrições gerais de vetores de expressão de plantas e genes repórteres, e transferência de genes mediada por *Agrobacterium* e *Agrobacterium*, podem ser encontradas em Gruber *et al.* (1993).

Os vetores de expressão contendo fragmentos genômicos ou sintéticos podem ser introduzidos em protoplastos ou tecidos ou células isoladas intactas. Preferivelmente, os vetores de expressão são introduzidos em tecido intacto. Os métodos gerais de cultivo de tecido de plantas são providos por exemplo por Maki *et al.*, (1993); e por Phillips *et al.* (1988). Preferivelmente, os vetores de expressão são introduzidos em milho ou outros tecidos de plantas usando um método de transferência de genes direto, como

entrega mediada por microprojéteis, injeção de DNA eletroporação, e outros. Mais preferivelmente vetores de expressão são introduzidos em tecidos de plantas usando a entrega mediada por microprojéteis com o dispositivo biolístico. Ver, por exemplo, Tomes et al. (1995). Os vetores da invenção
5 podem não somente ser usados para expressão de genes estruturais mas também podem ser usados em clonagem de exon-trap, ou procedimentos de promotor trap para detectar a expressão de genes diferencial em variedades de tecidos (Lindsey 1993; Auch & Reth 1990).

É articularmente preferido usar os vetores de tipo binário de
10 plasmídeos Ti e Ri de *Agrobacterium* spp. Os vetores derivados de Ti transformam uma ampla variedade de plantas superiores, incluindo plantas monocotiledôneas e dicotiledôneas, como soja, algodão, colza, tabaco, e arroz (Pacciotti 1985; Byrne 1987; Sukhapinda 1987; Lorz 1985; Potrykus, 1985; Park 1985; Hiei 1994). O uso de T-DNA para transformar células de planta
15 recebeu um estudo extensivo e é amplamente descrito (EP 120516; Hoekema, 1985; Knauf, 1983; e An 1985). Para introdução em plantas, os genes quiméricos da invenção podem ser inseridos em vetores binários, como descrito nos exemplos.

Outros métodos de transformação são disponíveis para os
20 versados na técnica, como absorção direta de construções de DNA estrangeiro (ver EP 295959), técnicas de eletroporação (Fromm 1986) ou bombardeio balístico de alta velocidade com partículas de metal revestidas com a construção de ácido nucleico (Kline 1987, e US 4,945,050). Uma vez transformadas, as células podem ser regeneradas pelos versados na técnica.

25 Os versados na técnica irão notar que a escolha do método pode depender do tipo de planta monocotiledônea marcada para transformação. Apropriados métodos de transformação de células de planta incluem, mas não são limitados a, microinjeção (Crossway 1986), eletroporação (Riggs 1986), transformação mediada por *Agrobacterium*,

transferência de genes direta (Paszkowski 1984), e aceleração de partículas balísticas usando dispositivos disponíveis de Agracetus, Inc., Madison, Wis. e BioRad, Hercules, Calif. (ver, por exemplo, US 4.945.050; e McCabe 1988). Também ver, Datta 1990; (arroz); Klein 1988 (milho); Klein 1988 (milho);
 5 Klein 1988 (milho); Fromm 1990 (milho); e Gordon-Kamm 1990 (milho); Koziel 1993 (milho); Shimamoto 1989 (arroz); Christou 1991 (arroz); Pedido de patente européia EP 0 332 581 (gramas de pomares e outras Pooideae); Vasil 1993 (trigo); Weeks 1993 (trigo), Li 1993 e Christou 1995 (arroz); Osjoda 1996 (milho via *Agrobacterium tumefaciens*), arroz (Hiei 1994), e
 10 milho (Gordon-Kamm 1990; Fromm 1990); todos sendo aqui incorporados por referência.

As células de *Agrobacterium tumefaciens* contendo um vetor compreendendo um cassete de expressão da presente invenção, em que o vetor compreende um plasmídeo Ti, são utilizáveis em métodos para fazer
 15 plantas transformadas. As células de plantas são infectadas com um *Agrobacterium tumefaciens* como descrito acima para produzir uma célula de planta transformada, e então uma planta é regenerada da célula de planta transformada. Numerosos sistemas de vetor de *Agrobacterium* utilizáveis na realização da presente invenção são conhecidos. Várias cepas de
 20 *Agrobacterium* podem ser empregadas, preferivelmente cepas desarmadas de *Agrobacterium tumefaciens* ou *rhizogenes*. Em uma forma de realização preferida, cepas de *Agrobacterium* para uso na prática da invenção incluem cepas octopina por exemplo, LBA4404 ou cepas agropina por exemplo, EHA101 ou EHA105. As cepas apropriadas de *A. tumefaciens* para
 25 transferência de DNA são por exemplo EHA101[pEHA101] (Hood 1986), EHA105[pEHA105] (Li 1992), LBA4404[pAL4404] (Hoekema 1983), C58C1[pMP90] (Koncz & Schell 1986), e C58C1[pGV2260] (Deblaere 1985). Outras cepas apropriadas são *Agrobacterium tumefaciens* C58, uma cepa de nopalina. Outras cepas apropriadas são *A. tumefaciens* C58C1 (Van

Larebeke 1974), A136 (Watson 1975) ou LBA4011 (Klapwijk 1980). Em outra forma de realização preferida a bactéria de origem no solo é uma variante desarmada de *Agrobacterium rhizogenes* cepa K599 (NCPBP 2659). Preferivelmente, estas cepas estão compreendendo uma variante desarmada de plasmídeo de um plasmídeo Ti ou Ri provendo as funções requeridas para a transferência de T-DNA em células de planta (por exemplo, os genes *vir*). Em uma forma de realização preferida, a cepa de *Agrobacterium* usadas para transformar um tecido de planta pre-cultivado com um composto fenólico de planta contém um Ti-plasmídeo de tipo L,L-succinamopina, preferivelmente desarmado, como pEHA101. Em outra forma de realização preferida, a cepa de *Agrobacterium* usada para transformar um tecido de planta pre-cultivado com um composto fenólico de planta contém um Ti-plasmídeo tipo octopina, preferivelmente desarmado, como pAL4404. Geralmente, quando usando Ti-plasmídeos tipo octopina ou plasmídeos auxiliares, prefere-se que o gene *virF* seja deletado ou inativado (Jarschow 1991).

O método da invenção também pode ser usado em combinação com cepas de *Agrobacterium* particulares para ainda aumentar a eficiência de transformação, como cepas *Agrobacterium* em que a expressão de gene *vir* e/ou indução da mesma é alterada devido à presença de genes *virA* ou *virG* mutantes ou quiméricos (por exemplo Hansen 1994; Chen e Winans 1991; Scheeren-Groot, 1994). Preferidas são outras combinações de cepas LBA4404 de *Agrobacterium tumefaciens* (Hiei 1994) com plasmídeos super-virulentos. Estes são preferivelmente vetores com base em pTOK246 (Ishida 1996).

Um vetor binário ou qualquer outro vetor pode ser modificado por técnicas de recombinação de DNA, multiplicadas em *E. coli*, e introduzidas em *Agrobacterium* por exemplo, técnicas de eletroporação ou outras técnicas de transformação (Mozo & Hooykaas 1991).

Agrobacterium é cultivado e usado em um modo similar ao descrito em Ishida (1996). O vetor compreendendo cepa de *Agrobacterium*

pode, por exemplo, ser cultivado durante 3 dias em meio YP (5 g/L extrato de levedura, 10 g/L peptona, 5 g/L NaCl, 15 g/L agar, pH 6.8) suplementado com os antibióticos apropriados (por exemplo, 50 mg/L espectinomicina). Bactérias são coletadas com um laço do meio sólido e recolocadas em suspensão. Em uma forma de realização preferida da invenção, culturas de *Agrobacterium* são iniciadas por uso de alíquotas congeladas a -80°C.

A transformação do tecido alvo (por exemplo, um embrião imaturo (para *Agrobacterium* pode ser realizada apenas contatando o tecido alvo com o *Agrobacterium*. A concentração de *Agrobacterium* usado para infecção e co-cultivação podem precisar ser variadas. Por exemplo, uma suspensão de células do *Agrobacterium* tendo uma densidade de população de aproximadamente de 10^5 - 10^{11} , preferivelmente 10^6 a 10^{10} , mais preferivelmente cerca de 10^8 células ou cfu / ml é preparada e o tecido alvo é imerso nesta suspensão durante cerca de 3 a 10 minutos. O tecido alvo resultante é então cultivado em um meio sólido durante vários dias junto com o *Agrobacterium*.

Preferivelmente, a bactérias é empregada em concentração de 10^6 a 10^{10} cfu/mL. Em uma forma de realização preferida a etapa de co-cultivo de cerca de 1 a 10 μ l da suspensão da bactéria de origem no solo (por exemplo, *Agrobactérias*) no meio de co-cultivo são aplicados diretamente a cada explante de tecido alvo e secados ao ar. Isto economiza tempo e trabalho e reduz o dano mediado por *Agrobacterium* não intencional por uso em excesso de *Agrobacterium*.

Para tratamento de *Agrobacterium*, as bactérias são recolocadas em suspensão em uma planta compatível com o meio de co-cultivo. A suplementação do meio de co-cultura com antioxidantes (por exemplo, nitrato de prata), compostos absorvendo fenol (como polivinil pirrolidona, Perl 1996) ou compostos tiol (por exemplo, ditiotreitól, L-cisteína, Olhoft 2001) que podem diminuir a necrose de tecido devido às

respostas de defesa da planta (como oxidação fenólica) podem ainda melhorar a eficiência da transformação mediada por *Agrobacterium*. Em outra forma de realização preferida, o meio de co-cultivo compreende pelo menos um composto tiol, preferivelmente selecionado dentre o grupo consistindo de tiolsulfato de sódio, ditionitriol (DTT) e cisteína. Preferivelmente a concentração está entre cerca de 1 mM e 10mM de L-Cisteína, 0,1 mM a 5 mM DTT, e/ou 0,1 mM a 5 mM tiolsulfato de sódio. Preferivelmente, o meio empregado durante o co-cultivo compreende de cerca de 1 µM a cerca de 10 µM de nitrato de prata e de cerca de 50 mg/L a cerca de 1,000 mg/L de L-Cisteína. Isto resulta em uma vulnerabilidade altamente reduzida contra o dando mediado por *Agrobacterium* (como necrose induzida) e melhora altamente a eficiência de transformação global.

Vários sistemas de vetor podem ser usados em combinação com *Agrobactérias*. Preferidos são sistemas de vetor binário. Os vetores binários comuns são baseados em plasmídeos de "ampla faixa de hospedeiro" como pRK252 (Bevan 1984) ou pTJS75 (Watson 1985) derivados plasmídeo RK2 tipo P. A maior parte destes vetores são derivados de pBIN19 (Bevan 1984). Vários vetores binários são conhecidos, alguns sendo comercialmente disponíveis como, por exemplo, pBI101.2 ou pBIN19 (Clontech Laboratories, Inc. USA). Os vetores adicionais forma melhorados com relação ao tamanho e manipulação (por exemplo pPZP; Hajdukiewicz 1994). Melhorados sistemas de vetor são descritos também em WO 02/00900.

Os métodos usando ou uma forma de uma transferência de genes direta ou transferência mediada por *Agrobacterium* geralmente, mas não necessariamente, são realizados com um marcador selecionável, que pode proporcionar resistência a um antibiótico (por exemplo, canamicina, hidromicina ou metotrexato) ou um herbicida (por exemplo, fosfinotricina). A escolha do marcador selecionável para transformação de planta não é, no entanto, crítica para a invenção.

Para algumas espécies de plantas, diferentes marcadores de antibiótico ou herbicida de seleção podem ser preferidos. Marcadores de seleção usados como rotina na transformação incluem o gene *nptII* que confere resistência a canamicina e antibióticos relacionados (Messing & Vierra, 1982; Bevan 1983), o gene *bar* confere resistência ao herbicida fosfinotricina (White 1990, Spencer 1990), o gene *hph* que confere resistência ao antibiótico hidromicina (Blochlinger & Diggelmann), e o gene *dhfr*, que confere resistência a metotrexato (Bourouis 1983).

5. Produção e Caracterização de Plantas Estavelmente Transformadas

As células de planta transgênica são então colocadas em um meio seletivo apropriado para a seleção de células transgênicas, que são então cultivadas em calos. Brotos são cultivados de calos. Plântulas são geradas a partir do broto por cultivo em meio de formação de raiz. As várias construções normalmente serão unidas a um marcador para seleção em células de planta. Convenientemente, o marcador pode ser resistência a um biocida (particularmente um antibiótico, como canamicina, G418, bleomicina, hidromicina, cloranfenicol, herbicida, ou outro). O marcador particular usado irá permitir a seleção de células transformadas como comparado a células faltando o DNA, que foi introduzido. Componentes de construções de DNA incluindo cassetes de transcrição desta invenção podem ser preparados a partir das sequências, que são nativas (endógenas) ou estrangeiras (exógenas) para o hospedeiro. Por "estrangeira", significa-se que a sequência não é encontrada no hospedeiro de tipo selvagem em que a construção é introduzida. As construções heterólogas irão conter pelo menos uma região que não é nativa para o gene do qual a região de transcrição-iniciação é derivada.

Para confirmar a presença dos transgenes em células e plantas transgênicas, vários testes podem ser realizados. Estes testes incluem, por exemplo, testes "biológicos moleculares" bem conhecidos dos versados na técnica, como Southern e Northern blotting, hibridização *in situ* e métodos de

amplificação com base em ácido nucleico como PCR ou RT-PCR ou TaqMan; testes "bioquímicos", como detectando a presença de produto de proteína, por exemplo, por meios imunológicos (ELISAs e Western blots) ou por função enzimática; testes de partes de plantas, como testes de sementes; e
5 também, por análise do fenótipo da planta completa regenerada, por exemplo, para resistência a doença ou pragas.

DNA pode ser isolado das linhagens de células ou partes de planta para determinar a presença de segmento de ácido nucleico pré-selecionado através de uso de técnicas bem conhecidas do versado na técnica.

10 Nota-se que as sequências intactas sem sempre estarão presentes, provavelmente devido a rearranjo ou deleção de sequências na célula.

A presença de elementos de ácido nucleico introduzidos através de métodos desta invenção pode ser determinada por reação de cadeia polimerase (PCR). Usando esta técnica, fragmentos discretos de ácido
15 nucleico são amplificados e detectados por eletroforese de gel. Este tipo de análise permite determinar se um segmento de ácido nucleico pré-selecionado está presente no transformante estável, mas não prova a integração do segmento de ácido nucleico pré-selecionado introduzido no genoma da célula hospedeira. Além disso, não é possível usando técnicas PCR determinar se os
20 transformantes tem genes exógenos introduzidos em diferente sítios no genoma, *isto é*, se os transformantes são de uma origem independente. Contempla-se que usando técnicas PCR seria possível clonar fragmentos do DNA genômico hospedeiro adjacente a um segmento de DNA pré-selecionado introduzido.

25 A prova positiva de integração de DNA no genoma do hospedeiro e as identidades independentes de transformantes podem ser determinadas usando a técnica de hibridização Southern. Usando esta técnica, as sequências de DNA específicas que foram introduzidas no genoma do hospedeiro e flanqueando as sequências de DNA hospedeiras podem ser

identificadas. Assim o padrão de hibridização Southern do dado transformante serve como uma característica de identificação do transformante. Além disso, é possível, através de hibridização Southern, demonstrar a presença dos segmentos de DNA pré-selecionados introduzidos em DNA de alto peso molecular, *isto é*, confirmar que o segmento de DNA pré-selecionado introduzido foi integrado no genoma da célula hospedeira. A técnica de hibridização Southern provê informação que é obtida usando PCR, por exemplo, a presença do segmento de DNA pré-selecionado, mas também demonstra a integração no genoma e caracteriza cada transformante individual.

Contempla-se que usando as técnicas de hibridização tipo dot ou slot blot que são modificações de técnicas de hibridização Southern, pode-se obter a mesma informação que é derivada de PCR, por exemplo, a presença do segmento de DNA pré-selecionado.

Ambas as técnicas de hibridização PCR e Southern podem ser usadas para demonstrar a transmissão do segmento de DNA pré-selecionado para a progênie, na maior parte dos casos, o padrão de hibridização Southern característico para um dado transformante irá segregar em progênie como um ou mais genes Mendelianos (Spencer 1992); Laursen 1994) indicando a herança estável do gene. A natureza não quimérica dos calos e os transformantes parentais (R_0) têm sugerido uma transmissão da linhagem germinal e os padrões idênticos de hibridização Southern blot e intensidades da transformação do DNA em calos, plantas R_0 e progênie R_1 que segregaram para o gene transformado.

Apesar de técnicas de análise de DNA poderem ser conduzidas usando DNA isolado de qualquer parte da planta, RNA pode somente ser expressado em tipos particulares de células ou tecido e assim será necessário preparar o RNA para análise a partir destes tecidos. As técnicas PCR também podem ser usadas para detecção e quantização de RNA produzido a partir de

segmentos de DNA pré-selecionados. Nesta aplicação de PCR é primeiro necessário transcrever de modo reverso o RNA em DNA, usando enzimas como transcriptase reversa, e então através de uso de técnicas PCR convencionais amplificar o DNA. Na maior parte dos casos de técnicas PCR, 5 apesar de utilizáveis, elas não irão demonstrar a integridade do produto de RNA. Outra informação sobre a natureza de um produto de RNA pode ser obtida por Northern blotting. Esta técnica irá demonstrar a presença de uma espécie de RNA e dar informação sobre a integridade do RNA. A presença ou ausência de uma espécie de RNA também pode ser determinada usando 10 hibridizações Northern de tipo dot ou slot blot. Estas técnicas são modificações de Northern blotting e irão somente demonstrar a presença ou ausência de espécie de RNA.

Apesar de Southern blotting e PCR poderem ser usados para detectar o segmento de DNA pré-selecionado em questão, eles não fornecem 15 informação de como o segmento de DNA pré-selecionado está sendo expressado. Expressão pode ser avaliada por especificamente identificando os produtos de proteína dos segmentos de DNA pré-selecionados introduzidos ou avaliando as mudanças fenotípicas ocasionadas por sua expressão.

