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

(21) **PI0607378-6 A2**

BRPI0607378A2

(22) Data de Depósito: 24/02/2006
(43) Data da Publicação: 23/03/2010
(RPI 2046)

(51) Int.Cl.:
C12N 15/82 (2010.01)
C12N 5/10 (2010.01)
A01H 5/00 (2010.01)
C12Q 1/68 (2010.01)

(54) Título: **CASSETES DE EXPRESSÃO, SEQÜÊNCIA NUCLEOTÍDICA ISOLADA, VETOR, CÉLULA HOSPEDEIRA TRANSGÊNICA OU ORGANISMO NÃO HUMANO, CÉLULA DE PLANTA OU PLANTA TRANSGÊNICA, MÉTODOS PARA IDENTIFICAR E/OU ISOLAR UMA SEQÜÊNCIA NUCLEOTÍDICA REGULADORA DA TRANSCRIÇÃO, PARA PROVER OU PRODUZIR UM CASSETE DE EXPRESSÃO TRANSGÊNICA, E PARA PROVER UMA SEQÜÊNCIA NUCLEOTÍDICA REGULADORA DE TRANSCRIÇÃO SINTÉTICA, E, SEQÜÊNCIA REGULADORA DE TRANSCRIÇÃO SINTÉTICA**

(30) Prioridade Unionista: 26/02/2005 EP 05004248.0

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

(72) Inventor(es): Elke Duwenig, Linda Patricia Loyall

(74) Procurador(es): Momsen, Leonardos & CIA.

(86) Pedido Internacional: PCT EP2006060266 de 24/02/2006

(87) Publicação Internacional: WO 2006/089950de 31/08/2006

(57) Resumo: CASSETES DE EXPRESSÃO, SEQUENCIA NUCLEOTIDICA ISOLADA, VETOR, CÉLULA HOSPEDEIRA TRANSGÊNICA OU ORGANISMO NÃO HUMANO, CÉLULA DE PLANTA OU PLANTA TRANSGÊNICA, MÉTODOS PARA IDENTIFICAR E/OU ISOLAR UMA SEQUENCIA NUCLEOTÍDICA REGULADORA DA TRANSCRIÇÃO, PARA PROVER OU PRODUZIR UM CASSETE DE EXPRESSÃO TRANSGÊNICA, E PARA PROVER UMA SEQÜÊNCIA NUCLEOTÍDICA REGULADORA DE TRANSCRIÇÃO SINTÉTICA, E, SEQÜÊNCIA REGULADORA DE TRANSCRIÇÃO SINTÉTICA. A presente invenção refere-se a cassetes de expressão compreendendo seqüéncias nucleotídicas reguladoras de transcrição com perfis de expressão preferenciais de semente ou específicos de semente em plantas obteníveis a partir do gene de Arabidopsis thaliana descrito pelo locus Atlg48130 ou seus ortólogos de outras espécies de plantas (como Linum usitatissimum). As seqüéncias nucleotídicas reguladoras de transcrição preferivelmente demonstram uma forte atividade de expressão especialmente em sementes e tecidos de semente.

PI0607378-6
PET: 020070140642

"CASSETES DE EXPRESSÃO, SEQÜÊNCIA NUCLEOTÍDICA ISOLADA, VETOR, CÉLULA HOSPEDEIRA TRANSGÊNICA OU ORGANISMO NÃO HUMANO, CÉLULA DE PLANTA OU PLANTA TRANSGÊNICA, MÉTODOS PARA IDENTIFICAR E/OU ISOLAR UMA SEQÜÊNCIA NUCLEOTÍDICA REGULADORA DA TRANSCRIÇÃO, PARA PROVER OU PRODUZIR UM CASSETE DE EXPRESSÃO TRANSGÊNICA, E PARA PROVER UMA SEQÜÊNCIA NUCLEOTÍDICA REGULADORA DE TRANSCRIÇÃO SINTÉTICA, E, SEQÜÊNCIA REGULADORA DE TRANSCRIÇÃO SINTÉTICA"

CAMPO DA INVENÇÃO

A presente invenção refere-se a cassetes de expressão compreendendo seqüências nucleotídicas reguladoras de transcrição com perfis de expressão preferencial de semente ou específica de semente em plantas obteníveis a partir do gene de *Arabidopsis thaliana* descrito pelo *locus* At1g48130 ou seus ortólogos de outras espécies de plantas (como *Linum usitatissimum*). As seqüências nucleotídicas reguladoras de transcrição preferivelmente demonstram uma atividade de expressão especialmente em sementes e tecidos de semente.

ANTECEDENTES INVENÇÃO

Manipulação de plantas para alterar e/ou melhorar as características fenotípicas (como produtividade ou qualidade) requer a expressão de genes heterólogos em tecidos de plantas. Esta manipulação genética baseia-se na disponibilidade de um meio para dirigir e controlar a expressão do gene como requerido. Por exemplo, manipulação genética baseia-se na disponibilidade e uso de promotores apropriados que são efetivos em plantas e que regulam expressão de gene de modo a dar os desejados efeito(s) na planta transgênica.

Os promotores preferenciais de semente ou específicos de semente são úteis para expressar genes assim como para produzir quantidades

grandes de proteína, para expressar óleos ou proteínas de interesse, por exemplo, anticorpos, genes para aumentar o valor nutricional da semente e semelhantes. É vantajoso assim escolher uma variedade de promotores diferentes de modo que o promotor mais apropriado possa ser selecionado para um gene em particular, construção, célula, tecido, planta ou meio ambiente. Além disso, o interesse crescente na co- transformação de plantas com unidades múltiplas de transcrição de plantas (PTU) e os problemas em potencial associados com o uso de seqüências regulatórias comuns para estes fins merecem ocasionar a disponibilidade de várias seqüências de promotor.

Assim, se tem uma grande necessidade na arte para a identificação de novas seqüências que podem ser usadas para a expressão de transgenes selecionados em plantas economicamente importantes. Assim é um objetivo da presente invenção prover novos e alternativos cassetes de expressão para expressão preferencial de semente ou específica de semente de transgenes em plantas. O objetivo é resolvido pela presente invenção.

SUMÁRIO DA INVENÇÃO

Uma primeira forma de realização da invenção refere-se a um cassete de expressão para regular a expressão preferencial de semente ou específica de semente em plantas compreendendo

i) pelo menos uma seqüência nucleotídica reguladora da transcrição de um gene codificando uma proteína peroxiredoxina (PER1), e funcionalmente ligada à mesma,

ii) pelo menos uma seqüência de ácido nucleico que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

Preferivelmente, o cassete de expressão da invenção compreende

a) pelo menos uma seqüência nucleotídica reguladora da transcrição de um gene de planta, referido gene de planta selecionado dentre o

grupo consistindo de

i) o gene descrito pelo *loci* do genoma de *Arabidopsis thaliana* do GenBank At1g48130, e

ii) genes ortólogos do gene descrito pelo *loci* do genoma de *Arabidopsis thaliana* do GenBank At1g48130,

e funcionalmente ligado ao mesmo

b) pelo menos uma seqüência de ácido nucleico que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

A seqüência nucleotídica reguladora da transcrição pode ser obtida ou é obtenível a partir de DNA genômico de planta de um gene codificando um polipeptídeo que

a) compreende pelo menos um motivo de seqüência selecionado dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL (SEQ ID NO: 37), preferivelmente MPG(I/L)T(L/I)GDTVPNLE (SEQ ID NO: 38),

ii) GDFTPVC (SEQ ID NO: 39), preferivelmente LFSHPGDFTPVCTTEL (SEQ ID NO: 40),

iii) VKLLGLS (SEQIDNO:41), preferivelmente RGVKLLGLSCDD (SEQ ID NO: 42),

iv) YPI(I/M)ADP (SEQ ID NO: 43), preferivelmente SKV(T/N)YPI(I/M)ADP (SEQ ID NO: 44),

v) KLSFLYP (SEQ ID NO: 45), preferivelmente (K/V)(I/V)KLSFLYPS (SEQ ID NO: 46),

vi) GRNMDEV (SEQ ID NO: 47), preferivelmente TGRNMDEV(L/V)RA (SEQ ID NO: 48),

vii) (I/V)ATP(V/A)NW (SEQ ID NO: 49), preferivelmente K(I/V)ATP(V/A)NW(K/N)P (SEQ ID NO: 50), e

viii) PSKKGYL (SEQ ID NO: 51), preferivelmente

LPSKKGYLR (SEQ ID NO: 52),

ou

b) tem pelo menos 50% identidade de seqüência de aminoácido para um polipeptídeo selecionado dentre o grupo descrito por SEQ ID NO: 13, 15, 17, 19, 21, e 23.

Preferivelmente, o cassete de expressão compreende a seqüência nucleotídica reguladora da transcrição, que é obtida ou é obtenível a partir de DNA genômico de planta de um gene codificando um polipeptídeo (preferivelmente uma proteína ortóloga) que

a) compreende pelo menos um motivo de seqüência selecionado dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL (SEQ ID NO: 37), preferivelmente MPG(I/L)T(L/I)GDTVPNLE (SEQ ID NO: 38),

ii) GDFTPVC (SEQ ID NO: 39), preferivelmente LFSHPGDFTPVCTTEL (SEQ ID NO: 40),

iii) VKLLGLS (SEQ ID NO: 41),

preferivelmente RGVKLLGLSCDD (SEQ ID NO: 42),

iv) YPI(I/M)ADP (SEQ ID NO: 43), preferivelmente SKV(T/N)YPI(I/M)ADP (SEQ ID NO: 44),

v) KLSFLYP (SEQ ID NO: 45), preferivelmente (K/V)(I/V)KLSFLYPS (SEQ ID NO: 46),

vi) GRNMDEV (SEQ ID NO: 47), preferivelmente TGRNMDEV(L/V)RA (SEQ ID NO: 48),

vii) (I/V)ATP(V/A)NW (SEQ ID NO: 49), preferivelmente K(I/V)ATP(V/A)NW(K/N)P (SEQ ID NO: 50), e

viii) PSKKGYL (SEQ ID NO: 51), preferivelmente LPSKKGYLR (SEQ ID NO: 52),

e

b) tem pelo menos 50% identidade de seqüência de

aminoácido para um polipeptídeo selecionado dentre o grupo descrito por SEQ ID NO: 13, 15, 17, 19, 21, e 23.

Preferivelmente, referida proteína ortóloga tem ainda a mesma atividade enzimática que a da proteína codificada por *locus* de *Arabidopsis thaliana* At1g48130.

Preferivelmente, a seqüência nucleotídica reguladora da transcrição é selecionada dentre o grupo de seqüências consistindo de

i) a seqüência descrita por SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11,

ii) um fragmento de pelo menos 50 bases consecutivas de uma seqüência sob i),

iii) uma seqüência nucleotídica tendo similaridade substancial com uma identidade de seqüência de pelo menos 50% para uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11;

iv) uma seqüência nucleotídica capaz de hibridizar (preferivelmente sob condições equivalentes à hibridização em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C) para uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11, ou o seu complemento;

v) uma seqüência nucleotídica capaz de hibridizar

(preferivelmente sob condições equivalentes à hibridização em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 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 de uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11, ou o seu complemento;

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

Em uma forma de realização preferidas, as seqüências especificadas sob ii), iii), iv) v) e vi) acima são capazes de modificar transcrição em uma célula de planta ou organismo, preferivelmente referidas seqüências tem substancialmente a mesma atividade reguladora de transcrição como a seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11.

Preferivelmente, uma seqüência nucleotídica tendo similaridade substancial a uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11 tem uma identidade de seqüência de pelo menos 50% ou 60%, preferivelmente pelo menos 70% ou 80%, mais preferivelmente pelo menos 90% ou 95%, o mais preferivelmente pelo menos 98% to uma seqüência descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11.

Preferivelmente, hibridização é realizada sob condições estringentes (incluindo condições de estringência baixa e elevada), mais preferivelmente sob condições de estringência elevada.

A seqüência nucleotídica heteróloga a ser expressada em preferivelmente ainda mais operativamente ligada regiões 3' não traduzidas, terminação transcrição e/ou sinais de poliadenilação. As regiões -não traduzidas 3' são apropriadas para estabilizar a expressão e estrutura de mRNA. Isto pode resultar em presença prolongada do mRNA e assim níveis de expressão melhorados. A terminação e sinais de poliadenilação são apropriados para estabilizar a expressão de mRNA, para assegurar um comprimento do transcrito de mRNA constante e para evitar a transcrição de leitura completa. Especialmente em construções de expressão de multigenes, isto é um aspecto importante. Além disso, a terminação correta de transcrição está ligada a uma re-iniciação de transcrição da seqüência de nucleotídeo 5' regulatória resultando em níveis de expressão melhorados. Os sinais acima mencionados podem ser qualquer sinal funcional em plantas e pode por exemplo ser isolado de genes de planta, genes de vírus de plantas, ou outros patógenos de plantas. No entanto, em uma forma de realização preferida, as regiões não traduzidas 3', terminação de transcrição e sinais de poliadenilação são de genes empregados como a fonte para os promotores desta invenção.

O cassete de expressão pode ser empregado para numerosos fins de expressão, como por exemplo expressão de uma proteína, ou expressão de um RNA anti-sentido, RNA sentido ou filamento duplo. Preferivelmente, a expressão da seqüência de ácido nucleico confere à planta um traço agronomicamente valioso.

Outras formas de realização da invenção referem-se a vetores compreendendo um cassete de expressão da invenção, e célula hospedeira transgênica ou organismo não humano compreendendo um cassete de expressão ou um vetor da invenção. Preferivelmente o organismo é uma

planta. Mais preferivelmente uma planta usada para produção de óleo como – por exemplo - *Brassica napus*, *Brassica juncea*, *Linum usitatissimum*, soja, ou girassol.

Outra forma de realização preferidos da invenção refere-se a um método para identificar e/ou isolar a seqüência nucleotídica reguladora da transcrição caracterizado em que referida identificação e/ou isolamento utiliza a seqüência de ácido nucleico codificando um polipeptídeo como descrito por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de pelo menos 15 nucleotídeos consecutivos de referida seqüência de ácido nucleico. Preferivelmente a seqüência de ácido nucleico utilizada para o isolamento é descrita por SEQ ID NO: 12, 14, 16, 18, 20, ou 22 ou uma parte de pelo menos 15 nucleotídeos consecutivos da mesma. Mais preferivelmente, referida identificação e/ou isolamento é realizado por um método selecionado dentre reação em cadeia da polimerase, hibridização, e triagem de base de dados.

Outra forma de realização da invenção refere-se a um método para prover ou produzir um cassete de expressão transgênica para expressão heteróloga em plantas compreendendo as etapas de:

I. isolar uma seqüência nucleotídica reguladora da transcrição de um gene de planta utilizando pelo menos uma seqüência de ácido nucleico codificando um polipeptídeo descrito por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de pelo menos 15 nucleotídeos consecutivos de referida seqüência de ácido nucleico, e

II. funcionalmente ligar referida seqüência nucleotídica reguladora da transcrição a outra seqüência nucleotídica de interesse, que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

Outra forma de realização da invenção refere-se a um método para prover um cassete de expressão transgênica para expressão preferencial

de semente ou específica de semente compreendendo as etapas de:

I. isolar de uma seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente utilizando pelo menos uma seqüência de ácido nucleico ou uma parte da mesma, em que referida seqüência está codificando um polipeptídeo descrito por SEQ ID NO: 8, 10,12, 14, 16, 18, ou 36, ou uma parte de pelo menos 15 nucleotídeos consecutivos do mesmo, e

II. funcionalmente ligar referida seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente a outra seqüência nucleotídica de interesse, que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente.

Preferivelmente, a seqüência nucleotídica utilizada para isolamento de referida seqüência nucleotídica reguladora da transcrição está codificando um polipeptídeo compreendendo pelo menos um motivo de seqüência selecionado dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL (SEQ ID NO: 37), preferivelmente MPG(I/L)T(L/I)GDTVPNLE (SEQ ID NO: 38),

ii) GDFTPVC (SEQ ID NO: 39), preferivelmente LFSHPGDFTPVCTTEL (SEQ ID NO: 40),

iii) VKLLGLS (SEQ ID NO: 41),

preferivelmente RGVKLLGLSCDD (SEQ ID NO: 42),

iv) YPI(I/M)ADP (SEQ ID NO: 43), preferivelmente SKV(T/N)YPI(I/M)ADP (SEQ ID NO: 44),

v) KLSFLYP (SEQ ID NO: 45), preferivelmente (K/V)(I/V)KLSFLYPS (SEQ ID NO: 46),

vi) GRNMDEV (SEQ ID NO: 47), preferivelmente TGRNMDEV(L/V)RA (SEQ ID NO: 48),

vii) (I/V)ATP(V/A)NW (SEQ ID NO: 49), preferivelmente K(I/V)ATP(V/A)NW(K/N)P (SEQ ID NO: 50), e

viii) PSKKGYL (SEQ ID NO: 51), preferivelmente LPSKKGYLR (SEQ ID NO: 52).

Algumas das seqüências reguladoras de transcrição aqui providas são novas como tal e não foram previamente descritas na arte. Assim, outra forma de realização da invenção refere-se a uma seqüência isolada nucleotídica (preferivelmente tendo atividade reguladora de transcrição, mais preferivelmente tendo atividade de promotor) selecionada dentre o grupo de seqüências consistindo de:

a) a seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e seqüências tendo uma identidade de pelo menos 99%, preferivelmente 99,5%, mais preferivelmente 99,8% para uma seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e seqüências compreendendo pelo menos 600 nucleotídeos consecutivos, preferivelmente pelo menos 800 nucleotídeo consecutivo, mais preferivelmente pelo menos 900 nucleotídeos consecutivos, mais preferivelmente pelo menos 1000 nucleotídeos consecutivos de uma seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e

b) a seqüência como descrita por SEQ ID NO: 9, 10, ou 11, e seqüências tendo uma identidade de pelo menos 40% ou 50%, preferivelmente 60% ou 70%, mais preferivelmente 80 ou 90%, o mais preferivelmente 95% ou 98% para uma seqüência como descrita por SEQ ID NO: 5 ou 26, e seqüências compreendendo pelo menos 15 ou 20 nucleotídeos consecutivos, preferivelmente pelo menos 50 ou 100 nucleotídeo consecutivo, mais preferivelmente pelo menos 250 nucleotídeos consecutivos, mais preferivelmente pelo menos 500 nucleotídeos consecutivos de uma seqüência como descrita por SEQ ID NO: 9, 10, ou 11.

DESCRIÇÃO DOS DESENHOS

Fig. 1: Alinhamento da proteína peroxiredoxina putativa

(PER1) (At1g48130 proteína) de *Arabidopsis* com seus ortólogos de várias espécies incluindo *Brassica napus, Triticum turgidum, Hordeum vulgare, Bromus secalinus e Linum usitatissimum.*

Fig 2: Seqüência da seqüência nucleotídica reguladora da transcrição de *Linum usitatissimum.* As três regiões de promotor com densidade diferentes de motivos de promotor estão em caixas (A, B, C). Para explanação detalhada dos motivos sublinhados, ver exemplo 8.

DEFINIÇÕES

Deve-se entender que esta invenção não é limitada à metodologia particular, protocolos, linhagens de células, espécies de plantas, ou gêneros, construções, e reagentes descritos como tal. Também deve-se entender que a terminologia usada aqui é para o fim de descrever formas de realização particulares somente, e não se destina a limitar o escopo da presente invenção que será limitada somente pelas reivindicações anexas. Deve-se notar que como usado aqui e nas reivindicações, as formas singulares "um" , "e" e "o" incluem referências no plural, salvo se o contexto ditar claramente de outro modo. Assim, por exemplo, referência a um "vetor" é uma referência a um ou mais vetores e inclui seus equivalentes bem conhecidos dos versados, 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 dados. Geralmente, o termo "cerca de" é usado para modificar um valor numérico acima e abaixo do valor dado por uma variância de 20 por cento, preferivelmente 10 por cento para cima ou para baixo (maior ou menor).

Como usado aqui, a palavra "ou" significa qualquer número de uma lista em 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 seqüências de codificação e/ou as seqüências 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 seqüências regulatórias. Os genes também incluem segmentos de DNA não expressados que, por exemplo, formam seqüências de reconhecimento para outras proteínas. Os genes podem ser obtidos de várias fontes, incluindo clonagem de uma fonte de interesse ou sintetizando de informação de seqüência conhecida ou prevista, e podem incluir seqüências projetadas para ter parâmetros desejados.

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

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

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

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

2) seqüências codificando partes de proteínas não naturalmente unidas, ou

3) partes de promotores que não são naturalmente unidas.

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

Um "transgene" se refere a um gene que foi introduzido no genoma por transformação e é estavelmente mantido. Os transgenes podem

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

Um "oligonucleotídeo" correspondendo a uma seqüência nucleotídica da invenção, por exemplo para uso em reações de amplificação ou sondagem, pode ter cerca de 30 ou menos nucleotídeos em comprimento (por exemplo, 9, 12, 15, 18, 20, 21, 22, 23, ou 24, ou qualquer número entre 9 e 30). Geralmente, os iniciadores específicos vão até 14 nucleotídeos no comprimento. Para uma especificidade ótima e eficácia de custo, os iniciadores de 16 a 24 nucleotídeos no comprimento podem ser preferidos. Os versados na arte são bem versados no projeto de iniciadores para uso nos processos como PCR. Se requerido, a sondagem pode ser feita com fragmentos de restrição completos do gene descrito aqui que pode ter 100's ou mesmo 1000's nucleotídeos em comprimento.

Os termos "polipeptídeo", "peptídeo", "oligopeptídeo", "polipeptídeo", "produto de gene ", "produto de expressão" e "proteína" são usados de modo interpermutável para fazer referência a um polímero ou oligômero de resíduos de aminoácidos consecutivos. Como usado aqui, o termo „seqüência de aminoácido" ou uma „seqüência de polipeptídeo" refere-se a uma lista de abreviaturas, letras, caracteres ou palavras representando resíduos de aminoácidos. Os aminoácidos podem ser referidos aqui por seus símbolos de três letras comumente conhecidos ou pelos símbolos de uma letra recomendados por IUPAC-IUB Biochemical Nomenclature Commission. As abreviaturas usadas aqui são os 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. Stcenteior, Biochemistry, 1988, W. H. Freeman and Company, novos York. A letra "x" como usada aqui dentro de uma seqüência de aminoácido pode fazer referência a qualquer resíduo de aminoácido.

"Seqüência de codificação" refere-se a uma seqüência de DNA ou RNA que codifica para uma seqüência de aminoácido específica e exclui as seqüências não de codificação. Ela pode constituir uma "seqüência de codificação não interrompida" , isto é, faltando um íntron, como em um cDNA ou pode incluir um ou mais íntrons ligados por junções de emenda apropriadas. Um "intron" é uma seqüência de RNA que está contida no transcrito primário mas que é removida através de 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 à seqüência de aminoácido codificada entre códons de iniciação de tradução e de terminação de uma seqüência de codificação. Os termos "códon de iniciação" e "códon de terminação" referem-se a uma unidade de três nucleotídeos adjacentes ("códon") em uma seqüência de codificação que especifica 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 a um produto resultante de transcrição catalisada por RNA polimerase de uma seqüência de DNA. Quando o transcrito de RNA está em uma cópia complemento perfeita da seqüência de DNA, ele é referido como o transcrito primário ou pode ser uma

seqüência de RNA derivada de processamento pós-transcripcional do transcrito primário e é referido como o RNA maduro. "RNA mensageiro" (mRNA) refere-se ao RNA que está sem íntrons 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.

"Seqüência nucleotídica reguladora da transcrição", "seqüências regulatórias ", e "seqüências regulatórias apropriadas", referem-se, cada, a seqüências de nucleotídeos influenciando a transcrição, processamento de RNA, ou estabilidade,e ou tradução da seqüência de nucleotídeo associada (ou funcionalmente ligada) a ser transcrita. A seqüência nucleotídica reguladora da transcrição pode ter várias localizações com a relação às seqüências nucleotídicas a serem transcritas. A seqüência nucleotídica reguladora da transcrição pode estar localizada a montante (seqüências não codificando 5') com, ou a jusante (seqüências não codificando 3') da base a ser transcrita (por exemplo, uma seqüência de codificação). As seqüências nucleotídicas reguladoras da transcrição podem ser selecionadas dentre o grupo consistindo de melhoradores, promotores, seqüências líder de tradução, íntrons, seqüências não traduzidas 5', seqüências não traduzidas 3', e seqüências de sinal de poliadenilação. Elas incluem seqüências naturais e sintéticas, assim como seqüências, que podem ser uma combinação de seqüências sintéticas e naturais. Como é notado acima, o termo "seqüência nucleotídica reguladora da transcrição" não é limitada aos promotores. No entanto, preferivelmente, uma seqüência nucleotídica reguladora da transcrição da invenção compreende pelo menos uma seqüência de promotor (por exemplo, uma seqüência localizada a montante do início de transcrição de um gene capaz de induzir a transcrição das seqüências a jusante). Em uma forma de realização preferida, a seqüência nucleotídica reguladora da transcrição da invenção compreende a seqüência 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 do referido gene também pode ser empregada. Como usado aqui, o termo "cis- elemento" ou "motivo de promotor" refere-se a um elemento regulador da transcrição cis-atuante que confere um aspecto de controle global da expressão do gene. 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 de um cis-elemento. Os promotores do processo contêm desejavelmente cis-elementos que podem conferir ou modular a expressão do gene. Os cis-elementos podem ser identificados por várias técnicas, incluindo análise de deleção, isto é, deleção de um ou mais nucleotídeos de terminação 5' ou interna para um promotor; análise de proteína de ligação a DNA, usando impressão digital DNAse I, interferência de metilação, testes de mobilidade – deslocamento de eletroforese, impressão digital genômica em vivo por PCR mediado por ligação, e outros testes convencionais, ou por análise de similaridade de seqüência de DNA com motivos cis-elementos conhecidos, por métodos de comparação de seqüência de DNA convencional. A estrutura fina de um cis-elemento pode ser ainda estudada por mutagênese (ou substituição) de um ou mais nucleotídeos ou 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 subseqüente.

"Seqüência de não codificação 5'" ou "seqüência não traduzida 5" ou "região" refere-se a uma seqüência nucleotídica localizada 5' (a montante) para a seqüência de codificação. Ela está presente no mRNA totalmente processado do códon de iniciação e pode afetar o processamento do transcrito primário para mRNA, estabilidade de mRNA ou eficácia de

tradução (Turner 1995).

"Seqüência de não codificação 3'" ou "seqüência não traduzida 3" ou "região" refere-se a uma seqüência nucleotídica localizada 3' (a jusante) a uma seqüência de codificação e inclui seqüências de sinal de poliadenilação e outras seqüências codificando sinais regulatórios capazes de afetar o processamento de mRNA ou expressão de genes. 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 de não codificação 3' é exemplificado por Ingelbrecht et al, 1989.

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

"Peptídeo de sinal" refere-se a uma extensão terminal mínima 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 "seqüência de sinal" refere-se a uma seqüência nucleotídica que codifica o peptídeo de sinal. O termo "peptídeo de trânsito" como usado aqui, refere-se a parte de um polipeptídeo expressado (preferivelmente para a 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 uma organela de célula (como os plastídeos (por exemplo, cloroplastos) ou mitocôndrias). O termo "seqüência de trânsito" refere-se a uma seqüência nucleotídica que codifica o peptídeo de trânsito.

"Promotor" refere-se a uma seqüência nucleotídica,

geralmente a montante (5') para sua seqüência de codificação, que controla a expressão da seqüência de codificação ao prover o reconhecimento para RNA polimerase e outros fatores requeridos para uma transcrição apropriada. "Promotor" inclui um promotor mínimo que é uma seqüência de DNA curta de uma caixa TATA e outras seqüências que servem para especificar o sítio de iniciação de transcrição, para a qual os elementos regulatórios são adicionados para controlar a expressão. "Promotor" também refere-se a uma seqüência nucleotídica que inclui um promotor mínimo mais elementos regulatórios que é capaz de controlar a expressão de uma seqüência de codificação ou RNA funcional. Este tipo de seqüência de promotor consiste de elementos a montante proximais e mais distais, os últimos elementos com freqüência referidos como melhoradores. Conseqüentemente, um "melhorador" é uma seqüência 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 um promotor. Ele é capaz de operar em ambas as orientações (normal ou movimentada) e é capaz de funcionar mesmo quando movimentada ou a montante ou a jusante do promotor. Ambos os melhoradores e outros elementos promotores a montante ligam as proteínas de ligação a DNA específicas da seqüência que mediam seus efeitos. Os promotores podem ser derivados em sua totalidade de um gene nativo, ou ser compostos de elementos diferentes, derivados de promotores diferentes encontrados na natureza, ou mesmo ser compostos de segmentos de DNA sintético. Um promotor também pode conter seqüências de DNA que estão envolvidas na ligação de fatores de proteína, que controlam a eficácia de iniciação de transcrição em resposta a condições fisiológicas ou de desenvolvimento.

O "sítio de iniciação" é a posição circundando o primeiro nucleotídeo que é parte da seqüência 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, outras seqüências codificando proteína na direção 3') são denominadas positivas, enquanto as seqüências a montante (principalmente das regiões de controle na direção 5') são denominadas negativas.

Os elementos promotores, particularmente um elemento TATA, que são inativos ou que tem uma 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 iniciação de transcrição, por exemplo, caixa TATA e/ou um iniciador.

"Expressão constitutiva" refere-se à expressão usando um promotor constitutivo ou regulado. "Expressão condicional" e "regulada" refere-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 plantas durante todos ou quase todos os estágios de desenvolvimento da planta. Cada um dos elementos de ativação de transcrição não demonstra uma especificidade de tecido absoluta, mas mediam a ativação de transcrição na maioria das partes de plantas 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 a promotores que dirigem a expressão de genes não constitutivamente, mas em um modo regulado de modo temporal e/ou espacial, e inclui tanto promotores específicos de tecido e indutíveis. Ele inclui seqüências naturais e sintéticas, assim como seqüências que podem ser uma combinação de seqüências sintéticas e naturais. Promotores diferentes podem dirigir a expressão de um gene em tecidos ou tipos de células diferentes, ou em estágios diferentes de desenvolvimento, ou

em resposta a condições ambientais diferentes. Novos promotores de vários tipos utilizáveis em células de plantas estão sendo constantemente descobertos, cujos numerosos exemplos podem ser encontrados na compilação por Okamuro et al (1989). Os promotores regulados típicos utilizáveis em plantas incluem, mas não são limitados a promotores indutíveis por agente de segurança, promotores derivados de 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 para 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 órgãos específicos (como folhas ou sementes), tecidos específicos (como embrião ou cotilédones) ou tipos de células específicas (como parênquima de folha ou células de armazenamento de sementes). Estes também incluem promotores que são temporariamente regulados, como em embriogênese prematura ou tardia, durante o amadurecimento das frutas nas sementes ou frutas em desenvolvimento, em folhas totalmente 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.

"Operativamente ligado" ou "funcionalmente ligado" refere-se preferivelmente à associação de seqüências de ácido nucleico em fragmento de ácido nucleico único de modo que a função de um é afetada pela outra. 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 expressão da seqüência de DNA de codificação (isto é, que a seqüência de codificação ou RNA funcional está sob o controle transcripcional do promotor). As seqüências de codificação podem ser ligadas operativamente a seqüências regulatórias em orientação sentido ou anti-sentido.

"Expressão" refere-se à 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, expressão pode fazer referência à transcrição do DNA anti-sentido somente. Além disso, expressão refere-se á transcrição e acúmulo estável de RNA sentido (mRNA) ou funcional. A expressão também pode fazer referência à produção de proteína.

"Expressão específica" é a expressão de produtos de genes, que é limitada a um ou poucos tecidos de plantas (limitação espacial) e/ou a um ou alguns estágios de desenvolvimento de plantas (limitação temporal). Sabe-se que dificilmente existe uma especificidade verdadeira: promotores parecem preferivelmente mudar em alguns tecidos, enquanto em outros tecidos pode-se não ter atividade ou apenas uma pequena. Este fenômeno é conhecido como expressão "avariada" . No entanto, com expressão específica nesta invenção significa-se a expressão preferível em um ou em alguns tecidos de plantas.

O "padrão de expressão" de um promotor (com ou sem melhorador) é o padrão de níveis de expressão, que mostra onde na planta e em que 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 de "estado uniforme" de um mRNA repórter transcrito padrão. Esta medida é indireta porque a concentração do mRNA repórter é

dependente não somente de sua taxa de síntese, mas também da taxa em que o mRNA é degradado. Assim, o nível de estado uniforme é o produto de taxas de síntese e taxas de degradação. A taxa de degradação pode, no entanto, ser considerada como prosseguindo em uma taxa fixa quando as seqüências transcritas são idênticas, e assim este vapor pode servir como uma medida de taxas de síntese. Quando promotores são comparados deste modo, técnicas disponíveis para os versados na arte são análise S1-RNAse hibridização, 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 níveis de expressão de mRNA. A análise de pontos de partida de transcrição em praticamente todos os promotores revelou que não se tem geralmente base única em que a transcrição começa, mas ao contrário um conjunto mais ou menos agrupado de sítios de iniciação, cada um que perfaz alguns pontos de partida do mRNA. Porque esta distribuição varia de promotor a promotor, as seqüências do mRNA repórter em cada uma das populações iria diferir de cada outro. Porque cada espécie de mRNA é mais ou menos inclinado à degradação, nenhuma taxa de degradação única pode ser esperada para mRNAs repórter diferentes. Foi mostrado para várias seqüências de promotor eucariótico que a seqüência circundando o sítio de iniciação ("iniciador") desempenha um papel importante na determinação do nível de expressão de RNA dirigida por este promotor específico. Isto também inclui parte das seqüências transcritas. A fusão direta de seqüências de promotor para repórter deve assim levar a níveis sub-ótimos de transcrição. Um procedimento comumente usado para analisar os padrões de expressão e níveis é através da determinação do nível "de estado uniforme" de acúmulo de proteína em uma célula. Os candidatos comumente usados para o gene repórter, bem conhecidos dos versados na arte, são beta-glucuronidase (GUS), cloranfenicol acetil transferase (CAT) e proteínas com propriedades fluorescentes, como

proteína fluorescente verde (GFP) de Aequora victoria. Em princípio, no entanto, muitas outras proteínas são apropriadas para este fim, desde que a proteína não interfira com funções de plantas essenciais. Para quantificação e determinação de localização, várias ferramentas são apropriadas. Os sistemas de detecção podem ser prontamente criados ou disponíveis com base em, por exemplo, imunoquímica, enzimático, detecção fluorescente e quantificação. Os níveis de proteína podem ser determinados em extratos de tecidos de plantas ou em tecido intacto usando análise no local de expressão de proteína. Geralmente, as linhagens transformadas individuais com uma construção de repórter de promotor quimérico irão variar em seus níveis de expressão do gene repórter. Também é observado com freqüência o fenômeno de que estes transformantes não expressam qualquer produto detectável (RNA ou proteína). A variabilidade em expressão é comumente dita como "efeitos de posição" apesar de mecanismos moleculares subjacentes a esta inatividade não serem geralmente evidentes.

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

"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 transgene.

"Silenciamento de gene" 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 à transcrição diminuída dos genes afetados, ou pós –transcripcional, quando a supressão é devido a aumentado rotação (degradação) de espécies de RNA homólogas para os genes afetados (English 1996). O silenciamento de genes inclui silenciamento de genes induzido por vírus (Ruiz et al, 1998).

Os termos "seqüência de DNA heteróloga" , "segmento de

DNA exógeno" ou "ácido nucleico heterólogo" , como usado aqui, referem-se, cada, a uma seqüência que se origina de uma fonte estranha à célula hospedeira particular ou, se da mesma fonte, é modificada de sua forma original. Assim, um gene heterólogo em uma célula hospedeira inclui um gene que é endógeno para a célula hospedeira particular mas foi modificado através de, por exemplo, o uso de embaralhamento de DNA. Os temos também incluem cópias múltiplas não de ocorrência natural, de uma seqüência de DNA de ocorrência natural. Assim, os termos referem-se a segmento de DNA que é estranho 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 seqüência de DNA "homóloga" é uma seqüência de DNA que é naturalmente associada com uma célula hospedeira em que ela é introduzida.

"Homólogo a" , no contexto da identidade de seqüência de nucleotídeo, refere-se à similaridade entre as seqüências de nucleotídeos de duas moléculas de ácido nucleico ou entre as seqüências de aminoácido de duas moléculas de proteína. Estimativas desta homologia são providas por ou hibridização DNA-DNA ou DNA-RNA sob condições de estringência como é bem conhecido pelos versados na arte (como descrito em Haines e Higgins (eds.), Nucleic Acid Hybridization, IRL Press, Oxford, U.K.), ou pela comparação de similaridade de seqüência entre dois ácidos nucleicos ou proteínas.

O termo "substancialmente similar" refere-se a seqüências de nucleotídeo e de aminoácido que representam equivalentes funcionais e/ou estruturais ou ortólogos de seqüências de *Arabidopsis thaliana* ou *Linum usitatissimum* descritas aqui.

Em seu sentido mais amplo, o termo "substancialmente similar", como usado aqui com relação a uma seqüência nucleotídica significa

que a seqüência 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 seqüência nucleotídica de referência, por exemplo a seqüência nucleotídica compreende um promotor de um gene que é o ortólogo do gene correspondendo à seqüência nucleotídica de referência, assim como seqüências de promotor que são estruturalmente relacionadas com as seqüências de promotor particularmente exemplificadas aqui, isto é, as seqüências de promotor substancialmente similares hibridizam para o complemento das seqüências de promotor exemplificadas aqui sob condições de estringência elevadas ou muito elevadas. Por exemplo, seqüências nucleotídicas alteradas, que simplesmente refletem a degenerescência do código genético mas mesmo assim codificam seqüências de aminoácidos que são idênticas a uma seqüência de aminoácido particular são substancialmente similares às seqüências particulares. O termo "substancialmente similar" também inclui seqüências nucleotídicas em que a seqüência foi modificada, por exemplo, para otimizar a expressão em células particulares, assim como seqüências nucleotídicas codificando um polipeptídeo variante, tendo uma ou mais substituições de aminoácidos com relação ao polipeptídeo (não modificado) codificado pela seqüência de referência, cuja(s) substituição (ões) não alteram a atividade do polipeptídeo variante com relação ao polipeptídeo não modificado.

Em seu sentido mais amplo, o termo "substancialmente similar" quando usado aqui com relação ao polipeptídeo significa que o polipeptídeo tem substancialmente a mesma estrutura e função que o polipeptídeo de referência. Além disso, seqüências de aminoácidos que são substancialmente similares a uma seqüência particular são as em que a identidade de aminoácido global é pelo menos 50% ou mais para as seqüências presentes. As modificações que resultam em seqüências de nucleotídeos ou de aminoácidos equivalentes são bem conhecidas dos

versados. A porcentagem de identidade de seqüência de aminoácido entre o polipeptídeo substancialmente similar e o de referência é de pelo mesmo, 50% por exemplo, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, e mesmo 90% ou more, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo menos 99%, em que o polipeptídeo de referência é um polipeptídeo (por exemplo, de *Arabidopsis thaliana* ou *Linum usitatissimum*) codificado por um gene com um promotor tendo qualquer uma dentre SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11, uma seqüência nucleotídica compreendendo uma matriz de leitura aberta tendo qualquer uma dentre SEQ ID NOs: 12, 14, 16, 18, 20, ou 22, que codifica uma dentre SEQ ID NOs: 8, 10,12, 14, 16, 18, ou 36. Uma indicação 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.

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

Além disso, uma seqüência nucleotídica que é "substancialmente similar" a uma seqüência nucleotídica de referência é dita ser "equivalente" à seqüência nucleotídica de referência. O versado na arte reconhecido que seqüências nucleotídicas equivalentes englobadas por esta invenção também podem ser definidas por sua capacidade de hibridizar, sob condições baixas, moderadas e/ou estringentes (por exemplo, 0,1 X SSC, 0,1% SDS, 65°C), com as seqüências 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 50%, por exemplo, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, e mesmo 90% ou mais, por exemplo, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, até pelo menos 99% da atividade de polinucleotídeo de comprimento completo ou polipeptídeo de comprimento completo.

"Gene de marcação" refere-se a um gene no replicon que expressa a seqüência de codificação de marcação desejada, RNA funcional ou proteína. O gene de marcação não é essencial para a replicação do replicon. Além disso, os genes de marcação podem compreender genes não virais nativos inseridos em um organismo não nativo, ou genes quiméricos, e estará sob o controle de seqüências regulatórias apropriadas. Assim, as seqüências regulatórias no gene de marcação podem vir de qualquer fonte, incluindo os vírus. Os genes de marcação podem incluir seqüências de codificação que são ou heterólogas ou homólogas aos genes de uma planta particular a ser transformada. No entanto, os genes de marcação não incluem genes virais nativos. Os genes de marcação típicos incluem, mas não são limitados a genes codificando uma proteína estrutural., uma proteína de armazenamento de sementes, uma proteína que transporta resistência a herbicida, e uma proteína que transporta resistência a insetos. As proteínas codificadas por genes de marcação são bem conhecidas como "proteínas estranhas" . A expressão de um gene de marcação 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 com relação ao

hospedeiro de planta de tipo selvagem ou não transgênico.

"Gene de replicação" refere-se a um gene codificando uma proteína de replicação viral. Além do ORF da proteína de replicação, o gene de replicação também pode conter outros ORF(s) de sobreposição ou não sobreposição, como encontrados em seqüências virais na natureza. Apesar de não essencial para replicação, estes ORFs adicionais podem melhorar o acúmulo de DNA viral e/ou replicação. Os exemplos destes ORFs adicionais são AC3 e AL3 em ACMV e geminivírus TGMV, respectivamente.

"Gene de replicação trans-atuante quimérico" refere-se ou a um gene de replicação em que a seqüência de codificação de uma proteína de replicação está sob o controle de um promotor de planta regulado diferente do gene de replicação viral nativo, ou um gene de replicação viral nativo modificado, por exemplo em que uma seqüência(s) específica de sítio é inserida na região 5' transcrita, mas não traduzida. Estes genes quiméricos também incluem inserção dos sítios conhecidos de ligação de proteína de replicação entre o promotor e o sítio de partida de transcrição que atenuam a transcrição de gene de proteína de replicação viral.

"Integrado cromossomicamente" refere-se à integração de um gene estranho ou construção de DNA no DNA hospedeiro por ligações covalentes. Onde os genes não são "integrados cromossomicamente" eles podem ser "transientemente expressados" . A expressão transiente de um gene refere-se a expressão de um gene que não é integrado no cromossomo hospedeiro mas funciona independentemente, ou como parte de um plasmídeo de replicação autônoma ou cassete de expressão, por exemplo ou como parte de outro sistema biológico, como um vírus.

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 totais podem ser regeneradas de células transgênicas por métodos bem conhecidos dos versados (ver, 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 integrada de modo estável no genoma geralmente bem conhecido na arte e é descrita em (Sambrook 1989; Innis 1995; Gelfand 1995; Innis & Gelfand 1999). Os métodos conhecidos de PCR incluem, mas não são limitados aos métodos usando iniciadores em pares, iniciadores abrigados, iniciadores específicos únicos, iniciadores degenerados, iniciadores específicos de genes, iniciadores específicos de vetores, iniciadores parcialmente desalinhados e outros. Por exemplo, plantas ou calos "transformados" "transformantes," e "transgênicos" tinham passado através dos processos de transformação e contém um gene estranho integrado em seu cromossomo. O termo "não transformado" refere-se a plantas normais que não tinham passado através do processo de transformação.

"Transientemente transformado" refere-se a células em que transgenes e DNA estranho foram introduzidos (por exemplo, por tais métodos como transformação mediada por Agrobacterium ou bombardeio biolístico) mas não selecionado para uma manutenção estável.

"Transformado estavelmente" refere-se às células que foram selecionadas e regeneradas em um meio de seleção após transformação.

"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 para sua manutenção estável.

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

"Transformante primário" e "geração T0" fazem referência a plantas transgênicas que são da mesma geração genética como 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 meiose e fertilização. 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.

Os termos "genoma" ou "DNA genômico" se referem 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 organelas celulares (por exemplo, mitocôndrias). Preferivelmente, os termos genoma ou DNA genômico fazem referência ao DNA cromossômico do núcleo.

O termo "DNA cromossômico" ou "seqüência 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 cromossomos ou cromatídeos, eles podem ser condensados ou não enrolado. Uma inserção no DNA cromossômico pode ser demonstrada

e analisada por vários métodos bem conhecidos dos versados na arte, por exemplo, análise de reação em cadeia da polimerase (PCR), análise Southern blot, hibridização em situ fluorescente, (FISH), e PCR em *situ*.

O termo "ácido nucléico" refere-se a deoxirribonucleotídeos ou ribonucleotídeos e seus polímeros em forma de filamento único ou duplo, compostos de monômeros (nucleotídeos) contendo um açúcar, fosfato ou base, que é ou 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 e são metabolizados em um modo similar aos nucleotídeos de ocorrência natural. Salvo especificado em contrário, uma seqüência de ácido nucleico particular também engloba implicitamente suas variantes conservativamente modificadas (por exemplo, substituição de códon degenerado) e seqüências complementares assim como a seqüência explicitamente indicada. Especificamente, substituições de códon degenerado podem ser obtidas por geração de seqüências em que a terceira posição de um ou mais códons selecionados (ou todos) é substituída com resíduos de base mista e/ou deoxiinosina (Batzer 1991; Ohtsuka 1985; Rossolini 1994). Um "fragmento de ácido nucleico "é uma fração de uma dada molécula de ácido nucléico. Em plantas superiores, ácido desoxirribonucleico (DNA) é o material genético enquanto ácido ribonucléico (RNA) está envolvido em transferência de informação contida dentro do DNA em proteínas. O termo "seqüência 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 em polímeros de DNA ou RNA. Os termos "ácido nucleico" ou "seqüência 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 proteína ou ácido

nucleico 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 nativo e, assim, não é 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 proteína ou molécula de ácido nucleico "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 isenta de precursores químicos ou outros produtos químicos quando quimicamente sintetizada. Preferivelmente, um ácido nucleico "isolado" é isento de seqüências (preferivelmente seqüências codificando proteína) que naturalmente flanqueiam o ácido nucleico (isto é, seqüências localizadas nas extremidades 5' e 3' do ácido nucléico) 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 que cerca de 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb, ou 0,1 kb de seqüências nucleotídicas que naturalmente flanqueiam a molécula de ácido nucleico no DNA genômico da célula da qual o ácido nucleico é derivado. Uma proteína que é substancialmente isenta de material celular inclui preparações de proteína ou polipeptídeo tendo menos 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 meio de cultura representa menos que cerca de 30%, 20%, 10%, ou 5% (por peso seco) de precursores químicos ou não proteína de produtos químicos de interesse. As seqüências nucleotídicas da invenção incluem tanto as seqüências de ocorrência natural, assim como formas mutantes (variantes). Estas variantes irão continuar a

possuir a atividade desejada, isto é, ou atividade de promotor ou atividade do produto codificado pela matriz de leitura aberta de seqüência nucleotídica não variante.

O termo "variante" com relação a uma seqüência (por exemplo, polipeptídeo ou seqüência de ácido nucleico como – por exemplo – a seqüência nucleotídica reguladora da transcrição da invenção) se destina a significar seqüências substancialmente similares. Para seqüências nucleotídicas compreendendo uma matriz de leitura aberta, variantes incluem as seqüências que, devido à degenerescência do código genético, codificam a seqüência idêntica de aminoácidos da proteína nativa. As variantes alélicas de ocorrência natural como estas podem ser identificadas com o uso de técnica de biologia molecular bem conhecidas, como, por exemplo, reação em cadeia da polimerase (PCR) e hibridização. Seqüências nucleotídicas variantes também incluem seqüências nucleotídicas sinteticamente derivadas, como as geradas, por exemplo, por uso de mutagênese dirigida ao sítio, 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. Em geral, as variantes de seqüência nucleotídica da invenção terão pelo menos 40, 50, 60, a 70%, por exemplo, preferivelmente 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, a 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% de identidade para seqüências nucleotídicas para a seqüência nucleotídica nativa (tipo selvagem ou endógeno).

"Variações conservativamente modificadas" de uma particular seqüência de ácido nucleico refere-se às seqüências de ácido nucleico que codificam seqüências de aminoácido idênticas ou essencialmente idênticas, ou onde a seqüência de ácido nucleico não codifica uma seqüência de aminoácido, para seqüências 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 em quaisquer dos códons correspondentes descritos sem alterar a proteína codificada. Estas variações de ácido nucleico são "variações silentes" que são uma espécie de "variações conservativamente modificadas". Cada seqüência de ácido nucleico descrita aqui, que codifica um polipeptídeo também descreve cada variação silente possível, exceto onde indicado ao contrário. Um versado 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. conseqüentemente, cada "variação silente" de um ácido nucléico, que codifica um polipeptídeo é implícita em cada seqüência descrita.

As moléculas de ácido nucleico da invenção podem ser "otimizadas" para melhorar a expressão em plantas de interesse (ver, por exemplo, WO 91/16432; Perlak 1991; Murray 1989). Deste modo, a matriz de leitura aberta em genes ou fragmentos de gene pode ser sintetizada usando códons preferidos pelas plantas (ver, por exemplo, Campbell & Gowri, 1990 para uma discussão de uso de códons preferidos do hospedeiro). Assim, as seqüências nucleotídicas podem ser otimizadas para expressão em qualquer planta. Reconhece-se que todo ou qualquer parte da seqüência do gene pode ser otimizada ou sintética. Isto é, as seqüências sintéticas ou parcialmente otimizadas podem ser também usadas. As seqüências nucleotídicas variantes e proteínas também englobam seqüências e proteínas derivadas de um procedimento mutagênico e recombinogênico, como embaralhamento de DNA. Com este procedimento, uma ou mais diferentes seqüências de codificação podem ser manipuladas para criar um novo polipeptídeo

possuindo as propriedades desejadas. Deste modo, bibliotecas de polinucleotídeos recombinantes são geradas a partir de uma população de polinucleotídeos de seqüência relacionada compreendendo regiões de seqüência que tem uma identidade de seqüência substancial e podem ser recombinadas de modo homólogo em vitro ou em vivo. As estratégias para este embaralhamento de DNA são bem conhecidas na arte (ver, por exemplo, Stemmer 1994; Stemmer 1994; Crameri 1997; Moore 1997; Zhang 1997; Crameri 1998; e US 5.605.7912, 14, 16, 18, 20, e 22,837,458).

Por polipeptídeo "variante" destina-se indicar um polipeptídeo derivado da proteína nativa por deleção (assim chamada truncagem) ou adição de um ou mais aminoácidos para a extremidade N-terminal e/ou C-terminal da proteína nativa; deleção ou adição de um ou mais aminoácidos em um ou mais sítios da proteína nativa; ou substituição de um ou mais aminoácidos em um ou mais sítios na proteína nativa. Estas variantes podem resultar de, por exemplo, polimorfismo genético ou de manipulação humana. Os métodos para esta manipulação são geralmente bem conhecidos na arte.

Assim, os polipeptídeos podem ser alterados em vários modos incluindo substituições de aminoácidos, deleções, truncagens e inserções. Os métodos para estas manipulações são geralmente bem conhecidos na arte. Por exemplo, as variantes de seqüência de aminoácido dos polipeptídeos podem ser preparadas por mutações no DNA. Os métodos para mutagênese e alterações de seqüência de nucleotídeos são bem conhecidos na arte (ver, por exemplo, Kunkel 1985; Kunkel 1987; US 4.873.192; Walker & Gaastra, 1983 e as referencias citadas). O guia para as substituições apropriadas de aminoácidos que não afetam a atividade biológica da proteína de interesse podem ser encontrados no modelo de Dayhoff et al. (1978). Substituições conservativas, como troca de um aminoácido por outro tendo propriedades similares são preferidas. As substituições, deleções ou adições individuais que alteram, adicionam ou deletam um único aminoácido ou uma porcentagem

pequena de aminoácidos (tipicamente menos que 5%, mais tipicamente menos que 1%) em uma seqüência codificada são "variações conservativamente modificadas," onde as alterações resultam na substituição de um aminoácido com um aminoácido quimicamente similar. As tabelas de substituído conservativa provendo aminoácidos funcionalmente similares são bem conhecidas na arte. Os seguintes cinco grupos contém, cada, aminoácidos que são substituições conservativas de um para o outro: Alifático: Glicina (G), Alanina (A), Valina (V), Leucina (L), Isoleucina (I); Aromático: Fenilalanina (F), Tirosina (I), 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, substituições individuais, deleções ou adições que alteram, adicionam ou deletam um aminoácido único ou uma porcentagem pequena de aminoácidos em uma seqüência codificada são também "variações conservativamente modificadas".

"Cassete de expressão "como usado aqui significa uma seqüência de DNA capaz de dirigir expressão de uma particular seqüência nucleotídica em uma célula hospedeira apropriada, compreendendo um promotor operativamente ligado a uma seqüência nucleotídica de interesse, que é –opcionalmente - ligada operativamente a sinais de terminação e/ou outros elementos regulatórios. Um cassete de expressão também pode compreender seqüências requeridas para uma tradução apropriado da seqüência nucleotídica. A região de codificação geralmente codifica para uma proteína de interesse mas também pode codificar para um RNA de interesse funcional, por exemplo RNA anti-sentido ou um RNA não traduzido, na direção sentido ou anti-sentido. O cassete de expressão compreendendo a seqüência 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 foi obtido em uma forma recombinante utilizável para expressão heteróloga. Um cassete de expressão pode ser montado de modo completamente extracelular (por exemplo, técnicas de clonagem recombinante). No entanto, um cassete de expressão também pode ser montado usando componentes endógenos em parte. Por exemplo, um cassete de expressão pode ser obtido por colocação (ou inserção) de uma seqüência de promotor a montante de uma seqüência endógena, que assim se torna funcionalmente ligada e controlada por referidas seqüências de promotor. Do mesmo modo, a seqüência de ácido nucleico a ser expressada pode ser colocada (ou inserida) a jusante de uma seqüência endógena de promotor assim formando um cassete de expressão. A expressão da seqüência nucleotídica no cassete de expressão pode também 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. No caso de um organismo multicelular, o promotor também pode ser específico para um tecido particular ou órgão ou estágio de desenvolvimento (por exemplo, promotores específicos de semente ou preferenciais de semente da invenção). Em uma forma de realização preferida, estes cassetes de expressão irão compreender a região de iniciação de transcrição da invenção ligada a uma seqüência nucleotídica de interesse. Este cassete de expressão é preferivelmente provido com uma pluralidade de sítios de restrição para inserção do gene de interesse para 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 transcripcional e translacional, uma seqüência de DNA de interesse, e uma região de terminação transcripcional e translacional funcional em plantas. A região de terminação pode ser nativa com a região de iniciação transcripcional, pode ser nativa com a seqüência de DNA de interesse, ou pode ser derivada de outra

fonte. As regiões de terminação convenientes são disponíveis de plasmídeo Ti de *A. tumefaciens*, como as regiões de terminação de octapina 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 como incluindo, inter alia, qualquer plasmídeo, cosmídeo, fago ou vetor binário de Agrobacterium em forma circular ou linear de filamento duplo ou único que pode ou não ser auto-transmissível ou mobilizável, e que pode transformar hospedeiro procariótico ou eucariótico ou por integração no genoma celular ou existir extracromossomicamente (por exemplo, plasmídeo de replicação autônoma com uma origem de replicação).

São especificamente incluídos os vetores de transporte o que significa um veículo de DNA capaz de, de modo natural ou por projeto, replicação em dois diferentes organismos hospedeiros, que podem ser selecionado dentre actinomicetes e espécies relacionadas, bactérias e eucarióticos (por exemplo, células de plantas superiores, mamíferos, levedura ou fungos).

Preferivelmente, o ácido nucleico no vetor está sob o controle de, e ligado operativamente a um promotor apropriado ou outros elementos regulatórios para transcrição em uma célula hospedeira, como microbiana, por exemplo, bacteriana, ou célula de planta. O vetor pode ser um vetor de expressão bi-funcional que funciona em múltiplos hospedeiros. 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 contêm um ou um número pequeno de sítios de reconhecimento de endonuclease de restrição em que as

seqüências de DNA estranho podem ser inseridas em um modo determinável sem perda da 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 provêem 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.

"Tecido de planta" inclui tecidos diferenciados e não diferenciados de plantas, incluindo, mas não limitados a raízes, hastes, brotos, folhas, pólen, sementes, tecido de tumor e várias formas de células e cultura como células únicas, protoplasto, embriões, e tecido de calos. O tecido da planta pode ser em plantas ou em um órgão, tecido ou cultura de células.

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

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

(b) Como usado aqui, "janela de comparação" faz referência a um segmento contíguo e especificado de uma seqüência de polinucleotídeos, em que a seqüência de polinucleotídeos na janela de comparação pode compreender adições ou deleções (isto é, espaços "*gaps*") comparada 'a seqüência de referência (que não compreende adições ou deleções) para

alinhamento ótimo das duas seqüências. Em geral, a janela de comparação tem pelo menos 20 nucleotídeos contíguos de comprimento, e, opcionalmente, pode ser 30, 40, 50, 100, ou mais longa. Os versados na arte compreendem que para evitar uma similaridade elevada em uma seqüência de referência devido à inclusão de espaços na seqüência de polinucleotídeos, uma penalidade de espaço é tipicamente introduzida e é subtraída do número de alinhamentos.

Os métodos de alinhamento de seqüências para comparação são bem conhecidos na arte. Assim, a determinação da identidade percentual entre quaisquer duas seqüências pode ser obtida usando um algoritmo matemático. Exemplos não limitantes, preferidos, de 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 pesquisa-para-local de Perason e Lipman (1988); o algoritmo de Karlin e Altschul (1990) modificado como em Karlin e Altschul (1993).

Implementações por computador destes algoritmos matemáticos podem ser utilizadas para comparação de seqüências de modo a determinar a identidade da seqüência. Estas implementações incluem, mas não estão limitadas a: CLUSTAL no programa PC/Gene (disponível de Intelligenetics, Mountain View, Califórnia); o programa ALIGN (versão 2.0) e GAP, BESTFIT, BLAST, FASTA e TFASTA no pacote de software Wisconsin Genetics, versão 8 (disponível de Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis. USA). Alinhamentos usando estes programas podem ser realizados usando os parâmetros de *default*. O programa CLUSTAL é bem descrito por Higgins et al. 1988, 1989; Corpet 1988, Huang et al. (1992) e Perason 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, mais de 2 seqüências) são preferivelmente realizados usando o algoritmo Clustal W Thompson 1994; por exemplo, no software VetorNTI™, versão 9; Invitrogen Inc.) com a matriz de classificação matriz BLOSUM62MT2 com as configurações de *default* (penalidade de abertura de espaço 15/19, penalidade de extensão de espaço 6.66/0,05; faixa de penalidade de separação de espaço 8; % identidade para o retardo de alinhamento 40; usando os espaços específicos de resíduo e os espaços de resíduo hidrofílicos).

Software para realizar as análises BLAST é acessível ao público através do National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). Este algoritmo envolve primeiro a identificação de pares de seqüência de elevado escore (HSPs) por identificação de palavras curtas de comprimento W na seqüência de busca, que ou se alinham ou atendem a um escore de limiar de valor positivo T quando alinhado com uma palavra do mesmo comprimento em uma seqüência de base de dados. T é referido como o limiar de escore de palavra vizinha (Altschul 1990). Estas dicas de palavra vizinha inicial atuam como sementes para a iniciação de buscas para encontrar HSPs mais longos contendo as mesmas. As dicas de palavras são então estendidas em ambas as direções junto de cada seqüência desde que o escore de alinhamento cumulativo possa ser aumentado. Os escores cumulativos são calculados usando, para seqüências nucleotídicas, os parâmetros M (escore de prêmio para um par de resíduos alinhados; sempre > 0) e N (escore de penalidade para resíduos desalinhados; sempre <0). Para seqüências de aminoácidos, uma matriz de escore é usada para calcular o escore cumulativo. A extensão das dicas de palavras em cada direção é parada quando o escore de alinhamento cumulativo fica fora da quantidade X de seu valor obtido máximo, o escore cumulativo via a zero ou abaixo devido ao acúmulo de um ou mais alinhamentos de resíduo de escore negativo, ou o final de ambas as

seqüências é alcançado.

Além de calcular a identidade de seqüência percentual, o algoritmo BLAST também realiza uma análise estatística da similaridade entre duas seqüências (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 um alinhamento entre duas seqüências de nucleotídeo ou de aminoácido deve ocorrer por acaso. Por exemplo, uma seqüência de ácido nucleico de teste é considerada similar à seqüência de referência se a menor probabilidade de soma em comparação com seqüência de ácido nucleico de teste para seqüência de ácido nucleico de referência for menos do que cerca de 0,1, mais preferivelmente menor que cerca de 0,01, e o mais preferivelmente menor que cerca de 0,001.

Para obter alinhamentos espaçados para fins de comparação Gapped BLAST (em BLAST 2.0) pode ser utilizado 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 moléculas. Ver Altschul et al., supra. Quando utilizando BLAST, Gapped BLAST, PSI-BLAST, os parâmetros de *default* dos respectivos programas (por exemplo, BLASTN para seqüências nucleotídicas, BLASTX para proteínas) podem ser usados. O programa BLASTN (para seqüências 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 seqüências de aminoácidos, o 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. Alinhamento também pode ser realizado manualmente por inspeção.

Para fins da presente invenção, comparação de seqüências

nucleotídicas para determinação de identidade de seqüência percentual para as seqüências de promotor descritas aqui é preferivelmente feita usando o programa BlastN (versão 1.4.7 ou posterior) com seus parâmetros de *default* ou qualquer programa equivalente. Por "programa equivalente "se destina qualquer programa de comparação de seqüência que, para quaisquer duas seqüências em questão, gera um alinhamento tendo alinhamentos de resíduos de nucleotídeo ou aminoácidos idênticos, e uma identidade de seqüência percentual idêntica quando comparado com o alinhamento correspondente gerado pelo programa preferido.

(c) Como usado aqui, "identidade de seqüência" ou "identidade" no contexto de duas seqüências de polipeptídeos ou de aminoácidos faz referência aos resíduos nas duas seqüências que são iguais quando alinhadas para correspondência máxima sobre uma janela de comparação especificada. Quando a percentagem da seqüência de identidade é usada em referência a proteínas, reconhece-se que as posições dos resíduos que não são idênticos freqüentemente diferem por substituições de aminoácidos conservativas, onde resíduos de aminoácidos são substituídos por outros resíduos de aminoácidos com propriedades químicas similares (por exemplo, carga ou hidrofobicidade) e, assim, não trocam as propriedades funcionais da molécula. Quando as seqüências diferem nas substituições conservativas, a identidade de seqüência percentual pode ser ajustada para cima para corrigir a natureza conservativa da substituição. As seqüências que diferem pelas substituições conservativas são referidas como tendo "similaridade de seqüência" ou "similaridade" . Meios para obter esta configuração são bem conhecidos dos versados na arte. Tipicamente, isto envolve classificar uma substituição conservativa como uma desalinhamento parcial em vez de um completo, aumentando assim a identidade de seqüência percentual. Assim, por exemplo, quando a um aminoácido idêntico é conferida uma classificação de 1 e a uma substituição não conservativa é

conferida uma classificação de zero, a uma substituição conservativa é conferida uma classificação entre zero e 1. A classificação de substituições conservativas é calculada, por exemplo, como implementado no programa PC/GENE (Intelligenetics, Mountain View, Califórnia).

(d) Como usado aqui, "porcentagem de identidade de seqüência" significa um valor determinado comparando duas seqüências otimamente alinhadas sobre uma janela de comparação, em que a porção da seqüência de polinucleotídeos na janela de comparação pode compreender adições ou deleções (isto é, espaços) como comparada à seqüência de referência (que não compreende adições ou deleções) para alinhamento ótimo das duas seqüências. A percentagem é calculada determinando o número de posições em que a base de ácido nucleico idêntica ou resíduo de aminoácidos ocorre em ambas as seqüências para dar o número de posições alinhadas, dividindo o número de posições alinhadas de modo correto pelo número total de posições na janela de comparação, e multiplicando o resultado por 100 para dar a identidade de seqüência percentual.

(e) (i) O termo "identidade substancial" de seqüências polinucleotídicas significa que um polinucleotídeo compreende uma seqüência que tem pelo menos 50%, 50% por exemplo, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 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 seqüência, comparado a uma seqüência de referência usando um dos programas de alinhamento descritos usando parâmetros padrões. Um versado na arte irá reconhecer que estes valores podem ser ajustados de modo apropriado para determinar a identidade correspondente de proteínas codificadas por duas seqüências nucleotídicas ao levar em conta a

degenerescência do códon, similaridade de aminoácido, posicionamento do quadro de leitura, e outros. A identidade substancial de seqüências de aminoácidos para estes fins normalmente significa identidade de seqüência de pelo menos 50% ou 60%, preferivelmente pelo menos 70% ou 80%, mais preferivelmente pelo menos 90%, 95%, e o mais preferivelmente pelo menos 98%.

Outra indicação de que as seqüências nucleotídicas são substancialmente idênticas é se duas moléculas hibridizam uma à outra em condições estringentes. (ver abaixo). Geralmente, as condições estringentes são selecionadas para ser 5°C mais baixas do que o ponto de fusão térmico (T_m) na seqüência específica em um pH e resistência iônica definidos. No entanto, as condições estringentes englobam temperaturas na faixa de cerca de 1°C a cerca de 20°C, dependendo do grau desejado de estringência como qualificado de outro modo aqui. Ácidos nucleicos que não hibridizam um ao outro em condições estringentes são ainda substancialmente idênticos se os polipeptídeos que eles codificam forem substancialmente idênticos. Isto pode ocorrer, por exemplo, quando uma cópia de um ácido nucleico é criada usando a degeneração de códons máxima permitida pelo código genético. Uma indicação de que duas seqüências de aminoácidos são substancialmente idênticas é quando o polipeptídeo codificado pelo primeiro ácido nucleico é imunologicamente reativo cruzado com o polipeptídeo codificado pelo segundo ácido nucleico.

(ii) O termo "identidade substancial" no contexto de um peptídeo indica que o peptídeo compreende uma seqüência com pelo menos 50%, por exemplo, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 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%, 83%, 84%, 85%, 86%, 87%, 88%, ou 89%, mais preferivelmente pelo menos 90%, 91%, 92%, 93%, ou 94%, ou mesmo mais preferivelmente, 95%, 96%, 97%,

98% ou 99%, de identidade de seqüência para a seqüência de referência sobre a janela de comparação especificada. Preferivelmente, alinhamento ótimo é conduzido usando o algoritmo de alinhamento de homologia de Needleman e Wunsch (1970). Uma indicação de que duas seqüências de peptídeos são substancialmente idênticas é 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 seqüências, tipicamente uma seqüência atua como uma seqüência de referência para a qual seqüências de teste são comparadas. Quando usando um algoritmo de comparação de seqüência, as seqüências de teste e referência são colocadas como entrada no computador, coordenadas subseqüentes são designadas, se necessário, e os parâmetros de programa de algoritmo de seqüência são designados. O algoritmo de comparação de seqüência então calcula a identidade de seqüência percentual para a(s) seqüência(s) de teste com relação à seqüência de referência com base nos parâmetros de programa designados.

Como notado acima, outra indicação de que duas seqüências de ácido nucleico são substancialmente idênticas é que as duas moléculas hibridizam uma na outra sob condições estringentes. A expressão "hibridizando especificamente para" refere-se à ligação, duplexagem, ou hibridização de uma molécula somente para uma seqüência nucleotídica particular sob condições estringentes quando esta seqüência está presente em um DNA ou RNA de mistura complexa (por exemplo, celular total). "Liga(m) substancialmente" refere-se a uma hibridização complementar entre um ácido nucleico de sonda e um ácido nucleico de marcação e engloba desalinhamentos menores que podem ser acomodados por redução da estringência do meio de hibridização para obter a detecção desejada da seqüência de ácido nucleico de marcação.

"Condições de hibridização estringentes" e "condições de lavagem de hibridização estringentes" no contexto de experiências de hibridização de ácido nucleico como hibridização Southern e Northern são dependentes de seqüência e são diferentes sob parâmetros ambientais diferentes. O T_m é a temperatura (sob pH e resistência iônica definidas) em que 50% da seqüência de marcação hibridizam para uma sonda perfeitamente alinhada. A especificidade é tipicamente a função de lavagens pós-hibridização, os fatores críticos sendo a resistência iônica e a temperatura da solução de lavagem final. Para híbridos DNA-DNA, o T_m pode ser aproximado da equação de Meinkoth e Wahl (1984):

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

em que M é a molaridade de cátions monovalentes, % GC é a percentagem de nucleotídeos guanosina e citosina no DNA, a % de forma é a percentagem de formamida na solução de hibridização, e L é o comprimento do híbrido nos pares de bases. O T_m é reduzido em cerca de 1°C para cada 1% de desalinhamentos; assim, as condições de T_m, de hibridização e/ou lavagem podem ser ajustadas para hibridizar em seqüências da identidade desejada. Por exemplo, se seqüências com ≥ 90% de identidade forem consideradas, o T_m pode ser diminuído em 10°C. Em geral, as condições estringentes são selecionadas de modo a serem de cerca de 5°C mais baixas do que o ponto de fusão térmico I para a seqüência específica e seu complemento em um pH e resistência iônica definidos. No entanto, condições severamente estringentes podem utilizar uma hibridização e/ou lavagem em 1, 2, 3 ou 4°C mais baixos do que o ponto de fusão térmico I; condições moderadamente estringentes podem utilizar uma hibridização e/ou lavagem em 6, 7, 8, 9 ou 10°C mais baixos do que o ponto de fusão térmico I; condições de baixa estringência podem utilizar uma hibridização e/ou lavagem em 11, 12, 13, 14, 15 ou 20°C mais baixos do que o ponto de fusão térmico I.Usando a equação, composições de hibridização e de lavagem e T desejados, os versados na arte

compreenderão que variações na estringência de soluções de hibridização e/ou lavagem são inerentemente descritas. Se o grau desejado de combinações incorretas 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 temperatura mais elevada pode ser usada. Um guia extenso sobre a hibridização de ácidos nucleicos é encontrado em Tijssen 1993. Geralmente, condições de hibridização altamente estringentes e lavagem são selecionadas para serem cerca de 5 °C menores do que o ponto de fusão térmico T_m para a seqüência específica em um pH e resistência iônica definidos.

Um exemplo de condições altamente estringentes de lavagem é 0,15 M NaCl a 72°C durante cerca de 15 minutos. Um exemplo de condições estringentes de lavagem é uma lavagem a 0,2 X SSC a 65°C durante 15 minutos (ver, Sambrook, infra, para uma descrição de tampão SSC). Com freqüência, uma lavagem de alta estringência é precedida por uma lavagem de baixa estringência para remover o sinal de sonda de fundo. Um exemplo de lavagem de estringência média para um duplex de, por exemplo, mais do que 100 nucleotídeos, é 1 X SSC a 45°C durante 15 minutos. Um exemplo lavagem de estringência baixa 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 estringentes tipicamente envolvem concentrações de sal de menos que cerca de 1,5 M, mais preferivelmente cerca de 0,01 a 1,0 M, concentração iônica de Na (ou) a pH 7,0 a 8,3, 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 estringentes também podem ser obtidas com a adição de agentes desestabilizantes como formamida. Geralmente, uma relação de sinal para ruído de 2 X (ou maior) do que observado para uma sonda não relacionada no teste particular de hibridização indica detecção de uma específica hibridização. Ácidos nucleicos que não hibridizam um em

cada outro sob condições estringentes são ainda substancialmente idênticos se as proteínas que eles codificam são 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.

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 no filtro em um blot Southern ou Northern 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 a 65°C. Condições exemplares de baixa estringência incluem hibridização com uma solução de 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. Condições exemplares de estringência moderada 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 usadas para clonar seqüências nucleotídicas ortólogas que são substancialmente idênticas às seqüências nucleotídicas de referência da presente invenção: uma seqüência nucleotídica de referência preferivelmente hibridiza na seqüência nucleotídica de referência em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais

preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 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 mudanças de seqüências 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 seqüências de DNA que são unidas juntas usando tecnologia de DNA recombinante e procedimentos usados para unir juntas as seqüências de DNA como descrito, por exemplo, em Sambrook et al., 1989.