Os testes para a produção e identificação de proteínas 20 específicas podem fazer uso de propriedades físico-químicas, estruturais, funcionais ou outras das proteínas. As propriedades físico-químicas e estruturais únicas permitem que as proteínas sejam separadas e identificadas por procedimentos eletroforéticos, como eletroforese de gel nativo ou desnaturante, ou focalização isoeletrica, ou por técnicas de cromatografia 25 como cromatografia de troca iônica ou exclusão de gel. As estruturas singulares de proteínas individuais oferecem oportunidades para uso de anticorpos específicos para detectar sua presença em formatos como teste ELISA. Combinações de abordagens podem ser empregadas com especificidade ainda maior como Western blotting em que os anticorpos são

usados para localizar produto de genes individuais que foram separados por técnicas eletroforéticas. As técnicas podem ser empregadas para confirmar absolutamente a identidade do produto de interesse como avaliação por sequenciamento do aminoácido após a purificação. Apesar destes estarem
5 dentre os mais comumente empregados, outros procedimentos também podem ser usados.

Os procedimentos de teste também podem ser usados para identificar a expressão de proteínas por sua funcionalidade, especialmente a capacidade de enzimas de catalisar reações químicas específicas envolvendo
10 substratos e produtos específicos. Estas reações podem ser seguidas ao prover e quantificar a perda de substratos ou a geração de produtos das reações para procedimentos físicos ou químicos. Os exemplos são como variadas como as enzimas a serem analisadas.

Com muita frequência a expressão de um produto de gene é
15 determinada por avaliação dos resultados fenotípicos de sua expressão. Estes testes também podem tomar muitas formas, incluindo mas não limitadas a análise da mudança em composição química, morfologia, ou propriedades fisiológicas da planta. As mudanças morfológicas podem incluir caules de maior estatura ou mais espessos. Com maior frequência, mudanças na
20 resposta das plantas ou partes de plantas aos tratamentos impostos são avaliadas sob condições controladas de modo cuidadoso chamadas biotestes. Duas ou mais gerações podem ser cultivadas para assegurar que a expressão preferida do tecido da característica fenotípica desejada sob condições de interesse seja estavelmente mantida e herdada.

25 6. Usos de Plantas transgênicas

Uma vez que o cassete de expressão da invenção foi transformado em uma espécie particular de planta, ele pode ser propagado na espécie ou movimentado em outras variedades da mesma espécie, particularmente incluindo variedades comerciais, usando técnicas de

cruzamento tradicionais. As plantas da invenção particularmente preferidas incluem as plantas agronomicamente utilizáveis ou culturas agronomicamente importantes, em particular as plantas de cereais listadas acima, em particular, se referidas plantas forem monocotiledôneas. As propriedades genéticas engenheiradas nas sementes e plantas transgênicas descritas acima são passadas adiante por reprodução sexual e podem assim ser mantidas e propagadas em plantas da progênie. A presente invenção também refere-se a célula, tecido, órgão semente ou parte de planta da planta transgênica obtida a partir de uma planta transgênica. Também estão incluídos na invenção os descendentes transgênicos da planta assim como as células, tecidos, órgãos, sementes e partes das plantas da planta transgênica obtida a partir dos descendentes.

Preferivelmente, o cassete de expressão em uma planta transgênica é sexualmente transmitido. Em uma forma de realização preferida, a sequência de codificação é sexualmente transmitida através de um ciclo sexual normal completo da planta R0 para a geração R1. Também preferido, o cassete de expressão é expressado nas células, tecidos, sementes ou planta de uma planta transgênica em que uma quantidade, que é diferente da quantidade nas células, tecidos, sementes ou planta da planta que somente difere no cassete de expressão, está ausente. As plantas transgênicas produzidas aqui são assim esperadas como sendo utilizáveis para uma ampla variedade de fins comerciais e de pesquisa. Plantas transgênicas podem ser criadas para uso em agricultura tradicional como possuindo traços benéficos para o agricultor (por exemplo, traços agronômicos como resistência ao déficit de água, resistência a pragas, ou rendimento aumentado), benéficas para o consumidor do grão coletado de uma planta (por exemplo, melhorado teor nutritivo em alimentos humanos ou ração animal; teor de vitamina, aminoácido, e antioxidante aumentados; a produção de anticorpos (imunização passiva e nutracêuticos), ou benéficas para o processador do

alimento (por exemplo, melhorados traços de processamento). Em tais usos, as plantas são geralmente cultivadas para uso de seu grão em alimentos humanos ou animais. Além disso, o uso de promotores específicos do embrião em plantas transgênicas pode proporcionar traços benéficos que estão
5 localizados nas sementes de plantas consumíveis (por animais e humanos) como não excluindo outros traços, como farinha, caroços, nozes ou óleo de semente. No entanto, outras partes das plantas, incluindo caules, cascas, partes vegetativas, e outros, também podem ter utilidade, incluindo o uso como parte d silagem para animais ou fins ornamentais. Com freqüência, os
10 constituintes químicos (por exemplo, óleos ou amidos) de milho e outras culturas são extraídos para uso em rações ou industrial e plantas transgênicas podem ser criadas que tem níveis melhorados ou modificados destes componentes.

Plantas transgênicas também podem encontrar uso na
15 fabricação comercial de proteínas ou outras moléculas, onde a molécula de interesse é extraída ou purificada de partes de plantas, sementes, e outros. As células ou tecido das plantas também podem ser cultivadas, cultivadas in vitro, ou fermentadas para a fabricação destas moléculas. As plantas transgênicas também podem ser usadas em programas de cruzamento
20 comerciais, ou podem ser cruzadas ou criadas em plantas de espécies de plantas relacionadas. As melhoras codificadas pelo cassete de expressão podem ser transferidas, por exemplo, das células de milho para células de outras espécies, por exemplo, por fusão do protoplasto.

As plantas transgênicas podem ter muitos usos em pesquisa ou
25 criação incluindo a criação de novas plantas mutantes através de mutagenese insercional, a fim de identificar mutantes benéficos que podem ser depois criados por mutação e seleção tradicionais. Um exemplo deve ser a introdução da sequência de DNA recombinante codificando um elemento transposável que pode ser usado para gerar variação genética. Os métodos da invenção

também podem ser usados para criar plantas tendo "sequências de assinatura" singulares ou outras sequências de marcador que podem ser usadas para identificar as linhagens ou variedades proprietárias.

Assim, as plantas transgênicas e sementes de acordo com a
5 invenção podem ser usadas em cruzamento de plantas, que visa o desenvolvimento de plantas com melhoradas propriedades conferidas pelo cassete de expressão, como tolerância a seca, doença, ou outros estresses. As várias etapas de cruzamento são caracterizadas por intervenção humana bem definida como a seleção das linhagens a serem cruzadas, dirigindo a
10 polinização das linhagens parentais, ou selecionando as plantas descendentes apropriadas. Dependendo das propriedades desejadas, diferentes medidas são tomadas. As relevantes técnicas são bem conhecidas na técnica e incluem mas não são limitadas a hibridização, cruzamento endogâmico, cruzamento retro-cruzado, cruzamento de múltiplas trilhas, mistura de variedades, hibridização
15 interespecífica, técnicas aneuplóides, etc. As técnicas de hibridização também incluem a esterilização de plantas para dar plantas estéreis masculinas ou femininas por meios mecânicos, químicos ou bioquímicos. A polinização cruzada de planta estéril masculina com pólen da linhagem diferente assegura que o genoma da planta estéril masculina mas fértil feminina irá obter
20 uniformemente propriedades de ambas as linhagens parentais. Assim, as sementes e plantas transgênicas de acordo com a invenção podem ser usadas para a criação de linhagens de plantas melhoradas que, por exemplo, aumenta a eficácia de métodos convencionais como tratamento com herbicida ou pesticida ou permitem dispensar os referidos métodos devido às suas
25 propriedades genéticas modificadas. Alternativamente novas culturas com melhorada tolerância a estresse podem ser obtidas que, devido ao "equipamento" genético otimizado, fornecem produto colhido de melhor qualidade do que os produtos que não foram capazes de tolerar condições de desenvolvimento adverso comparáveis.

Assim, a presente invenção refere-se em uma outra forma de realização ao uso da molécula de ácido nucleico do presente pedido para a expressão de um gene de interesse preferivelmente ou especificamente em células ou tecido embrionário.

5 Ainda, em outra forma de realização, a presente invenção refere-se ao uso da molécula de ácido nucleico da invenção para aumentar a transcrição de uma molécula de ácido nucleico em uma planta sob condições de estresse.

10 Em outra forma de realização, a presente invenção refere-se a um método para produzir uma planta com rendimento aumentado, e/ou tolerância ao estresse aumentada, e/ou qualidade nutricional aumentada, e/ou teor de óleo aumentado ou modificado de uma semente ou broto para a planta, em que o método compreende as etapas de

15 A) introduzir na planta a molécula de ácido nucleico da presente invenção, o cassete de expressão da presente invenção, ou o vetor de expressão da presente invenção, em que a molécula de ácido nucleico é operativamente ligada a pelo menos uma molécula de ácido nucleico cuja sequência é heterólogo em relação à referida primeira ou referida segunda sequência de ácido nucleico e é capaz de conferir à planta rendimento
20 aumentado, e/ou tolerância ao estresse aumentada, qualidade nutricional aumentada, e/ou teor de óleo aumentado ou modificado à planta; e

25 B) selecionar plantas transgênicas, em que as plantas tem rendimento aumentado e/ou tolerância ao estresse aumentada sob condições de estresse, e/ou qualidade nutricional e/ou teor de óleo aumentados aumentado ou modificado de uma semente ou um broto das plantas, como comparado com as plantas de tipo selvagem ou segregante nulo.

Sequências

SEQ ID NO: 1 Sequência de ácido nucleico codificando a sequência nucleotídica reguladora da transcrição de gene Cor78

de *Arabidopsis thaliana*

- 5
10
15
20
- SEQ ID NO: 2 Sequência de ácido nucleico de vetor binário de pBPSMM368
- SEQ ID NO: 3 Sequência de ácido nucleico codificando intron Os.BPSI.1
- SEQ ID NO: 4 Sequência de ácido nucleico codificando intron Zm.ubiquitina
- SEQ ID NO: 5 iniciador dianteiro promotor Cor78
5'-gcaagaatct caaacacgga gatctca-3'
- SEQ ID NO: 6 iniciador reverso de promotor Cor78
5'-atttgtgagt aaaacagagg agggctctca-3'
- SEQ ID NO: 7 iniciador BPSI.1-5'
5'-cccgggcaccctgcggagggttaagatccgatcacc-3'
- SEQ ID NO: 8 iniciador BPSI.1-3'
5'-cggaccggtacatcttgcattgtcatgtac-3'
- SEQ ID NO: 9 motivo CCAF/CCA1.01 localizado da posição 4-18 de t.cor78
gccgcaagaa tctca
- SEQ ID NOs: 9-138 como descrito em tabela 9
- SEQ ID NOs: 139-144 como descrito em Figura s 4a a 4c e 5

EXEMPLOS

E Métodos Gerais

25

Salvo indicado em contrário, os produtos químicos e reagentes nos Exemplos foram obtidos de Sigma Chemical Company (St. Louis, MO), endonucleases de restrição foram de New England Biolabs (Beverly, MA) ou Roche (Indianapolis, IN), oligonucleotídeos foram sintetizados por MWG Biotech Inc. (High Point, NC), e outras enzimas de modificação ou kits relacionados com os testes de biologia molecular e bioquímicos foram obtidos de Clontech (Palo Alto, CA), Pharmacia Biotech (Piscataway, NJ), Promega

Corporation (Madison, WI), ou Stratagene (La Jolla, CA). Materiais para os meios de cultura de células foram obtidos de Gibco/BRL (Gaithersburg, MD) ou DIFCO (Detroit, MI). As etapas de clonagem realizadas para os fins da presente invenção, como, por exemplo, clivagens de restrição, eletroforese de gel agarose, purificação de fragmentos de DNA, transferência de ácidos nucleicos para membranas de nitrocelulose e náilon, fragmento de DNA de ligação, transformação de células de *E. coli*, cultivo de bactérias, fagos de multiplicação e análise de sequência de DNA recombinante, são realizados como descrito por Sambrook (1989). O sequenciamento de moléculas de DNA recombinante é realizado usando sequenciador de DNA de fluorescência a laser seguindo o método of Sanger (Sanger 1977).

Para gerar plantas transgênicas, *Agrobacterium tumefaciens* (cepa C58C1 [pMP90]) é transformado com as várias construções de promotor:: vetor GUS (ver abaixo). As cepas de *Agrobacterium* resultantes são subsequentemente empregadas para obter plantas transgênicas. Para este fim uma colônia isolada transformada *Agrobacterium* é incubada em cultura 4 mL (Meio: meio YEB com 50 µg/mL Canamicina e 25 µg/mL Rifampicina) durante a noite a 28°C. Com esta cultura a 400 mL, cultura do mesmo meio é inoculada e incubada durante a noite (28 °C, 220 rpm). As bactérias são precipitadas por centrifugação (GSA-Rotor, 8000 U/min, 20 min) e o grânulo é recolocado em suspensão em meio de infiltração (1/2 MS-Meio; 0,5 g/L MES, pH 5,8; 50 g/L sacarose). A suspensão é colocada em uma caixa de planta (Duchefa) e 100 mL SILVET L-77 (Osi Special-ties Inc., Cat. P030196) são adicionados a uma final concentração de 0,02%. A caixa de planta com 8 a 12 plantas é colocada em um dessecador durante 10 a 15 min. Sob vácuo com subsequente ventilação espontânea (expansão). Este processo é repetido 2-3 vezes. A seguir todas as plantas são transferidas em vasos com solo úmido e cultivadas sob condições de tempo de dia longo (16 h luz temperatura do dia 22-24°C, temperatura da noite 19°C; 65% umidade

relativa). Sementes são coletadas após 6 semanas.

EXEMPLO 1. Construção de vetor

As construções de promotor cor78 de *Arabidopsis* foram feitas contendo íntron com (pBPSMM368; Figura 1) ou sem BPSI.1 (pBPSMM250) (Tabela 6). O íntron BPSI.1 foi substituído com íntron de ubiquitina de milho para o fim de comparação (pBPSMM346). Os íntrons conferindo (IME) melhora mediada por íntron como BPSI.1 e Zm.ubiquitina foram subclonados na região 5' não traduzida (UTR) do cassete de expressão.

Tabela 6. Construções quiméricas GUS

Vetor binário	Composição do cassete de expressão (promotor::íntron::gene repórter ::terminador)
pBPSMM250	At.promotor cor78::GUS::NOS3'
pBPSMM346	At.promotor cor78::Zm.ubiquitina íntron::GUS::NOS 3'
pBPSMM368	At.promotor cor78::BPSI.1 íntron::GUS::NOS3'

EXEMPLO 2. Transformação de planta monocotiledônea

A transformação de planta mediada por *Agrobacterium* usando técnicas padrões de transformação e regeneração também pode ser realizada para o fim de transformar as plantas de cultura (Gelvin 1995; Glick 1993).

A transformação de milho ou outras plantas monocotiledôneas pode ser realizada usando, por exemplo, uma técnica descrita em US 5,591,616.

A transformação of plantas usando bombardeio de partículas, absorção de DNA mediada por polietileno glicol, ou via a técnica de fibra de carbonato de silício é descrita, por exemplo, por Freeling & Walbot 1993.

EXEMPLO 3. Detecção da expressão do gene repórter

Para identificar as características do promotor e dos elementos essenciais do último, que ocasionam sua especificidade de tecido, é necessário colocar o próprio promotor e vários fragmentos do mesmo diante do que é conhecido como um gene repórter, o que permite a determinação da atividade de expressão. Um exemplo, que pode ser mencionado, é uma β -glucuronidase bacteriana (Jefferson 1987a). A atividade de β -glucuronidase pode ser

detectada *in-planta* por meio de um substrato cromogênico como ácido 5-bromo-4-cloro-3-indolil- β -D-glucurônico em uma coloração da atividade (Jefferson 1987b). Para estudar a especificidade do tecido, um tecido de planta é cortado, incrustado, colorido e analisado como descrito (por exemplo

5 Bäumlein 1991b).

Um segundo teste permite a determinação quantitativa da atividade GUS no tecido estudado. Para a determinação da atividade quantitativa, MUG (4-metilumbeliferil- β -D-glucuronídeo) é usado como substrato para β -glucuronidase, e o MUG é clivado em MU

10 (metilumbeliferona) e ácido glucurônico.

Para fazer isto, um extrato de proteína do tecido desejado é primeiro preparado e o substrato de GUS é então adicionado ao extrato. O substrato pode ser medido fluorimetricamente somente após o GUS ter sido reagido. As amostras que são subsequentemente medidas em um

15 fluorímetro são tomadas em vários pontos de tempo. Este teste pode ser realizado por exemplo com embriões de linhaça em vários estágios de desenvolvimento (21, 24 ou 30 dias após a florescência). Para esta finalidade, em cada caso, um embrião é triturado em um pó em um vaso de reação de 2 mL em nitrogênio líquido com a ajuda de um moinho de

20 trituração com vibração (Tipo : Retsch MM 2000). Após a adição de 100 μ L de tampão EGL (0,1 M KPO₄, pH 7,8; 1 mM EDTA; 5% glicerol; 1 M DTT), a mistura é centrifugada durante 10 minutos a 25°C e 14.000 x g. O sobrenadante é removido e recentrifugado. Novamente o sobrenadante é transferido para um novo vaso de reação e mantido em

25 gelo até outro uso. 25 μ L desta proteína extrato são tratados com 65 μ L de tampão EGL (sem DTT) e empregados no teste GUS. 10 μ L do substrato MUG (10 mM 4-metilumbeliferil- β -D-glucuronídeo) são agora adicionados e a mistura é submetida a vórtice, e 30 μ L são removidos imediatamente como valor zero e tratados com 470 μ L de tampão Stop

(0,2 M Na₂CO₃). Este procedimento é repetido para todas as amostras em um intervalo de 30 s. As amostras tomadas foram armazenadas em refrigerado até serem medidas. Outras leituras foram tomadas após 1 h e após 2 h. Uma série de calibração que continha concentrações de 0,1 mM a 10 mM MU (4-metilumbeliferona) foi estabelecida para a medida fluorimétrica. Se os valores de amostra estavam fora destas concentrações, menos extrato de proteína foi empregado (10 µL, 1 µL, 1 µL de uma diluição 1:10) e intervalos mais curtos foram medidos (0 h, 30 min, 1 h). A medida foi realizada em uma excitação de 365 nm e uma emissão de 445 nm em um aparelho Fluoroscan II (Labsystem). Como uma alternativa, o substrato de clivagem pode ser monitorado fluorimetricamente sob condições alcalinas (excitação a 365 nm, medida da emissão a 455 nm; Spectro Fluorimeter BMG Polarstar+) como descrito em Bustos (1989). Todas as amostras foram submetidas a uma determinação da concentração de proteína pelo método de Bradford (1976), assim permitindo uma identificação da atividade do promotor e resistência do promotor em vários tecidos e plantas.