A palavra "planta" refere-se a qualquer planta, particularmente a plantas agronomicamente úteis (por exemplo, plantas de semente), e "célula de planta" é uma unidade estrutural e fisiológica da planta, que compreende uma parede celular também pode se referir 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 uma 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 do desenvolvimento da planta. Estas estruturas incluem um ou mais órgãos de plantas incluindo, mas não são limitados a, fruta, 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, pétala, estames,

carpelos, anteros e óvulos), sementes (incluindo embrião, endosperma, e capa da semente) e frutas (o ovário maduro), tecidos de plantas (por exemplo, tecido vascular, tecido triturado e semelhantes) e células (por exemplo, células de guarda, células de ovo, tricomas e semelhantes), e sua progênie. A classe de plantas que pode ser usada no método da invenção é geralmente tão ampla como a classe de plantas superiores e inferiores que podem passar pelas técnicas de transformação, incluindo angiospermas (plantas monocotiledônea e dicotiledônea), gimnospermas, fetos, e algas multicelulares. Inclui plantas de uma variedade de níveis plóides, incluindo aneuplóide, poliplóide, diplóide, haplóide e hemizigótico. Estão incluídos no escopo da invenção os gêneros e espécies de plantas superiores e inferiores do reino vegetal. Estão ainda incluídas as plantas maduras, semente, brotos e mudinhas, e partes, material de propagação (por exemplo, sementes e fruta) e culturas, por exemplo culturas de célula, dai derivadas. Preferidos são plantas e materiais de planta das seguintes famílias de plantas: *Amaranthaceae, Brassicaceae, Carophyllaceae, Chenopodiaceae, Compositae, Cucurbitaceae, Labiatae, Leguminosae, Papilionoideae, Liliaceae, Linaceae, Malvaceae, Rosaceae, Saxifragaceae, Scrophulariaceae, Solanaceae, Tetragoniaceae.* Plantas anuais, perenes monocotiledôneas e dicotiledôneas são organismos hospedeiros preferidos para a geração de plantas transgênicas. O uso do sistema de recombinação ou método de acordo com a invenção é ainda vantajoso em plantas ornamentais, silvicultura, fruta, ou árvores ornamental, flores, flores de corte, arbustos ou gramas. Referida planta pode incluir –mas não deve ser limitado a - briófitos como, por exemplo, Hepaticae (hepática) e Musci (musgos); pteridófitos como fetos, rabo-de-cavalo, e flores-das-montanhas; gimnospermas como coníferas, cicádias, ginkgo e *Gnetaeae*; algas como *Clorophyceae, Faeophpyceae, Rhodophyceae, Myxophyceae, Xanthophyceae, Bacillariophyceae* (diatomáceas) e *Euglenophyceae*. Plantas para os fins da invenção podem compreender as famílias de Rosaceae como

rosas, Ericaceae como rododentros, azaléias, *Euphorbiaceae* como poinsetias e crotão, *Caryophyllaceae* como cravos, Solanaceae como petunias, *Gesneriaceae* como violeta africana, *Balsaminaceae* como não-me-toques, Orchidaceae como orquídeas, *Iridaceae* como gladíolo, íris, frésia e açafrão, *Compositae* como tagetes, *Geraniaceae* como gerânio, *Liliaceae* como *Drachaena, Moraceae* como ficus, *Araceae* como filodendros e muitas outras.

As plantas transgênicas de acordo com a invenção são ainda selecionadas em particular dentre as plantas de cultura dicotiledôneas como, por exemplo, das famílias de Leguminosae como ervilha, alfafa e soja; a família de *Umbelliferae*, particularmente o gênero *Daucus* (muito particularmente as espécies *carota* (cenoura)) e *Apium* (muito particularmente as espécies *graveolens* var. dulce (aipo)) e muitas outras; a família de Solanaceae, particularmente o gênero *Lycopersicon*, muito particularmente as espécies *esculentum* (tomate) e o gênero *Solanum*, muito particularmente as espécies *tuberosum* (batata) e *melongena* (berinjela), tabaco e muitas outras; e o gênero *Capsicum*, muito particularmente as espécies *annum* (pimenta) e muitas outras; a família de *Leguminosae*, particularmente o gênero *Glicina,* muito particularmente as espécies *max* (soja) e muitas outras; e a família de *Cruciferae*, particularmente o gênero *Brassica*, muito particularmente as espécies *napus* (colza de semente oleígena), *campestris* (beterraba), *oleracea* cv Tastie (repolho), *oleracea* cv Snowball Y (couve-flor) e *oleracea* cv Emperor (brócolis); e o gênero *Arabidopsis*, muito particularmente as espécies *thaliana* e muitas outras; a família de Compositae, particularmente o gênero *Lactuca*, muito particularmente as espécies *sativa* (alface) e muitas outras. As plantas transgênicas de acordo como a invenção podem ser selecionadas dentre plantas de culturas monocotiledôneas, como, por exemplo, cereais como trigo, cevada, sorgo e painço, centeio, triticale, milho, arroz ou aveia, e cana de açúcar. Ainda preferidas são as árvores como maçã, pêra, marmelo, ameixa, cereja, pêssego, nectarina, damasco, mamão papaya,

manga, e outras espécies lenhosas incluindo árvores coníferas e decíduas como álamo, pinheiro, sequóia, cedro, carvalho, *etc.* Especialmente preferidos são *Arabidopsis thaliana, Nicotiana tabacum*, colza de semente oleígena, soja, milho (milho), trigo, *Linum usitatissimum* (linhaça e linho), *Camelina sativa, Brassica juncea,* batata e tagetes.

Um "aumento significante" é um aumento que é maior do que a margem de erro inerente à técnica de medida, preferivelmente um aumento em cerca de 2 vezes ou mais.

"Significantemente menor" significa que a diminuição é maior do que a margem de erro inerente à técnica de medida, preferivelmente uma diminuição de cerca de 2 vezes ou mais.

DESCRIÇÃO DETALHADA DA INVENÇÃO

A presente invenção assim provê moléculas de ácido nucleico isoladas compreendendo uma seqüência nucleotídica de planta que dirige a transcrição preferencial de semente ou específica de semente de um fragmento de ácido nucleico operativamente ligado em uma célula de planta.

Especificamente, a presente invenção provê cassetes transgênicos de expressão para regular a expressão preferencial de semente ou específica de semente em plantas compreendendo

a) pelo menos uma seqüência nucleotídica reguladora da transcrição de um gene codificando uma proteína peroxiredoxina (PER1), e funcionalmente ligada à mesma

b) pelo menos uma seqüência de ácido nucleico que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

Preferivelmente, o cassete de expressão da invenção compreende

a) pelo menos uma seqüência nucleotídica reguladora da transcrição de um gene de planta, referido gene de planta selecionado dentre o

grupo consistindo de

i) o gene descrito pelos *loci* do genoma de *Arabidopsis thaliana* do GenBank At1g48130, e

ii) genes ortólogos do gene descrito pelos *loci* do genoma de *Arabidopsis thaliana* do GenBank At1g48130,

e funcionalmente ligado ao mesmo

b) pelo menos uma seqüência de ácido nucleico que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

Os promotores preferenciais de semente ou específicos de semente podem ser úteis para expressar genes assim como para produzir quantidades grandes de proteína, para expressar óleos ou proteínas de interesse, por exemplo, anticorpos, genes para aumentar o valor nutricional da semente e semelhantes.

O termo "semente" no contexto da invenção significa uma semente de uma planta em qualquer estágio de seu desenvolvimento, isto é partindo da fusão de pólen e oócitos, continuando sobre o estágio do embrião e o estágio da semente dormente, até a germinação da semente, terminando com os órgãos de mudinhas prematuras, como por exemplo cotilédones e hipocótilo.

"Transcrição específica de semente "no contexto desta invenção significa a transcrição de uma seqüência de ácido nucleico por um elemento regulador de transcrição em um modo em que a transcrição de referida seqüência de ácido nucleico em sementes contribui a mais do que 90%, preferivelmente mais do que 95%, mais preferivelmente mais do que 99% da quantidade completa do RNA transcrito de referida seqüência de ácido nucleico na planta completa durante qualquer um de seus estágios de desenvolvimento. Estas seqüências nucleotídicas reguladoras de transcrição especificamente descritas aqui são consideradas como sendo seqüências

nucleotídicas reguladoras de transcrição específicas de semente.

"Transcrição preferencial de semente "no contexto desta invenção significa a transcrição de uma seqüência de ácido nucleico por um elemento regulador de transcrição em um modo em que a transcrição de referida seqüência de ácido nucleico em sementes contribui para mais do que 50%, preferivelmente mais do que 70%, mais preferivelmente mais do que 80% da quantidade completa do transcrito de RNA da referida seqüência de ácido nucleico na planta completa durante qualquer um dos seus estágios de desenvolvimento.

"Homologo" é um termo genético usado na arte para indicar uma seqüência de polinucleotídeo ou polipeptídeo possuindo um grau elevado de parentesco de seqüência com uma seqüência de referência. Este parentesco pode ser quantificado por determinação do grau de identidade e/ou similaridade entre as duas seqüências como acima definido. Está dentro do significado deste termo genérico, os termos "ortólogo", e "parálogo". "Parálogo" refere-se a um polinucleotídeo ou polipeptídeo que está 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 ortólogo de gene significa preferivelmente um gene, que está codificando uma proteína ortóloga. Mais especificamente, o termo "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 seqüência dentre os ortólogos são o resultado de especiação. Mais preferivelmente, "gene ortólogo" ou "ortólogo" como usado aqui com relação ao gene de Arabidopsis descrito pelo *locus* At1g48130 (e/ou o gene de *Linum usitatissimum* compreendendo a seqüência como descrita por SEQ ID NO: 9, 10, 11 ou 20) significa um gene codificando uma proteína que tem pelo menos 50% ou 60%, preferivelmente pelo menos 70% ou 80%, mais preferivelmente pelo menos 80% ou 90%, o mais preferivelmente pelo menos

95% identidade de seqüência de aminoácido para um polipeptídeo descrito por qualquer uma das SEQ ID NOs: 13 ou 21. Mais preferivelmente, referida proteína ainda tem a propriedade enzimática de uma proteína peroxiredoxina (PER1). Várias destas proteínas (proteínas ortólogas) são descritas aqui, por exemplo, pela seqüência de polipeptídeos como descrita por SEQ ID NO: 13, 15, 17, 19, 21 ou 23. A identidade entre as proteínas ortólogas é descrita na Tabela 1 abaixo.

Tabela 1: Identidade entre as proteínas ortólogas (proteínas de comprimento completo)

	Arabidopsis thaliana	Brassica napus	Triticum turgidum	Hordeum vulgare	Bromus secalinus
Arabidopsis thaliana	100	90	71	72	66
Brassica napus		100	68	69	65
Triticum turgidum			100	98	90
Hordeum vulgare				100	90
Bromus secalinus					100

Tabela 2: Identidade entre as proteínas ortólogas em comparação com os fragmentos da proteína de *Linum usitatissimum* (comparação é limitada ao comprimento do fragmento e faixa de identidade é calculada por sua parte sobre o comprimento das seqüências parciais de *Linum usitatissimum* apenas)

	ARABIDOPSIS THALIANA	BRASSICA NAPUS	LINUM U. (part)	HORDEUM VULGARE	TRITICUM TURGIDUM	BROMUS SECALINUS
ARABIDOPSISTHALIANA	100	89	76	73	73	61
BRASSICANAPUS		100	73	68	68	56
LINUM U. (part)			100	73	73	58
HORDEUM VULGARE				100	100	82
TRITICUMTURGIDUM					100	82
BROMUSSECALINUS						100

Preferivelmente a seqüência nucleotídica reguladora da transcrição da invenção compreende pelo menos uma seqüência de promotor do gene respectivo (por exemplo, uma seqüência localizada a montante do início de transcrição do gene respectivo capaz de induzir a transcrição das seqüências a jusante). Referida seqüência nucleotídica reguladora da transcrição pode compreender a seqüência de promotor de referidos genes mas pode ainda compreender outros elementos como a seqüência 5'-não traduzida, melhorador, íntrons etc. Preferivelmente, referida seqüência de promotor dirige a transcrição preferencial de semente ou específica de semente de um segmento de ácido nucleico operativamente ligado em uma planta ou célula de planta por exemplo, um DNA de planta ligado compreendendo uma matriz de leitura aberta para um gene estrutural ou regulatório.

A seguinte Tabela 1 ilustra os genes dos quais os promotores da invenção são preferivelmente isolados, a função de referidos genes, o cDNA codificado pelos referidos genes, e a proteína (ORF) codificada pelos referidos genes.

Tabela 1: Genes, dos quais os promotores da invenção são preferivelmente isolados, função putativa de referidos genes, cDNA e a proteína codificada pelos referidos genes.

***Locus* Gene**	**Função Putativa**	**Promotor SEQ ID**	**ID *locus* mRNA SEQ ID cDNA**	**ID Proteína SEQ ID Proteína**
At1g48130	*Arabidopsis thaliana*proteína putativa peroxiredoxina (PER1)	SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8	BT003916 SEQ ID NO: 12	AAO41963.1 SEQ ID NO: 13
	Linum usitatissimum proteína putativa peroxiredoxina (PER1)	SEQ ID NO: 9, 10, 11	SEQ ID NO: 20	SEQ ID NO: 21

Algumas das seqüências reguladoras de transcrição providas aqui são novas e como tal não foram previamente descritas nas arte. Assim, outra forma de realização da invenção refere-se a uma seqüência isolada

nucleotídica (preferivelmente tendo atividade reguladora de transcrição, mais preferivelmente tendo atividade de promotor) selecionada dentre o grupo de seqüências consistindo de:

a) a seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e seqüências tendo uma identidade de pelo menos 99%, preferivelmente 99,5%, mais preferivelmente 99,8% para uma seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e seqüências compreendendo pelo menos 600 nucleotídeos consecutivos, preferivelmente pelo menos 800 nucleotídeo consecutivo, mais preferivelmente pelo menos 900 nucleotídeos consecutivos, mais preferivelmente pelo menos 1000 nucleotídeos consecutivos de uma seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e

b) a seqüência como descrita por SEQ ID NO: 9, 10, ou 11, e seqüências tendo uma identidade de pelo menos 40% ou 50%, preferivelmente 60% ou 70%, mais preferivelmente 80 ou 90%, o mais preferivelmente 95% ou 98% para uma seqüência como descrita por SEQ ID NO: 5 ou 26, e seqüências compreendendo pelo menos 15 ou 20 nucleotídeos consecutivos, preferivelmente pelo menos 50 ou 100 nucleotídeo consecutivo, mais preferivelmente pelo menos 250 nucleotídeos consecutivos, mais preferivelmente pelo menos 500 nucleotídeos consecutivos de uma seqüência como descrita por SEQ ID NO: 9, 10, ou 11.

Como especificado acima na seção de DEFINIÇÃO, identidades entre seqüências nucleotídicas são preferivelmente medidas pelo programa BLASTN usando parâmetros de *default* com 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 medir a identidade entre seqüências de aminoácido, o programa BLASTP é usado com parâmetros de *default* com um comprimento de palavra (W) de 3, uma expectativa (E) de 10, e a matriz de escore BLOSUM62 (ver Henikoff & Henikoff, 1989). O Programa BLAST versão 1.4.7 ou posterior é usado.

Preferivelmente, a seqüência nucleotídica reguladora da transcrição é selecionada dentre o grupo de seqüências consistindo de

i) a seqüência descrita por SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11,

ii) um fragmento de pelo menos 50 bases consecutivas de uma seqüência sob i),

iii) uma seqüência nucleotídica tendo similaridade substancial com uma identidade de seqüência de pelo menos 50% para uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11;

iv) uma seqüência nucleotídica capaz de hibridizar (preferivelmente sob condições equivalentes à hibridização em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 65°C) para uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11, ou o seu complemento;

v) uma seqüência nucleotídica capaz de hibridizar (preferivelmente sob condições equivalentes à hibridização em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em dodecil sulfato de sódio a

7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 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 de uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11, ou o seu complemento;

vi) uma seqüência nucleotídica que é o complemento ou complemento reverso de qualquer uma das seqüências nucleotídicas previamente mencionadas sob i) to v).

Em uma forma de realização preferida, as seqüências especificadas sob ii), iii), iv) v) e vi) acima são capazes de modificar a transcrição em uma célula de planta ou organismo, preferivelmente referidas seqüências têm substancialmente a mesma atividade reguladora de transcrição como a seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11.

Preferivelmente, uma seqüência nucleotídica tendo similaridade substancial a uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11 tem uma identidade de seqüência de pelo menos 50% ou 60%, preferivelmente pelo menos 70% ou 80%, mais preferivelmente pelo menos 90% ou 95%, o mais preferivelmente pelo menos 98% to uma seqüência descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11.

Preferivelmente, a hibridização é realizada sob condições estringentes (incluindo condições de estringência baixa e elevada), mais preferivelmente sob condições de estringência elevada.

A seqüência nucleotídica reguladora da transcrição pode ser obtida ou é obtenível a partir de DNA genômico de planta de um gene

codificando um polipeptídeo (preferivelmente de um gene ortólogo isto é, codificando uma proteína ortóloga) que

a) compreende pelo menos um motivo de seqüência selecionado dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL (SEQ ID NO: 37), preferivelmente MPG(I/L)T(L/I)GDTVPNLE (SEQ ID NO: 38),

ii) GDFTPVC (SEQ ID NO: 39), preferivelmente LFSHPGDFTPVCTTEL (SEQ ID NO: 40),

iii) VKLLGLS (SEQ ID NO:41),

preferivelmente RGVKLLGLSCDD (SEQ ID NO: 42),

iv) YPI(I/M)ADP (SEQ ID NO: 43), preferivelmente SKV(T/N)YPI(I/M)ADP (SEQ ID NO: 44),

v) KLSFLYP (SEQ ID NO: 45),

preferivelmente (K/V)(I/V)KLSFLYPS (SEQ ID NO: 46),

vi) GRNMDEV (SEQ ID NO: 47), preferivelmente TGRNMDEV(L/V)RA (SEQ ID NO: 48),

vii) (I/V)ATP(V/A)NW (SEQ ID NO: 49),

preferivelmente K(I/V)ATP(V/A)NW(K/N)P (SEQ ID NO: 50), e

viii) PSKKGYL (SEQ ID NO: 51), preferivelmente LPSKKGYLR (SEQ ID NO: 52),

ou

b) tem pelo menos 50% identidade de seqüência de aminoácido para um polipeptídeo selecionado dentre o grupo descrito por SEQ ID NO: 13, 15, 17, 19, 21, e 23.

Um equivalente funcional da seqüência nucleotídica reguladora da transcrição pode ser obtido ou é obtenível a partir de DNA genômico de planta de um gene codificando um polipeptídeo (preferivelmente de um gene ortólogo isto é, codificando uma proteína ortóloga) que é

substancialmente similar e preferivelmente tem pelo menos 50% ou 60%, preferivelmente pelo menos 70% ou 80%, mais preferivelmente pelo menos 90% ou 95%, o mais preferivelmente pelo menos 97% ou 98% identidade de seqüência de aminoácido para um polipeptídeo descrito por qualquer uma das SEQ ID NOs: 13, 15, 17, 19, 21, ou 23, respectivamente, ou um fragmento de referida seqüência nucleotídica reguladora da transcrição que demonstra atividade de promotor em um modo preferencial de semente ou específico de semente.

Os equivalentes funcionais adicionais à seqüência nucleotídica reguladora da transcrição da invenção são obteníveis de DNA genômico de planta de um gene codificando um polipeptídeo que compreende pelo menos dois ou três, preferivelmente pelo menos quatro ou cinco, mais preferivelmente seis, o mais preferivelmente todos os motivos de seqüência descritos por SEQ ID NO: 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51 ou 52 e – preferivelmente - demonstra atividade de promotor em plantas (preferivelmente em um modo preferencial de semente ou específico de semente).

Preferivelmente, o cassete de expressão compreende a seqüência nucleotídica reguladora da transcrição, que é obtida ou é obtenível a partir de DNA genômico de planta de um gene codificando um polipeptídeo (preferivelmente uma proteína ortóloga) que

a) compreende pelo menos um, preferivelmente pelo menos dois ou três, mais preferivelmente todos o(s) motivo(s) de seqüência selecionado (s) dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL (SEQ ID NO: 37), preferivelmente MPG(I/L)T(L/I)GDTVPNLE (SEQ ID NO: 38),

ii) GDFTPVC (SEQ ID NO: 39), preferivelmente LFSHPGDFTPVCTTEL (SEQ ID NO: 40),

iii) VKLLGLS (SEQ ID NO: 41),

preferivelmente RGVKLLGLSCDD (SEQ ID NO: 42),

iv) YPI(I/M)ADP (SEQ ID NO: 43), preferivelmente SKV(T/N)YPI(I/M)ADP (SEQ ID NO: 44),

v) KLSFLYP (SEQ ID NO: 45), preferivelmente (K/V)(I/V)KLSFLYPS (SEQ ID NO: 46),

vi) GRNMDEV (SEQ ID NO: 47), preferivelmente TGRNMDEV(L/V)RA (SEQ ID NO: 48),

vii) (I/V)ATP(V/A)NW (SEQ ID NO: 49), preferivelmente K(I/V)ATP(V/A)NW(K/N)P (SEQ ID NO: 50), e

viii) PSKKGYL (SEQ ID NO: 51), preferivelmente LPSKKGYLR (SEQ ID NO: 52),

e

b) tem pelo menos 50% identidade de seqüência de aminoácido para um polipeptídeo selecionado dentre o grupo descrito por SEQ ID NO: 13, 15, 17, 19, 21, e 23.

Preferivelmente, referida proteína ortóloga tem ainda a mesma atividade enzimática que a proteína codificada pelo *locus* de *Arabidopsis thaliana locus* At1g48130.

A atividade de uma seqüência nucleotídica reguladora da transcrição é considerada equivalente se a transcrição for iniciada em um modo preferencial de semente ou específico de semente (como definido acima). Este perfil de expressão é preferivelmente demonstrado usando genes repórteres operativamente ligados à referida seqüência nucleotídica reguladora da transcrição. Genes repórteres preferidos (Schenborn 1999) neste contexto são proteína fluorescente verde (GFP) (Chui 1996; Leffel 1997), cloranfenicol transferase, luciferase (Millar 1992), ß-glucuronidase ou β–galactosidase. Especialmente preferido é ß-glucuronidase (Jefferson 1987).

Além disso, a atividade reguladora de transcrição de uma função equivalente pode variar da atividade de sua seqüência parental,

especialmente com relação ao nível de expressão. O nível de expressão pode ser maior ou menor do que o nível de expressão da seqüência parental. Ambas as derivações podem ser vantajosas dependendo da seqüência de ácido nucleico de interesse a ser expressada. Preferidas são estas seqüências funcionais equivalentes, que – em comparação com sua seqüência parental– não deriva do nível de expressão de referida seqüência parental em mais do que 50%, preferivelmente 25%, mais preferivelmente 10% (como a ser preferivelmente julgado por ou expressão de mRNA ou expressão de proteína (por exemplo, gene repórter)). Ainda preferidas são seqüências equivalentes que demonstram uma expressão aumentada em comparação com sua seqüência parental, preferivelmente um aumento de pelo menos 50%, mais preferivelmente em pelo menos 100%, o mais preferivelmente em pelo menos 500%.

Preferivelmente, equivalente funcional da seqüência nucleotídica reguladora da transcrição pode ser obtido ou é obtenível a partir de DNA genômico de planta de um gene expressando um mRNA descrito por um cDNA que é substancialmente similar e preferivelmente tem pelo menos 50% ou 60%, preferivelmente pelo menos 70% ou 80%, mais preferivelmente pelo menos 90% ou 95%, o mais preferivelmente pelo menos 97% ou 98% de identidade de seqüência para uma seqüência descrita por qualquer uma dentre SEQ ID NOs: 12, 14, 16, 18, 20, ou 22, respectivamente, ou um fragmento de referida seqüência nucleotídica reguladora da transcrição que demonstra atividade de promotor em um modo preferencial de semente ou específico de semente.

Este equivalente funcional da seqüência nucleotídica reguladora da transcrição pode ser obtido a partir de outras espécies de planta usando as seqüências de promotor de *Arabidopsis thaliana* ou *Linum usitatissimum* preferenciais de semente ou específicas de semente, descritas aqui como sondas para triar genes estruturais homólogos em outras plantas

por hibridização sob condições de hibridização baixas, moderadas ou estringentes. Regiões das seqüências de promotor preferenciais de semente ou específicas de semente da presente invenção que são conservadas dentre as espécies também podem ser usadas como iniciadores de PCR para amplificar um segmento de uma espécie diferente de *Arabidopsis ou Linum usitatissimum*, e este segmento usado como uma sonda de hibridização (a última abordagem permitir uma triagem de maior estringência) ou em um teste de transcrição para determinar a atividade do promotor. Além disso, as seqüências de promotor preferenciais de semente ou específicas de semente podem ser empregadas para identificar seqüências estruturalmente relacionadas em uma base de dados usando algoritmos de computador.

Mais especificamente, com base nas seqüências de ácido nucleico da a presente invenção, ortólogos podem ser identificados ou isolados do genoma de qualquer organismo desejado, preferivelmente de outra planta, de acordo como técnicas bem conhecidas com base em sua similaridade de seqüência para as seqüências de ácido nucleico de *Arabidopsis* ou *Linum usitatissimum*, por exemplo, hibridização, PCR ou comparações de seqüência geradas por computador. Por exemplo, toda ou uma parte de uma seqüência de ácido nucleico particular de *Arabidopsis thaliana* ou *Linum usitatissimum* é usada como uma sonda que seletivamente hibridiza para as outras seqüências de genes presentes em uma população de fragmentos DNA genômico ou cDNA clonados (isto é, bibliotecas genômicas ou cDNA) de um organismo de fonte selecionado. Além disso, as bibliotecas de cDNA e genômicas apropriadas podem ser preparadas a partir de qualquer célula ou tecido de um organismo. Estas técnicas incluem triagem por hibridização de bibliotecas de DNA em placas (ou placas ou colônias; ver, por exemplo, Sambrook 1989) e amplificação por PCR usando iniciadores de oligonucleotídeo preferivelmente correspondendo a domínios de seqüência conservados dentre o polipeptídeo relacionado ou subseqüências das

seqüências nucleotídicas providas aqui (ver, por exemplo, Innis 1990). Estes métodos são particularmente bem apropriados para o isolamento de seqüências de genes de organismos intimamente relacionados com os organismos dos quais a seqüência de sonda é derivada. A aplicação destes métodos usando seqüências de *Arabidopsis thaliana* ou *Linum usitatissimum* como sondas é bem apropriada para o isolamento de seqüências de gene de qualquer organismo de fonte, preferivelmente outras espécies de plantas. Em uma abordagem de PCR, iniciadores de oligonucleotídeo podem ser projetados para uso em reações PCR para amplificar as seqüências de DNA correspondentes do cDNA ou DNA genômico extraído de qualquer planta de interesse. Métodos para projetar iniciadores de PCR e clonagem de PCR são geralmente conhecidos na arte.

Em técnicas de hibridização, toda ou parte de uma seqüência nucleotídica é usada como uma sonda que seletivamente hibridiza para outras seqüências nucleotídicas correspondentes presentes em uma população de fragmentos de DNA genômico clonados ou fragmentos de cDNA (isto é, bibliotecas genômica 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 hibridização podem ser feitas por rotulação de oligonucleotídeos sintéticos com base em uma seqüência da invenção. Métodos para a preparação de sondas para hibridização e para construção de bibliotecas de cDNA e genômica são geralmente conhecidos na arte e são descritos em Sambrook et al. (1989). Em geral, seqüências que hibridizam para as seqüências descritas aqui terão pelo menos cerca de 50% to 70% e mesmo cerca de 80% 85%, 90%, 95% a 9 8% ou mais de identidade com as seqüências descritas. Isto é, a similaridade de seqüência das seqüências pode estar na faixa, compartilhando pelo menos cerca de 50% to 70%, e mesmo

cerca de 80%, 85%, 90%, 95% a 98% de similaridade de seqüência.

As moléculas de ácido nucleico da invenção também podem ser identificadas por, por exemplo, uma busca de base de dados conhecidas para genes codificando polipeptídeos tendo uma seqüência especificada de identidade de aminoácido ou DNA tendo uma identidade de seqüência nucleotídica especificada. Métodos de alinhamento de seqüências para comparação são bem conhecidos na arte e são descritos aqui acima.

Assim, as moléculas de ácido nucleico isolado da invenção incluem os ortólogos das seqüências de *Arabidopsis thaliana* ou *Linum usitatissimum* descritas aqui, isto é, as seqüências nucleotídicas correspondentes em organismos diferentes de *Arabidopsis thaliana* ou *Linum usitatissimum* incluindo, mas não limitados a, plantas diferentes de *Arabidopsis thaliana* ou *Linum usitatissimum*, preferivelmente plantas dicotiledôneas, por exemplo, alfafa, girassol, soja, algodão, amendoim, tabaco, colza de semente oleígena ou beterraba de açúcar, mas também plantas de cereal como milho, trigo, centeio, gramas, sorgo, painço, cana de açúcar, cevada e banana. Um gene ortólogo é um gene de uma espécie diferente que codifica um produto tendo a mesma ou função similar, por exemplo, catalisando a mesma reação como um produto codificado por um gene de um organismo de referência. Assim, um ortólogo inclui polipeptídeos tendo menos que, por exemplo, 50% de identidade de seqüência de aminoácidos, mas cujo ortólogo codifica um polipeptídeo tendo a mesma ou função similar. Bases de dados como GenBank podem ser empregadas para identificar seqüências relacionadas com as seqüências de *Arabidopsis*, por exemplo, ortólogos em outras plantas dicotiledôneas. Alternativamente, técnicas de DNA recombinantes, como hibridização ou PCR, podem ser empregadas para identificar seqüências relacionadas com as seqüências de *Arabidopsis thaliana* ou *Linum usitatissimum* ou para clonar as seqüências equivalentes de DNAs diferentes.

As seqüências nucleotídicas reguladoras de transcrição da invenção ou seus equivalentes funcionais podem ser obtidas ou isoladas de qualquer fonte de planta ou não planta, ou produzidas sinteticamente por meios puramente químicos. As fontes preferidas incluem, mas não são limitadas às plantas, como definidas na seção de DEFINIÇÃO acima.

Assim, outra forma de realização preferida da invenção refere-se a um método para identificar e/ou isolar a seqüência nucleotídica reguladora da transcrição (preferivelmente com atividade reguladora de transcrição preferencial de semente ou específica de semente) caracterizado em que a referida identificação e/ou isolamento utiliza a seqüência de ácido nucleico codificando um polipeptídeo como descrita por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de referida seqüência de ácido nucleico. Preferidas são seqüências de ácido nucleico descritas por SEQ ID NO: 12, 14, 16, 18, 20, ou 22 ou partes das mesmas. "Parte" neste contexto significa uma seqüência de ácido nucleico de pelo menos 15 nucleotídeos consecutivos, preferivelmente pelo menos 25 nucleotídeos consecutivos, mais preferivelmente pelo menos 50 nucleotídeos consecutivos.

O método para identificação e/ou isolamento pode ser baseado em (mas não é limitado a) métodos descritos acima como reação em cadeia da polimerase, hibridização ou triagem de base de dados. Preferivelmente, este método da invenção é baseado em uma reação em cadeia da polimerase, em que referida seqüência de ácido nucleico ou sua parte é utilizada como um iniciador de oligonucleotídeo. O versado na arte conhece vários métodos para amplificar e isolar o promotor de um gene partindo de parte de sua seqüência de codificação (como, por exemplo, parte de um cDNA). Estes métodos podem incluir mas não são limitados a métodos como PCR inverso (" iPCR") ou "PCR interlaçado assimétrico térmico" (" TAIL PCR").

Assim, outra forma de realização da invenção refere-se a um método para prover ou produzir um cassete de expressão transgênica para

expressão heteróloga em plantas compreendendo as etapas de:

I. isolar de uma seqüência nucleotídica reguladora da transcrição um gene de planta utilizando pelo menos uma seqüência de ácido nucleico codificando um polipeptídeo descrito por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de pelo menos 15 nucleotídeos consecutivos de referida seqüência de ácido nucleico, e

II. funcionalmente ligar referida seqüência nucleotídica reguladora da transcrição a outra seqüência nucleotídica de interesse, que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

Ainda outra forma de realização da invenção refere-se a um método para prover um cassete de expressão transgênica para expressão preferencial de semente ou específica de semente compreendendo as etapas de:

I. isolar a seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente utilizando pelo menos uma seqüência de ácido nucleico ou uma parte da mesma, em que referida seqüência está codificando um polipeptídeo descrito por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de pelo menos 15 nucleotídeos consecutivos do mesmo, e

II. funcionalmente ligar referida seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente a outra seqüência nucleotídica de interesse, que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente.

Preferivelmente, a seqüência nucleotídica utilizada para isolamento de referida seqüência nucleotídica reguladora da transcrição está codificando um polipeptídeo (preferivelmente de um gene ortólogo isto é, codificando uma proteína ortóloga) compreendendo pelo menos um motivo

de seqüência selecionado dentre o grupo consistindo da seqüência de aminoácido

i) GDTVPNL (SEQ ID NO: 37), preferivelmente MPG(I/L)T(L/I)GDTVPNLE (SEQ ID NO: 38),

ii) GDFTPVC (SEQ ID NO: 39), preferivelmente LFSHPGDFTPVCTTEL (SEQ ID NO: 40),

iii) VKLLGLS (SEQ ID NO: 41),

preferivelmente RGVKLLGLSCDD (SEQ ID NO: 42),

iv) YPI(I/M)ADP (SEQ ID NO: 43), preferivelmente SKV(T/N)YPI(I/M)ADP (SEQ ID NO: 44),

v) KLSFLYP (SEQ ID NO: 45),

preferivelmente (K/V)(I/V)KLSFLYPS (SEQ ID NO: 46),

vi) GRNMDEV (SEQ ID NO: 47), preferivelmente TGRNMDEV(L/V)RA (SEQ ID NO: 48),

vii) (I/V)ATP(V/A)NW (SEQ ID NO: 49),

preferivelmente K(I/V)ATP(V/A)NW(K/N)P (SEQ ID NO: 50),

e

viii) PSKKGYL (SEQ ID NO: 51), preferivelmente LPSKKGYLR (SEQ ID NO: 52).

Preferivelmente, a seqüência de ácido nucleico empregada para o isolamento compreende pelo menos 15 nucleotídeos consecutivos, preferivelmente pelo menos 25 nucleotídeos consecutivos, mais preferivelmente pelo menos 50 nucleotídeos consecutivos de uma seqüência descrita por SEQ ID NO: 12, 14, 16, 18, 20, ou 22. Preferivelmente, o isolamento da seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente é realizado por uma reação em cadeia da polimerase utilizando referida seqüência de ácido nucleico como um iniciador. A ligação operacionável pode ser realizada por método de

clonagem padrão bem conhecido na arte, como clonagem mediada por ligação ou clonagem mediada por recombinação.

Preferivelmente, as seqüências nucleotídicas reguladoras de transcrição e promotores da invenção incluem um pedaço consecutivo de cerca de 25 a 2000, incluindo 50 a 500 ou 100 a 250, e até a 1000 ou 1500, nucleotídeos contíguos, por exemplo, 40 a cerca de 743, 60 a cerca de 743, 125 a cerca de 743, 250 a cerca de 743, 400 a cerca de 743, 600 a cerca de 743, de qualquer uma dentre SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, e 11, ou os seus promotores ortólogos, que incluem a região de promotor mínima.

Em uma forma de realização particular da invenção, referido pedaço consecutivo de cerca de 25 a 2000, incluindo 50 a 500 ou 100 a 250, e até 1000 ou 1500, nucleotídeos contíguos, por exemplo, 40 a cerca de 743, 60 a cerca de 743, 125 a cerca de 743, 250 a cerca de 743, 400 a cerca de 743, 600 a cerca de 743, tem pelo menos 50% ou 60%, preferivelmente pelo menos 70% ou 80%, mais preferivelmente pelo menos 90% e o mais preferivelmente pelo menos 95%, identidade de seqüência de ácido nucleico com um correspondente pedaço consecutivo de cerca de 25 a 2000, incluindo 50 a 500 ou 100 a 250, e até 1000 ou 1500, nucleotídeos contíguos, por exemplo, 40 a cerca de 743, 60 a cerca de 743, 125 a cerca de 743, 250 a cerca de 743, 400 a cerca de 743, 600 a cerca de 743, de qualquer um dentre SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, e 11, ou os seus promotores ortólogos, que incluem a região de promotor mínima. O pedaço acima definido de nucleotídeos contíguos preferivelmente compreendem um ou mais motivos de promotor selecionados dentre o grupo consistindo de caixa TATA, caixa GC-, caixa CAAT- e um sítio de partida de transcrição.

As seqüências nucleotídicas reguladoras de transcrição da invenção ou seus equivalentes funcionais são capazes de dirigir a expressão preferencial de semente ou específica de semente de uma seqüência de codificação em uma célula alvo, particularmente em uma célula de planta. As

seqüências de promotor e métodos descritos aqui são úteis na regulação da expressão preferencial de semente ou específica de semente, respectivamente, de qualquer seqüência nucleotídica heteróloga em uma planta hospedeira a fim de variar o fenótipo desta planta. Estes promotores podem ser usados com combinações de melhorador, elementos a montante, e/ou ativação de seqüências das regiões de flanco 5' de genes estruturais expressáveis em plantas Similarmente, o elemento a montante pode ser usado em combinação com várias seqüências de promotor de planta.

As seqüências nucleotídicas reguladoras de transcrição e promotores da invenção são úteis para modificar o fenótipo de uma planta. Varias mudanças no fenótipo de uma planta transgênica são desejáveis, isto é, a modificação da composição de ácido graxo em uma planta, alteração do teor de aminoácido de uma planta, alteração do mecanismo de defesa aos patógenos da planta, e semelhantes. Estes resultados podem ser obtidos ao prover a expressão de produtos heterólogos ou aumentada expressão de produtos endógenos em plantas. Alternativamente, os resultados podem ser obtidos ao prover uma redução de expressão de um ou mais produtos endógenos, particularmente enzimas ou cofatores na planta. Estas mudanças resultam em uma alteração no fenótipo da planta transformada.

Geralmente, as seqüências nucleotídicas reguladoras de transcrição e promotores da invenção 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 seqüência anti-sentido, uma seqüência codificando para uma seqüência de RNA filamento duplo, ou um transgene em plantas.

Uma ligação operativa pode - por exemplo – compreender um arranjo seqüencial da seqüência nucleotídica reguladora da transcrição da invenção (por exemplo, uma seqüência como descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11) com a seqüência de ácido nucleico a ser

expressada, e –opcionalmente –elementos regulatórios adicionais como por exemplo elementos de poliadenilação ou terminação de transcrição, melhoradores, íntrons etc, em um modo em que a seqüência nucleotídica reguladora da transcrição pode preencher sua função no processo de expressão da seqüência 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. Preferidos são os arranjos, em que a seqüência de ácido nucleico de interesse a ser expressada é colocada a jusante (isto é, na direção 3') da seqüência nucleotídica reguladora da transcrição da invenção em um modo, em que ambas as seqüências são covalentemente ligadas. Opcionalmente, seqüências adicionais podem ser inseridas entre as duas seqüências. Estas seqüências podem ser por exemplo sítios de clonagem de ligador ou múltiplos. Além disso, as seqüências podem ser inseridas codificando para partes de proteínas de fusão (no caso de uma proteína de fusão da proteína codificada pelo ácido nucleico de interesse ser destinada a ser expressada). Preferivelmente, a distância entre a seqüência de ácido nucleico de interesse a ser expressada e a seqüência nucleotídica reguladora da transcrição da invenção é não é maior do que 200 pares de base, preferivelmente não é maior que 100 pares de base, mais preferivelmente no é maior do que 50 pares de base.

Uma ligação operativa com relação a qualquer cassete de expressão ou da invenção pode ser realizada por vários métodos conhecidos na arte, compreendendo procedimentos tanto em *vitro* como em *vivo*. Assim, um cassete de expressão da invenção ou um vetor compreendendo este cassete de expressão pode ser realizado usando técnicas de recombinação e clonagem padrões bem conhecidas na arte (ver por exemplo, Maniatis 1989; Silhavy 1984; Ausubel 1987).

Um cassete de expressão também pode ser montado por inserção de uma seqüência nucleotídica reguladora da transcrição da invenção

(por exemplo, uma seqüência como descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11) no genoma da planta. Esta inserção irá resultar em uma ligação operativa em uma seqüência de ácido nucleico de interesse que, como tal, já existia no genoma. Esta inserção do ácido nucleico de interesse é expressada em um modo preferencial de semente ou específico de semente devido às propriedades reguladoras da transcrição da seqüência nucleotídica reguladora da transcrição. A inserção pode ser dirigida ou por acaso. Preferivelmente a inserção é dirigida e realizada por, por exemplo, recombinação homóloga. Por este procedimento, um promotor natural pode ser trocado contra a seqüência nucleotídica reguladora da transcrição da invenção, assim modificando o perfil de expressão de um gene endógeno. A seqüência nucleotídica reguladora da transcrição também pode ser inserida em um modo, em que o mRNA anti-sentido de um gene endógeno é expressado, assim induzindo o silenciamento do gene.

Similar, uma seqüência de ácido nucleico de interesse a ser expressada pode ser inserida no genoma da planta compreendendo a seqüência nucleotídica reguladora da transcrição em seu meio genômico i natural (isto é ligado ao seu gene natural) em um modo em que a seqüência inserida se torna operativamente ligada à seqüência nucleotídica reguladora da transcrição, assim formando um cassete de expressão da invenção.

O cassete de expressão pode ser empregado para numerosos fins de expressão como por exemplo expressão de uma proteína, ou expressão de um RNA anti-sentido, RNA sentido ou filamento duplo. Preferivelmente, a expressão da seqüência de ácido nucleico confere à planta um traço agronomicamente valioso.

A matriz de leitura aberta a ser ligada à seqüência nucleotídica reguladora da transcrição da invenção pode ser obtida a partir de um gene de resistência de inseto, um gene de resistência a doença como, por exemplo, gene de resistência a doença bacteriana, um gene de resistência a doença de

fungos, gene de resistência a doença viral, um gene de resistência a doença de nematódeos, um gene de resistência a herbicida, um gene afetando a composição ou qualidade dos grãos, um gene de utilização de nutrientes, um gene de redução de micotoxina, um gene de esterilidade masculina, um gene de marcador selecionável, um gene de marcador triável, um marcador selecionável negativo, marcador selecionável positivo, um gene afetando as características agronômicas da planta, isto é, rendimento, capacidade de ficar em pé, e semelhantes, ou um gene de resistência ao meio ambiente ou estresse, isto é, um ou mais genes que conferem resistência ou tolerância a herbicidas, resistência ou tolerância a insetos, resistência ou tolerância a doenças, (viral, bacteriana, fúngica, oomicetos, ou nematódeos), resistência ou tolerância a estresse, (como exemplificado por resistência ou tolerância a seca, calor, frio, congelamento, umidade excessiva, estresse de sal, ou estresse oxidativo), rendimentos aumentados, teor e compleição alimentícia, aparência física, esterilidade masculina, secura, capacidade de ficar em pé, fertilidade, propriedades ou quantidade de amido, quantidade e qualidade do óleo, composição de proteína ou aminoácido, e semelhantes. Por "resistente" significa-se uma planta, que não demonstra substancialmente mudanças fenotípicas como uma conseqüência de uma administração de agentes, infecção com um patógeno ou exposição a estresse. Por "tolerante", significa-se uma planta que, apesar de mostrar algumas mudanças fenotípicas como uma conseqüência de infecção, não tem uma capacidade reprodutiva substancialmente diminuída ou metabolismo substancialmente alterado.

As seqüências nucleotídicas reguladoras de transcrição preferenciais de semente ou específicas de semente (por exemplo, promotores) são úteis para expressar uma ampla variedade de genes incluindo os que alteram as vias metabólicas, conferem resistência a doenças, para a produção de proteínas, por exemplo, produção de anticorpos, ou para melhorar a absorção de nutrientes semelhantes. As seqüências nucleotídicas

reguladoras de transcrição preferenciais de semente ou específicas de semente (por exemplo, promotores) podem ser modificadas de modo a serem reguláveis, por exemplo, indutíveis. Os genes e seqüências nucleotídicas reguladoras de transcrição (por exemplo, promotores) descritos aqui acima podem ser usados para identificar genes ortólogos e suas seqüências nucleotídicas reguladoras de transcrição (por exemplo, promotores) que são também provavelmente expressados em um tecido e/ou modo de desenvolvimento particulares. Além disso, as seqüências nucleotídicas reguladoras de transcrição ortólogas (por exemplo, promotores) são úteis para expressar matrizes de leitura aberta ligadas. Além disso, por alinhamento das seqüências nucleotídicas reguladoras de transcrição (por exemplo, promotores) destes ortólogos, novos elementos cis podem ser identificados que são úteis para gerar seqüências nucleotídicas reguladoras de transcrição sintéticas (por exemplo, promotores).

A expressão regulando as seqüências nucleotídicas especificadas acima pode ser opcionalmente operativamente ligado a outras seqüências regulatórias apropriadas, por exemplo, uma seqüência de terminador de transcrição, operador, sítio de ligação de repressor, sítio de ligação de fator de transcrição e/ou um melhorador.

A presente invenção ainda provê um vetor recombinante contendo o cassete de expressão da invenção, e células hospedeiras compreendendo o cassete de expressão ou vetor, por exemplo, compreendendo um plasmídeo. O cassete de expressão ou vetor pode aumentar o genoma de uma planta 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 compreendida no DNA cromossômico do núcleo da planta. A presente invenção também provê uma planta transgênica preparada por este método, a

semente desta planta e plantas de progênie desta incluindo híbridos e cruzamentos endogâmicos. O cassete de expressão pode ser operativamente ligado a um gene estrutural, a sua matriz de leitura aberta, ou uma porção da mesma. O cassete de expressão pode ainda compreender um plasmídeo Ti e estar contido em uma célula de *Agrobacterium tumefaciens*; ele pode ser transportado em uma micropartícula, em que a micropartícula é apropriada para transformação balística de uma célula de planta; ou pode estar contido em uma célula de planta ou protoplasto. Além disso, o cassete de expressão ou vetor pode estar contido em uma planta transformada ou células da mesma, e a planta pode ser uma dicotiledônea ou monocotiledônea. Em particular, a planta pode ser uma planta dicotiledônea. Preferidas plantas transgênicas são milho, soja, cevada, alfafa, girassol, canola, soja, algodão, amendoim, sorgo, tabaco, beterraba de açúcar, arroz, trigo, centeio, gramas, painço, cana de açúcar, tomate, ou batata transgênicos.

A invenção também provê um método de cruzamento de plantas, por exemplo, para preparar uma planta transgênica fértil cruzada. 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, preparar a semente de uma planta transgênica fértil cruzada compreendendo o cassete de expressão particular. A semente é então plantada para obter uma planta transgênica fértil cruzada. A planta pode ser uma monocotiledônea ou a dicotiledônea. Em uma forma de realização particular, a planta é uma planta dicotiledônea. A planta transgênica fértil cruzada pode ter o cassete de expressão particular herdado através de um parental feminino ou através de um parental masculino. A segunda planta pode ser uma planta endogâmica. A planta transgênica fértil cruzada pode ser um híbrido. Também estão incluídas na presente invenção sementes de qualquer uma destas plantas transgênicas férteis cruzadas.

As seqüências nucleotídicas reguladoras de transcrição da invenção ainda compreendem seqüências que são complementares a uma (a seguir seqüência de "teste") que hibridiza sob condições estringentes com a molécula de ácido nucleico como descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11 assim como RNA que é transcrito da molécula de ácido nucleico. Quando a hibridização é realizada sob condições estringentes, ou a molécula de teste ou a de ácido nucleico da invenção é preferivelmente suportada, por exemplo, em uma membrana ou chip de DNA. Assim, ou uma molécula de teste desnaturada ou de ácido nucleico da invenção é preferivelmente primeiro ligada a um suporte e a hibridização é efetuada por um período de tempo especificado a uma temperatura de, por exemplo, entre 55 e 70°C, em solução salina tamponada com citrato de potência dupla (SC) contendo 0,1% SDS seguido por enxágüe do suporte na mesma temperatura mas com um tampão tendo uma concentração de SC reduzida. Dependendo do grau de estringência requerido, estes tampões de concentração reduzida são tipicamente de SC de potência única contendo 0,1% SDS, SC de potência média, contendo 0,1% SDS e SC de potência de um décimo, contendo 0,1% SDS. Mais preferivelmente, a hibridização é realizada sob condições de estringência elevada (como definido acima).

Virtualmente qualquer composição de DNA pode ser usada para a distribuição às células de planta recipientes, por exemplo, células de dicotiledônea, para produzir, por fim, 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 semelhantes, podem ser empregados. A construção de vetores, que podem ser empregados em conjunto com a presente invenção, serão bem conhecidos dos versados à luz da presente descrição (ver, por exemplo, Sambrook 1989; Gelvin 1990).

Vetores, plasmídeos, cosmídeos, YACs (cromossomos artificiais de leveduras) ACs (cromossomos artificiais bacterianos) e segmentos de DNA para uso na transformação destas células irão, como evidente, geralmente compreender o cDNA, gene ou genes que se deseja introduzidas nas células. Estas construções de DNA podem ainda incluir estruturas como promotores, melhoradores, poliligadores, ou mesmo genes regulatórios como desejado. O segmento de DNA, fragmento ou gene escolhido para introdução celular irá com freqüência codificar uma proteína que será expressada nas células recombinantes resultantes, como irão resultar em um traço triável ou selecionável 6 que irá conferir um fenótipo melhorado à planta regenerada. No entanto, este não pode ser sempre o caso, e a presente invenção também engloba plantas transgênicas incorporando transgenes não expressados.

Em algumas formas de realização, é contemplado que se pode desejar empregar vetores virais competentes para replicação na transformação de monocotiledôneas. Estes vetores incluem, por exemplo, vetores de transporte do vírus do ananismo do trigo (WDV) como pW1-11 e PW1-GUS (Ugaki 1991). Estes vetores são capazes de replicação autônoma em células de milho assim como E. coli, e como tal pode prover aumentada sensibilidade para detectar DNA distribuído nas células transgênicas. Um vetor de replicação também pode ser útil para distribuição de genes flanqueados por seqüências de DNA de elementos transposáveis, como Ac, Ds, ou Mu. Foi proposto (Laufs 1990) que a transposição destes elementos dentro do genoma do milho requer replicação de DNA. É também contemplado que elementos transposáveis devem ser úteis para introdução de segmentos de DNA ou fragmentos faltando elementos necessários para a seleção e manutenção do vetor de plasmídeo em bactérias, por exemplo, genes de resistência a antibióticos e origens de replicação de DNA. É também proposto que o uso de um transposável elemento como Ac, Ds, ou Mu deve ativamente promover a

ação do DNA desejado e assim aumentar a freqüência de células estavelmente transformadas. O uso de um elemento transposável como Ac, Ds, ou Mu pode ativamente promover a integração do DNA de interesse e assim aumentar a freqüência de células estavelmente transformadas. Elementos transposáveis podem ser úteis para permitir a separação de genes de interesse de elementos necessários para seleção e manutenção de um vetor plasmídeo em bactérias ou seleção de um transformante. Por uso de um elemento transposável, seqüências de DNA desejáveis e indesejáveis podem ser transposadas além de cada outra no genoma, como que através de segregação genética em progênie, pode-se identificar plantas com ou seqüências de DNA ou desejáveis ou indesejáveis.

A seqüência nucleotídica de interesse ligada a uma ou mais das seqüências nucleotídicas reguladoras de transcrição da invenção pode, por exemplo, codificar para um RNA ribossômico, um RNA anti-sentido ou qualquer outro tipo de RNA que não é traduzido em proteína. Em outra forma de realização preferida da invenção, referida seqüência nucleotídica de interesse é traduzida em um produto de proteína. A seqüência nucleotídica reguladora da transcrição e/ou seqüência nucleotídica de interesse ligada à mesma pode ser de origem homóloga ou heteróloga com relação à planta a ser transformada. A molécula recombinante de DNA útil para introdução em células de planta inclui a que foi derivada ou isolada de qualquer fonte, que pode ser subseqüentemente caracterizada com relação a estrutura, tamanho e/ou função, quimicamente alterada, e posteriormente introduzida em plantas. Um exemplo de uma seqüência nucleotídica ou segmento de interesse "derivado "de uma fonte, poderia ser uma seqüência nucleotídica ou segmento que é identificado como um fragmento utilizável dentro de um dado organismo, e que é então quimicamente sintetizado em forma essencialmente pura. Um exemplo desta seqüência nucleotídica ou segmento de interesse "isolado" de uma fonte, deveria ser uma seqüência nucleotídica ou segmento

que é excisado ou removido de referida fonte por meios químicos, por exemplo, pelo uso de endonucleases de restrição, de modo que pode ser ainda manipulada, por exemplo, amplificada, para uso na invenção, pela metodologia de engenharia genética. Esta seqüência nucleotídica ou segmento é comumente referida como "recombinante."

Assim, uma seqüência nucleotídica, segmento ou fragmento de interesse utilizável inclui DNA completamente sintético, DNA semi-sintético, DNA isolado de fontes biológicas, e DNA derivado de RNA introduzido. Geralmente, o DNA introduzido não é originalmente residente no genótipo da planta, que é o recipiente do DNA, mas está no escopo da invenção isolar um gene de um dado genótipo de planta, e subseqüentemente introduzir múltiplas cópias do gene no mesmo genótipo, por exemplo, para melhorar a produção de um dado produto de gene como uma proteína de armazenamento ou uma proteína que confere tolerância ou resistência a falta d'água.

A molécula recombinante introduzida de DNA inclui mas não é limitada a, DNA de genes de planta, e genes não de planta como os de bactérias, leveduras, animais ou vírus. O DNA introduzido pode incluir genes modificados, porções de genes, ou gene quiméricos, incluindo genes do mesmo ou diferentes genótipos. O termo "gene quimérico" ou "DNA quimérico" é definido como um gene ou uma seqüência de DNA ou segmento compreendendo pelo menos duas seqüências ou segmentos de DNA de espécies que não combinam DNA sob condições naturais, ou cujas seqüências ou segmentos de DNA são posicionadas ou ligadas em um modo que não ocorre normalmente no genoma nativo de planta não transformada.

A molécula recombinante introduzida de DNA usada para transformação aqui pode ser circular ou linear, filamento duplo ou filamento único. Geralmente, o DNA está na forma de DNA quimérico, como DNA plasmídeo que pode também conter regiões de codificação flanqueadas por seqüências regulatórias, que promovem a expressão do DNA recombinante

presente na planta resultante. Geralmente, a molécula recombinante introduzida de DNA será relativamente pequena, isto é, menor do que cerca de 30 kb para minimizar qualquer susceptibilidade a degradação física, química ou enzimática, que se sabe aumenta como o tamanho da molécula de nucleotídeo aumenta. Como notado acima, o número de proteínas, transcritos de RNA ou sua misturas, que é introduzido no genoma da planta, é preferivelmente pre-selecionado e definido por exemplo, dentre um a cerca de 5-10 de tais produtos do DNA introduzido que podem ser formados.

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

A obtenção de níveis suficientes de expressão de transgene nos tecidos de plantas apropriados é um aspecto importante na produção de culturas geneticamente engenheiradas. Expressão de seqüências de DNA heterólogas em um hospedeiro de planta é dependente da presença de um promotor ligado operativamente que é funcional dentro do hospedeiro de planta. A escolha da seqüência de promotor irá determinar quando e onde dentro do organismo a seqüência de DNA heteróloga é expressada.

É especificamente contemplado pelos inventores que se pode mutageneizar um promotor para potencialmente melhorar a utilidade dos

elementos para a expressão de transgenes em plantas. A mutagênese destes elementos pode ser realizada em seqüências de promotor aleatórias e mutageneizadas triadas para atividade por um procedimento de tentativa e erro. Alternativamente, seqüências particulares que provêem o promotor com características de expressão desejáveis, ou o promotor com atividade de melhora de expressão, pode ser identificado e estas ou seqüências similares introduzidas nas seqüências via mutação. É ainda contemplado que se pode mutageneizar estas seqüências a fim de melhorar sua expressão de transgenes em uma espécie particular.

Os meios para mutageneizar um segmento de DNA codificando a seqüência de promotor da presente invenção são bem conhecidos dos versados na arte. Como indicado, modificações do promotor ou outros elementos regulatórios podem ser feitas por procedimentos de mutagene aleatória ou sítio-específica. O promotor e outros elementos regulatórios podem ser modificados por alteração de sua estrutura através de adição ou deleção de um ou mais nucleotídeos de uma seqüência que codifica as correspondentes seqüências não modificadas.

Mutagênese pode ser realizada de acordo com qualquer uma das técnicas conhecidas na arte, como, e não limitados a, síntese de um oligonucleotídeo tendo uma ou mais mutações dentro de uma seqüência de uma região regulatória particular. Em particular, mutagene sítio-específica é uma técnica utilizável na preparação de mutantes de promotor, através de mutagênese específica do DNA subjacente. A técnica ainda provê uma pronta capacidade de preparar e testar variantes de seqüência, por exemplo, incorporando uma ou mais das acima considerações, por introdução de uma ou mais mudanças de seqüência nucleotídica no DNA. Mutagene sítio-específica permite a produção de mutantes através de uso de seqüências oligonucleotídicas específicas que codificam a seqüência de DNA da desejada mutação, assim como um suficiente número de nucleotídeos adjacentes,

prover uma seqüência de iniciador de suficiente tamanho e complexidade seqüência para formar um duplex estável em ambos os lados da junção de deleção sendo atravessada. Tipicamente, um iniciador de cerca de 17 a cerca de 75 nucleotídeos ou mais em comprimento é preferido, com cerca de 10 a cerca de 25 ou mais resíduos em ambos os lados da junção de uma seqüência sendo alterados.

Em geral, a técnica de mutagene sítio-específica é bem conhecida na arte, como exemplificado por várias publicações. Como será notado, a técnica tipicamente emprega um vetor de fago, que existe em tanto uma forma de filamento único como filamento duplo. Os vetores úteis típicos em mutagênese sítio-dirigida incluem vetores como o fago M13. Estes fagos são prontamente comercialmente disponíveis e seu uso é geralmente bem conhecidos versados na arte. Plasmídeos de filamento duplo também são empregados como rotina em mutagênese sítio-dirigida, que elimina a etapa de transferência do gene de interesse de um plasmídeo a um fago.

Em geral, mutagênese sítio-dirigida de acordo com o presente é realizada por primeiro obtenção de um vetor de filamento único ou afastamento por fusão dos dois filamentos de um vetor de filamento duplo, que inclui em sua seqüência, uma seqüência de DNA que codifica o promotor. O iniciador de oligonucleotídeo contendo a seqüência mudada desejada é preparado, geralmente sinteticamente. Este iniciador é então anelado com o vetor de filamento único, e submetido a enzimas de polimerização de DNA como fragmento Klenow de polimerase I de E. coli, a fim de completar a síntese do filamento contendo a mutação. Assim, um heteroduplex é formado em que um filamento codifica a seqüência original não mudada e o segundo filamento contendo a desejada mutação. Este vetor heteroduplex é então usado para transformar ou transfectar as células apropriadas, como células de E. coli, e células são selecionadas que incluem vetores recombinantes contendo o arranjo de seqüência mudado. Vetor DNA pode então ser isolado

destas células e usado para transformação da planta. Um esquema de seleção genético foi projetado por Kunkel et al. (1987) para enriquecer os clones incorporando oligonucleotídeos mutagênicos. Alternativamente, o uso de PCR com enzimas termoestáveis, comercialmente disponíveis, como Taq polimerase pode ser usado para incorporar um iniciador oligonucleotídeo mutagênico em um fragmento de DNA amplificado que pode então ser clonado em um vetor de clonagem ou expressão apropriado. Os procedimentos de mutagênese mediados por PCR de Tomic et al. (1990) e Upender et al. (1995) fornecem dois exemplos destes protocolos. Um PCR empregando uma ligase termoestável em adição a uma polimerase termoestável, também pode ser usado para incorporar um oligonucleotídeo mutagênico fosforilado em um fragmento de DNA amplificado que pode então ser clonado em um vetor de clonagem ou expressão apropriado. O procedimento de mutagênese descrito por Michael (1994) provê um exemplo de um tal protocolo.

A preparação de variantes de seqüência dos segmentos de DNA codificando promotores selecionados, usando mutagênese sítio-dirigida é provido como um meio de produzir espécies potencialmente úteis e não se significa ser limitativo, como existem outros modos em que as variantes de seqüências de DNA podem ser obtidas. Por exemplo, vetores recombinantes codificando a seqüência de promotor desejada podem ser tratados com agentes mutagênicos, como hidroxilamina, para obter variantes de seqüência.

Como usado aqui; o termo "procedimento de mutagênese dirigida para oligonucleotídeo "refere-se a processos dependentes de gabarito e propagação mediada por vetor que resultam em um aumento na concentração de uma molécula específica de ácido nucleico relativa à sua concentração inicial, ou em um aumento na concentração de um sinal detectável, como amplificação. Como usado aqui, o termo "procedimento de mutagênese dirigido a oligonucleotídeo" também se destina a fazer referência

a um processo que envolve uma extensão dependente de gabarito de uma molécula de iniciador. O termo processo dependente de gabarito refere-se à síntese de ácido nucleico de um RNA ou uma molécula de DNA em que a seqüência do filamento recentemente sintetizado de ácido nucleico é ditada por regras bem conhecidas de formação de pares de bases complementares (ver, por exemplo, Watson e Ramstad, 1987). Tipicamente, metodologias mediadas por vetor envolvem a introdução do fragmento de ácido nucleico em um vetor de DNA ou RNA, a amplificação clonal do vetor, e a recuperação do fragmento de ácido nucleico amplificado. Exemplos de tais metodologias são providos por patente US No. 4.233.224. Vários processos dependentes de gabarito são disponíveis para amplificar as seqüências de marcação de interesse presentes em uma amostra, estes métodos sendo bem conhecidos na arte e especificamente descritos aqui abaixo.

Onde um clone compreendendo um promotor foi isolado de acordo com a presente invenção, pode-se desejar delimitar as regiões de promotor essenciais dentro de um clone. Um meio marcado efetivo para preparar os promotores mutagenizantes se baseia na identificação de elementos regulatórios putativos dentro da seqüência de promotor. Isto pode ser iniciado por comparação com seqüências de promotor conhecidas como sendo expressadas em um modo específico de tecido ou de desenvolvimento singular. Seqüências, que são compartilhadas dentre promotores com padrões similares de expressão, são provavelmente candidatos para a ligação de fatores de transcrição e são assim provavelmente elementos que conferem padrões de expressão. Confirmação destes elementos regulatórios putativos pode ser obtida por análise de deleção de cada região regulatória putativa, seguido por análise funcional de cada construção de deleção por teste de um gene repórter, que é funcionalmente fixado a cada construção. Como tal, uma vez que a seqüência de promotor de partida é provida, qualquer um dentre um número de mutantes de deleção diferentes do promotor de partida pode ser

prontamente preparado.

Os fragmentos funcionalmente equivalentes de uma seqüência nucleotídica reguladora da transcrição da invenção podem também ser obtidos por remoção ou deleção de seqüências não essenciais sem a deleção da essencial. A limitação da seqüência nucleotídica reguladora da transcrição ao seus elementos mediadores de transcrição essenciais pode ser realizada in vitro por mutações de deleção de tentativa-e-seta, ou in silico usando rotinas de busca do elemento promotor. Regiões essenciais para a atividade do promotor com freqüência demonstram agrupamentos de alguns elementos de promotor conhecidos. Esta análise pode ser realizada usando algoritmos disponíveis de computador, como PLACE ("Plant Cis-acting Regulatory DNA Elements" ; Higo 1999), a base de dados de BIOBASE " Transfac" (Biologische Datenbanken GmbH, Braunschweig; Wingender 2001) ou a base de dados PlantCARE (Lescot 2002).

Preferivelmente, fragmentos equivalentes funcionais de uma das seqüências nucleotídicas reguladoras de transcrição da invenção compreende pelo menos 100 pares de base, preferivelmente, pelo menos 200 pares de base, mais preferivelmente pelo menos 500 pares de base de uma seqüência nucleotídica reguladora da transcrição como descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11. Mais preferivelmente, este fragmento começa a partir da extremidade 3' das seqüências indicadas.

Especialmente preferidos são os fragmentos equivalentes de seqüências nucleotídicas reguladoras de transcrição, que são obtidos por deleção da região codificando a região 5'-não traduzida do mRNA, assim somente provendo a região do promotor (não transcrita). A região 5'-não traduzida pode ser facilmente determinada por métodos conhecidos na arte (como análise 5'-RACE). Assim, algumas das seqüências nucleotídicas reguladoras de transcrição da invenção são fragmentos equivalentes de outros seqüências (ver Tabela 2 abaixo).

Tabela 2: Relações de seqüências nucleotídicas reguladoras de transcrição da invenção

Seqüência reguladora da transcrição	**Seqüência equivalente**	**Fragmento equivalente**
SEQ ID NO: 1 (1683 bp)	SEQ ID NO: 5 (1696 bp)	SEQ ID NO: 2 (1619 bp) SEQ ID NO: 3 (985 bp) SEQ ID NO: 4 (921 bp) SEQ ID NO: 6 (1632 bp) SEQ ID NO: 7 (989 bp) SEQ ID NO: 8 (925 bp)
SEQ ID NO: 9 (1727 bp)		SEQ ID NO: 10 (933 bp) SEQ ID NO: 11 (458 bp)

Como indicado acima, mutantes de deleção, mutantes de deleção do promotor da invenção também 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 a triagem da atividade consiste em fixar um promotor deletado ou construção de intron, que contém um segmento deletável para um marcador selecionável ou triável, e para isolar somente as células expressando o gene marcador. Deste modo, um número de diferentes construções 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.

Uma análise de motivo de promotor pode ser feita, por exemplo, por uso do software Genomatix MatInspector Release professional 7.3 (Agosto 2004) (ver Exemplo 8 para detalhes. De acordo com outra forma de realização da invenção, a seqüência reguladora de transcrição compreendida nos cassetes de expressão compreende pelo menos uma região de seqüência selecionada (preferivelmente 2 regiões, mais preferivelmente todas as 3 regiões; em que preferivelmente a região b) está seguindo a região a), e região c) se seguindo a região a) e/ ou região b)) dentre o grupo

consistindo de

a) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os seguintes motivos:

i) **TACT** (SEQ ID NO: 53), preferivelmente **an**TACTctt (SEQ ID NO: 77), mais preferivelmente **ag**TACTctttga (SEQ ID NO: 101),

ii) **AGAA** (SEQ ID NO: 54), preferivelmente **ga**nnnn**ac**n**AGAA** (SEQ ID NO: 78), mais preferivelmente a**gag**gag**ac**a**AGAA**a (SEQ ID NO: 102),

iii) **AAAG** (SEQ ID NO: 55), preferivelmente **ga**nnn**tgcAAAG**nnn**a** (SEQ ID NO: 79), mais preferivelmente **ga**ag**ttgcAAAG**ttg**ag** (SEQ ID NO: 103),

iv) **CCAT** (SEQ ID NO: 56), preferivelmente **CCATtccttgtagta** (SEQ ID NO: 80), mais preferivelmente **CCATtccttgtagta**ct (SEQ ID NO: 104),

v) **TAAA** (SEQ ID NO: 57), preferivelmente **TAAAnaatc** (SEQ ID NO: 81), mais preferivelmente **TAAAtaatc**tat (SEQ ID NO: 105),

vi) **CTAT** (SEQ ID NO: 58), preferivelmente **ccattgCTATcctngntagaaacta** (SEQ ID NO: 82), mais preferivelmente **tccattgCTATccttgttagaaacta**a (SEQ ID NO: 106),

vii) **ACGT** (SEQ ID NO: 59), mais preferivelmente **tnnnACGTa** (SEQ ID NO: 83), mais preferivelmente ttttac**ACGTa**cc (SEQ ID NO: 107),

em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo como a numeração indicada acima (isto é, número ii) está a jusante de número i)), b) uma região (tendo um comprimento de 50 a 1000 bp, preferivelmente 60 a 500 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais

preferivelmente todos os seguintes motivos

i) **AAAA** (SEQ ID NO: 60), preferivelmente **a**nnnn**aAAAAtgtta** (SEQ ID NO: 84), mais preferivelmente ac**a**aaac**aAAAAtgtta**atat (SEQ ID NO: 108),

ii) **GTTA** (SEQ ID NO: 61), preferivelmente **a**nn**anGTTA** (SEQ ID NO: 85), mais preferivelmente aca**a**aa**atGTTA**ata (SEQ ID NO: 109),

iii) **TTAA** (SEQ ID NO: 62), preferivelmente **TTAAna** (SEQ ID NO: 86), mais preferivelmente aaaaatg**TTAAta**tttt (SEQ ID NO: 110),

iv) **ACAT** (SEQ ID NO: 63), preferivelmente **gaaa**nn**a**nnn**ACAT** (SEQ ID NO: 87), mais preferivelmente atat**gaaa**at**a**tta**ACAT**ttt (SEQ ID NO: 111),

v) **TnCC** (SEQ ID NO: 64), preferivelmente **TnCC**nn**a**nnn**gt** (SEQ ID NO: 88), mais preferivelmente aat**TTCCatatttgt**aggaaa (SEQ ID NO: 112),

em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo como a numeração indicada acima (isto é, número ii) está a jusante de número i)),

c) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os seguintes motivos

i) **CATG** (SEQ ID NO: 65), preferivelmente **CATGca** (SEQ ID NO: 89), mais preferivelmente cccaagta**CATGca**aaatttaacggtt (SEQ ID NO: 113),

ii) **TTAT** (SEQ ID NO: 66), mais **TTATtataagtaaat** (SEQ ID NO: 90), **mais** preferivelmente **TTATtataagtaaat**tg (SEQ ID NO: 114),

iii) **TTTT** (SEQ ID NO: 67), mais **TTTTt** (SEQ ID NO: 91), mais preferivelmente at**TTTTt**atta (SEQ ID NO: 115),

iv) **ACGT** (SEQ ID NO: 68), preferivelmente **gtgg**nng**g**nnn**ACGT** (SEQ ID NO: 92), mais preferivelmente aaat**gtgg**ta**g**cgt**ACGT**gtg (SEQ ID NO: 116),

v) **AGAA** (SEQ ID NO: 69), preferivelmente **gg**nnnn**t**c**n**AGAA (SEQ ID NO: 93), mais preferivelmente t**gg**cggc**t**cc**AGAA**t (SEQ ID NO: 117),

vi) ***ACGT*** (SEQ ID NO: 70), preferivelmente cgt***ACGT***ggc (SEQ ID NO: 94), mais preferivelmente cttgcgt***ACGT***ggcggc (SEQ ID NO: 118),

vii) **ACGT** (SEQ ID NO: 71), preferivelmente **ACGTggc** (SEQ ID NO: 95), mais preferivelmente tgaagaga**ACGTggc**cctgag (SEQ ID NO: 119),

viii) **ACGT** (SEQ ID NO: 72), preferivelmente **g**nnn**ACGTg** (SEQ ID NO: 96), mais preferivelmente tgcatgaa**g**aga**ACGTg**gccctgagtc (SEQ ID NO: 120),

ix) **CATG** (SEQ ID NO: 73), preferivelmente **CATGca** (SEQ ID NO: 97), mais preferivelmente gttctctt**CATGca**aggatagtagaac (SEQ ID NO: 121),

x) **ACCC** (SEQ ID NO: 74), preferivelmente **c**A**CCCa**n**c** (SEQ ID NO: 98), mais preferivelmente actc**cACCCacc**tcc (SEQ ID NO: 122),

xi) **CCAT** (SEQ ID NO: 75), preferivelmente **ca**nn**CCATc** (SEQ ID NO: 99), mais preferivelmente ctttcc**ca**tc**CCATc**ca (SEQ ID NO: 123),

xii) **TTTT** (SEQ ID NO: 76), preferivelmente **TTTTt** (SEQ ID NO: 100), mais preferivelmente ca**TTTTt**atca (SEQ ID NO: 124),

em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo como a numeração indicada acima (isto é, número ii) está a jusante de número i)).

O termo "motivo" como usado acima significa que nem todos os nucleotídeos da seqüência indicada precisam estar presentes para qualificar a presença de um motivo. Os nucleotídeos individuais de uma matriz de motivo completa (por exemplo, atag**attaTTTA**gaa) são de relevância diferente. O motivo compreende uma região conservada (por exemplo, **attaTTTA**), – que consiste de uma região de núcleo (por exemplo, **TTA**; letras maiúsculas em negrito) e uma parte menos conservada (por exemplo, **att,** letras minúsculas em negrito). Os motivos descritos acima podem estar presentes em uma ou mais cópias na região respectiva (ver Exemplo 8 por exemplos).

Mais especificamente, a seqüência reguladora de transcrição compreendida nos cassetes de expressão compreende pelo menos uma região de seqüência (preferivelmente 2 regiões, mais preferivelmente todas as 3 regiões; em que preferivelmente região b) é seguinte região a), e região c) se seguindo a região a) e/ ou região b)) selecionada dentre o grupo consistindo de

a) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 53, 54, 55, 56, 57, 58, ou 59, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 54 está a jusante de SEQ ID NO: 53),

b) uma região (tendo um comprimento de 50 a 1000 bp, preferivelmente 60 a 500 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 60, 61, 62, 63 ou 64, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID

NO: 61 está a jusante de SEQ ID NO: 60),

c) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ou 76, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 71 está a jusante de SEQ ID NO: 70).

Preferivelmente, a seqüência reguladora de transcrição compreendida nos cassetes de expressão compreende pelo menos uma região de seqüência (preferivelmente 2 regiões, mais preferivelmente todas as 3 regiões; em que preferivelmente região b) é seguinte região a), e região c) se seguindo a região a) e/ ou região b)) selecionada dentre o grupo consistindo de

a) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 77, 78, 79, 80, 81, 82, ou 83, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 78 está a jusante de SEQ ID NO: 77),

b) uma região (tendo um comprimento de 50 a 1000 bp, preferivelmente 60 a 500 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 84, 85, 86, 87, ou 88, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 85 está a jusante de SEQ ID NO: 84),

c) uma região (tendo um comprimento de 50 a 1500 bp,

preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, ou 100, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 71 está a jusante de SEQ ID NO: 70).