EXEMPLO 4. Expressão específica de embrião em milho

A construção pBPSMM368 foi altamente expressada no embrião, especialmente em escutelo, enquanto a coloração no endosperma foi quase não detectável usando testes histoquímicos GUS (Figura 2). Esta expressão forte específica de embrião foi mantida durante a germinação. A construção pBPSMM368 não foi expressada em raízes ou folhas. No entanto esta expressão específica de embrião não foi detectada em combinação de At.promotor cor78 com construção de íntron Zm.ubiquitina (pBPSMM346). As plantas transgênicas de milho transformadas com pBPS346 mostram uma expressão constitutiva, mas no geral muito fraca. A construção pBPSMM250 não foi expressada em qualquer tecido analisado.

Tabela 7. Expressão GUS controlada por candidatos de promotor constitutivo

de monocotiledôneas

Tecidos / Estágios de desenvolvimento	Promotor (níveis expressão GUS)			
	pBPSMM232 ¹	pBPSMM250	pBPSMM346	pBPSMM368
3 dias após co-cultivo	+++++	ND	ND	-
Folhas em estágio de 5 folhas	+++++	-	-	-
Raízes em estágio de 5 folhas	+++++	-	-	-
Folhas em estágio de florescência	+++++	-	++	-
Caule	+++++	-	ND	-
Pre-polinização	+++++	-	ND	-
20 dias após polinização [DAP]				
Escutelo	+++++	-	ND	+++
Eixo do embrião	+++++	-	ND	+ ²
Endosperma	+++++	-	ND	-
30 DAP				
Escutelo	+++++	-	++++	++++
Eixo do embrião	+++++	-	++++	+ ²
Endosperma	+++++	-	+++	-
sementes secas				
Escutelo	++++	-	+++	+++
Eixo do embrião	++++	-	+++	+ ²
Endosperma	++++	-	+++	-
Embebição/germinação (72 h) ³				
Escutelo	+++++	-	++++	++++
Eixo do embrião	+++++	-	++++	+ ²
Endosperma	+++++	-	++++	-

¹ Controles positivos como um promotor constitutivo (pBPSMM232=Zm.ubiquitina promotor :: Zm.ubiquitina intron::GUS (PIV2)::NOS terminador; uma faixa de níveis de expressão GUS medidos por teste histoquímico (- a +++++), ND: não determinado ainda,

² Plantas transgênicas contendo pBPSMM368 mostrara expressão fraca na primeira área de nódulo intermodular do eixo do embrião, mas quase nenhuma expressão em outras regiões do eixo do embrião, que compreendia a raiz primária, plúmula, caule, folhas, e coleóptilos.

³ Três dias (72 h) após embebição (germinação), as plantas transgênicas contendo pBPSMM346 mostraram expressão em radícula e na semente completa, mas as contendo pBPSMM368 mostraram expressão limitada principalmente em escutelo.

EXEMPLO 5. Expressão indutível por seca

As plantas de milho transgênicas contendo pBPSMM250 (sem intron conferindo melhora mediada por íntron) não mostram a expressão indutível sob o estresse de seca (em estágio de 5 folhas, 4, 8, 13 dias após com água de manutenção). A fim de melhorar a expressão de transgene, intron Zm.ubiquitina foi adicionado no 5'UTR de construção de At.promotor cor78 para gerar a construção pBPSMM346. As plantas de milho transgênicas contendo pBPSMM346 em estágio de 5 folhas foram usadas para estresse de seca. A expressão indutível de seca foi detectada em plantas T0 usando RT-PCR quantitativo. Com base neste teste sensível, uma indução de aproximadamente 3,5 vezes foi detectada sob o estresse de seca. A indução

foi de até 3,5 vezes como determinado com PCR quantitativo (Figura 3).

EXEMPLO 6. Utilização de culturas transgênicas

Um gene repórter em pBPSMM368 pode ser substituído com um gene de interesse para expressar em um modo específico do embrião e conferir tolerância aos estresses ambientais bióticos e abióticos. As construções quiméricas são transformadas em plantas monocotiledôneas. Métodos padrões para transformação na técnica podem ser usados se requerido. As plantas transformadas são regeneradas usando métodos conhecidos. Vários fenótipos são medidos para determinar a melhora da biomassa, rendimento, composição de ácido graxo, elevado teor de óleo, tolerância a doença, ou quaisquer outros fenótipos que ligam a melhora ou estabilidade do rendimento. Os níveis de expressão de genes são determinados em diferentes estágios de desenvolvimento e em diferentes gerações (plantas T₀ a T₂ ou outras gerações). Resultados da avaliação em plantas leva a determinar os genes apropriados em combinação com este promotor para aumentar o rendimento, melhorar a tolerância a doença, e/ou melhorar a tolerância a estresse abiótico.

EXEMPLO 7. Expressão de gene de marcador selecionável em plantas monocotiledôneas

Um gene repórter em pBPSMM368 pode ser substituído com um gene de marcador selecionável e transformado em plantas monocotiledôneas como milho, trigo, arroz, centeio, centeio, milhete, sorgo, erva-castelhana ou coix, tritcale, cana de açúcar, ou aveias, mas não é limitado a estas espécies de plantas. Os métodos padrões para transformação na técnica podem ser usados se requerido. As plantas transformadas são selecionadas sob o agente de seleção de interesse e regeneradas usando métodos conhecidos. O esquema de seleção é examinado em estágios de desenvolvimento prematuro de tecidos ou cultura de células de tecido. Os níveis de expressão de genes podem ser determinados a diferentes estágios de

desenvolvimento e em diferentes gerações (plantas T_0 a T_2 ou outras gerações). Resultados da avaliação em plantas levam a determinar os genes apropriados em combinação com este promotor.

EXEMPLO 8. Análise de Deleção

5 O método de clonagem é descrito por Rouster (1997) e Sambrook (1989). O mapeamento detalhado destes promotores (*isto é*, limitando os segmentos de ácido nucleico relevantes para esta especificidade) é realizado por geração de vários vetores de expressão de gene repórter que primeiramente continha a região de promotor completa e segundo vários
10 fragmentos da mesma. Primeiro, a região de promotor completa ou fragmentos da mesma são clonados em um vetor binário contendo GUS ou outro repórter gene. Para este fim, fragmentos são empregados primeiro, que são obtidos usando enzimas de restrição para os sítios de clivagem de restrição internos na sequência de promotor de comprimento completo.
15 Segundo, os fragmentos de PCR são empregados que são providos com sítios de clivagem introduzidos por iniciadores. As construções GUS quiméricas contendo vários promotores deletados são transformadas em *Zea mays*, *Arabidopsis* e outras espécies de plantas usando métodos de transformação na técnica corrente. A atividade do promotor é analisada usando testes
20 histoquímicos GUS ou outros métodos apropriados em vários tecidos e órgãos em diferentes estágios de desenvolvimento. Modificação de uma sequência de promotores pode eliminar o vazamento com base nas necessidades.

EXEMPLO 9. Mutagenese *In vivo*

25 O versado é familiar com vários métodos para a modificação da atividade do promotor ou identificação de importantes elementos do promotor. Um destes métodos é baseado na mutação aleatória seguido por teste com os genes repórteres como descrito acima. A mutagenese *in vivo* de microorganismos pode ser obtida por passagem do DNA plasmídeo (ou de outro vetor) através de *E. coli* ou outros microorganismos (por exemplo

Bacillus spp. ou leveduras como *Saccharomyces cerevisiae*) em que a capacidade de manter a integridade da informação genética é rompida. As cepas mutator convencionais tem mutações nos genes para o sistema de reparo de DNA (por exemplo mutHLS, mutD, mutT e outros para referência, ver Rupp 1996). O versado está familiarizado com estas cepas. O uso destas cepas é ilustrado por exemplo por Greener (1994). A transferência de moléculas de DNA mudadas em plantas é preferivelmente efetuada após a seleção e teste dos microorganismos. Plantas transgênicas são geradas e analisadas como descrito acima.

10 EXEMPLO 10. Construção de vetor para superexpressão e experimentos de "nockout" de genes

10.1 Superexpressão

Vetores usados para a expressão de " genes candidatos " de comprimento completo de interesse em plantas (superexpressão) são projetados para superexpressar a proteína de interesse e são de dois tipos gerais, biolísticos e binários, dependendo do método de transformação de planta a ser usado.

Para transformação biolística (vetores biolísticos) as exigências são as seguintes:

1. uma estrutura dorsal com um marcador selecionável bacteriano (tipicamente, um gene resistência a antibiótico) e origem de replicação funcional em *Escherichia coli* (*E. coli* ; por exemplo, ColE1), e
2. uma porção específica de planta consistindo de:
 - a. um cassete de expressão de gene consistindo do promotor (por exemplo ZmUBIint MOD), o gene de interesse (tipicamente, cDNA de comprimento completo) e um terminador transcripcional (por exemplo, terminador nos *Agrobacterium tumefaciens*);
 - b. um cassete marcador de planta selecionável, consistindo do promotor apropriado, gene de marcador selecionável (por exemplo, D-

aminoácido oxidase; *dao1*) e terminador transcripcional (por exemplo terminador nos).

Os vetores projetados para transformação por *Agrobacterium tumefaciens* (*A. tumefaciens*; vetores binários) consistem de:

- 5 1. uma estrutura dorsal com um marcador selecionável bacteriano funcional em tanto *E. coli* como em *A. tumefaciens* (por exemplo, resistência a espectinomicina mediada pelo gene *aadA*) e duas origens de replicação, funcionais em cada um dos hospedeiros bacterianos mencionados, mais o gene *virG* de *A. tumefaciens*;
- 10 2. uma porção específica de planta, como descrito para vetores biolísticos acima, exceto neste caso esta porção é flanqueada por sequências de borda à direita e à esquerda de *A. tumefaciens* que mediam a transferência do DNA flanqueado para estas duas sequências para a planta.

10.2 Vetores de silenciamento de genes

- 15 Vetores projetados para reduzir ou abolir a expressão do gene único ou da família ou genes relacionados (vetores de silenciamento de genes) são também de dois tipos gerais correspondendo à metodologia usada para regular de modo negativo a expressão de gene: RNA interferência anti-sentido ou de filamento duplo (dsRNAi).

20 (a) Anti-sentido

- 25 Para vetores anti-sentido vetores, um fragmento de gene de comprimento completo ou parcial (tipicamente, uma porção do cDNA) pode ser usado nos mesmos vetores descritos para a expressão de comprimento completo, como parte do cassete de expressão do gene. Para a regulação negativa mediada por anti-sentido de expressão de genes, a região de codificação do gene ou fragmento do gene será na orientação oposta com relação ao promotor; assim, mRNA será tornado um filamento de não codificação (anti-sentido) em planta.

(b) dsRNAi

Para vetores dsRNAi, um fragmento de gene parcial (tipicamente, 300 a 500 pares de base de comprimento) é usado no gene cassete de expressão, e é expressado em ambas as orientações sentido e anti-sentido, separadas pela região de espaçador (tipicamente, um intron de planta, por exemplo intron 1 OsSH1, ou um marcador selecionável, por exemplo conferindo resistência a canamicina). Vetores deste tipo são projetados para formar um tronco mRNA de filamento duplo, resultante da formação de pares de base dos dois fragmentos de gene complementares *in planta*.

Os vetores biolísticos ou binários projetados para super-expressão ou nocaute podem variar em um número de diferentes modos, incluindo por exemplo os marcadores selecionáveis usados em planta e bactérias, os terminadores de transcrição usados em expressão de genes e os cassetes de marcador selecionável de plantas, e as metodologias usadas para clonagem em gene ou fragmentos de genes de interesse (tipicamente, clonagem à base em recombinase GatewayTM mediada por enzima de restrição).

EXEMPLO 11. Análise do sítio de ligação do fator de transcrição putativa para At.Cor78 promotor (SEQ ID NO: 1)

Com base nos resultados Genomatix dados abaixo, a caixa TATA em potencial pode ser localizada a partir do par de bases 790 a 804 de SEQ ID NO: 1.

Tabela 8 Agrupamentos de elementos de promotor identificados no At.promotor cor78 como descrito por SEQ ID NO: 1.

Familia/ Matriz	Outra informação	Opt.	Posição	Fil.	Nuc sim	Matrizsim.	Sequência (Red valor ci- > 60 maiúsculas: sequência de núcleo)
			de - para				
P\$AGP1/AGP1.01	proteína ligação AG motivo1	0,91	18 - 28	(+)	1,000	0,913	ggaGATCtcaa
P\$MYBL/MYBPH3,02	proteína de Petunia hibrida tipo Myb	0,76	55 - 71	(-)	0,817	0,847	gtaagtTTGTtttgagt
P\$MADS/AGL15.01	AGL15, Arabidopsis MADS-domínio proteína- tipo AGAMOUS 15	0,79	66 - 86	(+)	0,775	0,792	actTACGaaatttagtagaa
P\$MADS/AGL15.01	AGL15, Arabidopsis MADS-domínio proteína- tipo AGAMOUS 15	0,79	89 - 109	(-)	0,775	0,811	aatTACAatataatgtatata

Família/ Matriz	Outra informação	Opt.	Posição de - para	Fil.	Nuc sim	Matrizsim.	Sequência (Red valor ci- > 60 maiúsculas: sequência de núcleo)
P\$MADS/AGL15.01	AGL15, Arabidopsis MADS-domínio proteína -tipo AGAMOUS 15	0,79	90 - 110	(+)	0,775	0,823	ataTACAttatattgtaattt
P\$MYBL/MYBPH3.02	proteína de Petunia híbrida tipo Myb	0,76	110 - 126	(-)	0,817	0,799	aacattTTGTtacaaaa
P\$SEF4/SEF4.01	Soja fator embrião 4	0,98	123 - 133	(+)	1,000	0,997	tgTTTtatta
P\$AHBP/HAHB4.01	Girassol homeodomínio leucina-zipper proteína Hahb-4	0,87	127 - 137	(+)	1,000	0,923	tttattATTAt
P\$AHBP/HAHB4.01	Girassol homeodomínio leucina-zipper proteína Hahb-4	0,87	130 - 140	(+)	1,000	0,923	attattATTAt
P\$GTBX/SBF1.01	SBF-1	0,87	147 - 163	(+)	1,000	0,894	ttactggTTAAattaaa
P\$GAPB/GAP.01	Cis-elemento nos promotores GAPDH conferindo indutibilidade à luz	0,88	161 - 175	(+)	1,000	0,950	aaaaATGAatagaaa
P\$DOFF/PBOX.01	Caixa prolamina, conservada em cereal promotores de gene de proteína de armazenamento de semente	0,75	166 - 182	(+)	1,000	0,789	tgaatagaAAAAGgtgaa
P\$IBOX/IBOX.01	I-Caixa em genes rbcS e outros genes regulados em luz	0,81	167 - 183	(+)	0,750	0,817	gaataGAAAaggtgaat
P\$NCSI/NCSI.01	sequência de consenso 1 de Nodulina	0,85	172 - 182	(+)	1,000	0,856	gAAAAaggtgaa
P\$OCSE/OCSE.01	elementos tipo OCS	0,69	195 - 215	(-)	0,769	0,704	agaagaaaatgttACCTcct
P\$NCSI/NCSI.01	sequência de consenso 1 Nodulina	0,85	200 - 210	(-)	0,878	0,853	aAAATgtttac
P\$MADS/SQUA.01	MADS-caixa proteína SQUAMOSA	0,90	209 - 229	(+)	1,000	0,911	ttctctATTtttcatattt
P\$GAPB/GAP.01	Cis-elemento nos promotores GAPDH conferindo indutibilidade de luz	0,88	215 - 229	(-)	1,000	0,892	aaatATGAaaaaata
P\$MYBS/MYBST1.01	MybSt1 (Myb Solanum tuberosum 1) com uma única repetição de myb	0,90	226 - 242	(-)	1,000	0,945	taatttATCCtgaaat
P\$IBOX/GATA.01	fatores GATA Classe I	0,93	229 - 245	(+)	1,000	0,956	ttcagGATAaattattg
P\$AHBP/ATHB1.01	homeo caixa proteína 1 Arabidopsis thaliana	0,90	236 - 246	(+)	1,000	0,989	taaATTAttgt
P\$AHBP/ATHB5.01	proteína HDZip classe I ATHB5	0,89	236 - 246	(-)	0,829	0,940	acaATAAtta
P\$GTBX/GT3A.01	Trihélise fator de ligação de DNA GT-3a	0,83	239 - 255	(-)	0,750	0,839	taaactTTTAcataat
P\$NCSI/NCSI.01	sequência de consenso 1 Nodulina	0,85	246 - 256	(+)	1,000	0,898	tAAAAgtttac
P\$PSRE/GAAA.01	motivo GAAA envolvido em pólen ativação transcricional	0,83	253 - 269	(-)	1,000	0,836	aatgGAAAtctgttaa
P\$GTBX/S1F.01	S1F, fator de sítio de ligação 1 de rps1 promotor de espinafre	0,79	255 - 271	(-)	1,000	0,794	tcaaATGGaaacttgt
P\$WBXF/WRKY.01	fator associado com defesa de patógeno de tipo dedo de zinco específico de planta WRKY, caixa W	0,92	263 - 279	(+)	1,000	0,953	tccatTTGactagtgtta
P\$MYBS/TAMYB80.01	proteína MYB de trigo	0,83	279 - 295	(-)	1,000	0,980	gagaATATtctctatt