Mais preferivelmente, a seqüência reguladora de transcrição compreendida nos cassetes de expressão compreende pelo menos uma região de seqüência (preferivelmente 2 regiões, mais preferivelmente toda as 3 regiões; em que preferivelmente a região b) é seguinte região a), e região c) se seguindo a região a) e/ ou região b)) selecionada dentre o grupo consistindo de

a) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 101, 102, 103, 104, 105, 106, ou 107, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 78 está a jusante de SEQ ID NO: 77),

b) uma região (tendo um comprimento de 50 a 1000 bp, preferivelmente 60 a 500 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 108, 109, 110, 111, ou 112, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 85 está a jusante de SEQ ID NO: 84),

c) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais

preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, ou 124, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 114 está a jusante de SEQ ID NO: 115).

Os motivos essenciais dos promotores descritos aqui também podem ser empregados para construir um promotor sintético por fusão dos referidos motivos juntos por técnicas de clonagem padrão.

Assim, outra forma de realização da invenção refere-se a uma seqüência reguladora de transcrição sintética compreendendo pelo menos uma região de seqüência (preferivelmente 2 regiões, mais preferivelmente todas as 3 regiões; em que preferivelmente região b) é seguinte região a), e região c) se seguindo a região a) e/ ou região b)) selecionada dentre o grupo consistindo de

a) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 53, 54, 55, 56, 57, 58, ou 59, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 54 está a jusante de SEQ ID NO: 53),

b) uma região (tendo um comprimento de 50 a 1000 bp, preferivelmente 60 a 500 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 60, 61, 62, 63 ou 64, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 61 está a jusante de SEQ ID NO: 60),

c) uma região (tendo um comprimento de 50 a 1500 bp,

preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ou 76, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 71 está a jusante de SEQ ID NO: 70).

Preferivelmente, a seqüência reguladora de transcrição sintética compreende pelo menos uma região de seqüência (preferivelmente 2 regiões, mais preferivelmente todas as 3 regiões; em que preferivelmente região b) é a seguinte região a), e região c) se seguindo a região a) e/ ou região b)) selecionada dentre o grupo consistindo de

a) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 77, 78, 79, 80, 81, 82, ou 83, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 78 está a jusante de SEQ ID NO: 77),

b) uma região (tendo um comprimento de 50 a 1000 bp, preferivelmente 60 a 500 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 84, 85, 86, 87, ou 88, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 85 está a jusante de SEQ ID NO: 84),

c) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 89,

90, 91, 92, 93, 94, 95, 96, 97, 98, 99, ou 100, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 71 está a jusante de SEQ ID NO: 70).

Mais preferivelmente, a seqüência reguladora de transcrição sintética compreende pelo menos uma região de seqüência (preferivelmente 2 regiões, mais preferivelmente todas as 3 regiões; em que preferivelmente região b) é seguinte região a), e região c) se seguindo a região a) e/ ou região b)) selecionada dentre o grupo consistindo de

a) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 101, 102, 103, 104, 105, 106, ou 107, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 78 está a jusante de SEQ ID NO: 77),

b) uma região (tendo um comprimento de 50 a 1000 bp, preferivelmente 60 a 500 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 108, 109, 110, 111, ou 112, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 85 está a jusante de SEQ ID NO: 84),

c) uma região (tendo um comprimento de 50 a 1500 bp, preferivelmente 60 a 1000 bp) compreendendo pelo menos dois, preferivelmente três ou quatro, mais preferivelmente cinco de seis, o mais preferivelmente todos os motivos de promotor descritos por SEQ ID NO: 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, ou 124, em que referidos motivos são preferivelmente orientados em orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 114 está a jusante de SEQ ID

NO: 115).

Ainda outra forma de realização da invenção refere-se a um método para prover um seqüência nucleotídica reguladora de transcrição sintética caracterizado em que os motivos de promotor isolados ou agrupamentos de motivos de promotor são combinados (por exemplo, por técnicas de clonagem padrão, como ligação ou recombinação), estes motivos compreendendo pelo menos 5, preferivelmente pelo menos 8, mais preferivelmente pelo menos 10, o mais preferivelmente pelo menos 15 todos os motivos descritos por qualquer uma dentre SEQ ID NO: 53, 54, 55, 56, 57, 58, 59,: 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ou 76.

Preferivelmente, o método para prover seqüências nucleotídicas reguladoras de transcrição sintéticas, caracterizado em que os motivos de promotor isolados ou agrupamento de motivos de promotor são combinados (por exemplo, por técnicas de clonagem padrão, como ligação ou recombinação) estes motivos compreendendo pelo menos 5, preferivelmente pelo menos 8, mais preferivelmente pelo menos 10, o mais preferivelmente pelo menos 15 ou todos os motivos descritos por qualquer uma dentre SEQ ID NO: 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, ou 100.

Mais preferivelmente, o método para prover seqüências nucleotídicas reguladoras de transcrição sintéticas, é caracterizado em que os motivos de promotor isolados ou agrupamento de motivos de promotor são combinados (por exemplo, por técnicas de clonagem padrão, como ligação ou recombinação), estes motivos compreendendo pelo menos 5, preferivelmente pelo menos 8, mais preferivelmente pelo menos 10, o mais preferivelmente pelo menos 15 ou todos os motivos descritos por qualquer uma dentre SEQ ID NO: 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112,: 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, ou 124,

Preferivelmente, referidos motivos são orientados em

orientação 5'-3' de acordo com sua numeração de SEQ ID (isto é, SEQ ID NO: 114 está a jusante de SEQ ID NO: 115).

Um cassete de expressão da invenção pode compreender ainda elementos regulatórios. O termo neste contexto deve ser entendido em um significado amplo compreendendo todas as seqüências que podem influenciar a construção ou função do cassete de expressão. Elementos regulatórios podem, por exemplo, modificar a transcrição e/ou translação em organismo procariótico ou eucariótico. Em uma forma de realização preferida, o cassete de expressão da invenção compreendido a jusante (na direção 3') da seqüência de ácido nucleico a ser expressada em uma seqüência de terminação da transcrição e – opcionalmente elementos regulatórios adicionais– cada associado operativamente à seqüência de ácido nucleico a ser expressada (ou a seqüência nucleotídica reguladora da transcrição).

Os elementos regulatórios adicionais podem compreender promotores, promotores mínimos, ou elementos promotores adicionais, que podem modificar as propriedades de regulação da expressão. Por exemplo a expressão pode ser feita dependendo de alguns fatores de estresse, como estresse de água, abscisina (Lam 1991) ou estresse de calor (Schoffl 1989). Além disso, promotores ou elementos de promotor adicionais podem ser empregados, que podem realizar a expressão em outros organismos (como *E.coli* ou *Agrobacterium*). Estes elementos regulatórios podem ser encontrados nas seqüências de promotor ou bactérias como amy e SPO2 ou nas seqüências de promotor de promotores levedura ou fúngicos (como ADC1, MFa, AC, P-60, CYC1, GAPDH, TEF, rp28, e ADH).

Além disso, é contemplado que promotores combinando elementos de mais do que um promotor podem ser utilizáveis. Por exemplo, US 5.491.288 descreve combinar um promotor do vírus do mosaico da couve-flor com um promotor de histona. Assim, os elementos dos promotores descritos aqui podem ser combinados com elementos de outros promotores.

Promotores, que são úteis para expressão na planta de transgenes incluem os que são indutíveis, virais, sintéticos, constitutivos (Odell 1985), temporariamente regulados, espacialmente regulado, específicos para tecido e regulado espacial-temporariamente.

Onde a expressão em tecidos ou órgãos específicos é desejada, promotores específicos de tecido podem ser usados. Em contraste, onde expressão de gene em resposta a um estímulo é desejada, promotores indutíveis são os elementos regulatórios de escolha. Onde a expressão contínua é desejada em todas as células de uma planta, promotores constitutivos são utilizados. Seqüências regulatórias adicionais a montante e/ou a jusante da seqüência de núcleo de promotor podem ser incluídas em construções de expressão de vetores de transformação para ocasionar níveis variados de expressão de seqüências nucleotídicas heterólogas em uma planta transgênica.

A variedade de seqüências regulatórias 5' e 3' transcripcionais são disponíveis para uso na presente invenção. Terminadores transcripcionais são responsáveis pela terminação de transcrição e poliadenilação correta de mRNA. A seqüência de DNA regulatório 3' não traduzida 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 nucleotídeos e contém seqüências de terminação transcripcional e translacional da planta. Os terminadores transcripcionais 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 rbcS E9 de ervilha, 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 ou tomate, apesar de outros elementos 3'conhecidos dos versados na arte poderem ser também empregados. Alternativamente, pode-se também usar uma gama coixina oleosina3 ou outro terminador de gênero Coix.

Os elementos 3' preferidos incluem os do gene nopalina sintase de Agrobacterium tumefaciens (Bevan 1983), 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 ou tomate.

Como a seqüência de DNA entre o sítio de iniciação de transcrição e o início da seqüência de codificação, isto é, a seqüência líder não traduzida, pode influenciar a expressão de gene, pode-se também desejar empregar uma seqüência líder em particular. As seqüências líder preferidas são contempladas como incluindo as, que incluem seqüências, previstas para dirigir a expressão ótima do gene fixado, isto é, incluindo uma seqüência líder de consenso preferida, que pode aumentar ou manter a estabilidade do mRNA e evitar a iniciação de tradução inapropriada. A escolha de tais seqüências será conhecida dos versados na arte à luz da presente descrição. Seqüências que são derivadas de 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 seqüências que são verificadas como melhorando a expressão de gene em plantas transgênicas incluem seqüências de íntrons (por exemplo, de Adh1, bronze1, actina 1, actina 2 (WO 00/760067), ou o intron de sacarose sintase; ver: The Maize Handbook, Capítulo 116, Freeling e Walbot, Eds., Springer, New York (1994)) e seqüências líder viral (por exemplo, de TMV, MCMV e AMV; Gallie 1987). Por exemplo, várias seqüências líder não traduzidas derivadas de vírus são conhecidas como melhorando a expressão. Especificamente, seqüências líder do vírus do mosaico do tabaco (TMV), vírus de sarapintado clorótico do milho (MCMV), e vírus do mosaico da alfafa (AMV) foram demonstrados como sendo efetivos na melhora da expressão (por exemplo, Gallie 1987; Skuzeski 1990). Outros líderes conhecidos na arte incluem mas não são limitados a: líderes de Picornavirus,

por exemplo, líder de EMCV (região de não codificação de 5' de Encefalomiocardite) (Elroy-Stein 1989); líder de Potyvirus, por exemplo, líder de TEV (vírus do ataque do tabaco); líder de MDMV (vírus do mosaico do ananismo de milho); líder de proteína de ligação de cadeia pesada de imunoglobulina humana (BiP), (Macejak 1991); líder não traduzido do mRNA de proteína de capa do vírus mosaico de alfafa (AMV RNA 4), (Jobling 1987; líder de vírus mosaico do tabaco (TMV), (Gallie 1989; e líder de vírus de sarapintado clorótico do milho (MCMV) (Lommel 1991. ver também, Della-Cioppa 1987. Elementos regulatórios como intron Adh 1 (Callis 1987), intron sacarose sintase (Vasil 1989) ou elemento omega TMV (Gallie 1989), podem ainda ser incluídos onde desejado. São especialmente preferidas a região 5' não traduzida, íntrons e a região 3' não traduzida selecionadas dentre

a) o gene descrito pelo *loci* do genoma de *Arabidopsis thaliana* do GenBank At1g48130 e

b) genes ortólogos do gene descrito pelo *loci* do genoma de *Arabidopsis thaliana* do GenBanki At1g48130.

Mais preferidas são as seqüências 5'-não traduzidas compreendidas nas seqüências como descritas por SEQ ID NO: 1, 3, 5, 7, 9, 10, ou 11 (ver listagem de seqüência para especificação da localização).

Os preferidos elementos regulatórios adicionais são seqüências de melhoradores ou seqüências de poliadenilação.

A seqüência nucleotídica heteróloga a ser expressada é preferivelmente ainda operativamente ligada a regiões 3' não traduzidas, terminação de transcrição e/ou sinal de poliadenilação. As regiões 3'-não traduzidas são apropriadas para estabilizar a expressão e estrutura de mRNA. Isto pode resultar em presença prolongada do mRNA e assim níveis de expressão melhorados. Os sinais de terminação e poliadenilação são apropriados para estabilizar a expressão de mRNA, para assegurar o

comprimento do transcrito de mRNA constante e para evitar a transcrição através da leitura. Especialmente em construções de expressão de múltiplos genes, este é um aspecto importante. Além disso, a terminação de transcrição correta é ligada a uma re-iniciação de transcrição da seqüência nucleotídica 5 ′ regulatória, resultando em níveis de expressão melhorados. Os sinais acima mencionados podem ser qualquer sinal funcional em plantas e pode por exemplo ser isolado de genes de planta, genes de vírus de planta ou outros patógenos de planta. No entanto, em uma forma de realização preferida, as regiões 3' não traduzidas, terminação de transcrição e sinais de poliadenilação são dos genes empregados como a fonte para os promotores desta invenção. Ainda mais preferidos são as seguintes seqüências:

SEQ ID NO: 125 apresentando uma seqüência de 876 nucleotídeos seguindo a região de codificação (CDS) do gene peroxiredoxina (PER1) de *Arabidopsis thaliana* (At1g48130) compreendendo 3'-UTR usado como seqüência de terminação transcripcional

SEQ ID NO: 126 apresentando uma seqüência de 400 nucleotídeos seguindo a região de codificação (CDS) do gene peroxiredoxina (PER1) de *Arabidopsis thaliana* (At1g48130) compreendendo 3'-UTR usado como seqüência de terminação transcripcional (SEQ ID NO: 126 é um derivado truncado de SEQ ID NO: 125)

SEQ ID NO: 127 apresentando uma seqüência de 227 nucleotídeos apresentando a região 3'-não traduzida do gene peroxiredoxina (PER1) (At1g48130).

Conseqüentemente, em uma forma de realização preferida, o cassete de expressão da invenção ainda compreende funcionalmente associada à seqüência nucleotídica a ser expressada pelo menos uma seqüência selecionada dentre o grupo consistindo de:

a) as seqüências descritas por SEQ ID NO: 125, 126, e 127, e

b) uma seqüência compreendendo um fragmento de pelo

menos 25 (preferivelmente pelo menos 50 ou 75 ou 100, mais preferivelmente pelo menos 150 ou 200, o mais preferivelmente 250 ou 300) nucleotídeos consecutivos de uma seqüência descrita por SEQ ID NO: 125, 126, ou 127,

c) uma seqüência tendo uma identidade de pelo menos 50% (preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 90% ou 95%, o mais preferivelmente pelo menos 98%) para uma seqüência descrita por SEQ ID NO: 125, 126, ou 127

d) uma seqüência capaz de hibridizar (preferivelmente sob condições equivalentes à hibridização em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 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 (preferivelmente todos) nucleotídeos consecutivos das seqüência descritas por SEQ ID NO: 125, 126, ou 127, ou o seu complemento,

em que referida seqüência é capaz de mediar a estabilização de RNA, estabilização de expressão de RNA, terminação de transcrição de RNA e/ou poliadenilação de RNA.

As seqüências sozinhas são apropriadas e utilizáveis na expressão transgênica em plantas. Conseqüentemente, outra forma de realização da invenção refere-se a um cassete de expressão compreendendo:

i) pelo menos uma seqüência nucleotídica reguladora da transcrição capaz de mediar transcrição em plantas e funcionalmente ligada à

mesma,

ii) pelo menos uma seqüência de ácido nucleico, e funcionalmente ligada à mesma

iii) pelo menos uma seqüência selecionada dentre o grupo consistindo de:

a) as seqüências descritas por SEQ ID NO: 125, 126, e 127, e

b) uma seqüência compreendendo um fragmento de pelo menos 25 (preferivelmente pelo menos 50 ou 75 ou 100, mais preferivelmente pelo menos 150 ou 200, o mais preferivelmente 250 ou 300) nucleotídeos consecutivos de uma seqüência descrita por SEQ ID NO: 125, 126, ou 127,

c) uma seqüência tendo uma identidade de pelo menos 50% (preferivelmente pelo menos 60%, 70% ou 80%, mais preferivelmente pelo menos 90% ou 95%, o mais preferivelmente pelo menos 98%) para uma seqüência descrita por SEQ ID NO: 125, 126, ou 127

d) uma seqüência capaz de hibridizar (preferivelmente sob condições equivalentes à hibridização em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C, mais desejavelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 1 X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,5 X SSC, 0,1% SDS a 50°C, preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 0,1 X SSC, 0,1% SDS a 50°C, mais preferivelmente em dodecil sulfato de sódio a 7% (SDS), 0,5 M $NaPO_4$, 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 (preferivelmente todos) nucleotídeos consecutivos de as seqüência descrita por SEQ ID NO: 125, 126, ou 127, ou o seu complemento,

em que referida seqüência iii) é capaz de mediar a

estabilização de RNA, estabilização de expressão de RNA, terminação de transcrição de RNA e/ou poliadenilação de RNA;

em que ou referida seqüência reguladora de transcrição i) ou referida seqüência iii) ou ambos i e iii) são heterólogas com relação para referida seqüência nucleotídica ii).

Outras seqüências de poliadenilação preferidas são de genes de plantas ou genes T-DNA de Agrobacterium (como por exemplo as seqüência de terminador de genes OCS (octopina sintase) ou NOS (nopalina sintase)).

Exemplos de melhoradores incluem elementos do promotor de CaMV 35S, genes octopina sintase (Ellis et al., 1987), o gene actina I de arroz, o gene álcool desidrogenase de milho (Callis 1987), o gene shrunken I de milho (Vasil 1989), elemento Omega TMV (Gallie 1989) e promotores de eucariotes não de plantas (por exemplo, levedura; Ma 1988). Vetores para uso de acordo com a presente invenção podem ser construídos para incluir o elemento melhorador de ocs. Este elemento foi primeiro identificado como um melhorador palindrômico de 16 bp de gene octopina sintase (ocs) de ultilano (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, irão atuar para aumentar o nível de transcrição de promotores adjacentes quando aplicados no contexto de transformação de plantas.

Um cassete de expressão da invenção (ou um vetor derivado do mesmo) 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 origens de replicação (para permitir a replicação em bactérias; para o ORI de pBR322 ou o ori de P15A; Sambrook 1989), ou elementos requeridos para transferência de T-DNA de Agrobacterium (como

por exemplo as bordas esquerda e/ou direita do T-DNA).

Por fim, os segmentos de DNA os mais desejáveis para introdução em, por exemplo, um genoma de dicotiledôneas, podem ser genes homólogos ou famílias de genes que codificam um traço desejado (por exemplo, aumentado rendimento por acre) e que são introduzidos sob o controle de novos promotores ou melhoradores, etc., ou talvez mesmo homólogos ou promotores específicos para tecido (por exemplo, específicos para raiz, colar/bainha, verticilo, haste, haste da espiga, caroço, ou folha) ou elementos de controle. De fato, visa-se que um uso particular da presente invenção será na expressão de um gene em um modo preferencial de semente ou específico de semente.

Além disso, vetores podem ser construídos e empregados na marcação intracelular e um produto de gene específico dentro das células de uma planta transgênica ou em direção de uma proteína para o meio extracelular. Isto será geralmente obtido por junção de uma seqüência de DNA codificando uma seqüência de peptídeo de transito ou de sinal para a seqüência de codificação de um gene particular. O peptídeo de transito ou sinal resultante irá transportar a proteína a um destino intracelular ou extracelular particular, respectivamente, e será então pós -translacionalmente removido. Os peptídeos de transito ou sinal atuam por facilitação do transporte de proteínas através de membranas intracelulares, por exemplo, vacúolo, vesícula, plastídeo e membranas mitocondriais, enquanto peptídeos de sinal dirigem proteínas através da membrana extracelular.

Um exemplo particular deste uso diz respeito à direção de um gene de resistência a herbicida, como o gene EPSPS, a uma organela particular como o cloroplasto em vez do citoplasma. Isto é exemplificado pelo uso do peptídeo de trânsito rbcs que confere marcação específica de plastídeo de proteínas. Além disso, propõe-se que pode ser desejável marcar certos genes responsáveis por esterilidade masculina para o mitocôndrio, ou marcar

certos genes para resistência a organismos fitopatogênicos nos espaços extracelulares, ou marcar proteínas para o vacúolo.

Ao facilitar o transporte da proteína em compartimentos dentro e fora da célula, estas seqüências podem aumentar o acúmulo de produto de genes protegendo os mesmos da degradação proteolítica. Estas seqüências também permitem que seqüências de mRNA adicionais de genes altamente expressados sejam fixadas à seqüência de codificação dos genes. Uma vez que mRNA sendo traduzido por ribossomas é mais estável do que mRNA nu, a presença de mRNA traduzível em frente do gene pode aumentar a estabilidade global do transcrito de mRNA do gene e, assim, aumentar a síntese do produto de gene. Uma vez que seqüências de trânsito e de sinal são geralmente removidas pós- translacionalmente do produto de tradução inicial, o uso destas seqüências permite a adição de seqüências traduzidas extras que podem não aparecer no polipeptídeo final. A marcação de certas proteínas pode ser desejável a fim de melhorar a estabilidade da proteína (US 5.545.818).

Pode ser utilizável marcar o próprio DNA dentro de uma célula. Por exemplo, pode ser utilizável marcar o DNA introduzido no núcleo à medida que isto pode aumentar a freqüência de transformação. Dentro do próprio núcleo pode ser utilizável marcar um gene a fim de obter integração sítio-específica. Por exemplo, pode ser utilizável ter um gene introduzido através de transformação substituindo um gene existente na célula. Outros elementos incluem os que podem ser regulados por agentes endógenos ou exógenos, por exemplo, por proteínas dedo de zinco, incluindo proteínas dedo de zinco ocorrendo naturalmente ou proteínas 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 como myb. Por exemplo, uma proteína dedo de zinco quimérica pode incluir seqüências de aminoácidos, que se

ligam a uma seqüência de DNA específica (o dedo de zinco) e seqüências de aminoácidos que ativam (por exemplo, seqüências GAL 4) ou reprimem a transcrição das seqüências ligadas à seqüência de DNA específica.

É um dos objetos da presente invenção prover moléculas de DNA recombinantes compreendendo uma seqüência de nucleotídeo de acordo com a invenção operativamente ligada a um segmento de nucleotídeo de interesse.

Um segmento de nucleotídeo de interesse reflete os interesses e mercados comerciais envolvidos no desenvolvimento de culturas. As culturas e mercados comerciais mudam e, à medida que as nações em desenvolvimento abrem seus mercados ao mundo, novas culturas e tecnologias também irão surgir. Além disso, como a compreensão de traços e características agronômicas, como a produção e heterose aumentam, a escolha de genes para transformação irá mudar como uma conseqüência. As categorias gerais de nucleotídeos de interesse incluem, por exemplo, genes envolvidos em informação, como dedos de zinco, os envolvidos em comunicação, como quinases, e os envolvidos domésticos, como proteínas de choque térmico. As categorias de transgenes mais específicas, por exemplo, incluem genes codificando traços importantes para agronomia, resistência a inseto, resistência a doenças, resistência a herbicida, esterilidade, características de grão, e produtos comerciais. Os genes de interesse incluem, geralmente, os envolvidos metabolismo de nutrientes, amido, óleo, ou carboidrato, bem como os que afetam o tamanho de grão de cereal, carga de sacarose, proteínas dedo de zinco, 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, e outros.

Um versado na arte reconhece que o nível de expressão e regulação de um transgene em uma planta pode variar significativamente de linhagem para linhagem. Assim, tem-se que testar várias linhagens para

encontrar uma com o nível de expressão e regulação desejado. Uma vez que uma linhagem é identificada com a especificidade de regulação desejada de um transgene Cre quimérico, ela pode ser cruzada com linhagens possuindo diferentes replicons inativos ou transgene inativo para ativação.

Outras seqüências, que podem ser ligadas ao gene de interesse, que codifica um polipeptídeo, são as que podem marcar uma organela específico, por exemplo, no mitocôndrio, núcleo, ou plastídeo, em uma célula de planta. A marcação pode ser obtida provendo o polipeptídeo com uma seqüência de peptídeo de marcação apropriada, como peptídeo de sinal secretório (para secreção ou parede celular ou marcação de membrana, um peptídeo de trânsito plastídeo, um peptídeo de trânsito cloroplasto, por exemplo, a proteína de ligação a/b de clorofila, um peptídeo de marcação mitocondrial, um peptídeo de marcação de vacúolo, ou um peptídeo de marcação nuclear, e outros. Por exemplo, a subunidade pequena de peptídeo de trânsito de ribulose bisfosfato carboxilase, do peptído de trânsito EPSPS ou do peptídeo de trânsito de ácido dihidropicolínico sintase pode ser usada, para exemplos de seqüências de marcação de organela de plastídeo (ver WO 00/12732). Os plastídeos são uma classe de organelas de planta derivados de pró-plastídeos e incluem cloroplastos, leucoplastos, amiloplastos e cromoplastos. Os plastídeos são sítios principais de biossíntese em plantas. Além da fotossíntese no cloroplasto, os plastídeos também são sítios de biossíntese de lipídeos, redução de nitrato em amônio e armazenamento de amido. E apesar dos plastídeos conterem seu próprio genoma circular, a maior parte das proteínas localizadas nos plastídeos são codificadas pelo genoma nuclear e são importadas dentro da organela a partir do citoplasma.

Os transgenes usados com a presente invenção serão freqüentemente genes que dirigem a expressão de uma proteína particular ou produto de polipeptídeo, mas eles também podem ser segmentos de DNA não expressáveis,, por exemplo, transposons como Ds que não dirigem sua própria

transposição. Como usado aqui, um "gene expressável" é qualquer gene que é capaz de ser transcrito em RNA (por exemplo, mRNA, RNA anti-sentido, etc.) ou traduzido em uma proteína, expressado como um traço de interesse, ou outros, etc., e não está limitado a genes de marcador selecionável, triável ou não selecionável. A invenção também contempla que, quando um gene expressável, que não é necessariamente um gene marcador é empregado em combinação com um gene marcador, pode-se empregar os genes separados tanto em segmentos de DNA iguais como diferentes para transformação. No último caso, os vetores diferentes são liberados simultaneamente nas células recipientes de modo a maximizar a co-transformação.

A escolha dos segmentos de DNA particulares a serem liberados às células recipientes com freqüência dependerá da finalidade da transformação. Uma das principais finalidades de transformação de plantas de cultura é adicionar alguns traços comercialmente desejáveis, agronomicamente importantes, à planta. Estes traços incluem, mas não estão limitados a, resistência ou tolerância a herbicida, resistência ou tolerância a inseto, resistência ou tolerância a doenças (viral, bacteriana, fúngica, por nematódeos); resistência e/ou tolerância a estresse, como exemplificado por resistência ou tolerância aos estresses provocados por seca, calor, resfriamento, congelamento, umidade excessiva, sal; estresse oxidativo; produções aumentadas; composição e teor do alimento; aparência física; esterilidade masculina; secagem; capacidade de ficar em pé; fecundidade, propriedades de amido; quantidade e qualidade de óleo; e outros. Pode-se desejar incorporar um ou mais genes conferindo qualquer um destes traços ou traços desejáveis, como, por exemplo, um gene ou genes codificando resistência a patógenos.

Em algumas formas de realização, a presente invenção contempla a transformação de uma célula recipiente com mais do que um transgene vantajoso. Dois ou mais transgenes podem ser supridos em um

evento de transformação único usando qualquer um dos vetores de codificação de transgene distintos, ou usando um vetor único incorporando duas ou mais seqüências codificando o gene. Por exemplo, os plasmídeos contendo as unidades de expressão bar e aroA em qualquer orientação convergente, divergente ou co-linear, são considerados particularmente utilizáveis. Outras combinações preferidas são os de um gene de resistência a inseto, como um gene Bt, junto com um gene inibidor de protease como pinII, ou o uso de bar em combinação com qualquer um dos genes acima. Como evidente, qualquer um de dois ou mais transgenes de qualquer descrição, como os conferindo resistência a herbicida, inseto, doença (viral, bacteriana, fúngica, por nematódeo) ou a seca, esterilidade masculina, secura, capacidade de ficar em pé, fecundidade, propriedades de amido, qualidade ou quantidade de óleo, ou os aumentando a produção ou a qualidade nutricional podem ser empregados, como desejado.

1. Transgenes Exemplares

1.1 Resistência a Herbicida

Os genes codificando fosfinotricina acetiltransferase (bar e pat), genes EPSP sintase tolerantes a glifosato, gene gox de enzima degradativa de glifosato codificando glifosato oxidorredutase, genes deh (codificando uma enzima desalogenase que inativa dalapon), acetolactato sintase resistente a herbicida (por exemplo, sulfoniluréia e imidazolinona), e bxn (codificando uma enzima nitrilase que degrada bromoxinila) são bons exemplos de genes resistentes a herbicida para uso em transformação. Os genes bar e pat codificam uma enzima, fosfinotricina acetiltransferase (PAT), que inativa o herbicida fosfinotricina e previne que este composto iniba as enzimas glutamina sintetase. A enzima 5-enolpiruvilshikimato 3-fosfato sintase (EPSP sintase) é normalmente inibida pelo herbicida N-(fosfonometil) glicina (glifosato). No entanto, são conhecidos genes que codificam enzimas EPSP sintase resistentes a glifosato. O gene deh codifica a enzima dalapon

desalogenase e confere resistência ao herbicida dalapon. O gene bxn codifica uma enzima nitrilase específica que converte bromoxinila em um produto de degradação não herbicida.

1.2 Resistência a insetos

Um aspecto importante da presente invenção diz respeito à introdução de genes conferindo resistência a inseto em plantas. Os genes de resistência a inseto potenciais, que podem ser introduzidos, incluem, genes da toxina de cristal Bacillus thurigiensis ou genes Bt (Watrud 1985). Os genes Bt podem prover resistência a pragas de lepidópteros ou coleópteros como broca de milho europeu (ECB) e larva nociva da raiz de milho (CRW). Genes de toxina Bt preferidos para uso nestas formas de realização incluem os genes CrylA(b) e CrylA(c). Genes de endotoxinas de outras espécies de B. thurigiensis, que afetam o crescimento ou desenvolvimento de insetos, também podem ser empregados sob este aspecto. Inibidores de protease também podem proporcionar resistência (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 visado como sendo particularmente utilizável. Ainda mais vantajoso é o uso de um gene pinII em combinação com um gene de toxina Bt, cujo efeito combinado foi descoberto pelos presentes inventores como produzindo uma atividade inseticida sinergística. Outros genes, que codificam inibidores do sistema digestivo de insetos, ou os que codificam enzimas ou co-fatores 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 inseticidas adicionais ou alternativas. Lecitinas (originalmente denominadas fito-hemaglutininas) são proteínas de ligação de carboidratos multivalentes, que têm a capacidade de aglutinar glóbulos vermelhos de uma faixa de espécies. As lecitinas foram identificadas recentemente como agentes

inseticidas com atividade contra gorgulhos, ECB e larvas nocivas da raiz (Murdock 1990; Czapla & Lang, 1990). Os genes de lectina contemplados como sendo utilizáveis incluem, por exemplo, aglutinina de cevada e germe de trigo (WGA) e lecitinas de arroz (Gatehouse 1984), com WGA sendo preferido.

Os genes controlando a produção de polipeptídeos pequenos ativos 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 juvenil esterase, dirigido para pragas de insetos específicas, também podem resultar em atividade inseticida, ou talvez causar o término da metamorfose (Hammock 1990).

Plantas transgênicas expressando genes, que codificam enzimas que afetam a integridade da cutícula de insetos, formam ainda um outro aspecto da invenção. Estes genes incluem os codificando, por exemplo, quitinase, proteases, lípases e também genes para a produção de nik-comicina, um composto que inibe a síntese de quitina, sendo que a introdução de qualquer um destes é contemplada como produzindo plantas de milho resistentes a insetos. Os genes que codificam atividades que afetam a muda de insetos, como os que afetam a produção do ecdisteróise UDP-glucosil transferase, também estão dentro do escopo dos transgenes utilizáveis da presente invenção.

Os genes que codificam enzimas que facilitam a produção de compostos que reduzem a qualidade nutricional da planta hospedeira para as pragas de insetos também são englobados pela presente invenção. Pode ser possível, por exemplo, conferir atividade inseticida em uma planta alterando sua composição de esterol. Os esteróis são obtidos pelos insetos a partir de sua dieta e são usados para síntese de hormônios e estabilidade de membranas. Assim, alterações na composição de esterol da planta por

expressão de novos genes, por exemplo, os que promovem diretamente 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 insetos e, assim, dotar a planta com atividade inseticida. As lipoxigenases são enzimas de planta ocorrendo naturalmente que demonstram efeitos antinutricionais para os insetos e reduzem a qualidade nutricional de sua dieta. Assim, outras formas de realização da invenção referem-se a plantas transgênicas com atividade de lipoxigenase melhorada que podem ser resistentes à alimentação de insetos.

A presente invenção também provê métodos e composições pelos quais podem ser alcançadas mudanças qualitativas ou quantitativas em metabólitos secundários de plantas. Um exemplo refere-se à transformação de plantas para produzir DIMBOA que, como contemplado, irão conferir resistência à broca do milho europeu, larvas nocivas da raiz, e várias outras pragas de insetos de milho. Genes candidatos que são particularmente considerados para uso a este respeito incluem os genes no *locus* bx que se sabe estar envolvido na via sintética de DIMBOA (Dunn 1981). A introdução de genes que podem regular a produção de maisina, e genes envolvidos na produção de dhurrina em sorgo, também é contemplada como utilizável para facilitar a resistência a lagarta da espiga e larva nociva da raiz, respectivamente.

Tripsacum dactyloides é uma espécie de grama que é resistente a certos insetos, incluindo larva nociva da raiz do milho. Antecipa-se que os 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 de 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 usados como transgenes de acordo com o presente. Estes genes incluem, por exemplo, o inibidor de tripsina de feijão-de-vaca (CpTI; Hilder 1987), que pode ser usado como um deterrente de larva nociva da raiz; genes codificando avermectina (Campbell 1989; Ikeda 1987) que podem demonstrar ser particularmente utilizáveis como um deterrente de larva nociva da raiz de milho; genes de proteínas inativantes de ribossoma; e ainda genes que regulam estruturas de plantas. Milho transgênico incluindo genes de anticorpo antiinseto e genes que codificam para enzimas que podem abrigar um inseticida não tóxico (pró-inseticida) aplicado no exterior da planta dentro de um inseticida no interior da planta também são contemplados.

1.3 Resistência ao estresse ou ambiental

A melhora da capacidade de uma planta de tolerar vários estresses ambientais como, mas não limitadas a, seca, excesso de umidade, resfriamento, congelamento, temperatura elevada, sal e estresse 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 temperaturas de congelamento através da introdução de uma proteína "anticongelamento" como a de Winter Flounder (Cutler 1989) ou derivados de genes sintéticos da mesma. Tolerância melhorada a resfriamento também pode ser conferida através da expressão aumentada de glicerol-3-fosfato acetiltransferase em cloroplastos (Murata 1992; Wolter 1992). A resistência ao estresse oxidativo (freqüentemente exacerbada por condições como temperaturas de resfriamento em combinação com intensidades de muita luz) pode ser conferida por expressão de superóxido dismutase (Gupta 1993), e pode ser melhorada por glutationa redutase (Bowler 1992). Estas estratégias podem permitir tolerância a congelamento em campos de brotação recente, bem como estendendo as

zonas de maturidade posterior das variedades de maior produção para as de maturidade relativa anterior.

A expressão de novos genes que afetam favoravelmente o teor de água nas plantas, potencial de água total, potencial osmótico e turgidez pode melhorar a capacidade da planta de tolerar a seca. Como usado aqui, os termos "resistência à seca" e "tolerância a seca" são usados para fazer referência a uma resistência ou tolerância aumentada das plantas ao estresse induzido por uma redução na disponibilidade de água, como comparado em circunstâncias normais, e à capacidade da planta funcionar e sobreviver em ambientes com baixo teor de água, e apresentar um desempenho em um modo relativamente superior. Neste aspecto da invenção, é proposto, por exemplo, que a expressão de um gene codificando a biossíntese de solutos osmoticamente ativos pode conferir proteção contra seca. Dentro desta classe de genes estão os DNAs codificando desidrogenase manitol (Lee e Saier, 1982) e trehalose-6-fosfato sintase (Kaasen 1992). Através da ação subseqüente de fosfatases nativas na célula ou pela introdução e co-expressão de uma fosfatase específica, os genes introduzidos resultarão no acúmulo tanto de manitol como de trehalose, respectivamente, ambos os quais foram bem documentados como compostos protetores capazes de mitigar os efeitos do estresse. Verificou-se o acúmulo de manitol em tabaco transgênico e os resultados preliminares indicam que plantas expressando altos níveis deste metabólito são capazes de tolerar um estresse osmótico aplicado (Tarczynski 1992).

Similarmente, a eficácia de outros metabólitos na proteção tanto da função enzimática (por exemplo, alanopina ou ácido propiônico) como da integridade da membrana (por exemplo, alanopina) foi documentada (Loomis 1989) e, assim, a expressão do gene codificando a biossíntese destes compostos pode conferir resistência à seca de um modo similar a ou complementar à do manitol. Outros exemplos de metabólitos ocorrendo

naturalmente que são osmoticamente ativos e/ou provêem algum efeito protetor direto durante seca e/ou dessecação incluem açúcares e derivados de açúcar como frutose, eritritol (Coxson 1992), sorbitol, dulcitol (Karsten 1992), glicosilglicerol (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 glicina-betaína (Wyn-Jones e Storey, 1981). O crescimento continuado de canópia e a capacidade reprodutiva aumentada durante os períodos de estresse podem ser aumentados por introdução e expressão de genes como os controlando os compostos osmoticamente ativos discutidos acima e outros destes compostos, como representado em uma forma de realização exemplar pela enzima mioinositol 0-metiltransferase.

É contemplado que a expressão de proteínas específicas também pode aumentar a tolerância à seca. Três classes de proteínas embriogênicas tardias foram designadas com base em similaridades estruturais (ver Dure 1989). Todas as três classes destas proteínas foram demonstradas em sementes em maturação (isto é, dessecando). Nestes 3 tipos de proteínas, o tipo II (tipo dehidrina) estava geralmente implicado em tolerância à seca e/ou dessecação em partes de plantas vegetativas (por exemplo, Mundy e Chua, 1988; Piatkowski, 1990; Yamaguchi-Shinozaki, 1992). Recentemente, verificou-se que a expressão de um LEA tipo III (HVA-I) em tabaco influencia a altura da planta, maturidade e tolerância à seca (Fitspatrick, 1993). A expressão de genes estruturais de todos os três grupos pode, assim, conferir tolerância à seca. Outros tipos de proteínas induzidas durante o estresse provocado por água incluem tiol proteases, aldosales e transportadores de transmembrana (Guerrero 1990), que podem conferir várias funções protetores e/ou do tipo de reparo durante o estresse da seca. A expressão de um gene que efetua a biossíntese de lipídeos e, assim, a

composição da membrana também pode ser utilizável para conferir resistência à seca na planta.

Muitos genes que melhoram a resistência à seca têm modos complementares de ação. Assim, combinações destes genes podem ter efeitos aditivos e/ou sinergísticos na melhora de resistência à seca em milho. Muitos destes genes também melhoram a tolerância ao congelamento (ou resistência); os estresses físicos ocasionados durante congelamento e seca são similares por natureza e podem ser mitigados de modo similar. O benefício pode ser conferido via expressão constitutiva ou específica de tecido destes genes, mas os meios preferidos de expressar estes novos genes podem ser através do uso de um promotor induzido por turgor (como os promotores para os genes induzidos por turgor descritos em Guerrero et al. 1990 e Shagan 1993). Padrões de expressão espaciais e temporais destes genes podem possibilitar que o milho suporte melhor o estresse.

A expressão de genes que estão envolvidos com traços morfológicos específicos que permitem extrações de água aumentadas do solo secando pode ocasionar benefício. Por exemplo, a introdução e expressão de genes que alteram as características da raiz podem melhorar a absorção de água. A expressão de genes que melhoram a capacidade reprodutiva durante os tempos de estresse pode ser de valor significativo. Por exemplo, a expressão de DNAs que melhoram a sincronia da dispersão de pólen e receptividade das partes de flores femininas, isto é, estigmas, pode ser de benefício. Além disso, a expressão de genes que minimizam o aborto do caroço dos grãos de cereais durante os tempos de estresse pode aumentar a quantidade de grãos a ser coletados e, assim, representar valor. A regulação dos níveis de citoquinina em monocotiledôneas, como milho, por introdução e expressão de um gene isopentenil transferase com seqüências regulatórias apropriadas, pode melhorar o rendimento e resistência ao estresse em monocotiledôneas (Gan 1995).

Dado o papel global da água na determinação da produção, é contemplado que possibilitando às plantas utilizar água de modo mais eficiente, através da introdução e expressão de novos genes, será melhorado o desempenho global mesmo quando a disponibilidade de água no solo não é limitante. Introduzindo genes que melhoram a capacidade das plantas em maximizar o uso de água através de uma faixa total de estresses referentes à disponibilidade de água, a estabilidade do rendimento da produção ou consistência de seu desempenho pode ser alcançada.

Proteção melhorada de planta a fatores de estresses abióticos como seca, calor ou frio, também pode ser obtida – por exemplo, - super expressando polipeptídeos anticongelamento de Myoxocefalus Scorpius (WO 00/00512), Myoxocefalus octodecemspinosus, o ativador de transcrição de Arabidopsis thaliana CBF1, glutamato desidrogenases (WO 97/12983, WO 98/11240), genes de proteína quinase dependentes de cálcio (WO 98/26045), calcineurinas (WO 99/05902), caseína quinase de leveduras (WO 02/052012), farnesiltransferases (WO 99/06580; Pei ZM et al. (1998) Science 282: 287-290), ferritina (Deak M et al. (1999) Nature Biotechnology 17: 192-196), oxalato oxidase (WO 99/04013; Dunwell JM (1998) Biotechn Genet Eng Rev 15: 1-32), fator DREB1A ("elemento B 1A de resposta a desidratação" ; Kasuga M et al. (1999) Nature Biotech 17: 276-286), genes de síntese de manitol ou trehalose como trehalose-fosfato sintase ou trehalose-fosfato fosfatase (WO 97/42326) ou inibindo genes como trehalase (WO 97/50561).

1.4 Resistência às doenças

É proposto que resistência a doenças pode ser realizada através da introdução de genes no ciclo das 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.

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

É proposto que resistência aumentada às doenças causadas por bactérias e fungos pode ser alcançada através da introdução de novos genes. É contemplado que genes codificando os denominados "antibióticos peptídicos" , proteínas relacionadas a patogênese (PR), resistência a toxina, e proteínas afetando interações hospedeiro- patógeno como características morfológicas serão utilizáveis. Antibióticos peptídicos são seqüências de polipeptídeos, que são inibidoras para 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 ingestão do patógeno em uma planta hospedeira e devem ser divididos em pelo menos cinco classes de proteínas (Bol 1990). Incluídas entre as proteínas PR estão beta-1,3-glucanases, quitinases e osmotina e outras proteínas que, acredita-se, funcionam em resistência de plantas a organismos doentes. Outros genes foram identificados que têm propriedades antifúngicas, por exemplo, UDA

(lectina de urtiga picante) e heveína (Broakgert 1989; Barkai-Golan 1978). Sabe-se que certas doenças de plantas são causadas pela produção de fitotoxinas. 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. Os novos genes de expressão que alteram interações entre a planta hospedeira e patógeno podem ser utilizáveis para reduzir a capacidade do organismo doente de invadir os tecidos da planta hospedeira, por exemplo, um aumento no caráter ceroso da cutícula da folha ou outras características morfológicas.

Nematódeos parasíticos de plantas são uma das causas de doenças em muitas plantas. É proposto que pode ser possível tornar uma planta resistente a estes organismos através da expressão de novos genes. Antecipa-se que o controle de infestações de nematódeos pode ser realizado alterando-se a capacidade do nematódeo de reconhecer ou atacar uma planta hospedeira ou possibilitar à planta produzir compostos nematicidas, incluindo mas não limitados a proteínas.

Além disso, uma resistência a fungos, insetos, nematódeos e doenças pode ser obtida pelo acúmulo marcado de alguns 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 da planta e reação ao estresse como são induzidas quando as plantas são feridas ou atacadas por micróbios, ou quimicamente, por exemplo, por ácido salicílico, ácido jasmônico ou etileno, ou lisozinas de fontes não plantas como, por exemplo, T4-lisozima ou lisozima de uma variedade de mamíferos, proteínas inseticidas como endotoxina de *Bacillus thurigiensis*, inibidor de a-amilase ou inibidores de protease (inibidor de tripsina de feijão-de-vaca), lecitinas como aglutinina de germe de trigo, RNAses ou ribozimas. Outros exemplos são ácidos nucleicos

que codificam a proteína chit42 endoquitinase de Trichoderma harzianum (GenBank Número de acesso S78423) ou citocromo multi-funcional, não hidroxilante P-450 (CYP79) de sorgo bicolor (GenBank Número de Acesso U32624), ou equivalentes funcionais destas. O acúmulo de glucosinolatos como proteção de pragas (Rask L et al. (2000) Plant Mol Biol 42: 93-113; Menard R et al. (1999) Fitochemistry 52: 29-35), a expressão de endotoxinas de *Bacillus thurigiensis* (Vaeck et al. (1987) Nature 328: 33-37) ou a proteção contra ataque por fungos, por expressão de quitinases, por exemplo, de feijões (Broglie et al. (1991) Science 254: 1194-1197), é vantajosa. Resistência a pragas como, por exemplo, a praga de arroz Nilaparvata lugens em plantas de arroz pode ser obtida por expressão de aglutinina lectina de anêmola (*Galanthus nivalis*) (Rao et al. (1988) Plant J 15 (4): 469-77). A expressão de genes crylA(b) e crylAc sintéticos, que codificam endotoxinas D de *Bacillus thurigiensis* específico de lepidópteros pode ocasionar uma resistência a pragas de insetos em várias plantas (Goyal RK et al. (2000) Crop Protection 19 (5): 307-312). Outros genes de marcação que são apropriados para a defesa contra patógenos compreendem "proteína de inibição de poligalacturonase" (PGIP), taumatina, invertase e peptídeos antimicrobianos como lactoferrina (Lee TJ et al. (2002) J Amer Soc Horticult Sci 127 (2): 158-164). Outras seqüências de ácido nucleico que podem ser vantajosamente usadas aqui incluem traços de controle de insetos (patentes US 6.063.597; 6.063.756; 6.093.695; 5.942.664; e 6.110.464), resistência a doença fúngica (patentes US 5.516.671; 5.773.696; 6.121.436; 6.316.407; e 6.506.962), resistência a vírus (patentes US 5.304.730 e 6.013.864), resistência a nematódeos (patente US 6.228.992), e resistência a doença bacteriana (patente US 5.516.671).

1.5 Redução/ Eliminação de Micotoxina

A produção de micotoxinas, incluindo aflatoxina e fumonisina, por fungos associados com plantas é um fator significativo para tornar o grão não utilizável. Estes organismos fúngicos não causam sintomas de doença

e/ou interferem com o crescimento da planta, mas produzem produtos químicos (micotoxinas) que são tóxicos para os animais. A inibição do crescimento destes fungos pode reduzir a síntese destas substâncias tóxicas e, assim, reduzir perdas de grãos devido à contaminação por micotoxinas. Os novos genes podem ser introduzidos em plantas que podem inibir a síntese da micotoxina sem interferir com o crescimento de fungos. A expressão de um novo gene, que codifica uma enzima capaz de tornar a micotoxina não tóxica, pode ser utilizável a fim de obter contaminação de grão por micotoxina reduzida. O resultado de qualquer um dos mecanismos acima pode ser uma presença reduzida de micotoxina no grão.

1.6 Composição ou Qualidade do Grão

Os genes podem ser introduzidos em plantas, 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.

Por exemplo, o uso principal de grão de milho é voltado para alimentos ou ração. A introdução de genes que alterem a composição do grão pode melhorar muito o valor do alimento ou da ração. Os componentes primários de grão de milho são amido, proteína e óleo. Cada um destes componentes primários de grão de milho pode ser melhorado alterando seu nível de composição. Vários exemplos podem ser mencionados para fins ilustrativos mas de modo algum prover uma lista exaustiva de possibilidades.

A proteína de muitos grãos de cereais é subótima para fins de ração ou de alimento, especialmente quando consumida por porcos, aves domésticas e humanos. A proteína é deficiente em vários aminoácidos que são essenciais na dieta destas espécies, requerendo a adição de suplementos ao grão. Aminoácidos essenciais limitantes podem incluir lisina, metionina, triptofano, treonina, valina, arginina e histidina. Alguns aminoácidos tornam-

se limitantes somente após o grão ser suplementado com outros insumos para formulações de ração. Por exemplo, quando o grão é suplementado com farinha de soja para atender aos requisitos de lisina, metionina torna-se limitante. Os níveis destes aminoácidos essenciais em sementes e grãos podem ser elevados por mecanismos que incluem, mas não estão limitados à 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 às sementes ou grãos.

Um mecanismo para aumentar a biossíntese dos aminoácidos é introduzir genes que desregulam as vias biossintéticas de aminoácidos de modo que a planta não mais pode controlar adequadamente os níveis que são produzidos. Isto pode ser feito desregulando ou contornando etapas na via biossintética de aminoácidos que são normalmente regulados por níveis do produto final de aminoácido da via. Exemplos incluem a introdução de genes que codificam versões desreguladas das enzimas aspartoquinase ou ácido dihidrodipicolínico (DHDP)- sintase para aumentar a produção de lisina e treonina, e antranilato sintase para aumentar a produção de triptofano. A redução do catabolismo dos aminoácidos pode ser obtida por introdução de seqüências de DNA que reduzem ou eliminam a expressão de genes codificando enzimas que catalisam etapas nas vias catabólicas como a enzima lisina-cetoglutarato redutase.

A composição da proteína do grão pode ser alterada para melhorar o equilíbrio de aminoácidos em uma variedade de modos, incluindo elevar a expressão de proteínas nativas, diminuir a expressão dos com composição fraca, mudar a composição de proteínas nativas, ou introduzir genes codificando proteínas totalmente novas possuindo composição superior. O DNA pode ser introduzido, que diminui a expressão de membros da família de zeína de proteínas em armazenamento. Este DNA pode codificar ribozimas ou seqüências anti-sentido dirigidas para conferir expressão de proteínas zeína

ou expressão de reguladores da expressão de zeína como o produto do gene opaco-2. A composição de proteína do grão pode ser modificada através do fenômeno de co-supressão, isto é, inibição da expressão de um gene endógeno através da expressão de um gene estrutural idêntico ou fragmento de gene introduzido através de transformação (Goring 1991). Além disso, o DNA introduzido pode codificar enzimas, que degradam zeínas. Os decréscimos na expressão de zeína que são obtidos podem ser acompanhados por aumentos em proteínas com composição de aminoácidos mais desejáveis ou aumentos em outros constituintes principais da semente como amido. Além disso, um gene quimérico pode ser introduzido que compreende uma seqüência de codificação para uma proteína nativa de composição de aminoácidos adequada, tal como para uma das proteínas globulina ou zeína de 10 kD e um promotor ou outra seqüência regulatória designada para elevar a expressão de referida proteína. A seqüência de codificação de referido gene pode incluir códons adicionais ou de substituição para aminoácidos essenciais. Além disso, uma seqüência de codificação obtida de outras espécies, ou, uma seqüência parcialmente ou completamente sintética codificando uma seqüência de peptídeo completamente única designada para melhorar a composição de aminoácidos da semente pode ser empregada.

A introdução de genes que alteram o teor de óleo do grão pode ser de grande valor. Aumentos no teor de óleo podem resultar em aumentos no teor de energia metabolizável e densidade das sementes para usos em ração e alimentos. Os genes introduzidos podem codificar enzimas que removem ou reduzem limitações de taxa ou etapas reguladas na biossíntese de ácido graxo ou lipídeo. Estes genes podem incluir, mas não estão limitados aos que codificam acetil-CoA carboxilase, ACP-aciltransferase, beta-cetoacil-ACP sintase, mais outras atividades biossintéticas de ácido graxo bem conhecidas. Outras possibilidades são genes que codificam proteínas que não possuem atividade enzimática como proteína de transporte de acila. Outros exemplos

incluem 2-acetiltransferase, complexo piruvato de oleosina desidrogenase, acetil CoA sintetase, citrato de ATP liase, ADP- glicose pirofosforilase e genes de transporte carnitina-CoA-acetil-CoA. Antecipa-se que a expressão de genes relacionada à biossíntese de óleo será marcada no plastídeo, usando uma seqüência de peptídeo de trânsito plastídeo e, preferivelmente, expressada no embrião da semente. Os genes podem ser introduzidos os quais alteram o equilíbrio de ácidos graxos presentes no óleo proporcionando uma alimentação mais saudável ou nutritiva. O DNA introduzido também pode codificar seqüências que bloqueiam a expressão de enzimas envolvidas na biossíntese de ácidos graxos, alterando as proporções de ácidos graxos presentes no grão como descrito abaixo.

Genes podem ser introduzidos os quais melhoram o valor nutritivo do componente amido do grão, por exemplo, aumentando o grau de ramificação, resultando na utilização melhorada do amido em vacas ao retardar seu metabolismo.

Além de afetar os constituintes principais do grão, genes podem ser introduzidos, os quais afetam uma variedade de outros aspectos nutritivos, de processamento, ou outra qualidade do grão como usado para ração ou alimento. Por exemplo, a pigmentação do grão pode ser aumentada ou diminuída. A melhora e estabilidade da pigmentação amarela é desejável em algumas rações animais e pode ser obtida por introdução de genes que resultam em produção melhorada de xantofilas e carotenos eliminando as etapas de limitação de taxas em sua produção. Estes genes podem codificar formas alteradas das enzimas fitoeno sintase, fitoeno desaturase ou licopeno sintase. Além disso, milho branco não pigmentado é desejável para produção de muitos produtos alimentícios e pode ser produzido pela introdução de DNA, que bloqueia ou elimina etapas nas vias de produção de pigmento.

Ração ou alimentos compreendendo alguns grãos de cereais possui quantidades insuficientes de vitaminas e deve ser suplementada para

proporcionar um valor nutritivo adequado. A introdução de genes que melhoram a biossíntese de vitaminas em sementes pode ser considerada incluindo, por exemplo, vitaminas A, E, B_{12}, colina, e outras. Por exemplo, grãos de milho também não possuem um teor de mineral suficiente para um valor nutritivo ótimo. Os genes que afetam o acúmulo ou disponibilidade de compostos contendo fósforo, enxofre, cálcio, manganês, zinco e ferro dentre outros podem ser valiosos. Um exemplo pode ser a introdução de um gene que reduz a produção de ácido ftálico ou codifica a enzima fitase, de modo a melhora a quebra de ácido fítico. Estes genes podem aumentar os níveis de fosfato disponíveis na dieta, reduzindo a necessidade de suplementação com fosfato mineral.

Numerosos outros exemplos de melhora de cereais para fins de ração e alimentos podem ser descritos. As melhoras podem mesmo não envolver necessariamente o grão, mas podem, por exemplo, melhorar o valor do grão para silagem. A introdução de DNA para realizar isto pode incluir seqüências que alteram a produção de lignina como as que resultam no fenótipo de "nervura central marrom de folhas" associado com o valor superior de ração para gado.

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

A melhora do valor de produtos de moagem a úmido pode incluir alterar a quantidade ou qualidade de 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 de relações na biossíntese de amido ou diminuindo os níveis dos outros componentes do grão resultando em aumentos proporcionais de amido. Um exemplo do primeiro pode ser a introdução de genes codificando enzimas ADP-glicose pirofosforilase com atividade regulatória alterada ou que são expressados em nível maior. Exemplos das últimas podem incluir inibidores seletivos de, por exemplo, biossíntese de proteína ou de óleo expressados durante os últimos estágios do desenvolvimento de grão de cereal.

As propriedades do amido podem ser beneficamente alteradas por mudança da relação de amilose para amilopectina, o tamanho das moléculas de amido, ou seu padrão de ramificação. Através destas mudanças, uma ampla faixa de propriedades pode ser modificada as quais incluem, mas não estão limitadas a mudanças em temperatura de gelatinização, calor de gelatinização, transparência das películas e pastas, propriedades teológicas, e outras. Para realizar estas mudanças em propriedades, os genes que codificam atividade de amido sintase solúvel ou ligada a grânulos ou atividade da enzima de ramificação podem ser introduzidos sozinhos ou em combinação. DNA como construções anti-sentido também pode ser usado para diminuir os níveis de atividade endógena destas enzimas. Os genes ou construções introduzidos podem possuir seqüências regulatórias que sincronizam sua expressão em intervalos específicos em biossíntese de amido e desenvolvimento de grânulos de amido. Além disso, pode ser aconselhável introduzir e expressar genes que resultam na derivatização *in vivo*, ou outra modificação, das porções de glicose da molécula de amido. A fixação covalente de qualquer molécula pode ser visada, limitada somente pela existência de enzimas que catalisem as derivatizações e a acessibilidade de substratos apropriados no grânulo de amido. Exemplos de derivações importantes podem incluir a adição de grupos funcionais como grupos aminas, carboxilas ou fosfato, que provêem sítios de derivatizações *in vitro* subseqüentes ou afetam as propriedades do amido através da introdução de

cargas iônicas. Exemplos de outras modificações podem incluir mudanças diretas das unidades de glicose como perda de grupos hidroxila ou sua oxidação em grupos aldeído ou carboxila.

Óleo é um outro produto de moagem a úmido de milho ou outros grãos, cujo valor pode ser melhorado pela introdução e expressão de genes. A quantidade de óleo que pode ser extraída por moagem a úmido pode ser elevada por abordagens, como descrito para ração e alimentos acima. As propriedades do óleo também podem ser alteradas para melhorar seu desempenho na produção e uso de óleo de cozinha, gorduras para bolos, lubrificantes ou outros produtos derivados de óleo ou melhora de seus atributos para a saúde quando usados nas aplicações relacionadas a alimentos. Novos ácidos graxos também podem ser sintetizados que, quando da extração, podem servir como materiais de partida para síntese de produtos químicos. As mudanças em propriedades do óleo podem ser obtidas alterando o tipo, nível, ou arranjo de lipídeos dos ácidos graxos presentes no óleo. Isto, por sua vez, pode ser realizado pela adição de genes que codificam enzimas que catalisam a síntese de novos ácidos graxos e os lipídeos possuindo os mesmos ou aumentando os níveis de ácidos graxos nativos enquanto possivelmente reduzindo os níveis dos precursores. Alternativamente, seqüências de DNA podem ser introduzidas as quais tornam mais lentas ou bloqueiam as etapas na biossíntese de ácido graxo resultando no aumento em intermediários precursores de ácido graxo. Genes que podem ser adicionados incluem dessaturases, epoxidases, hidratases, desidratases, e outras enzimas que catalisam reações envolvendo intermediários de ácido graxo. Exemplos representativos de etapas catalíticas que podem ser bloqueadas incluem as dessaturações de ácido esteárico em oleico e ácido oleico em linolênico resultando nos respectivos acúmulos de ácidos esteárico e oleico.

Melhoras nos outros produtos principais de moagem a úmido de cereais, farinha de glúten e ração de glúten também podem ser obtidas pela

introdução de genes para obter novas plantas. Possibilidades representativas incluem, mas não estão limitadas às descritas acima para melhora do valor da ração ou alimentação.

Além disso, pode ser ainda considerado que a planta seja usada para a produção ou fabricação de compostos biológicos utilizáveis apropriados que ou não foram produzidos de todo, ou não foram produzidos no mesmo nível, na planta anteriormente. As novas plantas produzindo estes compostos são tornadas possíveis pela introdução e expressão de genes por métodos de transformação. As possibilidades incluem, mas não estão limitadas a qualquer composto biológico que é atualmente produzido por qualquer organismo como proteínas, ácidos nucleicos, metabólitos primários e intermediários, polímeros carboidratos, etc. Os compostos podem ser produzidos pela planta, extraído quando da colheita e/ou processamento, e usados para qualquer fim utilizável atualmente reconhecido como fármacos, fragrâncias, enzimas industriais, para citar apenas alguns.

Outras possibilidades para exemplificar a faixa de traços ou propriedades do grão potencialmente codificadas por genes introduzidos em plantas transgênicas incluem grão com menor susceptibilidade à quebra para fins de exportação ou tamanho de grita maior quando processados por moagem a seco através de introdução de genes que melhoram a síntese de gama-zeína, milho de pipoca com melhorado estouro, qualidade e volume de expansão através dos genes que aumentam a espessura do pericarpo, milho com grão mais branco para uso em alimentos através de introdução de genes que efetivamente bloqueiam a expressão de enzimas envolvidas nas vias de produção de pigmentos, e melhorada qualidade de bebidas alcoólicas ou milho dose através de introdução de genes que afetam o aroma como o gene shrunken (codificando sacarose sintase) para milho doce.

1.7 Composição ou qualidade do tubérculo ou semente

Vários traços podem ser expressados com vantagem,

especialmente em sementes ou tubérculos para melhorar a composição ou qualidade. As seqüências de ácido nucleico úteis que podem ser combinadas com a seqüência de ácido nucleico promotora da presente invenção e prover traços no produto final melhorados, incluem, sem limitação, as codificando proteínas de armazenamento em semente, enzimas de vias de ácido graxo, enzimas biossintéticas de tocoferol, enzimas biossintéticas de aminoácidos, e enzimas de ramificação de amido. Uma discussão de DNAs heterólogos exemplares úteis para a modificação de fenótipos de planta pode ser encontrada em, por exemplo, patentes US 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 PCT Publicação WO 02/057471, cada sendo especificamente incorporada 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 rações, por exemplo de fitases e celulases. Especialmente preferidos são ácidos nucleicos como o cDNA artificial, que codifica uma fitase microbiana (GenBank Número de Acesso: A19451) ou seus equivalentes funcionais.

- Expressão de genes, que ocasionam um acúmulo de produtos químicos finos, como de tocoferóis, tocotrienóis ou carotenóides. Um exemplo, que pode ser mencionado, é fitoeno dessaturase. Preferidos são ácidos nucleicos, que codificam a fotoeno dessaturase de Narcissus pseudonarcissus (GenBank Número de Acesso: X78815) ou seus equivalentes funcionais. 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., Seed Sci. Res., 1:209:219 (1991); Keegstra, Cell, 56(2):247-53 (1989); Nawrath et al., Proc. Natl. Acad. Sci. USA, 91:12760-12764 (1994); Xia et al., J. Gen. Microbiol., 138:1309-1316 (1992); Lois et al., Proc. Natl.

Acad. Sci. USA, 95 (5):2105-2110 (1998); Takahashi et al., Proc. Natl. Acad. Sci. USA, 95(17):9879-9884 (1998); Norris et al., Plant Physiol., 117:1317-1323 (1998); Bartley e Scolnik, Plant Physiol., 104:1469-1470 (1994); Smith et al., Plant J., 11:83-92 (1997); WO 00/32757; WO 00/10380; Saint Guily et al., Plant Physiol., 100(2):1069-1071 (1992); Sato et al., J. DNA Res., 7(1):31-63 (2000), todos incorporados aqui por referência.

- produção de amido (patente US Nos. 5.750.876 e 6.476.295), produção elevada de proteínas (patente US No. 6.380.466), amadurecimento de frutas (patente US No. 5.512.466), nutrição animal e humana melhorada (patente US Nos. 5.985.605 e 6.171.640), biopolímeros (patente US No. 5.958.745 e Publicação de Patente U.S. No. 2003/0028917), resistência ao estresse ambiental (patente US No. 6.072.103), peptídeos farmacêuticos (patente US No. 6.080.560), traços de processamento melhorados (patente US No. 6.476.295), digestibilidade melhorada (patente US No. 6.531.648), baixo teor de rafinose (patente US No. 6.166.292), produção industrial de enzimas (patente US No. 5.543.576), aroma melhorado (patente US No. 6.011.199), fixação de nitrogênio (patente US No. 5.229.114), produção de sementes híbridas (patente US No. 5.689.041), e produção de biocombustível (patente US No. 5.998.700), os elementos genéticos e transgenes descritos nas patentes listadas acima são aqui incorporados por referência. As enzimas de ramificação de amido preferidas (para modificação de propriedades de amido) incluem as especificadas em patentes US Nos. 6.232.122 e 6.147.279; e PCT Publicação WO 97/22703, todos incorporados aqui por referência.

- produção de óleos modificados (patente US No. 6.444.876), produção de alto teor de óleo (patentes US Nos. 5.608.149 e 6.476.295), ou teor de ácido graxo modificado (patente US No. 6.537.750). Preferidas enzimas de via de ácido graxo incluem tioesterases (patentes us 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 (patentes US 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) todos incorporados aqui por referência.

- Enzimas biossintéticas preferidas de aminoácidos incluem antranilato sintase (patente US No. 5,965,727 e publicações PCT WO 97/26366, WO 99/11800, WO 99/49058), triptofano decarboxilase (PCT Publicação WO 99/06581), treonina decarboxilase (patentes US Nos. 5,534,421 e 5,942,660; Publicação PCT WO 95/19442), treonina deaminase (Publicações PCT WO 99/ 02656 e WO 98/55601), ácido dihidrodipicolínico sintase (patente US No. 5,258,300), e aspartato quinase (patentes US Nos. 5.367.110; 5.858.749; e 6.040.160) todos incorporados aqui por referência.

- Produção de nutracêuticos, como, por exemplo, ácidos graxos poliinsaturados, (por exemplo, ácido araquidônico, ácido eicosapentaenóico, ou ácido docosahexaenóico) por expressão de ácido graxo elongases e/ou desaturases, ou produção de proteínas com valor nutricional melhorado como, por exemplo, com um teor elevado de aminoácidos essenciais (por exemplo, o gene de albumina de elevado teor de metionina 2S da castanha-do-pará). Preferidos são ácidos nucleicos que codificam a albumina de elevado teor de metionina 2S de Bertholletia excelsa (GenBank Número de Acesso: AB044391), Δ6-acil-lipídeo desaturase de Physcomitrella patens (GenBank Número de Acesso: AJ222980; Girke et al. (1998) Plant J 15:39-48), Δ6-desaturase de Mortierella alpina (Sakuradani et al. 1999 Gene 238:445-453), Δ5-desaturase de Caenorhabditis elegans (Michaelson et al. 1998, FEBS Letters 439:215-218), Δ5-ácido graxo desaturase de Caenorhabditis elegans (des-5) (GenBank Número de Acesso: AF078796),

Δ5-desaturase de Mortierella alpina (Michaelson et al. JBC 273:19055-19059), Δ6-elongase de Caenorhabditis elegans (Beaudoin et al. 2000, PNAS 97:6421-6426), Δ6-elongase de Physcomitrella patens (Zank et al. 2000, Biochemical Society Transactions 28:654-657), ou seus equivalentes funcionais.

- Produção de proteínas e enzimas de elevada qualidade para fins industriais (por exemplo, enzimas, como lipases) ou como fármacos (como, por exemplo, anticorpos, fatores de coagulação de sangue, interferons, linfocinas, fator de estimulação de colônias, ativadores de plasminogene, hormônios ou vacinas, como descrito por Hood EE, Jilka JM (1999) Curr Opin Biotechnol 10(4): 382-6; Ma JK, Vine ND (1999) Curr Top Microbiol Immunol 236:275-92). Por exemplo, foi possível produzir avidina recombinante a partir de albume de galinha e β-glucuronidase bacteriano (GUS) em uma grande escala em plantas de milho transgênico (Hood et al. (1999) Adv Exp Med Biol 464:127-47. Review).

- A obtenção de uma aumentada capacidade de armazenamento em células que normalmente compreendem menos proteínas de armazenamento ou lipídeos de armazenamento, com o fim de aumentar a produção 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 Número de Acesso: L25042), ou seus equivalentes funcionais. Alternativamente, em alguns cenários, uma teor de proteína de armazenamento elevado pode ser vantajoso para a produção de produto com elevado teor de Protewin. Preferidas proteínas de armazenamento de semente incluem zeínas (patentes US 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; Publicações PCT 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 (patentes US Nos. 5.003.045 e 5.576.203),

proteína de castanha-do-pará (patente US No. 5.850.024), proteínas isentas de fenilalanina (PCT Publicação WO 96/17064), albumina (Publicação PCT WO 97/35023), b-conglicinina (Publicação PCT WO 00/19839), 11S (patente US No. 6.107.051), alfa-hordotionina (patente US Nos. 5.885.802 e 5.88.5801), proteínas de armazenamento de semente de arcelina (patente US No. 5.270.200), lectinas (patente US No. 6.110.891), e glutenina (patente US Nos. 5.990.389 e 5.914.450) todos incorporados aqui por referência.

- Redução dos níveis de atividade de enzimas fosforilase de tubérculo tipo α- glucano L (GLTP) ou fosforilase de tubérculo tipo α-glucano H (GHTP) preferivelmente dentro do tubérculo de batata (ver US 5.998.701). A conversão de amidos em açúcar em tubérculos de batata, particularmente quando armazenados em temperaturas abaixo 7°C, é reduzida em tubérculos demonstrando atividade de enzima GLTP ou GHTP reduzida. A redução do adoçamento no frio de batatas permite o armazenamento das batatas em temperaturas mais frias, resultando em dormência prolongada, reduzida incidência de doenças e aumentada vida de armazenamento. A redução da atividade de GLTP ou GHTP dentro do tubérculo de batata pode ser obtida por tais técnicas como a supressão de expressão de gene usando RNA homólogo anti-sentido ou filamento duplo, o uso de seqüências de co-supressão, silenciamento regulatório. Uma planta de batata tendo características de armazenamento no frio melhoradas, compreendendo uma planta de batata transformada com um cassete de expressão tendo uma seqüência de promotor TPT operativamente ligada a uma seqüência de DNA compreendendo pelo menos 20 nucleotídeos de um gene codificando uma α glucano fosforilase selecionada dentre o grupo consistindo de fosforilase de tubérculo de tipo α glucano L- (GLTP) e fosforilase de tipo α glucano H (GHTP).

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.

1.8 Características Agronômicas de Plantas

Dois dos fatores determinando onde as plantas podem ser cultivadas são a temperatura média diária durante a estação de crescimento e a extensão de tempo entre as geadas. Dentro das áreas, onde é possível cultivar uma planta particular existem limitações variadas sobre o tempo máximo que se permite cultivar até a maturidade e ser colhida. A planta a ser cultivada em uma área selecionada por sua capacidade de amadurecer e secar para ter um teor de umidade que pode ser colhida dentro do período de tempo requerido com uma produção máxima possível. Assim, plantas de variadas maturidades foram desenvolvidas para diferentes locais de cultivo. Além da necessidade de uma secagem suficiente para permitir a colheita, temos a desejabilidade de ter a ocorrência de uma secagem máxima no campo de modo a minimizar a quantidade de energia requerida para a secagem adicional após a colheita. Também, quanto mais prontamente o grão pode secar, mais tempo se tem disponível para o crescimento e o enchimento do caroço. Os genes que influenciam a maturidade e/ou secagem podem ser identificados e introduzidos nas linhagens de plantas usando técnicas de transformação para criar novas variedades adaptadas para diferentes locais de cultivo ou o mesmo local de cultivo, mas tendo uma relação de produção para umidade melhorada na colheita. A expressão de genes que estão envolvidos na regulação do desenvolvimento da planta pode ser especialmente utilizável, por exemplo, os genes de bainha áspera e sem lígula, que foram identificados em plantas.

Os genes podem ser introduzidos em planta que irão melhorar a capacidade de ficar em pé e outras características do crescimento das plantas. Por exemplo, a expressão de novos genes, que conferem caules mais fortes, melhorados sistemas de raiz, ou evitam ou reduzem a queda de espigas podem ser de grande valor para o fazendeiro de cultivo de milho. A introdução e expressão de genes que aumentam a quantidade total de foto

assimilado disponível, por exemplo, por aumento da distribuição de luz e/ou intersecção podem ser vantajosas. Além disso, a expressão de genes que aumentam a eficácia de fotossíntese e/ou a canópia das folhas pode ainda aumentar os ganhos em produtividade. Estas abordagens podem permitir aumentadas populações de plantas no campo.

O retardo da senescência vegetativa de estação tardia pode aumentar o fluxo de assimilados no grão e, assim, aumentar a produção. A super-expressão de genes dentro de plantas que estão associados com "permanecer verde" ou a expressão de qualquer gene que retarda a senescência pode ser vantajoso. Por exemplo, um mutante de não amarelecimento foi identificado em *Festuca pratensis* (Davies 1990). Expressão deste gene assim como de outros pode evitar a ruptura prematura de clorofila e, assim manter a função da canópia.

1.9 Utilização de Nutrientes

A capacidade de utilizar os nutrientes e minerais disponíveis pode ser um fator limitante no crescimento de muitas plantas. Foi proposto que seria possível alterar a absorção de nutrientes, tolerar extremos de pH, mobilização através da planta, reuniões de armazenamento, e disponibilidade para atividades metabólicas pela introdução de novos genes. Estas modificações podem permitir a uma planta utilizar de modo mais eficiente os nutrientes disponíveis. Foi contemplado que um aumento na atividade de, por exemplo, uma enzima que está normalmente presente na planta e envolvida na utilização de nutriente iria aumentar a disponibilidade de um nutriente. Um exemplo desta enzima seria uma fitase. Também é contemplado que a expressão de um novo gene pode tornar uma fonte de nutriente disponível a qual não era previamente acessível, por exemplo, uma enzima que libera um componente de valor nutriente de uma molécula mais complexa, talvez uma macromolécula.