Família/ Matriz	Outra informação	Opt.	Posição de - para	Fil.	Nuc sim	Matrizsim.	Sequência (Red valor ci- > 60 maiúsculas: sequência de núcleo)
P\$MYBS/TAMYB80.01	proteína MYB de trigo	0,83	284 - 300	(+)	1,000	0,941	aggaATATtctctagta
P\$AHBP/HAHB4.01	Girassol homeodominio leucina-zipper proteína Hahb-4	0,87	300 - 310	(+)	1,000	0,920	aagatcATTAt
P\$LREM/ATCTA.01	Motivo envolvido em biossíntese de carotenóide e tocoferol e na expressão de genes relacionados com a fotossíntese	0,85	312 - 322	(+)	1,000	0,859	tcATCTacttc
P\$IBOX/GATA.01	Fatores GATA Classe I	0,93	318 - 334	(-)	1,000	0,975	tagaaGATAaaagaagt
P\$HEAT/HSE.01	elemento choque térmico	0,81	319 - 333	(-)	1,000	0,858	agaagataaaAGAAg
P\$MADS/AGL1.01	AGL1, Arabidopsis MADS-domínio proteína -tipo AGAMOUS 1	0,84	329 - 349	(-)	0,975	0,864	ttaTTCCtctactgtagaag
P\$MADS/AGL1.01	AGL1, Arabidopsis MADS-domínio proteína -tipo AGAMOUS 1	0,84	330 - 350	(+)	0,995	0,867	ttcTACCtagtaggaataa a
P\$MYBS/TAMYB80.01	proteína MYB de trigo	0,83	337 - 353	(-)	0,750	0,831	ttgtTTATtctctact
P\$DOFF/PBOX.01	caixa prolamina conservada em promotores de cereal de gene de proteína de armazenamento de semente	0,75	362 - 378	(+)	0,776	0,796	tcctttgtAAATacaaa
P\$GTBX/SBF1.01	SBF-1	0,87	402 - 418	(-)	1,000	0,897	cgtaaaaTTAAaattga
P\$NACF/TANAC69.01	fator ligação DNA domínio NAC de trigo	0,68	409 - 431	(-)	1,000	0,723	cttttttttaTACGtaaaatta
P\$HMGF/HMG_IY.01	grupo alta mobilidade proteínas tipo I/Y	0,89	417 - 431	(-)	1,000	0,944	ctttTATTttatagc
P\$NCSI/NCSI.01	sequência de consenso 1 Nodulina	0,85	426 - 436	(+)	1,000	0,969	tAAAAgatcat
P\$AHBP/WUS.01	Homeodominio proteína WUSCHEL	0,94	448 - 458	(-)	1,000	0,963	ctcctTAATcg
P\$GTBX/GT1.01	fatores ligação GT1- Caixa com a trihélice domínio ligação DNA	0,85	520 - 536	(-)	0,843	0,881	catgtgGTCAttttacg
P\$TCPF/ATTCP20.01	fator transcrição TCP classe I (Arabidopsis)	0,94	534 - 546	(-)	1,000	0,956	attgGCCCatcat
P\$DREB/CRT_DRE.01	Elemento resposta repetição C- desidratação	0,89	551 - 565	(+)	1,000	0,913	atggaCCGActacta
P\$MADS/AGL15.01	AGL15, Arabidopsis MADS-domínio proteína AGAMOUS-tipo 15	0,79	557 - 577	(-)	1,000	0,887	actTACTattattagtagtcg
P\$MADS/AGL15.01	AGL15, Arabidopsis MADS-domínio proteína AGAMOUS-tipo 15	0,79	558 - 578	(+)	1,000	0,896	gacTACTaataatagtaagt t
P\$AHBP/HAHB4.01	Girassol homeodominio leucina-zipper proteína Hahb-4	0,87	563 - 573	(-)	1,000	0,941	actattATTAg
P\$SPF1/SP8BF.01	DNA-proteína ligação de batata doce que liga a seq. SP8a (ACTGTGTA) e SP8b (TACTATT) de esporamina e genes beta- amilase	0,87	564 - 576	(-)	1,000	0,919	ctTACTattatta
P\$GTBX/GT3A.01	fator ligação DNA GT-3a Trihélice	0,83	570 - 586	(+)	1,000	0,889	tagtaaGTTAcatttta
P\$EINL/TEIL.01	TEIL (tabaco tipo EIN3-)	0,92	574 - 582	(-)	1,000	0,933	aTGTAactt
P\$GTBX/SIF.01	SIF, fator ligação sítio 1 de promotor rps1 de espinafre	0,79	585 - 601	(+)	1,000	0,792	taggATGGAataaatat

Família/ Matriz	Outra informação	Opt.	Posição	Fil.	Nuc sim	Matrizsim.	Sequência (Red valor ci- > 60 maiúsculas: sequência de núcleo)
			de - para				
P\$DREB/CRT_DRE.01	elemento de resposta repetição C - desidratação	0,89	601 - 615	(+)	1,000	0,940	tcataCCGAcacgag
P\$DOFF/PBF.01	PBF (MPBF)	0,97	617 - 633	(+)	1,000	0,979	ttgaaagaAAAGggaaa
P\$TEFB/TEF1.01	TEF elementos cis atuantes em tanto promotores dependentes de RNA polimerase II- como sequências espaçadoras de rDNA	0,76	624 - 644	(+)	0,956	0,817	aaAAGGgaaaaaagaaa aaa
P\$NCS1/NCS1.01	sequência de consenso 1 Nodulina	0,85	649 - 659	(+)	1,000	0,909	tAAAAgatata
P\$DREB/CRT_DRE.01	elemento de resposta repetição C - desidratação	0,89	658 - 672	(+)	1,000	0,986	tactaCCGAcacgag
P\$DOFF/DOF3.01	Dof3 - fator de transcrição dedo de zinco único	0,99	671 - 687	(+)	1,000	0,996	agttccaaAAAGcaaaa
P\$NCS1/NCS1.01	sequência de consenso 1 Nodulina	0,85	688 - 698	(+)	1,000	0,966	aAAAAgatcaa
P\$DREB/CRT_DRE.01	elemento de resposta repetição C - desidratação	0,89	695 - 709	(+)	1,000	0,903	tcaagCCGAcacaga
P\$CE3S/CE3.01	elemento copulação 3 (CE3), não-ACGT ABRE	0,77	703 - 721	(-)	1,000	0,853	ttcttaCGCGtgctgtgt
P\$CGCG/ATSR1.01	Arabidopsis thaliana gene 1 respondente a sinal 1, Ca2+/ homólogo de proteína ligação calmodulina para NtER1 (gene respondente a etileno de tabaco prematuro)	0,84	708 - 718	(-)	1,000	0,919	ctaCGCGgtgc
P\$CGCG/ATSR1.01	Arabidopsis thaliana gene 1 respondente a sinal 1, homólogo de proteína ligação calmodulina para NtER1 (gene respondente a etileno de tabaco prematuro)	0,84	709 - 719	(+)	1,000	0,878	acaCGCGtaga
P\$GBOX/TGA1.01	Arabidopsis leucina zipper proteína TGA1	0,90	727 - 747	(-)	1,000	0,982	gtgggtTGACgtcaagtcac
P\$GBOX/TGA1.01	Arabidopsis leucina zipper proteína TGA1	0,90	728 - 748	(+)	1,000	0,961	tgactTGACgtcacaccacg
P\$WBXF/WRKY.01	WRKY fator tipo dedo de zinco específico de planta associado com defesa de patógeno, caixa W	0,92	728 - 744	(+)	1,000	0,934	tgactTTGAcgtcacac
P\$OPAQ/RITA1.01	Ativador 1 transcrição de arroz (RITA), proteína zipper leucina básica, altamente expressada durante o desenvolvimento da semente	0,95	730 - 746	(+)	1,000	0,958	actttgACGTcacacca
P\$SALT/ALFIN1.01	proteína dedo de zinco em raízes alfafa, regula tolerância a sal	0,93	738 - 752	(-)	1,000	0,948	ttttcGTGGgtgtgac
P\$GBOX/GBF1.01	fator ligação proteína bZIP G-Caixa	0,94	757 - 777	(+)	1,000	0,968	cgcttcacACGTgtccctta
P\$ABRE/ABRE.01	elementos resposta ABA	0,82	758 - 774	(+)	1,000	0,971	gcttcacACGTgtccct
P\$OCSE/OCSE.01	elementos tipo OCS	0,69	762 - 782	(-)	1,000	0,778	agagataaaggacACGTatg
P\$GAGA/GAGABP.01	proteínas de ligação(GA)n/(CT)n (GBP, soja; BBR, centeio)	0,75	764 - 788	(-)	1,000	0,797	actgagAGAGataaagg acacgta
P\$IBOX/GATA.01	Fatores GATA classe I	0,93	768 - 784	(-)	1,000	0,939	agagaGATAaagggaca
P\$GAGA/GAGABP.01	proteínas de ligação(GA)n/(CT)n (GBP, soja; BBR, centeio)	0,75	772 - 796	(-)	0,750	0,785	atagagAGACTgagagag ataaagg

Família/ Matriz	Outra informação	Opt.	Posição	Fil.	Nuc sim	Matrizsim.	Sequência (Red valor ci- > 60 maiúsculas: sequência de núcleo)
			de - para				
P\$GAGA/GAGABP.01	proteínas de ligação(GA)n/(CT)n (GBP, soja; BBR, centeio)	0,75	774 - 798	(-)	1,000	0,786	ttatagAGAGactgagagagataaa
P\$GAGA/GAGABP.01	proteínas de ligação(GA)n/(CT)n (GBP, soja; BBR, centeio)	0,75	776 - 800	(-)	1,000	0,782	gtttatAGAGactgagagagata
P\$TBP/TATA.01	Caixa TATA de planta	0,88	790 - 804	(+)	1,000	0,909	ttctTATAaacttag

Tabela 9 Motivos de promotor e motivos de promotor de núcleo identificados em At.cor78 (SEQ ID NO: 1) com SEQ ID NOS

Nome do motivo do promotor	Posição a At.cor78	Sequência	Filamento	SEQ ID NO:
AGPI/AGPI.01 motivo	18-28	ggaGATCtaa	(+)	9
AGPI/AGPI.01 motivo de núcleo		nnngatcnnn n		10
MYBL/MYBPH3,02 motivo1	55-71	gtaagtTTGTttgagt	(-)	11
MYBL/MYBPH3,02 motivo2	110-126	aacattTTGTtcaaaa	(-)	18
MYBL/MYBPH3,02 motivo de núcleo		nnnnnttgt nnnnnnn		12
MADS/AGL15.01 motivo1	66-86	actTACGaaatttagtagaa	(+)	13
MADS/AGL15.01 motivo de núcleo1		nnntacgnnn nnnnnnnnnn n		14
MADS/AGL15.01 motivo2	89-109	aatTACAatataatgtatata	(-)	15
MADS/AGL15.01 motivo3	90-110	ataTACAttatattgtaattt	(+)	16
MADS/AGL15.01 motivo de núcleo2		nnntacannn nnnnnnnnnn n		17
MADS/AGL15.01 motivo4	557-577	actTACTattatttagtagtcg	(-)	89
MADS/AGL15.01 motivo de núcleo4		nnntactnnn nnnnnnnnnn n		90
MADS/AGL15.01 motivo5	558-578	gacTACTaataatagtaagtt	(+)	91
SEF4/SEF4.01 motivo	123-133	tgTTTTtatta	(+)	19
SEF4/SEF4.01 motivo de núcleo		nnntttnnnn n		20
AHBP/HAHB4.01 motivo1	127-137	tttattATTAt	(+)	21
AHBP/HAHB4.01 motivo de núcleo		nnnnnnatta n		22
AHBP/HAHB4.01 motivo2	130-140	attattATTAt	(+)	23
AHBP/HAHB4.01 motivo3	300-310	aagatcATTAt	(+)	61
AHBP/HAHB4.01 motivo4	563-573	actattATTAg	(-)	92
GTBX/SBF1.01 motivo1	147-163	ttactggTTAAattaaa	(+)	24
GTBX/SBF1.01 motivo de núcleo		nnnnnnntta annnnnn		25
GTBX/SBF1.01 motivo2	402-418	cgtaaaaTTAAaattga	(-)	75
GAPB/GAP.01 motivo1	161-175	aaaaATGAatagaaa	(+)	26
GAPB/GAP.01 motivo de núcleo		nnnnatgann nnnnn		27
GAPB/GAP.01 motivo2	215-229	aaatATGAaaaaata	(-)	40
DOFF/PBOX.01 motivo1	166-182	tgaatagaAAAGgtgaa	(+)	28
DOFF/PBOX.01 motivo de núcleo1		nnnnnnnnaa agnnnnn		29
DOFF/PBOX.01 motivo2	362-378	tcctttgtAAATacaaa	(+)	73
DOFF/PBOX.01 motivo de núcleo2		nnnnnnnnaa atnnnnn		74
IBOX/IBOX.01 motivo	167-183	gaataGAAAaggtgaat	(+)	30
IBOX/IBOX.01 motivo de núcleo		nnnnngaaan nnnnnnn		31
NCSI/NCSI.01 motivo6	172-182	gAAAAGgtgaa	(+)	32
NCSI/NCSI.01 motivo de núcleo1		naaaannnnn n		33
NCSI/NCSI.01 motivo1	246-256	tAAAAGtttac	(+)	51
NCSI/NCSI.01 motivo3	426-436	tAAAAGatcat	(+)	80
NCSI/NCSI.01 motivo4	649-659	tAAAAGatata	(+)	105
NCSI/NCSI.01 motivo5	688-698	aAAAAGatcaa	(+)	109
NCSI/NCSI.01 motivo2	200-210	aAAATgtttac	(-)	36
NCSI/NCSI.01 motivo de núcleo2		naaatnnnnn n		37
OCSE/OCSL.01 motivo1	195-215	agaagaaaatgtttACCTcct	(-)	34
OCSE/OCSL.01 motivo de núcleo1		nnnnnnnnnn nnnnacctnn n		35
OCSE/OCSL.01 motivo2	762-782	agagataaaggacACGTatg	(-)	128
OCSE/OCSL.01 motivo de núcleo2		nnnnnnnnnn nnnnacgtnn n		129
MADS/SQUA.01 motivo	209-229	ttcttctATTTttcataattt	(+)	38
MADS/SQUA.01 motivo de núcleo		nnnnnnnatt tnnnnnnnnn n		39

Nome do motivo do promotor	Posição a	Sequência	Filamento	SEQ ID NO:
MYBS/MYBST1.01 motivo	226-242	taatttATCCtgaataat	(-)	41
MYBS/MYBST1.01 motivo de núcleo		nnnnnnatcc nnnnnnn		42
IBOX/GATA.01 motivo1	229-245	ttcagGATAaattattg	(+)	43
IBOX/GATA.01 motivo de núcleo		nnnnngatan nnnnnnn		44
IBOX/GATA.01 motivo2	318-334	tagaaGATAaaagaagt	(-)	64
IBOX/GATA.01 motivo3	768-784	agagaGATAaagggaaca	(-)	132
AHBP/ATHB1.01 motivo	236-246	taaATTAttgt	(+)	45
AHBP/ATHB1.01 motivo de núcleo		nnnattannn n		46
AHBP/ATHB5.01 motivo	236-246	acaATAAttta	(-)	47
AHBP/ATHB5.01 motivo de núcleo		nnnataannn n		48
GTBX/GT3A.01 motivo1	239-255	taaactTTTAcaataat	(-)	49
GTBX/GT3A.01 motivo de núcleo1		nnnnnnnttta nnnnnnn		50
GTBX/GT3A.01 motivo2	570-586	tagtaaGTTAcatttta	(+)	95
GTBX/GT3A.01 motivo de núcleo2		nnnnnnngtta nnnnnnn		96
PSRE/GAAA.01 motivo	253-269	aaatgGAAAtcttcttaa	(-)	52
PSRE/GAAA.01 motivo de núcleo		nnnnngaaan nnnnnnn		53
GTBX/S1F.01 motivo1	255-271	tcaaATGGaaatcttgt	(-)	54
GTBX/S1F.01 motivo de núcleo		nnnnnatgenn nnnnnnn		55
GTBX/S1F.01 motivo2	585-601	taggATGGaataaatat	(+)	99
WBXF/WRKY.01 motivo1	263-279	tcctATGActagtgtta	(+)	56
WBXF/WRKY.01 motivo de núcleo		nnnnnttgann nnnnnnn		57
WBXF/WRKY.01 motivo2	728-744	tgactTTGAcgtcacac	(+)	119
MYBS/TAMYB80.01 motivo1	279-295	gagaATATtctctatt	(-)	58
MYBS/TAMYB80.01 motivo de núcleo1		nnnnnatatnn nnnnnnn		59
MYBS/TAMYB80.01 motivo2	284-300	aggaATATtctctagta	(+)	60
MYBS/TAMYB80.01 motivo3	337-353	ttgtTTATtctctact	(-)	71
MYBS/TAMYB80.01 motivo de núcleo3		nnnnnttatnn nnnnnnn		72
LREM/ATCTA.01 motivo	312-322	tcATCTacttc	(+)	62
LREM/ATCTA.01 motivo de núcleo		nnatctnnnn n		63
HEAT/HSE.01 motivo	319-333	agaagataaaAGAAg	(-)	65
HEAT/HSE.01 motivo de núcleo		nnnnnnnnnnn agaan		66
P\$MADS/AGL1.01 motivo	329-349	ttaTTCtctactggtagaag	(-)	67
P\$MADS/AGL1.01 motivo de núcleo		nnnttcennn nnnnnnnnnnn n		68
MADS/AGL1.01 motivo	330-350	ttcTACCagtagaggaataaa	(+)	69
MADS/AGL1.01 motivo de núcleo		nnntaccennn nnnnnnnnnnn n		70
NACF/TANAC69.01 motivo	409-431	ctttattttaTACGtaaaatta	(-)	76
NACF/TANAC69.01 motivo de núcleo		nnnnnnnnnnn ntacgnnnnn nnn		77
HMGF/HMG IY.01 motivo	417-431	ctttTATTtatacg	(-)	78
HMGF/HMG IY.01 motivo de núcleo		nnnttatnnn nnnnn		79
AHBP/WUS.01 motivo	448-458	ctctTAATcg	(-)	81
AHBP/WUS.01 motivo de núcleo		nnnnntaatn n		82
GTBX/GT1.01 motivo	520-536	catgtgGTCAtttacg	(-)	83
GTBX/GT1.01 motivo de núcleo		nnnnnnngtca nnnnnnn		84
TCPF/ATTCP20.01 motivo	534-546	attgGCCcatcat	(-)	85
TCPF/ATTCP20.01 motivo de núcleo		nnnnngccenn nnn		86
DREB/CRT_DRE.01 motivo1	551-565	atggaCCGActacta	(+)	87
DREB/CRT_DRE.01 motivo de núcleo		nnnnnccgan nnnnn		88
DREB/CRT_DRE.01 motivo2	601-615	tcataCCGAcatacag	(+)	100
DREB/CRT_DRE.01 motivo3	658-672	tactaCCGAcatacag	(+)	106
DREB/CRT_DRE.01 motivo4	695-709	tcaagCCGAcacaga	(+)	110
SPF1/SP8BF.01 motivo	564-576	ctTACTattatta	(-)	93
SPF1/SP8BF.01 motivo de núcleo		nnntactnnnn nnn		94
EINL/TEIL.01 motivo	574-582	aTGTAactt	(-)	97
EINL/TEIL.01 motivo de núcleo		ntgtannnn		98
DOFF/PBF.01 motivo	617-633	ttgaaagaAAAGggaaa	(+)	101
DOFF/PBF.01 motivo de núcleo		nnnnnnnnnaa agnnnnn		102
TEFB/TEF1.01 motivo	624-646	aaAAGGgaaaaaagaaaaaa	(+)	103
TEFB/TEF1.01 motivo de núcleo		nnaaaggnnnn nnnnnnnnnnn n		104
DOFF/DOF3.01 motivo	671-687	agttccaaAAAGcaaaa	(+)	107
DOFF/DOF3.01 motivo de núcleo		nnnnnnnnnaa agnnnnn		108