1.10 Esterilidade masculina

A esterilidade masculina é utilizável na produção de semente híbrida. Foi proposto que a esterilidade masculina pode ser produzida através da expressão de novos genes. por exemplo, foi mostrado que a expressão de genes que codificam proteínas que interferem com o desenvolvimento da inflorescência e/ou gametófito masculino resultam em esterilidade masculina. Os genes de ribonuclease quiméricos que expressam nos anteros de tabaco transgênico e colza de semente oleígena foram demonstrados como levando a esterilidade masculina (Mariani 1990). Por exemplo, um número de mutações foram descobertas em milho que confere esterilidade masculina citoplásmica. Uma mutação em particular, referida como citoplasma T, também correlaciona com sensibilidade para a mangra da folha do milho Southern. Uma seqüência de DNA, designada TURF-13 (Levings 1990), foi identificada que correlaciona com citoplasma T. Isto seria possível através da introdução de TURF-13 via transformação para separar a esterilidade masculina de sensibilidade à doença. Como é necessário ser possível restaurar a esterilidade masculina para fins de cruzamento e para a produção de grãos, foi proposto que os genes codificando restauração de esterilidade masculina também podem ser introduzidos.

1.11. Seqüências Não Expressando Proteínas

1.11.1 Expressando RNA

DNA pode ser introduzido em plantas para o fim de expressar transcritos de RNA que funcionam para afetar o fenótipo de planta que ainda não foram traduzidos em proteína. Dois exemplos são RNA anti-sentido e RNA com atividade ribozima. Ambos podem servir a possíveis funções na redução ou eliminação da expressão de genes de planta nativos ou introduzidos.

Genes podem ser construídos ou isolados, que quando transcritos, produzem RNA anti-sentido ou RNA filamento duplo que é complementar a todo ou parte(s) de RNA(s) mensageiro marcado. O RNA

anti-sentido reduz a produção do produto polipeptídeo do RNA mensageiro. O produto polipeptídeo pode ser qualquer proteína codificada pelo genoma da planta. Os genes acima mencionados serão referidos como genes anti-sentido. Um genes anti-sentido pode assim ser introduzido em uma planta por métodos de transformação para produzir uma nova planta transgênica com expressão reduzida de uma proteína de interesse selecionada. Por exemplo, a proteína pode ser uma enzima que catalisa uma reação na planta. Redução da atividade de enzima pode reduzir ou eliminar os produtos da reação que incluem qualquer um composto enzimaticamente sintetizado na planta, como ácidos graxos, aminoácidos, carboidratos, ácidos nucleicos e semelhantes. Alternativamente, a proteína pode ser uma proteína de armazenamento, como uma zeína, ou uma proteína estrutural, cuja expressão diminuída pode levar a mudanças na composição do aminoácido da semente ou mudanças morfológicas da planta, respectivamente. As possibilidades citadas acima são providas somente a título de exemplo e não representam a faixa completa de aplicações.

Expressão de RNA anti-sentido ou RNA filamento duplo por um dos cassetes de expressão da invenção é especialmente preferida. Também expressão de RNA sentido pode ser empregada para silenciamento de genes (co-supressão). Este RNA é preferivelmente um RNA não traduzível. A regulação do gene por RNA filamento duplo ("RNA interferência filamento duplo"; dsRNAi) é bem conhecido na arte e descrito para vários organismos incluindo plantas (por exemplo, Matzke 2000; Fire A et al 1998; WO 99/32619; WO 99/53050; WO 00/68374; WO 00/44914; WO 00/44895; WO 00/49035; WO 00/63364).

Genes também podem ser construídos ou isolados, que quando transcritos produzem enzimas de RNA, ou ribozimas, que podem atuar como endoribonucleases e catalisam a clivagem de moléculas de RNA com seqüências selecionadas. A clivagem de RNA's mensageiros selecionados

pode resultar na produção reduzida de seus produtos polipeptídeos codificados. Estes genes podem ser usados para preparar novas plantas transgênicas, que possuem os mesmos.

As plantas transgênicas podem possuir níveis reduzidos de polipeptídeos incluindo mas não limitados aos polipeptídeos citados acima que podem ser afetados por RNA anti-sentido.

É também possível que genes podem ser introduzidos para produzir novas plantas transgênicas, que tem expressão reduzida de um produto de gene nativo, por um mecanismo de cossupressão. Foi demonstrado em tabaco, tomate, e petúnia (Goring 1991; Smith 1990; Napoli 1990; van der Krol 1990) que a expressão do transcrito sentido de um gene nativo irá reduzir ou eliminar a expressão do gene nativo em um modo similar ao observado para genes anti-sentido. O gene introduzido pode codificar toda ou parte da proteína nativa introduzida mas sua tradução pode não ser requerida para redução de níveis desta proteína nativa.

1.11.2 Não Expressando RNA

Por exemplo, elementos DNA incluindo os de elementos transposáveis como Ds, Ac, ou Mu, podem ser inseridos em um gene e causar mutações. Estes elementos de DNA podem ser inseridos a fim de inativar (ou ativar) em um gene e assim "etiquetar" um traço particular. Neste caso, o elemento transposável não causa instabilidade da mutação etiquetada, porque a utilidade do elemento não depende de sua capacidade de se movimentar no genoma. Uma vez etiquetado o traço desejado, a seqüência de DNA introduzida pode ser usada para clonar o correspondente gene, por exemplo, usando a seqüência de DNA introduzido como um iniciador PCR junto com técnicas PCR de clonagem de genes (Shapiro, 1983; Dellaporta 1988). Uma vez identificado(s), o (s) gene(s) completo (s) para o traço particular, incluindo regiões de controle ou regulatórias onde desejado pode (m) ser isolado(s), clonado(s) e manipulado(s) como desejado. A utilidade dos

elementos de DNA introduzidos em um organismo para o fim de etiquetar genes é independente da seqüência de DNA e não depende de qualquer atividade biológica da seqüência de DNA, isto é, transcrição em RNA ou tradução em proteína. A única função do elemento de DNA é romper a seqüência de DNA de um gene.

É contemplado que seqüências de DNA não expressadas, incluindo seqüências sintética novas, podem ser introduzidas em células como "rótulos" de propriedade destas células e plantas e suas sementes. Não seria necessário rotular para um elemento de DNA para interromper a função de um gene endógeno ao organismo hospedeiro, como a única função deste DNA seria identificar a origem do organismo. Por exemplo, seria possível introduzir uma seqüência de DNA singular em uma planta e este elemento de DNA iria identificar todas as células, plantas, e progênie destas células como tendo surgido da fonte rotulada. É proposto que a inclusão de DNAs de rótulos iria permitir distinguir um germplasma de propriedade ou germplasma derivado de tal, de um germplasma não rotulado..

Outro elemento possível, que pode ser introduzido, é um elemento de região de fixação de matriz (MAR), como elemento de lisozima A de galinha (Stief 1989), que pode ser posicionado em torno de um gene de interesse expressável para efetuar um aumento em expressão global do gene e diminuir os efeitos dependentes de posição quando da incorporação no genoma da planta (Stief 1989; Phi-Van 1990).

Outras seqüências nucleotídicas de interesse que podem ser contempladas para uso dentro do escopo da presente invenção em ligação operativa com as seqüências de promotor de acordo como a invenção são moléculas ácido nucleico isoladas, por exemplo, DNA ou RNA, compreendendo a seqüência nucleotídica da planta de acordo como a invenção, compreendendo uma matriz de leitura aberta que é preferivelmente expressada em um tecido específico, isto é, tecido de semente, raiz, tecido

verde (folha e caule), panícula, ou pólen, ou é expressada constitutivamente.

2. Genes Marcadores

A fim de melhorar a capacidade de identificar transformantes, pode-se desejar empregar um gene marcador selecionável ou triável como, ou em adição a, o gene de interesse expressável. "Genes Marcadores" são genes que conferem um fenótipo distinto a células expressando o gene marcador e assim permite que estas células transformadas sejam distintas das células que não tem o marcador. Estes genes podem codificar ou um marcador selecionável ou triável, dependendo de se o marcador confere um traço que se "selecionar" por meios químicos, isto é, através do uso de um agente seletivo (por exemplo, um herbicida, antibióticos, ou semelhantes), ou se for simplesmente um traço que se pode identificar através de observação ou teste, isto é, por `triagem ` (por exemplo, traço do *locus* R, a proteína fluorescente verde (GFP)). Como evidente, muitos exemplos de genes marcadores apropriados são conhecidos na arte e podem ser empregados na prática da invenção.

Estão incluídos dentro dos termos de genes marcadores selecionáveis ou triáveis, também os genes que codificam um "marcador secretável" cuja secreção pode ser detectada como um meio de identificação ou seleção para células transformadas. Os exemplos incluem marcadores, que codificam um antígeno secretável que pode ser identificado por interação de anticorpo, ou mesmo enzimas secretáveis, que podem ser detectadas por sua atividade catalítica. As proteínas secretáveis caem em várias classes, incluindo proteínas pequenas, difusíveis, 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 seqüência líder como que encontrada na unidade de expressão de extensina ou PR-S de tabaco).

Com relação a marcadores secretáveis selecionáveis, o uso de um gene que codifica a proteína que se torna seqüestrada na parede celular, e cuja proteína inclui um epítopo único, é considerado como sendo particularmente vantajoso. Este marcador de antígeno secretado poderia idealmente empregar uma seqüência de epítopo que iria prover um fundo baixo no tecido da planta, uma seqüência líder-promotor, que iria conferir uma expressão eficiente e marcação através da membrana de plasma, e iria produzir proteína que é ligada na parede celular e ainda acessível a anticorpos. Uma proteína de parede normalmente secretada modificada para incluir um epítopo único iria atender a estas tais exigências.

Um exemplo de uma proteína apropriada para modificação de 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 uma variedade de proteínas de parede ricas em ultilano e/ou glicina (Keller 1989) pode ser modificada pela adição de sítio antigênico para criar um marcador triável.

Uma forma de realização exemplar de um marcador triável secretável refere-se ao uso de uma seqüência de milho codificando a proteína HPRG de parede, modificada para incluir um epítopo 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 dentre a variedade extremamente ampla de combinações de antígeno -anticorpo conhecidas dos versados na arte. O epítopo extracelular único pode então ser detectado diretamente usando rotulação de anticorpos em conjunto com adjuvantes cromogênicos ou fluorescentes.

Elementos da presente descrição podem ser exemplificados em detalhes através do uso dos genes bar e/ou GUS, e também através do uso de vários outros marcadores. Como evidente, à luz da descrição, numerosos

outros genes marcador selecionáveis e/ou triáveis possíveis serão evidentes para os versados na arte além dos especificados aqui abaixo. Assim, será entendido que a seguinte discussão é exemplar em vez de exaustiva. Á luz das técnicas aqui descritas e as técnicas recombinantes gerais que são bem conhecidas na arte, a presente invenção torna possível a introdução de qualquer gene, incluindo genes marcadores, em uma célula recipiente de modo a gerar uma planta transformada.

2.1 Marcadores Selecionáveis

Vários marcadores selecionáveis são conhecidos na arte apropriados para transformação de planta. Estes marcadores podem incluir mas não são limitados a:

2.1.1 Marcadores de seleção negativa

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

- Fosfinotricina acetiltransferases (PAT; também chamado resistência a Bialofos ®; bar; de Block 1987; Vasil 1992, 1993; Weeks 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 a glutamina sintetase,

(Murakami 1986; Twell 1989) causando um rápido acúmulo de amônia e morte celular.

- 5-enolpiruvilshikimato-3-fosfato sintase alterada (EPSPS) conferindo resistência a Glifosate® (N-(fosfonometil)glicina) (Hinchee 1988; Shah 1986; Della-Cioppa 1987). Onde se emprega um gene EPSP sintase mutante, benefício adicional pode ser realizado através da incorporação de um peptídeo de transito de cloroplasto apropriado, CTP (EP-A1 0 218 571).

- enzimas degradando Glifosate® (Glifosate® oxidoreductase; gox),

- dehalogenases inativando 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 Bromoxinil® (bxn; Stalker 1988)

- genes de resistência a canamicina ou geneticina (G418) (NPTII; NPT ou neo; Potrikus 1985) codificando por exemplo, para neomicina fosfotransferases (Fralei 1983; Nehra 1994)

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

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

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

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

Os genes marcadores selecionáveis negativos adicionais que conferem resistência a antibióticos incluem o gene aadA, que confere resistência ao antibiótico espectinomicina, gentamicina acetil transferase,

estreptomicina fosfotransferase (SPT), aminoglicosídeo-3-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; Erikson 2004). Especialmente preferidos marcadores de seleção negativos neste contexto são o gene *dao*l (EC: 1.4. 3.3: GenBank Número de acesso: U60066) da levedura *Rhodotorula gracilis* (*Rhodosporidium toruloides*) e o gene *dsd*A de *E. coli* (D-serina dehidratase (D-serina deaminase) [EC: 4.3. 1.18; GenBank número de acesso: J01603).

O material de planta transformada (por exemplo, células, embriões, tecidos ou plântulas) que expressam tais genes marcadores são capazes de desenvolver na presença de concentrações de um composto de seleção correspondente (por exemplo, antibiótico ou herbicida) que suprime crescimento de um tecido tipo selvagem não transformado. As plantas resultantes pode ser cruzadas e hibridizadas de modo comum. Duas ou mais gerações devem ser cultivadas a fim de assegurar que a integração genômica seja estável e hereditária. Os métodos correspondentes 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 de fase de indução de calo 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. É além disso possível (apesar de não explicitamente preferido) fazer a seleção também através de um esquema completo de regeneração incluindo a formação da raiz.

Por exemplo, com o gene de resistência a fosfinotricina (bar) como o marcador seletivo, fosfinotricina 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 incluída no meio. As concentrações típicas para seleção são de 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 de 20 a 40 mg/l.

2.1.2 Marcador de seleção positivo

Além disso, marcador de seleção positivo pode ser empregado. Genes como isopenteniltransferase de *Agrobacterium tumefaciens* (cepa:PO22; Genbank Número de acesso: AB025109) podem –como uma enzima chave da biossíntese de citoquinina – facilitar a regeneração de plantas transformadas (por exemplo, por seleção em meio isento de citoquinina). Os métodos correspondentes de seleção são descritos (Ebinuma 2000a,b). Os marcadores de seleção positivos adicionais, que conferem uma vantagem de crescimento a uma planta transformada em comparação com uma não transformada, são descritos por exemplo, em EP-A 0 601 092. Os marcadores de seleção de estimulação do crescimento podem incluir (mas não devem ser limitados a) β -glucuronidase (em combinação com por exemplo, um glucuronídeo de citoquinina), mannose-6-fosfato isomerase (em combinação com mannose), UDP-galactose-4-epimerase (em combinação com por exemplo, galactose), em que mannose-6-fosfato isomerase em combinação com mannose é especialmente preferida.

2.1.3 Marcador de contra-seleção

Os marcadores de contra-seleção são especialmente apropriados para selecionar organismos como seqüências deletadas definidas compreendendo o referido marcador (Koprek 1999). Exemplos para marcador contra- seleção compreendem timidina quinases (TK), citosina deaminases (Gleave 1999; Perera 1993; Stougaard 1993), proteínas citocromo P450 (Koprek 1999), haloalcano desalogenases (Naested 1999), produtos de genes iaaH (Sundaresan 1995), citosina deaminase codA (Schlaman & Hooykaas 1997), produtos de gene tms2 (Fedoroff & Smith 1993), ou α-naftaleno acetamida (NAM; Depicker 1988). Os marcadores de contra seleção podem ser úteis na construção de linhagens de etiqueta transposon. Por exemplo, ao marcar um elemento transposável autônomo como Ac, Master Mu, ou En/Spn com um marcador de contra seleção, pode-se selecionar transformantes em que o elemento autônomo não está estavelmente integrado no genoma. Isto seria desejável, por exemplo, quando a expressão transiente do elemento autônomo é desejada para ativar em trans a transposição de um elemento transposável defeituoso, como Ds, mas a integração estável do elemento autônomo não é desejada. A presença do elemento autônomo pode não ser desejada a fim de estabilizar o elemento defectivo, isto é, evitar que ele ainda passe por transposição. No entanto, é proposto que se a integração estável de um elemento transposável autônomo for desejada em uma planta, a presença de um marcador selecionável negativo pode tornar possível eliminar o elemento autônomo durante o processo de cruzamento.

2.2. Marcadores triáveis

Os marcadores triáveis que podem ser empregados incluem, mas não são limitados a, gene beta-glucuronidase (GUS) ou uidA que codifica uma enzima para os vários substratos cromogênicos são conhecidos; um gene R-*locus*, que codifica um produto que regula a produção de pigmentos de antocianina (cor vermelha) em tecidos de plantas (Dellaporta 1988); gene beta-lactamase (Sutcliffe 1978), que codifica uma enzima para a qual os

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 (Katz 1983) que codifica uma enzima capaz de oxidar tirosina em DOPA e dopaquinona que por sua vez condensa para formar o composto melamina facilmente identificável, gene β - galactosidase, que codifica uma enzima para a qual existem substratos cromogênicos; um gene luciferase (lux) (Ow 1986), que permite para detecção de bioluminescência; ou mesmo um gene aequorina (Prasher 1985), que pode ser empregado em detecção de bioluminescência sensível a cálcio, ou um gene de proteína fluorescente verde (Niedz 1995).

Genes do complexo de genes R do milho são contemplados por serem particularmente úteis como marcadores triável. O complexo de genes R em milho codifica uma proteína que atua para regular a produção de pigmentos antocianina na maior parte da semente e tecido de planta. Um gene do complexo de genes R foi aplicado em 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 de um pigmento vermelho e, se estavelmente incorporado, pode ser visualmente classificado como um setor vermelho. Se uma linhagem de milho for dominante para genes codificando os intermediários enzimáticos na via biossintética de antocianina (C2, A1, A2, Bz1 e Bz2), mas transporta um alelo recesso no *locus* R, transformação de qualquer célula da qual a linhagem com R irá resultar em uma 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.

Foi ainda proposto que as regiões de gene regulatório R podem ser empregadas em construções quimérica a fim de prover mecanismos para

controlar a expressão de gene quiméricos. Uma maior diversidade de expressão fenotípica é conhecida no *locus* R do que em qualquer um dos outros *locus* (Coe 1988). É contemplado que regiões regulatórias obtidas de regiões 5' para o gene R estrutural seriam valiosas na direção da expressão de genes, por exemplo, resistência a insetos, resistência a seca, tolerância a herbicida ou outras regiões codificando proteínas. Para os fins da presente invenção, acredita-se que qualquer um dos vários membros da família do gene R pode ser empregado 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 genes R e é funcionalmente similar aos *loci* R e B *em* que Sn controla a deposição específica de tecido dos pigmentos de antocianina em algumas células de mudinhas e 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 pelo 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, câmeras de vídeo de luz fraca, câmaras de contagem de fótons ou luminometria de múltiplas cavidades. É também visado que este sistema pode ser desenvolvido para uma triagem populacional para bioluminescência, como em placas de cultura de tecidos, ou mesmo para uma triagem da planta completa. Onde o uso de um gene marcador triável como lux ou GFP é desejado, benefício pode ser realizado por criação de uma fusão de gene entre o gene marcador triável e um gene marcador selecionável, por exemplo, um gene de fusão GFP-NPTII. Isto pode permitir, por exemplo, seleção de células transformadas seguido por triagem de plantas transgênicas ou sementes.

3. Moléculas de DNA exemplares

A invenção provê uma molécula de ácido nucleico isolada, por exemplo, DNA ou RNA, compreendendo uma seqüência nucleotídica de

plantas, compreendendo uma matriz de leitura aberta que é preferivelmente expressada em um tecido de planta específico, isto é, em sementes, raízes, tecido verde (folha e caule), panículas ou pólen, ou é expressada constitutivamente, ou um seu promotor.

Estes promotores incluem, mas não são limitados a, constitutivos, indutíveis, regulados temporalmente, regulados em desenvolvimento, regulados espacialmente, regulados quimicamente, respondentes a estresse, específicos de tecido, promotores virais e sintéticos. Seqüências de promotor são conhecidas como sendo fortes ou fracas. Um promotor forte provê um nível elevado de expressão de gene, enquanto um promotor fraco provê um nível muito baixo de expressão de gene. Um promotor indutível é um promotor que provê ligar ou desligar a expressão de gene em resposta a um agente adicionado exogenamente, ou a um estímulo ambiental ou de desenvolvimento. Um promotor bacteriano, como o promotor P_{tac}, pode ser induzido em variados níveis de expressão de gene dependendo do nível de isotiopropilgalactosídeo adicionado às células bacterianas transformadas. Uma seqüência isolada de promotor que é um promotor forte para ácido nucleico heterólogo é vantajosa porque ela provê um nível suficiente de expressão de gene para permitir uma detecção e seleção fáceis de células transformadas e provê um nível elevado de expressão de genes, quando desejado.

Dentro de uma região de promotor de planta, existem vários domínios que são necessários para a função completa do promotor. O primeiro de destes domínios esta imediatamente a montante do gene estrutural e forma a "região de promotor de núcleo "contendo seqüências de consenso, normalmente 70 pares de base imediatamente a montante do gene. A região de promotor de núcleo contém as caixas características CAAT e TATA mais as seqüências circundantes, e representa a seqüência de iniciação de transcrição que define o início do ponto de transcrição para o gene estrutural.

A presença da região de promotor de núcleo define uma seqüência como sendo um promotor: se a região estiver ausente, o promotor é não funcional. Além disso, a região de promotor de núcleo é insuficiente para proporcionar uma atividade de promotor completa. Uma série de seqüências regulatórias a montante do núcleo constitui o restante do promotor. As seqüências regulatórias determinam o nível de expressão, padrão espacial e temporal de expressão e, para um sub-conjunto importante de promotores, a expressão sob condições indutivas (regulação por fatores externos, como luz, temperatura, produtos químicos, hormônios).

A expressão regulada da proteína de replicação viral de transatuação quimérica pode ser ainda regulada por outras estratégicas genéticas. Por exemplo, ativação de genes mediada por Cre, como descrito por Odell et al. 1990. Assim, um fragmento de DNA contendo seqüência regulatória 3' ligada por sítios lox entre o promotor e a proteína de replicação codificando a seqüência que bloqueia a expressão de um gene de replicação quimérica do promotor pode ser removido por excisão mediada por Cre e resultar na expressão do gene de replicação trans-atuando. Neste caso, o gene Cre quimérico, o gene de replicação trans-atuando quimérico, ou ambos, podem estar sob o controle de promotores indutíveis ou específicos de tecido e desenvolvimento. Uma estratégia genética alternada é o uso de gene supressor de tRNA. Por exemplo, a expressão regulada de um gene supressor de tRNA pode controlar de modo condicional a expressão de uma seqüência codificando proteína de replicação trans-atuando contendo um códon de terminação apropriado, com descrito por Ulmasov et al. 1997. Novamente, nem o gene supressor de tRNA quimérico, nem o gene de replicação trans-atuando quimérico, ou ambos, pode estar sob o controle de promotores indutíveis ou específicos de tecido e desenvolvimento.

Freqüentemente é desejável ter uma expressão contínua ou indutível de uma seqüência de DNA em todas as células de um orgânica em

um modo independente de tecido.. Por exemplo, resistência aumentada de uma infecção de planta t6 por patógenos de origem no solo e ar, pode ser obtida por manipulação genética do genoma da planta de modo a compreender um promotor contínuo ligado operativamente a um gene de resistência a patógeno heterólogo, com as proteínas de resistência a patógeno são continuamente expressadas em todos os tecidos da planta.

Alternativamente, pode ser desejável inibir a expressão de uma seqüência de DNA nativa dentro das sementes de uma planta de modo a obter um fenótipo desejado. Neste caso, esta inibição pode ser obtida com transformação da planta para compreender um promotor ligado operativamente a uma seqüência nucleotídica anti-sentido, de modo que a expressão preferencial de semente ou específica de semente da seqüência anti-sentido produz um transcrito de RNA que interfere com a tradução de mRNA da seqüência de DNA nativo.

Para definir uma região de promotor mínima, um segmento de DNA representando a região de promotor é removido da região 5' do gene de interesse e operativamente ligado à seqüência codificando um gene marcador (repórter) por técnicas de DNA recombinantes bem conhecidas na arte. O gene repórter é operativamente ligado a jusante do promotor, de modo que os transcritos se iniciando no promotor prosseguem através do gene repórter. Os genes repórter geralmente codificam proteínas, que são facilmente medidas, incluindo, mas não limitadas a, cloranfenicol acetil transferase (CAT), beta-glucuronidase (GUS), proteína fluorescente verde (GFP), beta-galactosidase (beta-GAL), e luciferase.

A construção contendo o gene repórter sob o controle do promotor é então introduzida em um tipo de célula apropriado por técnicas de transfecção bem conhecidas na arte. Para estar a proteína repórter, lisados de célula são preparados e testes apropriados, que são bem conhecidos na arte, para a proteína repórter são realizados. Por exemplo, se CAT for o gene

repórter de escolha, os lisados de células transfectadas com construções contendo CAT sob o controle de um promotor sob estudo são misturados com cloranfenicol e acetil-coenzima A (acetil-CoA) isotopicamente rotulados. A enzima CAT transfere o grupo acetila de acetil-CoA para a posição 2- ou 3 de cloranfenicol. A reação é monitorada por cromatografia de camada fina, que separa o cloranfenicol acetilado de material não reagido. Os produtos de reação são então visualizados por auto-radiografia.

O nível de atividade de enzima corresponde à quantidade de enzima que foi feita, que por sua vez revela o nível de expressão do promotor de interesse. Este nível de expressão pode ser comparado a outros promotores para determinar a resistência relativa do promotor sob estudo. A fim de se certificar que o nível de expressão é determinado pelo promotor, em vez de pela estabilidade do mRNA, o nível do mRNA repórter pode ser medido diretamente, como por análise Northern blot.

Uma vez detectada a atividade, análises mutacionais e/ou delecionais podem ser empregadas para determinar a região mínima e/ou seqüências requeridas para iniciar a transcrição. Assim, as seqüências podem ser deletados na extremidade 5' da região do promotor e/ou na extremidade 3' da região de promotor, e as substituições de nucleotídeos introduzidas. Estas construções são então introduzidas nas células e sua atividade determinada.

Em uma forma de realização, o promotor pode ser um promotor gama zeína, um promotor ole16 de oleosina, um promotor de globulinas, um promotor I de actina, um promotor cl de actina, um promotor de sacarose sintetase, um promotor de INOPS, um promotor de EXM5, um promotor de globulina2, um b-32, um promotor ADPG-pirofosforilase, um promotor LtpI, um promotor 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 pólen, um promotor de proteína específico de antero, um promotor de RTS2

de gene específico de antero, um promotor de gene específico de pólen, um promotor de gene específico de *tapeturn*, um promotor RAB24 de gene específico de *tapeturn*, um promotor de subunidade alfa de antranilato sintase, um promotor de alfa zeína, um promotor de subunidade beta de antranilato sintase, um promotor de dihidrodipicolinato sintase, um promotor Thil, um promotor de álcool desidrogenase, um promotor de proteína de ligação cab, um promotor de H3C4, um promotor de enzima de ramificação de amido RUBISCO SS, um promotor de ACCase, um promotor de actina3, um promotor de actina7, um promotor GF14-12 de proteína regulatória, um promotor L9 de proteína ribossômica, um promotor de enzima biossintética celulose, um promotor de S-adenosil-L-homocisteína hidrolase, um promotor de superóxido dismutase, um promotor de receptor de C-quinase, um promotor de fosfoglicerato mutase, um promotor de mRNA RCc3 específico de raiz, um promotor de glicose-6 fosfato isomerase, um promotor de pirofosfato-fructose 6-fosfatolfosfotransferase, um promotor de ubiquitina, um promotor de beta-cetoacil-ACP sintase, um promotor 11de foto sistema de 33 kDa, um promotor de proteína envolvendo oxigênio, um promotor de subunidade ATPase vacuolar de 69 kDa, um promotor de proteína semelhante a metalotioneína, um promotor de gliceraldeído-3-fosfato desidrogenase, um promotor de proteína semelhante a ABA e indutível do amadurecimento, um promotor de fenilalanina amônia liase, um promotor de adenosina trifosfatase S-adenosil-L-homocisteína hidrolase, um de promotor a-tubulina, um promotor cab, um promotor PEPCase, um promotor de gene R, um promotor de lectina, um promotor de complexo de colheita de luz, 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 proteína de ligação a auxina, um promotor de gene UDP glicose flavonóide glicosil -transferase, um promotor NTI, um promotor de actina, um promotor opaco 2, um promotor b70, um promotor de oleosina, um promotor CaMV

35S, um promotor CaMV 34S, um promotor CaMV 19S, um promotor de histona, um promotor indutível do turgor, um promotor de carboxilase RuBP de subunidade pequena de ervilha, um promotor de Ti plasmídeo manopina sintase, um promotor de Ti plasmídeo nopalina sintase, um promotor de calcona isomerase de petúnia, um promotor de proteína I rica em glicina de feijão, um promotor de transcrito de CaMV 35S, um promotor de patatina de batata, ou um promotor de carboxilase RuBP de subunidade pequena de S-E9.

4. Plantas transformadas (transgênicas) da invenção e métodos de preparação

As espécies de plantas podem ser transformadas com a construção de DNA da presente invenção pela transformação mediada por DNA de protoplastos de células de planta e subseqüente regeneração da planta a partir dos protoplastos transformados, de acordo com procedimentos bem conhecidos na arte.

Qualquer tecido de planta capaz de propagação clonal subseqüente, seja por organogenese ou embriogênese, pode ser transformado com um vetor da presente invenção. O termo "organogenese," como usado aqui, significa um processo pelo qual os brotos e raízes são desenvolvidos seqüencialmente de centros meristemáticos; o termo "embriogênese," como usado aqui, significa um processo pelo qual os brotos e raízes se desenvolvem juntos em um modo combinado (não seqüencialmente), seja a partir de células somáticas ou gametas. O tecido escolhido particular irá variar dependendo dos sistemas de propagação clonal disponíveis, e melhor apropriados para uma espécie em particular sendo transformada. As marcações exemplares de tecido incluem discos de folha, pólen, embriões, cotilédones, hipocótilos, megagametófitos, tecido de calos, tecido meristemático existente (por exemplo, eritemas apicais, brotos axilares, e meristemas de raiz), e tecido de meristema induzido, (por exemplo, meristema de cotilédone e meristema de ultilano).

Plantas da presente invenção podem adquirir um variedade de formas. As plantas pode ser quimeras de células transformadas 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 (por exemplo, um estoque de raiz transformada enxertado em um rebento não transformado em espécies de cítricos). As plantas transformadas podem ser propagadas por uma variedade de meios, como por técnicas de propagação clonal ou cruzamentos clássicos. Por exemplo, a planta de primeira geração (ou T1) transformada pode ser auto-cruzada para dar a segunda geração de plantas transformadas homozigóticas (ou T2), e as plantas T2 ainda propagadas através de técnicas de cruzamento clássicas. Um marcador dominante selecionável (como npt II) pode ser associado com o cassete de expressão para auxiliar no cruzamento.

Assim, a presente invenção provê uma célula de planta transformada (transgênica), *in planta* ou *ex planta*, incluindo um plastídeo transformado ou outra organela, por exemplo, núcleo, mitocôndria ou cloroplasto. A presente invenção pode ser usada para transformação de qualquer uma dentre as espécies de plantas, incluindo, mas não limitadas a, células de uma espécie de planta especificada acima na seção de DEFINIÇÃO. Preferivelmente, as plantas transgênicas da presente invenção são plantas de cultura e em particular cereais (por exemplo, milho, alfafa, girassol, arroz, Brassica, canola, soja, cevada, soja, beterraba de açúcar, algodão, cártamo, amendoim, sorgo, trigo, painço, tabaco, Linum usitatissimum (linhaça e linho), Camelina sativa, Brassica juncea, etc.), e mesmo mais preferivelmente milho, arroz e soja. Outras formas de realização da invenção estão relacionadas com células, culturas de célula, tecidos, partes (como órgãos de plantas, folhas, raízes, etc.) e material de propagação (como sementes) destas plantas.

O cassete de expressão transgênica da invenção pode não somente estar compreendido em plantas ou células de plantas mas pode também estar contido com vantagem em outros organismos, como por exemplo bactérias. Assim, outra forma de realização da invenção refere-se a células transgênicas ou organismos transgênicos não humanos, compreendendo um cassete de expressão da invenção. Preferidos são os organismos procarióticos e eucarióticos. Ambos os microorganismos e organismos superiores estão compreendidos. Os microorganismos preferidos são bactérias, levedura, algas, e fungos. As bactérias preferidas são as do gênero *Escherichia, Erwinia, Agrobacterium, Flavobacterium, Alcaligenes, Pseudomonas, Bacillus ou Cyanobacterim* como – por exemplo - *Synechocystis* e outras bactérias descritas em Brock Biology of Microorganisms 8a. Edição (páginas A-8, A-9, A10 e A11). O mais preferivelmente, as células transgênicas ou organismos transgênicos não humanos, compreendendo um cassete de expressão da invenção, são uma célula de planta ou planta (como definido acima), mais preferivelmente a planta usada para a produção de óleo como – por exemplo - *Brassica napus, Brassica juncea, Linum usitatissimum*, soja, Camelina ou girassol.

São especialmente preferidos os microorganismos capazes de infectar plantas e transferir o DNA em seu genoma, especialmente bactérias do gênero Agrobacterium, preferivelmente Agrobacterium tumefaciens e rizogenes. Preferidas leveduras são Candida, Saccharomyces, Hansenula e Pichia. Preferidos fungos são Aspergillus, Trichoderma, Ashbya, Neurospora, Fusarium, e Beauveria. Os mais preferidos são os organismos de plantas, como definido acima.

A transformação de plantas pode ser empreendida com uma única molécula de DNA ou múltiplas moléculas de DNA (isto é, co-transformação), e ambas as técnicas são apropriadas para uso com os cassetes de expressão da presente invenção. Numerosos vetores de transformação são

disponíveis para a transformação de plantas, e os cassetes de expressão desta invenção podem ser usados em conjunto com qualquer um destes vetores. A seleção de vetor irá depender das técnicas de transformação preferidas e as espécies de marcação para a transformação.

Várias técnicas se encontram disponível e conhecidas dos versados na arte para a introdução de construções em um hospedeiro de célula 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 semelhantes (Ver, por exemplo, EP 295959 e EP 138341) (ver abaixo). No entanto, células diferentes das células de planta podem ser transformadas com os cassetes de expressão da invenção. As descrições gerais de vetores de expressão de plantas e genes repórter, e a transferência de genes por Agrobacterium e mediada por Agrobacterium, podem ser encontradas em Gruber et al. (1993).

Os vetores de expressão contendo fragmentos genômicos ou sintéticos podem ser introduzidos em protoplastos ou em tecidos intactos ou células isoladas. Preferivelmente, os vetores de expressão são introduzidos em tecido intacto. Os métodos gerais de cultivo de tecidos de plantas são providos, por exemplo, por Maki et al., (1993); e por Phillips et al. (1988). Preferivelmente, vetores de expressão são introduzidos em milho ou outros tecidos de plantas usando um método de transferência de genes direto, distribuição mediada por microprojéteis, injeção de DNA,, eletroporação e semelhantes. Mais preferivelmente, os vetores de expressão são introduzidos em tecidos de plantas usando a distribuição por meio de microprojéteis, com o dispositivo biolístico. Ver, por exemplo, Tomes et al. (1995). Os vetores da invenção podem não somente ser usados para expressão de genes estruturais, como também podem ser usados em clonagem de exon-*trap*, ou

procedimentos *trap* de promotor para detectar a expressão diferencial de genes em vários tecidos (Lindsey 1993; Auch & Reth 1990).

É particularmente preferido usar os vetores de tipo binário de 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 tem recebido um estudo extensivo sendo amplamente descrito (EP 120516; Hoekema, 1985; Knauf, 1983; e An 1985). Para a introdução em plantas, os genes quiméricos da invenção podem ser inseridos em vetores binários, como descrito nos exemplos.

Outros métodos de transformação são disponíveis dos versados na arte, como absorção direta de construções de DNA estranhas (ver EP 295959), técnicas de eletroporação (Fromm 1986) ou bombardeio balístico em alta velocidade com partículas de metal revestidas com as construções de ácido nucleico (Kline 1987, e US 4.945.050). Uma vez transformadas, as células pode ser regeneradas pelos versados na arte. São de particular relevância, os métodos recentemente descritos para transformar os genes estranhos em culturas comercialmente importantes, como colza (De Block 1989), girassol (Everett 1987), soja (McCabe 1988; Hinchee 1988; Chee 1989; Christou 1989; EP 301749), arroz (Hiei 1994), e milho (Gordon-Kamm 1990; Fromm 1990).