Nome do motivo do promotor	Posição a	Sequência	Filamento	SEQ ID NO:
CE3S/CE3.01 motivo	703-721	tctctaCGCGTgtctgtgt	(-)	111
CE3S/CE3.01 motivo de núcleo		nnnnnnncgcg nnnnnnnnn		112
CGCG/ATSR1.01 motivo1	708-718	ctaCGCGgtgc	(-)	113
CGCG/ATSR1.01 motivo de núcleo		nnncgcggnnn n		114
CGCG/ATSR1.01 motivo2	709-719	acaCGCGtaga	(+)	115
GBOX/TGA1.01 motivo1	727-747	gtggtgTGACgtcaaagtc	(-)	116
GBOX/TGA1.01 motivo de núcleo		nnnnnnntgac nnnnnnnnnn n		117
GBOX/TGA1.01 motivo2	728-748	tgactTGACgtcacaccacg	(+)	118
OPAQ/RITA1.01 motivo	730-746	actttgACGTcacacca	(+)	120
OPAQ/RITA1.01 motivo de núcleo		nnnnnnnactg nnnnnnn		121
SALT/ALFIN1.01 motivo	738-752	tttcGTGGgtgac	(-)	122
SALT/ALFIN1.01 motivo de núcleo		nnnnngtgcg nnnnn		123
GBOX/GBF1.01 motivo	757-777	cgctcatACGTgtcccttta	(+)	124
GBOX/GBF1.01 motivo de núcleo		nnnnnnnnnac gtnnnnnnnn n		125
ABRE/ABRE.01 motivo	758-774	gcttcatACGTgtccct	(+)	126
ABRE/ABRE.01 motivo		nnnnnnnnacg tnnnnnn		127
GAGA/GAGABP.01 motivo1	764-788	actgagAGAGataaaggacacgta	(-)	130
GAGA/GAGABP.01 motivo de núcleo1		nnnnnnnagag nnnnnnnnnn nnnnn		131
GAGA/GAGABP.01 motivo2	772-796	atagagAGACtgagagagataaagg	(-)	133
GAGA/GAGABP.01 motivo de núcleo2		nnnnnnnagac nnnnnnnnnn nnnnn		138
GAGA/GAGABP.01 motivo3	774-798	ttatagAGAGactgagagagataaa	(-)	134
GAGA/GAGABP.01 motivo4	776-800	gtttatAGAGagactgagagagata	(-)	135
TBPF/TATA.01 motivo	790-804	tctcTATAaacttag	(+)	136
TBPF/TATA.01 motivo de núcleo		nnntatann nnnnn		137

EXEMPLO 12. Resistência melhorada contra pelo menos um fator de estresse, qualidade nutricional de uma semente ou um broto, rendimento, ou frequência de excisão do marcador de seleção

Um gene repórter em pBPSMM368 pode ser substituído com

- 5 (1) genes de resistência a estresse abiótico (14-3-3 proteína e fosfolipase específica de C fosfoinositida: WO0177355 e US6720477),
- (2) genes envolvidos em biossíntese de vitamina E (tirosina aminotransferase (BT000782: WO02072848), porfobilinogênio deaminase putativa, omega-3 ácido graxo desaturase putativa [NM185577])
- 10 (3) genes de resistência a estresse biótico (proteína resistência tipo I2C-5 de *Oryza sativa* *Fusarium* [NM194161], expressor constitutivo de genes relacionados com patogenicidade 5 (cpr5: NM185577), GTPase [WO03020939], Despolimerização de actina Fator 3 [WO2004035798], interatuante de t-SNARE de ROR2 e pintaxina, interatuante de SNAP34
- 15 [WO2004081217],
- (4) gene endonuclease *homing* (por exemplo a sequência codificando a endonuclease I-SceI *homing*) a ser expressada em embrião

durante a germinação, assim melhorando— por exemplo — a tolerância a estresses ambientais abióticos, vigor prematuro resultando em melhora potencial do rendimento, a quantidade de vitamina E, tolerância a estresses bióticos e a frequência de excisão do marcador. As construções quiméricas são transformadas em plantas monocotiledôneas. Métodos padrões para a transformação na técnica podem ser usados se requerido. As plantas transformadas são regeneradas usando métodos conhecidos. Vários fenótipos são medidos para determinar a melhora de biomassa, rendimento, composição de ácido graxo, elevado teor de óleo, tolerância a doença, ou quaisquer outros fenótipos que indicam melhora do rendimento ou estabilidade do rendimento. Os níveis de expressão de genes são determinados em diferentes estágios de desenvolvimento e em diferentes gerações (plantas T0 a T2 ou outras gerações). Resultados da avaliação em plantas levam à identificação de apropriados genes em combinação com este promotor que aumentam o rendimento, melhoram a tolerância a doença, melhoram tolerância a estresse abiótico e/ou aumentam a qualidade nutricional de semente ou broto.

EXEMPLO 13. Expressão de transgene para melhorar os traços de rendimento de alimentos, rações em plantas monocotiledôneas

Um gene repórter em pBPSMM368 pode ser substituído com um gene de interesse para superexpressar principalmente em embrião para melhorar o valor de nutrição em embrião quando transformado em plantas monocotiledôneas como arroz, centeio, milho, trigo, ou erva-castelhana, mas não é limitado às espécies da planta. O gene de interesse pode ser de sequência não de codificação (por exemplo precursor de miRNA, ou ta-siRNA) para regular de modo negativo um ou mais genes alvos. Métodos padrões para transformação na técnica podem ser usados se requerido. As plantas transformadas são selecionadas sob o agente de interesse de seleção e regeneradas usando métodos conhecidos. O esquema de seleção é examinado nos estágios de desenvolvimento prematuros de tecidos ou células de cultura

de tecido. Os níveis de expressão de genes podem ser determinados em diferentes estágios de desenvolvimento e em diferentes gerações (plantas T0 a T2 ou outras gerações). Resultados da avaliação em plantas levam a determinar os genes apropriados em combinação com este promotor.

5 Referências

1. Abel *et al.*, Science, 232:738 (1986).
2. Agrawal A *et al.* (1998) Nature 394(6695):744-451
3. Altschul *et al.*, J. Mol. Biol., 215:403 (1990).
4. Altschul *et al.*, Nucleic Acids Res., 25:3389 (1997).
- 10 5. Amara JF *et al.* (1997) Proc Natl Acad Sci USA
94(20):10618-1623
6. An *et al.*, EMBO J., 4:277 (1985).
7. Angrand PO *et al.* (1998) Nucl. Acids Res. 26(13):3263-3269
8. Argast GM *et al.* (1998) J Mol Biol 280: 345-353
- 15 9. Auch & Reth, Nucleic Acids Research, 18:6743 (1990).
- 10 10. Ausubel *et al.* (1987) Current Protocols in Molecular
Biology, Greene Publishing Assoc. and Wiley Interscience (1987).
11. Ballas *et al.*, Nucleic Acids Res., 17:7891 (1989).
12. Barkai-Golan *et al.*, Arch. Microbiol., 116:119 (1978).
- 20 13. Barker *et al.*, (1983) Plant Molec. Biol. 2: 335-50.
14. Bartel D 2004, Cell 116, 281-297
15. Bartley and Scolnik (1994) Plant Physiol., 104:1469-1470
16. Batzer *et al.*, Nucleic Acid Res., 19:5081 (1991).
17. Bäumlein *et al.* Mol Gen Genet 225:121-128 (1991)
- 25 18. Beall EL, Rio DC (1997) Genes Dev. 11(16):2137-2151
19. Beaudoin *et al.* 2000, PNAS 97:6421-6426
20. Becker *et al.* (1994) Plant J., 5:299-307,
21. Beerli RR *et al.* (1998) Proc Natl Acad Sci USA
95(25):14628-14633

22. Beerli RR *et al.* (2000) J Biol Chem 275(42):32617-32627
23. Beerli RR *et al.* (2000) Proc Natl Acad Sci U S A. 97
(4):1495-1500
24. Belfort M and Roberts RJ (1997) Nucleic Acids Res 25:
5 3379-3388
25. Bell-Pedersen *et al.* (1990) Nucleic Acids Res 18:3763-3770
26. Bernal-Lugo and Leopold, Plant Physiol., 98:1207 (1992).
27. Bevan *et al.*, Nature, 304:184 (1983).
28. Bevan *et al.*, Nucl. Acids Res., 11:369 (1983).
- 10 29. Bevan, Nucl. Acids Res., 12:8711 (1984).
30. Bibikova M *et al.* (2001) Mol Cell Biol. 21:289-297
31. Blackman *et al.*, Plant Physiol., 100:225 (1992).
32. Blanc V *et al.* (1996) Biochimie 78(6):511-517
33. Blochlinger & Diggelmann, Mol Cell Biol, 4:2929 (1984).
- 15 34. Bol *et al.*, Ann. Rev. Phytopath., 28:113 (1990).
35. Bouchez *et al.*, EMBO J., 8:4197 (1989).
36. Bourouis *et al.*, EMBO J., 2:1099 (1983).
37. Bowler *et al.*, Ann. Rev. Plant Physiol., 43:83 (1992).
38. Bradford, Anal.Biochem. 72:248-254 (1976)
- 20 39. Branson and Guss, Proc. North Central Branch
Entomological Society of America (1972).
40. Broakgert *et al.*, Science, 245:110 (1989).
41. Broglie *et al.* (1991) Science 254:1194-1197
42. Bustos *et al.* (1989) Plant Cell 1:839-853
- 25 43. Byrne *et al.* Plant Cell Tissue and Organ Culture, 8:3 (1987).
44. Callis *et al.*, Genes and Develop., 1:1183 (1987).
45. Campbell and Gowri, Plant Physiol., 92:1 (1990).
46. Campbell, ed. Ivermectin and Abamectin, Springer-Verlag,
New York, (1989).

47. Cecchini E *et al.* (1998) *Mutat Res* 401(1-2):199-206
48. Chee *et al.* *Plant Physiol.*, 91:1212 (1989).
49. Cheikh-N *et al.* (1994) *Plant Physiol.* 106(1):45-51
50. Chen and Winans J. *Bacteriol.* 173: 1139-1144 (1991).
- 5 51. Chen *et al.*, (1988) *EMBO J.*, 6:3559-3564
52. Christou *et al.* (1995) *Annals of Botany* 75:407-413
53. Christou *et al.* *Proc. Natl. Acad. Sci USA*, 86:7500 (1989).
54. Christou *et al.*, *Biotechnology*, 9:957 (1991).
55. Christou *et al.*, *Plant Physiol.*, 87:671 (1988).
- 10 56. Chu *et al.* (1990) *Proc Natl Acad Sci USA* 87:3574-3578
57. Chui *et al.* *Curr Biol* 6:325-330 (1996).
58. Coe *et al.*, In: *Corn and Corn Improvement*, Sprague *et al.* (eds.) pp. 81-258 (1988).
59. Corneille S *et al.* (2001) *Plant J* 27:171-178
- 15 60. Corpet *et al.* *Nucleic Acids Res.*, 16:10881 (1988).
61. Cote V *et al.* (1993) *Gene* 129:69-76
62. Coxson *et al.*, *Biotropica*, 24:121 (1992).
63. Crameri *et al.*, *Nature Biotech.*, 15:436 (1997).
64. Crameri *et al.*, *Nature*, 391:288 (1998).
- 20 65. Crossway *et al.*, *BioTechniques*, 4:320-334 (1986).
66. Cuozzo *et al.*, *Bio/Technology*, 6:549 (1988).
67. Cutler *et al.*, *J. Plant Physiol.*, 135:351 (1989).
68. Czako M & Marton L (1994) *Plant Physiol* 104:1067-1071
69. Czaplá and Lang, *J. Econ. Entomol.*, 83:2480 (1990).
- 25 70. Dale EC and Ow DW (1991) *Proc Natl Acad Sci USA* 88:10558-10562
71. Datta *et al.*, *Bio/Technology*, 8:736-740 (1990).
72. Davies *et al.*, *Plant Physiol.*, 93:588 (1990).
73. Dayhoff *et al.*, *Atlas of Protein Sequence and Structure*, Natl.

Biomed. Res. Found., Washington, C. D. (1978).

74. De Blaere *et al.*, Meth. Enzymol., 143:277 (1987).

75. De Block *et al.* Plant Physiol., 91:694 (1989).

76. De Block *et al.*, EMBO Journal, 6:2513 (1987).

5 77. Deak M *et al.* (1999) Nature Biotechnology 17:192-196

78. Deblaere *et al.* Nucl Acids Res 13:4777-4788 (1985)

79. Della-Cioppa *et al.* Bio/Technology 5:579-584 (1987)

80. Della-Cioppa *et al.*, Plant Physiology, 84:965-968 (1987).

81. Dellaporta *et al.*, in Chromosome Structure and Function,

10 Plenum Press, 263-282 (1988).

82. Depicker AG *et al.* (1988) Plant Cell rep 104:1067-1071

83. Depicker *et al.*, Plant Cell Reports, 7:63 (1988).

84. Dixon M & Kleppe *Biochim. Biophys. Acta* 96 (1965) 383-

389

15 85. Dixon M & Kleppe K *Biochim. Biophys. Acta* 96 (1965) 368-

382

86. Dixon M & Kleppe K. *Biochim. Biophys. Acta* 96 (1965)

357-367

87. Dunn *et al.*, Can. J. Plant Sci., 61:583 (1981).

20 88. Dunwell JM (1998) Biotechn Genet Eng Rev 15:1-32

89. Dure *et al.*, Plant Mol. Biol., 12:475 (1989).

90. Ebinuma *et al.* Proc Natl Acad Sci USA 94:2117-2121

(2000a).

91. Ebinuma *et al.* Selection of Marker-free transgenic plants

25 using the oncogenes (*ipt*, *rol A*, *B*, *C*) of *Agrobacterium* as selectable markers, In
Molecular Biology of Woody Plants. Kluwer Academic Publishers (2000b).

92. Eddy *et al.* (1991) Genes Dev. 5:1032-1041

93. Eichholtz *et al.* Somatic Cell and Molecular Genetics 13, 67-

76 (1987).

94. Ellis *et al.*, (1984) Mol. Gen. Genet. 195:466-73
95. Elroy-Stein *et al.*, Proc. Natl. Acad. Sci. U.S.A., 86:6126
(1989).
96. English *et al.*, Plant Cell, 8:179 (1996).
- 5 97. Erdmann *et al.*, J. Gen. Microbiol., 138:363 (1992).
98. Erikson *et al.* Nat Biotechnol. 22(4):455-8 (2004).
99. Everett *et al.*, Bio/Technology, 5:1201(1987).
100. Fedoroff & Smith Plant J 3:273- 289 (1993).
101. Fedoroff NV & Smith DL (1993) Plant J 3:273-289
- 10 102. Fire *et al.* Nature 391:806-811 (1998).
103. Fitzpatrick, Gen. Engineering News, 22:7 (1993).
104. Fox *et al.* (1992) Plant Mol. Biol. 20:219-33
105. Fraley *et al.* Proc Natl Acad Sci USA 80:4803 (1983).
106. Freeling & Walbot (1993) "The maize handbook" ISBN 3-
15 540-97826-7, Springer Verlag New York)
107. Fromm *et al.*, Bio/Technology, 8:833-839 (1990).
108. Fromm *et al.*, Nature (London), 319:791 (1986).
109. Gabler M *et al.* (2000) Enzyme Microb. Techno. 27:605–
611
- 20 110. Galbiati *et al.* Funct. Integr Genozides, 20 1:25-34 (2000).
111. Gallego ME (1999) Plant Mol Biol 39(1):83-93
112. Gallie *et al.* Nucl Acids Res 15:8693-8711 (1987).
113. Gallie *et al.*, Nucleic Acids Res., 15:3257 (1987).
114. Gallie *et al.*, The Plant Cell, 1:301 (1989).
- 25 115. Gan *et al.*, Science, 270:1986 (1995).
116. Gatehouse *et al.*, J. Sci. Food Agric., 35:373 (1984).
117. Gelfand, eds., PCR Strategies Academic Press, New York
(1995).
118. Gilmour *et al.* Plant Mol Biol 17 :1233-1240 (1991)

119. Gelvin *et al.*, Plant Molecular Biology Manual, (1990).
120. Girke *et al.* (1998) Plant J 15:39-48
121. Gleave AP *et al.* (1999) Plant Mol Biol 40(2):223-235
122. Gleave *et al.* Plant Mol Biol. 40(2):223-35 (1999)
- 5 123. Gordon-Kamm *et al.*, Plant Cell, 2:603 (1990).
124. Goring *et al.*, PNAS, 88:1770 (1991).
125. Goyal RK *et al.* (2000) Crop Protection 19(5):307-312
126. Gruber, *et al.*, Vectors for Plant Transformation, in:
Methods in Plant Molecular Biology & Biotechnology" in Glich *et al.*, (Eds. pp.
10 89-119, CRC Press, 1993).
127. Guerineau *et al.*, Mol. Gen. Genet., 262:141 (1991).
128. Guerrero *et al.*, Plant Mol. Biol., 15:11 (1990).
129. Guo *et al.* (1997) EMBO J 16: 6835- 6848
130. Guo *et al.* (2000) Science 289:452- 457
- 15 131. Gupta *et al.*, PNAS, 90:1629 (1993).
132. Haines and Higgins (eds.), Nucleic Acid Hybridization, IRL
Press, Oxford, U.K.
133. Hajdukiewicz *et al.*, Plant Mol Biol 25:989-994 (1994).
134. Hammock *et al.*, Nature, 344:458 (1990).
- 20 135. Hansen *et al.* *Proc. Natl. Acad. Sci. USA* 91:7603-7607
(1994).
136. Hare P & Chua NH (2002) Nat. Biotechnol. 20, 575–580.
137. Haren L *et al.* (1999) Annu Rev Microbiol. 1999;53:245-
281
- 25 138. Haseloff *et al.* (1997) Proc Natl Acad Sci USA 94(6):2122-
2127
139. Hayford *et al.* Plant Physiol. 86:1216 (1988)
140. Hemenway *et al.*, EMBO Journal, 7:1273 (1988).
141. Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA*, 89:10915

(1989).

142. Hiei *et al.* Plant J 6: 271-282 (1994)
143. Higgins *et al.*, Gene, 73:237 (1988).
144. Higo *et al.* (1999) Nucl Acids Res 27(1): 297-300
- 5 145. Hilder *et al.*, Nature, 330:160 (1987).
146. Hille *et al.* Plant Mol. Biol. 7:171 (1986)
147. Hinchee *et al.* Bio/Technology 6:915 (1988).
148. Hoekema *et al.* Nature 303:179-181 (1983).
149. Hoekema, In: The Binary Plant Vector System. Offset-
10 drukkerij Kanters B.V.; Alblasterdam (1985).
150. Hood EE and Jilka JM (1999) Curr Opin Biotechnol
10(4):382-6
151. Hood *et al.* (1999) Adv Exp Med Biol 464:127-47. Review
152. Hood *et al.* J Bacteriol 168:1291-1301 (1986).
- 15 153. Horvath *et al* (1993) Plant Physiol. 103:1047-1053
154. Huang B *et al.* (1996) J Protein Chem 15(5):481-9
155. Huang *et al.*, CABIOS, 8:155 (1992).
156. Ikeda *et al.*, J. Bacteriol., 169:5612 (1987).
157. Ikuta *et al.*, Biotech., 8:241 (1990).
- 20 158. Imai T *et al.* (2001) Proc Natl Acad Sci USA 98(1):224-
228)
159. Ingelbrecht *et al.*, Plant Cell, 1:671 (1989).
160. Innis and Gelfand, eds., PCR Methods Manual (Academic
Press, New York) (1999).
- 25 161. Innis *et al.*, eds., PCR Protocols: A Guide to Methods and
Applications (Academic Press, New York (1995).
162. Innis *et al.*, PCR Protocols: A Guide to Methods and
Applications, Academic Press, Inc., San Diego, Calif. (1990).
163. Ishida *et al.* Nature Biotech 745-750 (1996).

164. Janssen DB (1989) *J Bacteriol* 171(12):6791-9
165. Janssen DB *et al.* (1994) *Annu Rev Microbiol* 48:163-191
166. Jasin M (1996) *Trends Genet.* 12:224-228
167. Jefferson *et al.* (1987) *EMBO J* 6:3901-3907
- 5 168. Jefferson *et al.* *EMBO J* 6:3901-3907 (1987).
169. Jefferson *et al.* *Plant Mol Biol Rep* 5:387-405 (1987).
170. Jené *et al.* *Techniques for Gene Transfer*, in: *Recombinant Plants*, Vol. 1, Engineering and Utilization, edited by SD Kung and R Wu, Academic Press, pp. 128-143 (1993)
- 10 171. Jicks GR and Raikhel NV (1995) *Annu. Rev. Cell Biol.* 11:155-188
172. Jobling *et al.*, *Nature*, 325:622 (1987).
173. Johnson *et al.*, *PNAS USA*, 86:9871 (1989).
174. Jones *et al.* *Mol. Gen. Genet.*, 210:86 (1987).
- 15 175. Joshi *et al.*, *Nucleic Acid Res.*, 15:9627 (1987).
176. Kaasen *et al.*, *J. Bacteriol.*, 174:889 (1992).
177. Kang JS and Kim JS (2000) *J Biol Chem* 275(12):8742-8748
178. Karlin and Altschul, *Proc. Natl. Acad Sci. USA*, 87:2264
- 20 (1990).
179. Karlin and Altschul, *Proc. Natl. Acad. Sci. USA*, 90:5873 (1993).
180. Karlin-Neumann GA *et al.* (1991) *Plant Cell* 3:573-582
181. Karsten *et al.*, *Botanica Marina*, 35:11 (1992).
- 25 182. Kasuga *et al.*, (1999) *Nature Biotechnology* 17(3):287-291
183. Kasuga M *et al.* (1999) *Nature Biotech* 17:276-286
184. Katz *et al.*, *J. Gen. Microbiol.*, 129:2703 (1983).
185. Kaufman PD and Rio DC (1992) *Cell* 69(1):27-39
186. Kawasaki *et al.* (1991) *J Biol Chem* 266:5342-5347

187. Keegstra (1989) *Cell*, 56(2):247-53
188. Keenan T *et al.* (1998) *Bioorg Med Chem.* 6(8):1309-1335
189. Keller *et al.*, *EMBO Journal*, 8:1309 (1989).
190. Keller *et al.*, *Genes Dev.*, 3:1639 (1989)
- 5 191. Kilby NJ *et al.* (1995) *Plant J* 8:637-652
192. Kim JS *et al.* (1997) *Proc Natl Acad Sci USA* 94(8):3616-3620
193. Klapwijk *et al.* *J. Bacteriol.*, 141,128-136 (1980).
194. Klein *et al.*, *Bio/Technology*, 6:559-563 (1988).
- 10 195. Klein *et al.*, *Plant Physiol.*, 91:440-444 (1988).
196. Klein *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:4305-4309(1988).
197. Klug A (1999) *J Mol Biol* 293(2):215-218
198. Knauf, *et al.*, *Genetic Analysis of Host Range Expression by Agrobacterium In: Molecular Genetics of the Bacteria-Plant Interaction*, Puhler, A. ed., Springer-Verlag, New York, 1983.
- 15 199. Kobayashi T *et al.* (1995) *Jpn J Genet* 70(3):409-422
200. Komro *et al.*, (1985) *Plant Mol. Biol.* 4:253-63
201. Koncz & Schell *Mol Gen Genet* 204:383-396 (1986).
- 20 202. Kononowicz *et al.*, (1992) *Plant Cell* 4:17-27.
203. Koprek *et al.* *Plant J* 19(6): 719-726 (1999).
204. Koprek T *et al.* (1999) *Plant J* 19(6):719-726
205. Koster and Leopold, *Plant Physiol.*, 88:829 (1988).
206. Koziel *et al.*, *Biotechnology*, 11:194 (1993).
- 25 207. Kridl *et al.*, (1991) *Seed Sci. Res.*, 1:209-219
208. Kuiper HA *et al.* (2001) *Plant J.* 27, 503–528
209. Kunkel *et al.*, *Methods in Enzymol.*, 154:367 (1987).
210. Kunkel, *Proc. Natl. Acad. Sci. USA*, 82:488 (1985).
211. Lam and Chua , *J Biol Chem*; 266(26):17131-17135 (1991).

212. Laufs *et al.*, PNAS, 87:7752 (1990).
213. Lawton *et al.*, Mol. Cell Biol., 7:335 (1987).
214. Lee and Saier, J. Bacteriol., 153 (1982).
215. Lee TJ *et al.* (2002) J Amer Soc Hortic Sci 127(2):158-
- 5 164
216. Leffel *et al.* Biotechniques 23(5):912-8 (1997).
217. Leffel SM *et al.* (1997) Biotechniques 23(5):912-8
218. Lescot *et al.* Nucleic Acids Res 30(1):325-7 (2002).
219. Leung *et al.*, (1991) Mol. Gen. Genet. 230:463-74
- 10 220. Levings, Science, 250:942 (1990).
221. Li *et al.* (1993) Plant Cell Reports 12:250-255
222. Li *et al.* Plant Mol Biol 20:1037-1048 (1992).
223. Lindsey *et al.*, Transgenic Research, 2:3347 (1993).
224. Liu *et al.*, Plant J. 8, 457-463 (1995)
- 15 225. Logie C and Stewart AF (1995) Proc Natl Acad Sci USA
92(13):5940-5944
226. Lois *et al.* (1998) *Proc. Natl. Acad. Sci. USA*, 95 (5):2105-
2110
227. Lommel *et al.*, Virology, 181:382 (1991).
- 20 228. Loomis *et al.*, J. Expt. Zool., 252:9 (1989).
229. Lorz *et al.*, Mol. Gen. Genet., 199:178 (1985).
230. Lyznik LA *et al.* (1996) Nucleic Acids Res 24:3784-3789
231. Ma *et al.*, Nature, 334 :631 (1988).
232. Ma JK and Vine ND (1999) Curr Top Microbiol Immunol
25 236:275-92
233. Ma, Q. H. *et al.*, (1998) Australian Journal of Plant
Physiology 25(1):53-59
234. Macejak *et al.*, Nature, 353:90 (1991).
235. Maki *et al.*, Methods in Plant Mol. Biol. & Biotechnol,

Glich *et al.*, 67-88 CRC Press, (1993).

236. Maniatis *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY), (1989).

237. Mapp AK *et al.* (2000) Proc Natl Acad Sci USA

5 97(8):3930-3935

238. Mariani *et al.*, Nature, 347:737 (1990).

239. Marshall (1991) Gene 104:241-245

240. Massey V *et al.* *Biochim. Biophys. Acta* 48 (1961) 1-9

241. Matzke *et al.* (2000) Plant Mol Biol 43:401-415 (2000).

10 242. McBride *et al.*, PNAS USA, 91:7301 (1994).

243. McCabe *et al.*, Bio/Technology, 6:923 (1988).

244. McKnight SL *et al.* (1980) Nucl Acids Res 8(24):5931-

5948

245. McKnight SL *et al.* (1980) Nucl Acids Res 8(24):5949-

15 5964

246. Meinkoth and Wahl, Anal. Biochem., 138:267 (1984).

247. Meister A & Wellner D Flavoprotein amino acid oxidase.

In: Boyer, P.D., Lardy, H. and Myrbäck, K. (Eds.), *The Enzymes*, 2nd ed., vol. 7, Academic Press, New York, 1963, p. 609-648

20 248. Menard R *et al.* (1999) Phytochemistry 52:29-35

249. Messing and Vierra, Gene, 19:259 (1982).

250. Michael *et al.*, J. Mol. Biol., 26 :585 (1990)

251. Michaelson *et al.* (1998) FEBS Letters 439:215-218

252. Michaelson *et al.* JBC 273:19055-19059

25 253. Millar *et al.* (1992) Plant Mol Biol Rep 10:324-414

254. Millar *et al.* Plant Mol Biol Rep 10:324-414 (1992).

255. Miyano M *et al.* (1991) J Biochem 109:171-177

256. Mogen *et al.*, Plant Cell, 2:1261 (1990).

257. Mohr *et al.* (2000) Genes & Development 14:559-573

258. Monnat RJ Jr *et al.* (1999) *Biochem Biophys Res Com*
255:88-93
259. Moore *et al.*, *J. Mol. Biol.*, 272:336 (1997).
260. Mozo and Hooykaas, *Plant Mol. Biol.* 16:917-918 (1991).
- 5 261. Mullen CA *et al.* (1992) *Proc Natl Acad Sci USA* 89(1):33-
37
262. Mundy and Chua, *EMBO J.*, 7:2279 (1988).
263. Munroe *et al.*, *Gene*, 91:151 (1990).
264. Murakami *et al.*, *Mol. Gen. Genet.*, 205:42 (1986).
- 10 265. Murata *et al.*, *FEBS Lett.*, 296:187 (1992).
266. Murdock *et al.*, *Phytochemistry*, 29:85 (1990).
267. Murray *et al.*, *Nucleic Acids Res.*, 17:477 (1989).
268. Muthuswamy SK *et al.* (1999) *Mol Cell Biol* 19(10):6845-
685
- 15 269. Myers and Miller, *CABIOS*, 4:11 (1988).
270. Naested H *et al.* (1999) *Plant J* 18(5):571-576
271. Naested, *Plant J* 18:571-576 (1999).
272. Napoli *et al.*, *Plant Cell*, 2:279 (1990).
273. Nawrath *et al.* (1994) *Proc. Natl. Acad. Sci. USA*,
20 91:12760-12764
274. Needleman and Wunsch, *J. Mol. Biol.*, 48:443-453 (1970).
275. Negri A *et al.* (1992) *J Biol Chem.* 267:11865-11871
276. Nehra *et al.* *Plant J.* 5:285-297 (1994)
277. Ni M *et al.* (1995) *Plant J* 7(4):661-676
- 25 278. Niedz *et al.*, *Plant Cell Reports*, 14:403 (1995).
279. Nordin *et al.* *Plant Mol Biol* 21 :641-653 (1991)
280. Norris *et al.* (1998) *Plant Physiol.*, 117:1317-1323
281. O'Keefe DP (1991) *Biochemistry* 30(2):447-55
282. O'Keefe DP *et al.* (1994) *Plant Physiol* 105:473-482

283. Odell *et al.*, Mol. Gen. Genet., 113:369 (1990).
284. Odell *et al.*, Nature, 313:810 (1985).
285. Ohtsuka *et al.*, J. Biol. Chem., 260:2605 (1985).
286. Olhoft *et al.* Plant Cell Rep 20: 706–711 (2001).
- 5 287. Onouchi H *et al.*(1995) Mol Gen Genet 247:653-660
288. Osborne BI *et al.* (1995) Plant J. 7:687-701
289. Osjoda *et al.* (1996) Nature Biotechnology 14:745-750
290. Ow DW and Medberry SL (1995) Crit Rev in Plant Sci
14:239-261
- 10 291. Ow *et al.*, Science, 234:856 (1986).
292. Pacciotti *et al.*, Bio/Technology, 3:241 (1985).
293. Parizotto E A *et al.*, 2004, Genes & Development, 18:
2237-2242
294. Park *et al.*, J. Plant Biol., 38:365 (1985).
- 15 295. Paszkowski *et al.*, EMBO J., 3:2717-2722 (1984).
296. Pearson and Lipman, Proc. Natl. Acad. Sci., 85:2444
(1988).
297. Pearson *et al.*, Meth. Mol. Biol., 24:307 (1994).
298. Pei ZM *et al.* (1998) Science 282:287-290
- 20 299. Perera *et al.* Plant Mol. Biol 23(4): 793-799 (1993).
300. Perera RJ *et al.* (1993) Plant Mol Biol 23(4):793-799
301. Perlak *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:3324 (1991).
302. Phillips *et al.*, In Corn & Corn Improvement, 3rd Edition 10
Sprague *et al.* (Eds. pp. 345-387)(1988).
- 25 303. Phi-Van *et al.*, Mol. Cell. Biol., 10:2302 (1990).
304. Piatkowski *et al.*, Plant Physiol., 94:1682 (1990).
305. Pilone MS (2000) Cell. Mol. Life. Sci. 57:1732–1740
306. Potrykus *et al.*, Mol. Gen. Genet., 199:183 (1985).
307. Potrykus, Trends Biotech., 7:269 (1989).

308. Prasher *et al.*, Biochem. Biophys. Res. Comm., 126:1259
(1985).
309. Preston *et al.* (1981) J Virol 38(2):593-605
310. Proudfoot, Cell, 64:671 (1991).
- 5 311. Rao *et al.* (2000) Prog Nucleic Acid Res Mol Biol 64:1-63
312. Rao *et al.* (1998) Plant J 15(4):469-77
313. Rask L *et al.* (2000) Plant Mol Biol 42:93-113
314. Reed *et al.*, J. Gen. Microbiol., 130:1 (1984).
315. Reichel *et al.* (1996) Proc Natl Acad Sci USA 93(12):5888-
10 5893
316. Riggs *et al.*, *Proc. Natl. Acad. Sci. USA*, 83:5602-5606
(1986).
317. Risseuw E (1997) Plant J 11(4):717-728
318. Roeckel, P. *et al.*, (1997) Transgenic Research 6(2):133-141
- 15 319. Rossolini *et al.*, Mol. Cell. Probes, 8:91 (1994).
320. Ruiz, Plant Cell, 10:937 (1998).
321. Russell SH *et al.* (1992) Mol Gene Genet 234:49-59.
322. Saint Guily *et al.* (1992) Plant Physiol., 100(2):1069-1071
323. Sakuradani *et al.* (1999) Gene 238:445-453
- 20 324. Salomon S & Puchta H (1998) EMBO J 17(20):6086-6095
325. Sambrook *et al.*, Molecular Cloning: A Laboratory Manual
(2d ed., Cold Spring Harbor Laboratory Press, Plainview, N.Y.) (1989).
326. Sanfacon *et al.*, Genes Dev., 5:141 (1991).
327. Sanford *et al.*, Particulate Science and Technology, 5:27
25 (1987).
328. Sanger *et al.*, (1990) Plant Mol. Biol. 14: 433-43
329. Sarguiel B *et al.* (1990) Nucleic Acids Res 18:5659-5665
330. Sarguiel B *et al.* (1991) Mol Gen Genet. 255:340-341
331. Sato *et al.* (2000) J. DNA Res., 7(1):31-63

332. Scheeren-Groot *et al.*, J. Bacteriol 176: 6418-6426 (1994).
333. Schenborn and Groskreutz, Mol Biotechnol 13(1): 29-44 (1999).
334. Schenborn E, Groskreutz D. (1999) Mol Biotechnol 13(1):29-44
335. Schlaman and Hooykaas, Plant J 11:1377-1385 (1997).
336. Schlaman HRM & Hooykaas PFF (1997) Plant J 11:1377-1385
337. Schoffl *et al.* Mol Gen Genetics 217(2-3):246-53 (1989).
338. Schultz LW and Clardy J (1998) Bioorg Med Chem Lett. 8(1):1-6
339. Segal DJ and Barbas CF 3rd., Curr Opin Chem Biol (2000) 4(1):34-39
340. Serino G (1997) Plant J 12(3):697-701
341. Shagan *et al.*, Plant Physiol., 101:1397 (1993).
342. Shah *et al.* Science 233: 478 (1986).
343. Shapiro, Mobile Genetic Elements, Academic Press, N.Y. (1983).
344. Sharrocks AD *et al.* (1997) Int J Biochem Cell Biol 29(12):1371-1387
345. Sheen *et al.* (1995) Plant J 8(5):777-784
346. Shimamoto *et al.*, Nature, 338:274 (1989).
347. Silhavy *et al.*, Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), (1984).
348. Skuzeski *et al.*, Plant Molec. Biol. 15: 65-79 (1990).
349. Smith *et al.* (1997) Plant J., 11:83-92
350. Smith *et al.*, Adv. Appl. Math., 2:482.(1981).
351. Smith *et al.*, Mol. Gen. Genet., 224:447 (1990).
352. Spencer *et al.*, Theor. Appl. Genet, 79:625 (1990).

353. St. Clair *et al.* (1987) *Antimicrob Agents Chemother* 31(6):844-849
354. Stalker *et al.*, *Science*, 242:419 (1988).
355. Staub *et al.*, *EMBO J.*, 12:601 (1993).
- 5 356. Staub *et al.*, *Plant Cell*, 4:39 (1992).
357. Steifel *et al.*, *The Plant Cell*, 2:785 (1990).
358. Stemmer (1994) *Nature* 370:389-391
359. Stemmer (1994) *Proc Natl Acad Sci USA* 91:10747-10751
360. Stemmer, *Nature*, 370:389 (1994).
- 10 361. Stemmer, *Proc. Natl. Acad. Sci. USA*, 91:10747 (1994).
362. Stief *et al.*, *Nature*, 341:343 (1989).
363. Stougaard J (1993) *Plant J* 3:755-761); EP-A1 595 837
364. Stougaard, *Plant J* 3:755-761 (1993)
365. Sugita Ket *et al.* (2000) *Plant J.* 22:461-469
- 15 366. Sukhapinda *et al.*, *Plant Mol. Biol.*, 8:209 (1987).
367. Sundaresan *et al.* *Gene Develop* 9: 1797-1810 (1995).
368. Sundaresan V *et al.* (1995) *Gene Develop* 9:1797-1810
369. Sutcliffe, *PNAS USA*, 75:3737 (1978).
370. Svab *et al.*, *Plant Mol. Biol.* 14:197 (1990).
- 20 371. Svab *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:8526 (1990).
372. Svab *et al.*, *Proc. Natl. Acad. Sci. USA*, 90:913 (1993).
373. Szybalski *et al.* (1991) *Gene* 100:13-26
374. Takahashi *et al.* (1998) *Proc. Natl. Acad. Sci. USA*, 95(17):9879-9884
- 25 375. Tarczyński *et al.*, *PNAS USA*, 89:2600 (1992).
376. Teeri *et al.*, (1989) *EMBO J.*, 8: 343-50
377. Thillet *et al.*, *J. Biol. Chem.*, 263:12500 (1988).
378. Thompson *et al.*, *NAR* 22(22):4673-4680 (1994).
379. Thykjaer T *et al.* (1997) *Plant Mol Biol* 35(4):523-530

380. Tian *et al.* (1997) Plant Cell Rep 16:267-271; WO 97/41228
381. Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, Elsevier, N.Y. (1993).
- 5 382. Tissier AF *et al.* (1999) Plant Cell 11:1841-1852
383. Tomes *et al.*, Plant Cell, Tissue and Organ Culture: Fundamental Methods, Springer Verlag, Berlin (1995).
384. Tomic *et al.*, NAR, 12:1656 (1990).
385. Tsai SY *et al.* (1998) Adv Drug Deliv Rev 30(1-3):23-31
- 10 386. Turmel M *et al.* (1993) J Mol Biol 232: 446-467
387. Turmel M *et al.* (1995a) Nucleic Acids Res 23:2519-2525
388. Turmel M *et al.* (1995b) Mol. Biol. Evol. 12, 533-545
389. Turner *et al.*, Molecular Biotechnology, 3:225 (1995).
390. Twell *et al.*, Plant Physiol., 91:1270 (1989).
- 15 391. Ugaki *et al.*, Nucl. Acids Res., 19:371 (1991).
392. Ulmasov *et al.*, Plant Mol. Biol., 35:417 (1997).
393. Umhau S. *et al.* (2000) *Proc. Natl. Acad. Sci. USA* 97:12463–12468
394. Upadhyaya NM *et al.* (2000) Plant Mol Biol Rep 18:227-223
- 20 395. Upender *et al.*, Biotechniques, 18:29 (1995).
396. van der Krol *et al.*, Plant Cell, 2:291 (1990).
397. Vanden Elzen *et al.* Plant Mol Biol. 5:299 (1985).
398. Vasil *et al.* Bio/Technology, 10:667-674 (1992).
- 25 399. Vasil *et al.* Bio/Technology, 11:1153-1158 (1993).
400. Vasil *et al.*, Mol. Microbiol., 3:371 (1989).
401. Vasil *et al.*, Plant Physiol., 91:1575 (1989).
402. Vernon and Bohnert, EMBO J., 11:2077 (1992).
403. Wagner *et al.* (1981) Proc Natl Acad Sci USA 78(3):1441-

1445

404. Walker and Gastra, eds., *Techniques in Molecular Biology*, MacMillan Publishing Company, New York (1983).
405. Wan & Lemaux *Plant Physiol.*, 104:3748 (1994).
- 5 406. Wang *et al.*, *Mol. Cell. Biol.*, 12:3399 (1992).
407. Wang J *et al.* (1997) *Nucleic Acids Res* 25: 3767-3776
408. Waterman, *Introduction to Computational Biology: Maps, sequences and genomes*. Chapman & Hall. London (1995).
409. Watrud *et al.*, in *Engineered Organisms and the Environment* (1985).
- 10 410. Watson *et al.* *J. Bacteriol* 123, 255-264 (1975)
411. Watson *et al.*, *Corn: Chemistry and Technology* (1987).
412. Weeks *et al.* *Plant Physiol* 102:1077-1084 (1993)
413. Weissinger *et al.*, *Annual Rev. Genet.*, 22:421 (1988).
- 15 414. Wernette CM (1998) *Biochemical & Biophysical Research Communications* 248(1):127-333
415. White *et al.*, *Nucl Acids Res*, 18, 1062 (1990).
416. Wigler M *et al.* (1977) *Cell* 11(1):223-232
417. Wingender *et al.* *Nucleic Acids Res* 29(1):281-3 (2001).
- 20 418. Wolter *et al.*, *EMBO Journal*, 11:4685 (1992).
419. Wyn-Jones and Storey, *Physiology and Biochemistry of Drought Resistance in Plants*, Paleg *et al.* (eds.), pp. 171-204 (1981).
420. Xia *et al.* (1992) *J. Gen. Microbiol.*, 138:1309-1316
421. Xiao YL and Peterson T (2000) *Mol Gen Genet* 263(1):22-
- 25 29
422. Xiaohui Wang H *et al.* (2001) *Gene* 272(1-2): 249-255
423. Yamaguchi-Shinozaki *et al.*, *Plant Cell Physiol.*, 33:217 (1992).
424. Yurimoto H *et al.* (2000) *Yeast* 16:1217-1227

425. Zank *et al.* 2000, Biochemical Society Transactions 28:654-

657

426. Zhang *et al.*, *Proc. Natl. Acad. Sci. USA*, 94:4504 (1997).

427. Zhang L *et al.* (2000) *J Biol Chem* 275(43):33850-33860

5

428. Zinselmeier *et al.* (1995) *Plant Physiol.* 107(2):385-391

429. Zubko E *et al.* (2000) *Nat Biotechnol* 18:442-445

430. Zubko *et al.* (2000) *Nature Biotech* 18(4):442-445

431. Zukowsky *et al.*, *PNAS USA*, 80:1101 (1983).

10 Todas as publicações patentes e pedidos de patente são
incorporados aqui por referência. Apesar de no relatório acima esta invenção
ter sido descrita com relação a algumas formas de realização preferidas da
mesma, e muitos detalhes terem sido especificados para fins de ilustração,
será evidente para os versados na técnica que a invenção é susceptível a
formas de realização adicionais e que alguns dos detalhes descritos aqui
15 podem ser variados consideravelmente sem sair dos princípios básicos da
invenção.

REIVINDICAÇÕES

1. Molécula de ácido nucleico isolada, caracterizada pelo fato de compreender uma sequência reguladora da transcrição de plantas, em que a sequência reguladora da transcrição compreende:

5 i) uma primeira sequência de ácido nucleico compreendendo a sequência de promotor cor78 correspondente a SEQ ID NO:1, e

operacionalmente ligada à mesma,

 ii) uma segunda sequência de ácido nucleico compreendendo o primeiro intron de um gene de planta codificando a Metalotionina 1 (MET1)
10 correspondente a SEQ ID NO:3.

2. Molécula de ácido nucleico de acordo com a reivindicação 1, caracterizada pelo fato de compreender uma sequência espaçadora de 1 pb a 100 pb que está localizada entre a sequência do promotor cor78 e o primeiro nucleotídeo do referido intron.

15 3. Molécula de ácido nucleico de acordo com a reivindicação 1 ou 2, caracterizada pelo fato de compreender um 5'UTR que está localizado entre a sequência do promotor cor78 e o primeiro nucleotídeo do referido intron.

20 4. Molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1 a 3, caracterizada pelo fato de que o referido primeiro intron de MET1 está localizado na sequência de um intron de uma sequência nucleotídica transcrita sob o controle da sequência nucleotídica reguladora da transcrição.

25 5. Molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1 a 4, caracterizada pelo fato de que o referido primeiro intron é derivado de um MET1 de uma planta monocotiledônea.

6. Molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1 a 5, caracterizada pelo fato de que o promotor cor78 é derivado de planta dicotiledônea.

7. Molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1 a 6, caracterizada pelo fato de que a sequência reguladora da transcrição regula a expressão específica em embrião de uma sequência de ácido nucleico operacionalmente ligada.

5 8. Molécula de ácido nucleico de acordo com qualquer uma das reivindicações 1 a 7, caracterizada pelo fato de que a sequência reguladora da transcrição regula a expressão indutível por estresse de uma sequência de ácido nucleico operacionalmente ligada.

10 9. Cassete de expressão, caracterizado pelo fato de compreender

i) uma molécula de ácido nucleico como definida em qualquer uma das reivindicações 1 a 8,

ii) e operacionalmente ligada à mesma uma ou mais moléculas de ácido nucleico.

15 10. Vetor de expressão, caracterizado pelo fato de compreender a molécula de ácido nucleico como definida em qualquer uma das reivindicações 1 a 8 ou o cassete de expressão como definido na reivindicação 9.

20 11. Método para excisão de sequências alvo de uma planta, caracterizado pelo fato de compreender as etapas de

A) construir um cassete de expressão por ligar operacionalmente a molécula de ácido nucleico como definida em qualquer uma das reivindicações 1 a 8 a pelo menos uma molécula de ácido nucleico que é heteróloga em relação à referida primeira ou referida segunda sequência
25 de ácido nucleico como definida em qualquer uma das reivindicações 1 a 8 e que é capaz de conferir a excisão de uma sequência alvo de uma planta ou célula de planta, e

B) introduzir referido cassete de expressão estável ou transiente diretamente ou indiretamente em uma célula de planta ou em uma

planta compreendendo pelo menos uma sequência alvo excisável pelo produto de expressão de uma molécula de ácido nucleico que é heteróloga e, em que a referida célula de planta ou planta expressa a referida sequência de ácido nucleico que é heteróloga, e

5 C) selecionar plantas transgênicas que demonstram a excisão de referida sequência alvo.

12. Método de acordo com a reivindicação 11, caracterizado pelo fato de que referida molécula de ácido nucleico, que é heteróloga e que é operativamente ligada à referida molécula de ácido nucleico como definida
10 em qualquer uma das reivindicações 1 a 8, é uma molécula de ácido nucleico conferindo a expressão de uma recombinase sítio específica.

13. Método de acordo com a reivindicação 12, caracterizado pelo fato de que a referida molécula de ácido nucleico, que é heteróloga e que é operacionalmente ligada à referida molécula de ácido nucleico como
15 definida em qualquer uma das reivindicações 1 a 8 é uma molécula de ácido nucleico conferindo expressão de uma endonuclease sítio específica capaz de induzir uma ruptura de filamento duplo de DNA específica *para* a sequência alvo e em que a sequência alvo a ser deletada é flanqueada por sequências tendo uma orientação, um comprimento suficiente e uma homologia para cada
20 outra de modo a permitir a recombinação homóloga entre as mesmas.

14. Método para produzir uma planta com rendimento aumentado, e/ou tolerância a estresse aumentada, e/ou qualidade nutricional aumentada, e/ou teor de óleo aumentado ou modificado de uma semente ou broto para a planta, caracterizado pelo fato de compreender as etapas de

25 A) introduzir na planta a molécula de ácido nucleico como definida em qualquer uma das reivindicações 1 a 8, o cassete de expressão como definido na reivindicação 9, ou o vetor de expressão como definido na reivindicação 10, em que a molécula de ácido nucleico é operacionalmente ligada a pelo menos uma molécula de ácido nucleico cuja sequência é

heteróloga em relação à referida primeira ou referida segunda sequência de ácido nucleico e é capaz de conferir à planta rendimento aumentado, e/ou tolerância ao estresse aumentada, qualidade nutricional aumentada, e/ou teor de óleo aumentado ou modificado à planta; e

5 B) selecionar plantas transgênicas, em que as plantas têm rendimento aumentado e/ou tolerância ao estresse aumentada sob condições de estresse, e/ou qualidade nutricional aumentada e/ou teor de óleo aumentado ou modificado de uma semente ou um broto das plantas, como comparado com as plantas de tipo selvagem ou segregante nulo.

10 15. Uso da molécula de ácido nucleico como definida em qualquer uma das reivindicações 1 a 8, caracterizado pelo fato de ser para a expressão de um gene de interesse preferivelmente ou especificamente em células ou tecido embrionário.

15 16. Uso da molécula de ácido nucleico como definida em qualquer uma das reivindicações 1 a 8, caracterizado pelo fato de ser para aumentar a transcrição de uma molécula de ácido nucleico em uma planta sob condições de estresse.

FIG.1

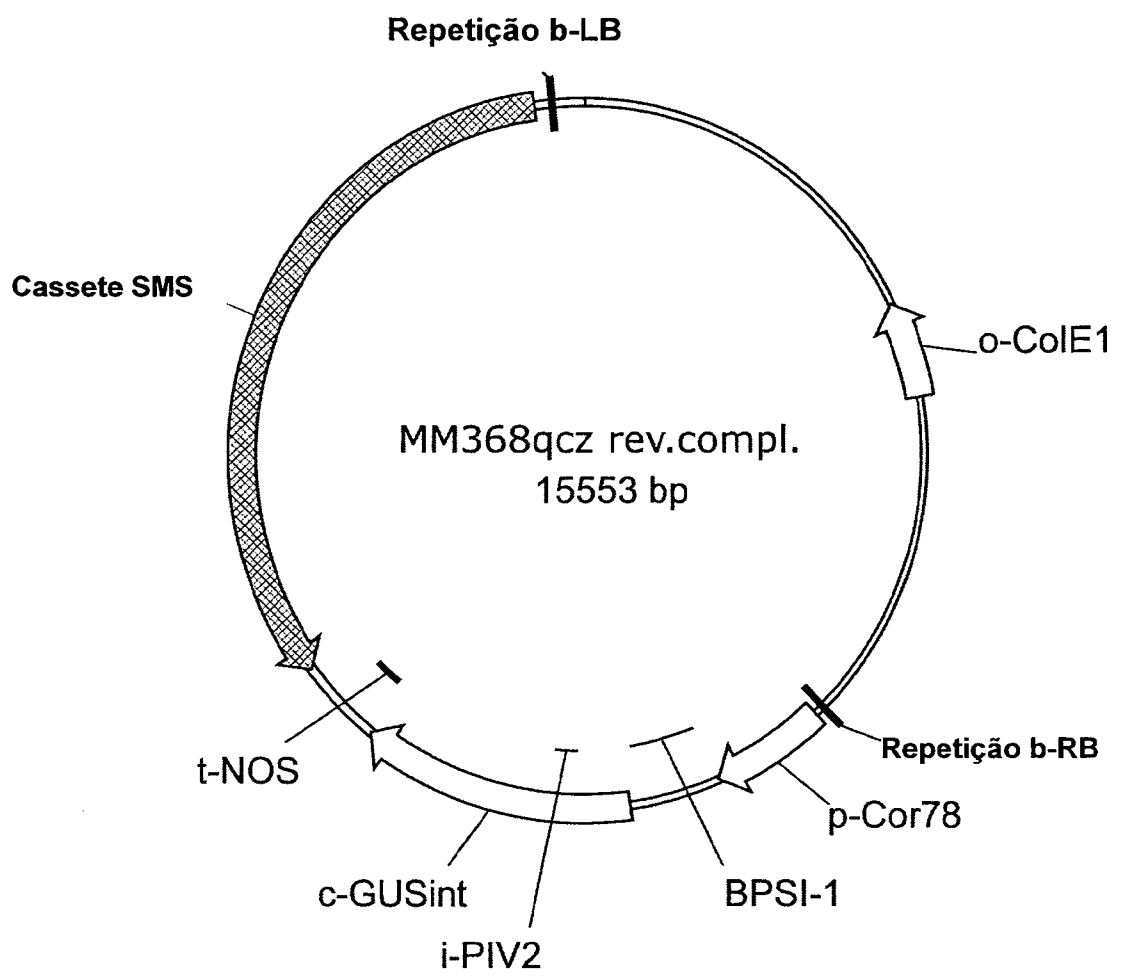


FIG.2

Coloração GUS histoquímica de construção de promotores At. Cor78

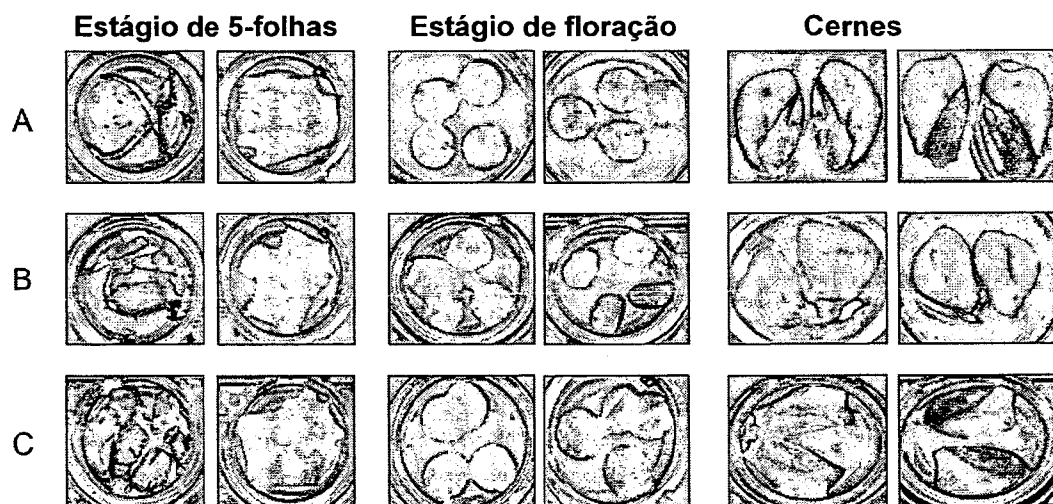


FIG.3

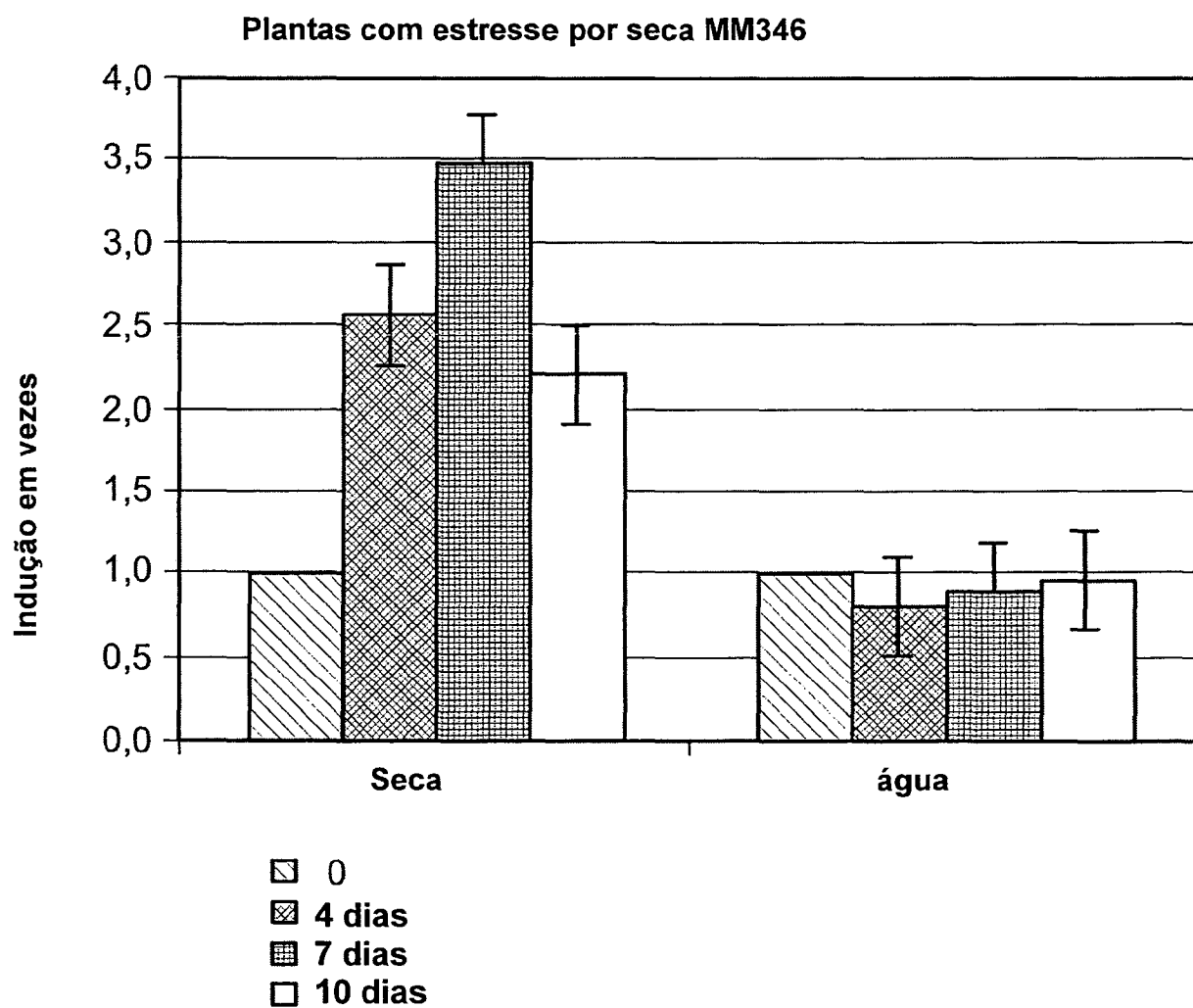


FIG.4A

Seqüência de gene RD29A

```

atggagtcac agttgacacg tccttatggt catgagcaag cagaagaacc aatcagaatt
caccatccag aagaagagca tcatgagaag ggagcatcca aagtgttgaa gaaagtaaaa
gaaaaggcta agaaaatcaa gaacagtctc actaaacatg gaaatggta tgatcacgat
gtggaagatg atgatgatga gtatgacgag caagaccag aagttcacgg cgcaccagtg
tatgaatcct ctgccgtgag aggtgggtgta acgggtaaac ctaagtctct tagtcatgcc
ggagaaacta atgttccggc atcggaggag attgttcctc cagggacaaa agtttttctt
gtcgtgtctt ctgaccacac caaaccatt gagcctgtat cattacaaga tacctcttac
ggacatgagg cactggctga tcctgtaaga acgacggaaa catcgactg ggaagcgaaa
agagaggcac cgactcatta tcctctcggg gtgtcagaat tttcagacag aggagagagc
agagaggctc atcaagagcc attgaacact cctgtgtctc tgctttcagc aacagaggac
gtgactagga cgtttgctcc tgggtggta gatgactatc tcggtggta acggaaagtc
aacgtcgaga cgccaaaacg tttggaggaa gatccggctg ctccaggagg aggatcggat
tatctcagtg gtgtatctaa ttatcagtcc aaagttactg atcccacgca taaaggtgga
gaagctggag taccagagat tgctgagctc cttggtagaa tgaaagtac tgatgagctt
cctgatcaga aatcaagaca aggacgcgaa gaagactttc cgacgagaag ccatgagttt
gatctgaaga agaatctga tatcaacaag aattctccgg caagatttg aggggaatca
aaagctggga tggaggaaga ttttccgaca agaggtgatg tgaaagtaga gattggattg
ggaagagact taccgacggg aactcatgat cagttctcac cagaactatc tcgtcccaaa
gagagagatg attctgagga aaccaaagat gattcgacac atgagacaaa accaagcacc
tacacagagc agttagcttc agctacatca gccataacta acaaagctat agccgcaag
aacgtcgttg cctcaaagct aggttacacc ggagagaatg gcggcgggca aagcgagagc
cctgtaaaaag atgaaactcc gagatctgtt actgcttacg ggcagaaaag ggcgggaact
gttgctgaga agttgactcc ggtttacgaa aaagtcaaag aaacaggatc aacggtgatg
acaaagctac ctctctccgg aggtggaagt ggagtgaagg agacgcaaca aggggaagag
aaaggtgtga cggctaaaaa ttatatatca gagaagctga aacctggaga agaggacaaa
gctttatcgg aatgatagc tgagaaactt cattttggag gaggaggaga gaagaagaca
acggctacaa aggaggtgga agtgacggtt gagaagatac cttccgacca gatagcggag
gggaaaggac atggtgaggc ggttgacagag gaaggaaaag gtggagaagg aatggtgggg
aaagttaaag gagcggtcac ttcttggtc ggtggtaaac cgaagtcgcc acggtccggt
gaagagtctc cacaatcact tggcaccacc gttgggacta tggggttttc ggattccggt

```

Seqüência de proteína Rd29A

```

MESQLTRPYGHEQAEPIRIHHPEEEHHEKGASKVLKKVKEKAK
KIKNSLTKHNGNHDHVEDDDDEYDEQDPEVHGAPVYESSAVRGGVTGKPKSLSHAGE
TNVPASEEIVPPGTVKFPVSSDHTKPIEPVSLQDTSYGHEALADPVRTTETSDWEAK
REAPTHYPLGVSEFSRDRGESREAHQEPLNTPVSLLSATEDVTRTFAPGGEDDYLGQR
KVNVTGPKRLEEDPAAPGGGSDYLSGVSNYQSKVTDPTHKGGEAGVPEIAESLGRMKV
TDESPDQSRQGREEDFPTRSHFDLKKESDINKNSFARFGGESKAGMEEDFPTRGDV
KVESGLGRDLPTGTHDQFSPELSRPKERDDSEETKDESTHETKTPSTYTEQLASATSAI
TNKAI AAKNVVASKLGYTGENGGSQSESPVKDETPRSVTAYGQKVAGTVAEKLTPVYE
KVKETGSTMVKLPLSGGSGVKETQQGEEKGVTAKNYISEKLKPGEEKKALSEMIAE
KLHFGGGGEKKTATKEVEVTVEKIPSDQIAEGKGHGEAVAEKGKGGMVGKVKGAV
TSWLGGKPKSPRSVEESPQSLGTTVGTMGFSDSGGSELGGSGGKGKVQDSGN

```

FIG.4B

Promotor de gene cor78 de proteína induzida por baixa temperatura

CGCGCCGCAAGAATCTCAAACACGGAGATCTCAAAGTTTGAAAGAAAATTTATTTCTTCGACTCAAAACAAACTT
 ACGAAATTTAGGTAGAACTTATATACATTATATTGTAATTTTTTGTAACAAAATGTTTTATTATTATTATAGAA
 TTTTACTGGTTAAATTAAAAATGAATAGAAAAGGTGAATTAAGAGGAGAGAGGAGGTAAACATTTTCTTCTATTT
 TTTCATATTTTCAGGATAAATTATGTAAAAGTTTACAAGATTTCCATTTGACTAGTGTAATGAGGAATATTCT
 CTAGTAAGATCATTATTTTCATCTACTTCTTTTATCTTCTACCAGTAGAGGAATAACAATATTTAGCTCCTTTGT
 AAATACAAATTAATTTTCGTTCTTGACATCATTCAATTTTAATTTTACGTATAAAATAAAAGATCATACCTATTA
 GAACGATTAAGGAGAAAATACAATTGGAATGAGAAGGATGTGCCGCTTGTTATAATAAACAGCCACACGACGTAAA
 CGTAAAATGACCACATGATGGGCCAATAGACATGGACCGACTACTAATAATAGTAAGTTACATTTTAGGATGGAA
 TAAATATCATACCGACATCAGTTTGAAAGAAAAGGGAAAAAAGAAAAATAAATAAAAGATATACTACCGACAT
 GAGTTCCAAAAGCAAAAAAAGATCAAGCCGACACAGACACGCGTAGAGAGCAAAATGACTTTGACGTCACAC
 CACGAAAACAGACGCTTCATACGTGTCCCTTTATCTCTCTCAGTCTCTCTATAAACTTAGTGAGACCCTCCTCTG
 TTTTACTCACAAATTTAAT

Seqüência 78 de proteína induzida por baixa temperatura

MDQTEEPPLNTHQQHPPEVEHHENGATKMFRKVKARAKKFKNSL
 TKHGQSNEHEQDHDLVEEDDDDELEPEVIDAPGVTGKPRETNVPASEEIIPPGTKVF
 PVVSSDYTKPTESVPVQEAISYGHDAHAHSVRTTFTSDKEEKRDVPIHHPLSELSDREE
 SRETHHESLNPVSVLLSGTEDVTSTFAPSGDDEYLDGQRKVNVPITILEEESAVSDY
 LSGVSNYQSKVTDPTKEETGGVPEIAESFGNMEVTDDESPDQKPGQFERDLSTRSKEFK
 EFDQDFDSVLGKDSPAIFPGESGVVFPVGFGEDESGAELEKDFPTRSHDFDMKTETGMD
 TNSPSRSHEFDLKTESGNDKNSPMGFGSESGAELEKEFDQKNDSGRNEYSPESDGGLG
 APLGGNFPPVRSHELDLKNESDIDKDVPTGFDGEPDFLAKGRPGYGEASEEDKFPARSD
 DVEVETELGRDPKTETLDQFSPELSHPKERDEFKESRDDFEETRDEKTEEPKQSTYTE
 KFASMLGYSGEIPVGDQTQVAGTVDEKLTVPNEKDQETESAVTTKLPISGGGSGVVEEQ
 RGEDKSVSGRDYVAEKLTTTEEDKAFSDMVAEKLQIGGEEBKETTTKEVEKISTEKA
 ASEEGEAVEEEVKGGGMVGRIKGFVGGGATDEVKPESPHSVVEEAPKSSGWFGGGATE
 EVKPKSPHSVEESPQSLGSTVVPVQKEL

FIG.4C: Gene cor78 proteína induzida por baixa temperatura

```

1 gatctcaaaag tttgaaagaa aattttatctc ttcgactcaa aacaaactta cgaaatttag
61 gtagaactta tatacattat attgtaattt tttgtaacaa aatgttttta ttattattat
121 agaattttac tggttaaatt aaaaatgaat agaaaagggtg aattaagagg agagaggagg
181 taaacatttt cttctatctt ttcataatctt caggataaat tattgtaaaa gtttacaaga
241 tttccatttg actagtgtaa atgaggaata ttctctagta agatcattat ttcactact
301 tcttttatct tctaccagta gaggaataaa caatatcttag ctctttgtta aatacaaat
361 aattttcgtt cttgacatca ttcaatttta attttctgtat aaaataaaag atcataccta
421 ttagaacgat taaggagaaa tacaattcga atgagaagga tgtgccgttt gttataataa
481 acagccacac gacgtaaacg taaaatgacc acatgatggg ccaatagaca tggaccgact
541 actaataata gtaagttaca ttttaggatg gaataaatat cataccgaca tcagtttgaa
601 aagaaaaagg gaaaaaaaga aaaaataaat aaaagatata ctaccgacat gaggttccaaa
661 aagcaaaaaa aaagatcaag ccgacacaga cagcgctaga gagcaaaatg actttgacgt
721 cacaccacga aaacagacgc ttcatatcgtg tccctttatc tctctcagtc tctctataaa
781 cttagtgtaga ccctcctctg ttttactcac aaatatgcaa actagaaaac aatcatcagg
841 aataaagggt ttgattactt ctattggaaa gaaaaaaact ttggaaaatg gatcaaacag
901 aggaaccacc actcaacaca caccagcagc acccaggtag attctaattt cagaaactta
961 tatttttttt aagtgacaat cctctgaatt tacttaaaact tattgtgatt tatggatata
1021 gaagaagttg aacatcatga gaatggtgag actaagatgt ttaggaaagt aaaggctaga
1081 gctaagaagt tcaagaacag tctcactaaa catggacaaa gcaatgagca tgagcaagat
1141 catgatttgg ttgaagaaga tgatgatgat gacgagctag aacctgaagt gatcgatgca
1201 ccaggttact ttcttttgta gtttcattca acttatgata taaaaactat tggttattca
1261 atttttgcgtg acattaacgg tttggtatct gtatatgcag cgtaacaggt aaacctagag
1321 aaactaatgt tccagcatcg gaggaatata ttccaccagg gacaaaagggtg tttcctgtcg
1381 tgtcttccga ttacacccaa cccactgaat ctgtaccagt acaaggaggcc tcttacggac
1441 acgatgcacc ggctcattct gtaaggacga cgtttacatc ggacaaggaa gagaaaagag
1501 atgtaccgat tcatcatcct ctgtccgaat tgtcagacag agaagagagt agagagactc
1561 atcatgagtc attgaacact ccggtctctc tgctttctgg aacagaggat gtaacgagta
1621 cgtttgcctcc aagtgggtgat gatgaataatc ttgatgggtc acggaaggtc aacgtcgaga
1681 ccccgataac gttggaggaa gagtcggctg tttcagacta tcttagtggg gtatctaatt
1741 atcagtccaa agttactgat cccacccaaag aaggtaagaa ctttgacctt ttaagattgt
1801 gtttttcttt agtgattatg aatatgtaat aactctgtta cgttgtgttt ggtttagaaa
1861 ctggaggagt accggagatt gctgagtcct ttggtaatat ggaagtgact gatgagctc
1921 ctgacagaaa gcccagacaa tttgaaagag acttgtcgac gagaagcaaa gaattccaa
1981 agtttgatca ggactttgac tctgttctcg gtaaggattc gccggcgaaa tttccagggtg
2041 aatcaggagt tgttttcccg gtgggctttg gtgacgagtc aggagctgag ctggaaaaag
2101 attttccgac gagaagtcac gatttttgata tgaagactga aactggaatg gacacgaatt
2161 ctccatcaag aagccatgaa tttgatctga agactgaatc tggaaacgac aagaattctc
2221 cgatgggctt tggtagtgaa tcaggagctg agctggaaaa agaatttgat cagaagaacg
2281 attctggaa gaaacagtat tcgcccgaat ctgacggcgg tttaggagct ccggtcgga
2341 gaaattttcc ggtgagaagt catgagttgg atctgaagaa cgaatctgat atcgacaagg
2401 atgtgccgac gggatttgac ggagaaccag attttctggc gaagggaaga cctggatacg
2461 gtgaggcatc agaaggagat aaatttccgg cgagaagtga tgatgtggaa gtagagactg
2521 agctgggaag agaccctaa gacggagactc ttgatcaatt ctaccggaa ctttctcatc
2581 ctaaaagaaa agatgagttt aaggagtcca gagatgattt tgaggagacg agagatgaga
2641 aaacagagga gccaaaacag agcacttaca cagagaagtt tgcttcaatg ctaggttact
2701 ccggagaaat tccggtggga gatcaaaactc aagtggcggg aactgttgat gagagggtga
2761 ctccggtcaa tgagaaggat caagaaacag agtctgccgt gacgacgaag ttacctatct
2821 ccggagggtg aagtggagta gaggagcaac gaggggaaga taaaagtgtg tccggtagag
2881 attatgtggc ggagaaactg acaactgaag aagaagacaa agccttttct gatatggttg
2941 ccgagaaact tcagattgga ggagaagaag agaagaagga aacgacgaca aaggaaagtg
3001 agaagatctc taccgagaag gcagcatcgg aggagggtga ggcggtgga gagggaagtga
3061 aaggaggagg aggaatggtt gggaggatta aaggatggtt cgggtggtgt gcgactgatg
3121 aggtgaagcc agaatcgcca cattctgttg aaggagctcc aaaatcatct ggctgggttg
3181 gtgggtggtg gacggaggag gtgaagccaa aatcgctca ttcggttgaa gagtctccac
3241 aatcacttgg ctccactgtt gttccggtgc agaaggagct ttaagaatat gagaactgag
3301 attttcaagt ttacttttg atgtttatgt gtgttttgtt tgactctttt gatgtattat
3361 ggtataattc cttgtttgtg tgaaaaaagg acatttggtt aataaattgt tccgcttttg
3421 attaagaagt tctccatac cagctactag gtctaagtg ggtaaaatca ttggatttat
3481 tcccttcaaa gttcttagaa ttattcacag gatttacatt atgagctagt agt

```

FIG.5

Seqüência de gene MET1

```
atgtcgtgct gcggaggaaa ctgcgggtgc ggatccggct gccagtgcgg cagcggctgc  
ggaggatgca agatgtaccc ggagatggct gaggaggtga ccactacca gactgtcatc  
atgggtgttg caccttcaa gggtcatgcc gaggggttg aggccggcgc cgccgccgga  
gcaggagcag agaacgggtg caagtgcggc gacaactgca cctgcaacc ctgcaactgc  
ggcaagtga
```