Os versados na arte irão notar que a escolha do método pode depender do tipo de planta, isto é, monocotiledônea ou dicotiledônea, marcada para transformação. Os métodos apropriados de transformação de células de planta incluem, mas não são limitados a, microinjeção (Crossway 1986), eletroporação (Riggs 1986), transformação mediada por Agrobacterium (Hinchee 1988), transferência direta de genes (Paszkowski

1984), e aceleração balística de partículas usando dispositivos disponíveis de Agracetus, Inc., Madison, Wis. e BioRad, Hercules, Calif. (ver, por exemplo, US 4,945,050; e McCabe 1988). Ver, também, Weissinger 1988; Sanford 1987 (cebola); Christou 1988 (soja); McCabe 1988 (soja); Datta 1990 (arroz); Klein 1988 (milho); Klein 1988 (milho); Klein 1988 (milho); Fromm 1990 (milho); e Gordon-Kamm 1990 (milho); Svab 1990 (cloroplasto de tabaco); Koziel 1993 (milho); Shimamoto 1989 (arroz); Christou 1991 (arroz); Pedido de patente européia EP 0 332 581 (gramas de pomares e outros Pooideae); Vasil 1993 (trigo); Weeks 1993 (trigo).

Em outra forma de realização, uma seqüência nucleotídica da presente invenção é diretamente transformada no genoma do plastídeo. A tecnologia de transformação do plastídeo é extensivamente descrita em US 5.451.513, 5.545.817, e 5.545.818, em pedido PCT no. WO 95/16783, e em McBride et al., 1994. A técnica básica para transformação de cloroplastos envolve a introdução de regiões de DNA de plastídeo clonado flanqueando um marcador selecionável junto com o gene de interesse em um tecido de marcação apropriado, por exemplo, usando biolística ou transformação de protoplastos (por exemplo, transformação mediada por PEG ou cloreto de cálcio). As regiões de flanco 1 a 1,5 kb, chamadas seqüências de marcação, facilitam a recombinação ortóloga com o genoma do plastídeo e assim permitem a substituição ou modificação de regiões específicas do plastoma. Inicialmente, mutações por pontos nos genes 16S rRNA e rps12 do cloroplasto conferindo resistência a espectinomicina e/ou estreptomicina são utilizados como marcadores selecionáveis para transformação (Svab 1990; Staub 1992). Isto resultou em transformantes homoplásmicos estáveis a uma freqüência de aproximadamente um por 100 bombardeios de folhas alvo. A presença de sítios de clonagem entre estes marcadores permitiu a criação de um vetor de marcação de plastídeo para introdução de genes estranhos (Staub 1993). Aumentos substanciais em freqüência de transformação são obtidos

por substituição dos genes de resistência a antibióticos de proteína-r ou rRNA recessivos com um marcador selecionável dominante, o gene aadA bacteriano codificando a enzima destoxificando espectinomicina aminoglicosídeo-3N-adeniltransferase (Svab 1993). Outros marcadores selecionáveis úteis para a transformação de plastídeos são conhecidos na arte e englobados dentro do escopo da invenção. Tipicamente, aproximadamente 15-20 ciclos de divisão de células seguindo a transformação são requeridos para alcançar um estado homoplastídico. A expressão de plastídeos, em que genes são inseridos por recombinação ortóloga em todas dos vários milhares de cópias do genoma de plastídeo circular presentes em cada célula de planta, obtém vantagem do número de cópias enorme com relação aos genes expressados nucleares de modo a permitir níveis de expressão que podem prontamente exceder 10% da proteína de planta solúvel total. Em uma forma de realização preferida, uma seqüência nucleotídica da presente invenção é inserida em um vetor de marcação de plastídeo e transformada no genoma de plastídeo de um hospedeiro de planta desejado. As plantas homoplásticas para genomas de plastídeos contendo uma seqüência nucleotídica da presente invenção são obtidas, e são b preferivelmente capazes de uma expressão elevada da seqüência nucleotídica.

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 úteis em métodos de fabricação de plantas transformadas. Células de planta são infectadas com um *Agrobacterium tumefaciens*, como descrito acima, de modo a produzir uma célula de planta transformada, e então a planta é regenerado a partir da célula de planta transformada. Numerosos sistemas de vetor de Agrobacterium utilizáveis para realizar a presente invenção são conhecidos.

Várias cepas de *Agrobacterium* podem ser empregadas, preferivelmente cepas de *Agrobacterium tumefaciens* desarmado ou

rhizogenes. Em uma forma de realização preferida, cepas de *Agrobacterium* para uso na prática da invenção incluem cepas de octopina, por exemplo, LBA4404 ou cepas de agropina, por exemplo, EHA101 ou EHA105. As cepas apropriadas de *A. tumefaciens* para 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 (NCPPB 2659). Preferivelmente, estas cepas estão compreendendo uma variante de plasmídeo desarmada em 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* usada para transformar o tecido de planta pre-cultivado com composto fenólico de planta contém um plasmídeo Ti de tipo L,L-succinamopina, preferivelmente desarmado, como pEHA101. Em outra forma de realização preferida, a cepa de *Agrobacterium* usada para transformar o tecido de planta pre-cultivado com um composto fenólico de planta contém um plasmídeo Ti de tipo octopina, preferivelmente desarmado, como pAL4404. Geralmente, quando usando plasmídeos Ti tipo octopina ou plasmídeos auxiliares, é preferido 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 eficácia de transformação, como as cepas de *Agrobacterium* em que a expressão de gene vir e/ou sua indução é alterada devido à presença de genes virA ou virG mutantes ou quiméricos (por exemplo, Hansen 1994; Chen e Winans 1991;

Scheeren-Graiz, 1994). Preferidas são ainda combinações de *Agrobacterium* tumefaciens cepa LBA4404 (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 comuns, multiplicado em E. coli, e introduzido em *Agrobacterium* por, por exemplo, 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 o antibiótico apropriado (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 de marcação (por exemplo, um embrião imaturo) por *Agrobacterium* pode ser realizada por apenas contato do tecido de marcação com o *Agrobacterium*. A concentração de *Agrobacterium* usada para infecção e co-cultivo pode precisar ser variada. Por exemplo, uma suspensão de células de *Agrobacterium* tendo uma densidade de população de aproximadamente 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 de marcação é imerso nesta suspensão durante cerca de 3 a 10 minutos. O tecido de marcação resultante é então cultivado em um meio sólido durante vários dias junto com *Agrobacterium*.

Preferivelmente, a bactéria é empregada em concentração de 10^6 a 10^{10} cfu/ml. Em uma forma de realização preferida para a etapa de co-cultivo, cerca de 1 a 10 µl de uma suspensão de bactérias de origem no solo (por exemplo, *Agrobactérias*) no meio de co-cultivo são diretamente

aplicados a cada explante de tecido de marcação e secados ao ar. Isto economiza mão de obra e tempo e está reduzindo o dano mediado por Agrobacterium por pretendido pelo uso em excesso de Agrobacterium.

Para o tratamento por *Agrobacterium*, as bactérias são recolocadas em suspensão em um meio de co-cultivo compatível com a planta. A. Suplementação do meio de co-cultivo com antioxidantes (por exemplo, nitrato de prata), composto absorvendo fenol (como polivinilpirrolidona, Perl 1996) ou compostos tiol (por exemplo, ditiotreitol, L-cisteína, Olhoft 2001) que podem diminuir a necrose de tecido devido às respostas de defesa da planta (como a oxidação fenólica) pode 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, ditiotrietol (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 do tecido de marcação contra o dano mediado por Agrobacterium (como necrose induzida) e melhora altamente a eficiência de transformação global.

Vários sistemas de vetor podem ser usados em combinação com Agrobacteria. Preferidos são sistemas de vetor binário. Os vetores binários comuns são baseados em plasmídeos com "uma faixa de hospedeiro ampla" como pRK252 (Bevan 1984) ou pTJS75 (Watson 1985) derivados do plasmídeo RK2 tipo P. A maior parte deste 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). Vetores adicionais foram melhorados com

relação ao tamanho e manipulação (por exemplo, pPZP; Hajdukiewicz 1994). Sistemas de vetor melhorados são descrito também em WO 02/00900.

Métodos usando ou uma forma de transferência de genes direta ou transferência mediada por Agrobacterium, geralmente, mas não necessariamente, são empreendidos com um marcador selecionável, que pode prover resistência a um antibiótico (por exemplo, canamicina, higromicina ou metotrexato) ou um herbicida (por exemplo, fosfinotricina). A escolha de um marcador selecionável para transformação de plantas não é, no entanto, crítica para a invenção.

Para algumas espécies de plantas, diferentes marcadores de seleção de antibióticos ou herbicidas podem ser preferidos. Os marcadores de seleção usados como rotina em transformação incluem o gene nptII que confere resistência a canamicina e antibióticos relacionados (Messing & Vierra, 1982; Bevan 1983), o gene bar que confere resistência ao herbicida fosfinotricina (White 1990, Spencer 1990), o gene hph que confere resistência ao antibiótico higromicina (Blochlinger & Diggelmann), e o gene dhfr, que confere resistência a metotrexato (Bourouis 1983).

5. Produção e Caracterização de plantas estavelmente transformadas

As células de plantas transgênicas são então colocadas em um meio seletivo apropriado para a seleção de células transgênicas, que são então cultivadas até a formação do calo. Brotos são cultivados a partir dos calos. As plântulas são geradas 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, higromicina, cloranfenicol, herbicida, ou semelhantes). 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 de seqüências, que são nativas (endógenas) ou estranhas (exógenas) ao hospedeiro. Por "estranho" significa-se que a seqüência 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-região é derivada.

Para confirmar a presença dos transgenes em células e plantas transgênicas, uma variedade de testes pode ser realizada. Estes testes incluem, por exemplo, testes de "biologia molecular "bem conhecidos dos versados na arte, 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 um produto de proteína, por exemplo, por meios imunológicos (ELISAs e Western blots) ou por função enzimática, testes de partes das plantas, como testes de sementes; e também, por análise do fenótipo da planta regenerada completa, por exemplo, para resistência a doenças ou pragas.

DNA pode ser isolado das linhagens de células ou qualquer uma das partes da planta para determinar a presença do segmento de ácido nucleico pré-selecionado através do uso de técnicas bem conhecidas dos versados na arte. Nota-se que as seqüências intactas nem sempre estarão presentes, provavelmente devido ao rearranjo ou deleção de seqüências na célula.

A presença de elementos de ácido nucleico introduzidos através de métodos desta invenção pode ser determinada por reação em cadeia da polimerase (PCR). Usando esta técnica, fragmentos discretos de ácido nucleico são amplificados e detectados por gel eletroforese. Este tipo de a análise permite determinar se um segmento de ácido nucleico pré-selecionado está presente em um transformante estável, mas não fornece a integração do segmento de ácido nucleico pré-selecionado introduzido no genoma da célula

hospedeira. Além disso, não é possível usando técnicas de PCR determinar se os transformantes tem genes exógenos introduzidos em diferentes sítios no genoma, isto é, se os transformantes são de origem independente. É contemplado que usando as técnicas PCR, seria possível clonar os fragmentos do DNA genômico hospedeiro adjacente a um segmento de DNA pré-selecionado introduzido.

A prova positiva de integração do DNA no genoma hospedeiro e as identidades independentes de transformantes podem ser determinadas usando a técnica de hibridização Southern. Usando esta técnica, seqüências de DNA específicas, que foram introduzidas no genoma hospedeiro e estão flanqueando as seqüências de DNA hospedeiro podem ser identificadas. Assim, o padrão de hibridização Southern de um dado transformante serve como uma característica de identificação deste transformante. Além disso, é possível através de hibridização Southern, demonstrar a presença de segmentos de DNA pré-selecionados introduzidos em DNA de peso molecular elevado, isto é, confirmar que o segmento de DNA introduzido pré-selecionado foi integrado no genoma da célula hospedeira. A técnica de hibridização Southern provê informação que é obtida usando PCR, por exemplo, a presença de um segmento de DNA pré-selecionado, mas também demonstra a integração no genoma e caracteriza cada transformante individual.

É contemplado que usando as técnicas de hibridização "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 de um segmento de DNA pré-selecionado.

Ambas as técnicas de hibridização PCR e Southern podem ser usadas para demonstrar a transmissão de um 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 uma herança estável do gene. A natureza não quimérica do calo e dos transformantes parentais (R_0) foi sugerida por transmissão da linhagem germinal e os padrões idênticos de hibridização Southern blot e a intensidades do DNA transformante em calos, plantas R_0 plantas e progênie R_1 que segregaram para o gene transformado.

Apesar das técnicas de análise de DNA poderem ser conduzidas usando DNA isolado de qualquer parte de uma 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 quantificação de RNA produzido a partir dos segmentos de DNA pré-selecionados introduzidos. Nesta aplicação de PCR, é primeiro necessário reverter o RNA transcrito em DNA, usando enzimas como transcriptase reversa, e então, através de uso de técnicas PCR convencionais, amplificar o DNA. Na maior parte dos casos, as técnicas de PCR, apesar de utilizáveis, não irão demonstrar a integridade do produto RNA. Outra informação sobre a natureza do produto 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 pode também ser determinada usando hibridizações Northern "dot" ou "slot blot". Estas técnicas são modificações de Northern blotting e somente irão demonstrar a presença ou ausência de uma 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 informação como com relação ao segmento de DNA pré-selecionado sendo expressado. Expressão pode ser avaliada por especificamente identificando os produtos de proteína dos segmentos de DNA pré-selecionados introduzidos ou avaliação das mudanças fenotípicas ocasionadas por sua expressão.

Testes para a produção e identificação de proteínas específicas podem fazer usas propriedades físico químicas, estruturais, funcionais, ou outras das proteínas. As propriedades físico químicas ou estruturais singulares permitem que as proteínas sejam separadas e identificadas por procedimentos eletroforéticos, como eletroforese de gel nativo ou desnaturante ou focalização isoelétrica, ou por técnicas de cromatografia 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 testes ELISA. As combinações de abordagens podem ser empregadas com uma especificidade ainda maior, como Western blotting em que os anticorpos são usados para localizar produtos de genes individuais que foram separados por técnicas eletroforéticas. As técnicas adicionais podem ser empregadas para confirmar absolutamente a identidade do produto de interesse como avaliação por seqüenciamento de aminoácidos após purificação. Apesar destes estarem 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 substratos e produtos específicos. Estas reações podem ser seguidas ao prover e quantificar a perda de substratos ou a geração de produtos das reações por procedimentos físicos ou químicos. Os exemplos são tão variados como as enzimas a serem analisadas.

Com muita freqüência, a expressão de um produto de gene é determinada por avaliação dos resultados fenotípicos de sua expressão. Estes testes podem tomar muitas formas, incluindo, mas não limitadas a análise de mudanças em composição química, morfologia, ou propriedades fisiológicas da planta. As mudanças morfológicas podem incluir uma estatura maior ou

talos mais espessos. Com muita freqüência, mudanças em resposta de plantas ou partes de plantas a serem impostas, os tratamentos são avaliados sob condições cuidadosamente controladas, chamadas biotestes.

6. Uso de plantas transgênicas

Uma vez que o cassete de expressão da invenção foi transformado em uma espécie de planta particular, ele pode ser propagado nas espécies ou movimentado para outras variedades das mesmas espécies, particularmente incluindo variedades comerciais, usando técnicas de criação tradicionais. As plantas particularmente preferidas da invenção incluem as culturas agronomicamente importantes listadas acima. As propriedades genéticas engenheiradas nas sementes e plantas transgênicas descritas acima são passadas adiante por reprodução sexual e podem, assim, ser mantidas e propagadas nas plantas da progênie. A presente invenção também refere-se a uma célula, tecido, órgão, semente ou parte de planta de uma planta transgênica, obtidos da planta transgênica. Também estão incluídos na invenção os descendentes transgênicos da planta, assim como células, tecidos, órgãos, sementes e partes das plantas de plantas transgênicas obtidos partir dos descendentes.

Preferivelmente, o cassete de expressão na planta transgênica é sexualmente transmitido. Em uma forma de realização preferida, a seqüência de codificação é sexualmente transmitida através de um ciclo sexual normal completo da planta R0 para a geração R1. Adicionalmente preferido, o cassete de expressão é expressado nas células, tecidos, sementes ou planta de uma planta transgênica em uma quantidade que é diferente da quantidade nas células, tecidos, sementes ou planta de uma planta, que somente difere em que o cassete de expressão está ausente. As plantas transgênicas produzidas aqui são assim esperadas como sendo úteis para uma variedade de fins comerciais e de pesquisa. Plantas transgênicas podem ser criadas para uso em agricultura tradicional de modo a possuir traços benéficos para o agricultor (por exemplo,

traços agronômicos como resistência a falta d'água, resistência a pragas, resistência a herbicidas ou produção aumentada), benéficos para o consumidor do grão coletado de um planta (por exemplo, teor nutritivo melhorado em ração animal ou alimentação humana; teor de vitamina, aminoácidos e antioxidantes aumentado; a produção de anticorpos (imunização passiva) e nutracêuticos), ou benéficos para os processadores de ração (por exemplo, traços de processamento aumentados). Em tais usos, as plantas são geralmente cultivadas para o uso de seu grão em rações animais ou alimentos para humanos. Adicionalmente, o uso de promotores específicos de raiz em plantas transgênicas pode proporcionar traços benéficos que estão localizados nas raízes consumíveis (por animais e humanos) de plantas como cenouras, pastinacas, e beterrabas. No entanto, outras partes das plantas, incluindo talos, cascas, partes vegetativas, e semelhantes, também podem ter utilidade, incluindo uso como parte de silagem para animais ou fins ornamentais. Com freqüência, constituintes químicos (por exemplo, óleos ou amidos) de milho e outras culturas são extraídos para rações ou uso industrial e plantas transgênicas podem ser criadas tendo níveis melhorados ou modificados de tais componentes.

Plantas transgênicas também podem encontrar uso na fabricação comercial de proteínas ou outras moléculas, onde a molécula de interesse é extraída ou purificada de partes de plantas, sementes, e semelhantes. Células ou tecidos de um planta também podem ser cultivados, levados a crescer in vitro, ou fermentados para fabricar tais moléculas. As plantas transgênicas também podem ser usadas em programas de cruzamento comercias, ou podem ser cruzadas ou criadas em plantas de espécie de cultura relacionada. As melhoras codificadas pelo cassete de expressão podem ser transferidas, por exemplo, de células de milho para células de outras espécies, por exemplo, por fusão de protoplastos.

As plantas transgênicas podem ter muitos usos em pesquisa ou

cruzamentos, incluindo a criação de novas plantas mutantes através de mutagênese insercional, a fim de identificar os mutantes benéficos que podem ser depois criados por mutação e seleção tradicional. Um exemplo deve ser a introdução de uma seqüência 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 "seqüências de assinatura" singulares, ou outras seqüências de marcador que podem ser usadas para identificar variedades e linhagens de propriedade de uma empresa.

Assim, as plantas e sementes transgênicas de acordo com a invenção podem ser usadas no cruzamento de plantas, que visa o desenvolvimento de plantas com melhoradas propriedades conferidas pelo cassete de expressão, como tolerância de seca, doença, ou outros estresses. As ´varias etapas de cruzamento são caracterizadas por intervenção humana bem definida, como seleção das linhagens a serem cruzadas, dirigindo a polinização das linhagens parentais, ou selecionando as plantas descendentes apropriadas. Dependendo das propriedades desejadas, diferentes medidas de cruzamento são tomadas. As técnicas relevantes são bem conhecidas na arte e incluem, mas não são limitadas a hibridização, cruzamento endogâmico, cruzamento retro-cruzado, cruzamento de múltiplas trilhas, mistura de variedades, hibridização 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 uma planta estéril masculina com pólen de uma linhagem diferente assegura que o genoma da planta estéril masculina diferente da planta fértil feminina irá alcançar uniformemente propriedades de ambas as linhagens parentais. Assim, as sementes e plantas transgênicas de acordo com a invenção podem ser usadas para o cruzamento de linhagens de plantas melhoradas, que, por exemplo, aumentam a eficácia

de métodos convencionais como tratamento de pesticida ou herbicida ou permitem prescindir destes referidos métodos devido às suas propriedades genéticas modificadas. Alternativamente, novas culturas com melhorada tolerância ao estresse podem ser obtidas que, devido a seu "equipamento" genético otimizado fornecem um produto coletado de maior qualidade do que os produtos que não foram capazes de tolerar condições de desenvolvimento adversas comparáveis.

EXEMPLOS

Materiais e métodos gerais

Salvo indicado em contrário, produtos químicos e reagentes nos Exemplos foram obtidos de Sigma Chemical Company (St. Louis, MO), endonucleases de restrição foram obtidos 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 com relação a produtos bioquímicos e testes biológicos moleculares foram de Clontech (Palo Alto, CA), Farmacia Biotech (Piscataway, NJ), Promega Corporation (Madison, WI), ou Stratagene (La Jolla, CA). Materiais para 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 em membranas de nitrocelulose e náilon, ligação de fragmentos de DNA, transformação de células de *E. coli*, cultivo de bactérias, multiplicação de fagos e análises de seqüência de DNA recombinante são realizados como descrito por Sambrook (1989). O seqüenciamento de moléculas recombinantes de DNA é realizado usando sequenciador de DNA fluorescência a laser de ABI seguindo o método de Sanger (Sanger 1977).

Exemplo 1: Geração de plantas transgênicas

1.1 Geração de plantas transgênicas de *Arabidopsis*

thaliana

Para gerar plantas transgênicas de *Arabidopsis*, *Agrobacterium tumefaciens* (cepa C58C1[pMP90]) é transformado com as várias construções de promotor::GUS vetor (ver abaixo). As cepas resultantes de *Agrobacterium* são subseqüentemente empregadas para obter plantas transgênicas. Para este fim, uma colônia de *Agrobacterium* isolada transformada é incubada em 4 ml de cultura (meio: meio YEB com 50 µg/ml canamicina e 25 µg/ml Rifampicina) durante a noite a 28°C. Com esta cultura, uma cultura de 400 ml do mesmo meio é inoculada e incubada durante a noite (28 °C, 220 rpm). As bactérias precipitaram por centrifugação (Rotor GSA, 8.000 U/min, 20 min) e a pelota é recolocada 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 concentração final de 0,02%. Uma caixa de planta com 8 a 12 plantas é colocada em um exsicador durante 10 a 15 min, sob vácuo, com ventilação espontânea subseqüente, (expansão). Este processo é repetido 2-3 vezes. A seguir, todas as plantas são transferidas em sacos com solo úmido e cultivadas sob condições tempo diurno longo (16 h luz; temperatura do dia 22-24°C, temperatura da noite 19 °C; 65% umidade relativa). As sementes são coletadas após 6 semanas.

1.2 Geração de linhaça transgênica

As plantas de linhaça transgênica podem ser geradas por exemplo pelo método de Bell et al., 1999, In Vitro Cell. Dev. Biol.-Plant. 35(6): 456-465 por meio de bombardeio de partículas. As transformações mediadas por Agrobacteria podem ser geradas por exemplo pelo método de Mlynarova et al. (1994), Célula de planta Report 13: 282-285.

1.3 Transformação mediada por *Agrobacterium* de *Brassica napus*

Colza de semente oleígena pode ser transformada por pecíolo

cotiledonário ou transformação de hipocótilos (Moloney 1989; De Block 1989). O uso de antibióticos para a seleção de *Agrobacteria* e plantas depende do vetor binário e da cepa de *Agrobacterium* usada para a transformação. A seleção de colza de semente oleígena é geralmente realizada usando canamicina como marcador de planta selecionável.

Mais especificamente, colza de semente oleígena pode ser transformada como a seguir: cepas de *Agrobacterium* transformadas com o plasmídeo de interesse cultivadas em 50 mL meio YEB a 28°C durante a noite. A solução de *Agrobacterium* é misturada com meio líquido de co-cultivo (sais MSB5 concentrados duplos (Duchefa), 30 g/L sacarose (Duchefa), 3.75 mg/Ll BAP (6-benzilamino purina, Duchefa), 0,5 g/L MES (Duchefa), 0,5 mg/L GA3 (ácido giberélico, Duchefa); pH5.2) até OD_{650} de 0,5 é alcançado. Pecíolos de mudinhas de 4 dias de *Brassica napus* cv. Westar cultivadas em meio de crescimento B (sais MSB5 (Duchefa), 3% sacarose (Duchefa), 0.8% oxoidagar (Oxoid GmbH); pH 5,8) são cortados. Os pecíolos são imersos durante 2-3 segundos na solução de *Agrobacterium* e depois colocados em meio sólido para co-cultivo (meio de co-cultivo suplementado com 1.6% Oxoidagar). O co-cultivo dura 3 dias (a 24°C e cerca de 50 µMol/m^2s intensidade de luz). Depois, os pecíolos são transferidos para o meio de co-cultivo suplementado com o agente de seleção apropriado (18 mg/L canamicina (Duchefa) para plantas compreendendo o marcador nptII de canamicina para plantas transportando o marcador de resistência *nptII*, ou 0.3 a 30 mM D-aminoácidos; como descrito abaixo) para plantas compreendendo um cassete de expressão para o gene *dao1* de *Rhodotorula gracilis*) e 300 mg/L Timentina (Duchefa)

Os pecíolos transformados são incubados no meio de seleção durante quatro semanas a 24°C. Esta etapa é repetida até os brotos aparecerem. Os brotos são transferidos para meio A6 (sais MS (Sigma Aldrich), 20 g/L sacarose, 100 mg/L mio-inositol (Duchefa), 40 mg/L adenina

sulfato (Sigma Aldrich), 500 mg/L MES, 0,0025 mg/L BAP (Sigma), 5 g/L oxoidagar (Oxoid GmbH), 150 mg/L timetina (Duchefa), 0,1 mg/L IBA (ácido indol butírico, Duchefa); pH 5,8) suplementado com o agente de seleção apropriado (18 mg/L canamicina (Duchefa) para plantas compreendendo o marcador *nptII* de canamicina para plantas transportando o marcador de resistência *nptII*, ou 0.3 a 30 mM D-aminoácidos; como descrito abaixo) até eles serem alongados. Os brotos alongados são cultivados em meio A7 (meio A6 sem BAP) para a formação de raiz. As plantas com raiz são transferidas para solo e cultivadas na estufa.

EXEMPLO 2: Condições de crescimento para plantas para a análise de expressão específica de tecido

Para obter mudinhas de 4 e 7 dias de idade, cerca de 400 sementes (*Arabidopsis thaliana* ecotipo Columbia) são esterilizadas com a 80% (v/v) solução etanol:água durante 2 minutos, tratadas com solução de hipoclorito de sódio (0,5% v/v) durante 5 minutos, lavadas três vezes com água destilada e incubadas a 4°C durante 4 dias para assegurar uma germinação padronizada. Subseqüentemente, as sementes são incubadas em discos de Petri com meio MS (Sigma M5519) suplementado com 1% sacarose, 0,5 g/l MES (Sigma M8652), 0.8% Difco-BactoAgar (Difco 0140-01), ajustado a pH 5.7. As mudinhas são cultivadas sob um ciclo de 16 h luz / 8 h escuro (luz branca Philips 58W/33) a 22°C e coletadas após 4 ou 7 dias, respectivamente.

Para obter o tecido de raiz, 100 sementes são esterilizadas como descrito acima, incubadas a 4°C durante 4 dias, e transferidas para frascos de 250ml com meio MS (Sigma M5519) suplementado com 3% sacarose e 0,5 g/l MES (Sigma M8652) adicionais, ajustado a pH 5.7 durante o posterior crescimento. As mudinhas são cultivadas a um ciclo de 16 h luz / 8 h escuro (luz branca Philips 58W/33) a 22°C e 120 rpm e coletadas após 3 semanas. Para todos os outros órgãos de plantas empregados, sementes são

semeadas em solo padrão (Tipo VM, Manna-Italia, Via S. Giacomo 42, 39050 San Giacomo/ Laives, Bolzano, Itália), incubadas durante 4 dias a 4°C para assegurar uma germinação uniforme, e subseqüentemente cultivadas sob um regime de 16 h luz / 8 escuro (OSRAM Lumi-lux Dayluz 36W/12) a 22°C. As folhas de roseta jovem são coletadas no estágio de 8 folhas (após cerca de 3 semanas), folhas de roseta madura são coletadas após 8 semanas brevemente antes da formação do caule. Os ápices dos caules saindo dos brotos são coletados brevemente após a brotação. Caule, folhas do caule, e botões de flor são coletados em estágio de desenvolvimento 12 (Bowmann J (ed.), *Arabidopsis*, Atlas of Morphology, Springer New York, 1995) antes do desenvolvimento do estame. As flores abertas são coletadas em estágio de desenvolvimento 14 imediatamente após o desenvolvimento do estame. As flores murchando são coletadas em estágio 15 a 16. Os brotos verdes e amarelos usados para a análise têm um comprimento de 10 a 13 mm.

As plantas transgênicas de colza e linhaça regeneradas são testadas em cultura de tecidos para funcionalidade parcial prematura. Para as análises detalhadas, 3 plantas individuais por linhagem de inserção única ou múltipla (5-15 linhagens em total por construção) são analisadas na geração T1 a T3 com relação à expressão em potencial dos candidatos de promotor em todos os tecidos não semente, assim como em diferentes fases de desenvolvimento da semente:

- sementes maduras
- mudinhas de 3 dias de idade: base da raiz, ponta da raiz, raiz principal, raiz lateral, outras áreas da raiz, cotilédone, hipocótilo
- mudinhas de 10 dias de idade: base da raiz, ponta da raiz, raiz principal, raiz lateral, outras áreas da raiz, cotilédone, hipocótilo, folhas primárias, seguintes folhas
- mudinhas de 17 dias de idade: base da raiz, ponta da raiz, raiz principal, raiz lateral, outros áreas da raiz, cotilédone, hipocótilo, folhas

primárias, seguintes folhas

- Planta adulta: raiz, folhas jovens, folhas maduras, caule

- Flor 0, cápsula /síliquas/fruta 1 daf, 3 daf, semente/embrião 3 daf, cápsula/síliquas/fruta 6 daf, semente/embrião6 daf, cápsula/síliquas/fruta 9 daf, semente/embrião 9 daf, cápsula/síliquas/fruta 12 daf, semente/embrião 12 daf, cápsula/síliquas/fruta 15 daf, semente/embrião 15 daf, cápsula/síliquas/fruta 18 daf, semente/embrião 18 daf, cápsula/síliquas/fruta 21 daf, semente/embrião 21 daf, cápsula/síliquas/fruta 24 daf, semente /embrião 24 daf (dias após fertilização)

- alternativamente, embriões de, por exemplo, cápsula de linhaça e síliquas de semente de colza são isolados de diferentes estágios de desenvolvimento de fruta com base em parâmetros visuais e classificados nos seguintes estágios de desenvolvimento do embrião: prematuro, jovem, médio, tardio e maduro.

Os promotores são também verificados por sua indutibilidade por estresse biótico e abiótico via pulverização de ABA sobre as folhas.

EXEMPLO 3: Demonstração de perfil de expressão

Para demonstrar e analisar as propriedades reguladoras de transcrição de um promotor da ligação útil a operativamente, o promotor ou seus fragmentos para um gene repórter, que pode ser empregado para monitorar sua expressão tanto qualitativamente como quantitativamente. Preferivelmente ß-glucuronidase bacteriano é usado (Jefferson 1987). A atividade ß-glucuronidase pode ser monitorada *in planta* com substratos cromogênicos como ácido 5-bromo-4-cloro-3-indolil-ß-D-glucurônico durante testes de atividade correspondentes (Jefferson 1987). Para determinação de atividade de promotor e especificidade de tecido, tecido de planta é dissecado, incrustado, colorido e analisado como descrito (por exemplo, Bäumlein 1991).

Para atividade quantitativa de ß-glucuronidase, análise MUG

(metilumbeliferil glucuronídeo) é usado como um substrato, que é convertido em MU (metilumbeliferone) e ácido glucurônico. Sob condições alcalinas, esta conversão pode ser quantitativamente monitorada fluorometricamente (excitação a 365 nm, medida a 455 nm; SpectroFluorimeter Thermo Life Sciences Fluoroscan) como descrito (Bustos 1989).

EXEMPLO 4: Clonagem dos fragmentos do promotor

Para isolar os fragmentos do promotor descritos por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, e 11, DNA genômico é isolado de *Arabidopsis thaliana* (ecotipo Columbia) ou *Brassica napus* como descrito (Galbiati 2000). O genômico DNA isolado é empregado como DNA matriz para uma amplificação mediada por reação em cadeia da polimerase (PCR) usando os iniciadores de oligonucleotídeos e protocolos indicados abaixo (Tabela 3).

Tabela 3: Iniciadores de oligonucleotídeos PCR para amplificação das várias seqüências nucleotídicas reguladoras de transcrição e enzimas de restrição para modificação dos produtos PCR resultantes

SEQ ID	**Promotor**	**Iniciador dianteiro**	**Iniciador reverso**	**Enzima de restrição**
SEQ ID NO:1	AtPER-1683-Lo322	AtPER-1683-Lo322 SEQ ID NO: 24	AtPER-1683-Lo322 SEQ ID NO: 25	SalI/NdeI
SEQ ID NO:2	AtPER-1619-Lo322	AtPER-1619-Lo322 SEQ ID NO: 24	AtPER-1619-Lo322 SEQ ID NO: 26	SalI/HindII
SEQ ID NO:3	AtPER-985-Lo322	AtPER-985-Lo322 SEQ ID NO: 27	AtPER-985-Lo322 SEQ ID NO: 28	SpeI/NdeI
SEQ ID NO:4	AtPER-921-Lo322	AtPER-921-Lo322 SEQ ID NO:27	AtPER-921-Lo322 SEQ ID NO: 29	SpeI/NdeI
SEQ ID NO:5	AtPER-1696-bac	AtPER-1696-bac SEQ ID NO: 24	AtPER-1696-bac SEQ ID NO: 25	SalI/NdeI
SEQ ID NO:6	AtPER-1632-bac	AtPER-1632-bac SEQ ID NO: 24	AtPER-1632-bac SEQ ID NO: 30	SalI/HindIII
SEQ ID NO:7	AtPER-989-bac	AtPER-989-bac SEQ ID NO: 27	AtPER-989-bac SEQ ID NO: 31	SpeI/KpnI
SEQ ID NO:8	AtPER-925-bac	Se AtPER-925-bac SEQ ID NO: 27	Se AtPER-925-bac SEQ ID NO: 32	SpeI/NdeI
SEQ ID NO:9	LuPER –Lo383 – 1727 bp	LuPER –Lo383 SEQ ID NO: 33	LuPER –Lo383 SEQ ID NO: 34	SalI/SpeI

SEQ ID NO:10	LuPER –933bp	LuPER –933bp SEQ ID NO: 35	LuPER –933bp SEQ ID NO: 34	SalI/SpeI
SEQ ID NO:11	LuPER –458bp	LuPER –458bp SEQ ID NO: 36	LuPER –458bp SEQ ID NO: 34	SalI/SpeI

A amplificação é realizada como a seguir:

100 ng DNA genômico

1X PCR tampão

2,5 mM $MgCl_2$,

200 μM cada de dATP, dCTP, dGTP e dTTP

10 pmol de cada um dos iniciadores oligonucleotídeo

2,5 Unidades Pfu DNA Polimerase (Stratagene)

em um volume final de 50 μl

O seguinte programa de temperatura é empregado para as várias amplificações (BIORAD Thermocycler).

1. 95°C durante 5 min

2. 54°C durante 1 min, seguido por 72°C durante 5 min e 95°C durante 30 segundos. Repetido 25 vezes

3. 54°C durante 1 min, seguido por 72°C durante 10 min.

4. armazenamento a 4°C

Os produtos PCR resultantes são digeridos com as endonucleases de restrição especificadas na Tabela acima (Tabela 3) e clonados no vetor pSUN0301 (SEQ ID NO: 147) (pre-digerido com as mesmas enzimas) a montante e em ligação operativa ao gene glucuronidase (GUS). Seguindo a transformação estável de cada uma destas construções em *Arabidopsis thaliana*, a especificidade do tecido e perfil de expressão foram analisados para um teste GUS histoquímico e quantitativo, respectivamente.

EXEMPLO 5: Perfil de expressão das várias construções promotor::GUS em plantas de *A. Thaliana*, colza e linhaça estavelmente transformadas

O padrão de expressão de glucuronidase (GUS) dos promotores peroxiredoxina (PER1) de Arabidopsis (pAtPer; SEQ ID NO: 1,

2, 3, 4, 5, 6, 7, e 8) e Linum usitatissimum (pLuPer; SEQ ID NO: 9, 10, e 11) foram investigados em Arabidopsis, colza e linhaça.

Resultados gerais: Expressão GUS fraca a média - forte foi detectada em embriões prematuros e jovens. Expressão GUS média - forte a forte foi encontrada em embriões médios, tardios e maduros. Não foi verificada funcionalidade parcial.

Caracterização em Arabidopsis: expressão GUS forte dos promotores pAtPer e pLuPer foi detectada em embriões durante todos os estágios de desenvolvimento. Não foi verificada funcionalidade parcial.

Caracterização em colza: expressão fraca do promotor pAtPer foi detectada em embriões prematuros e jovens. A expressão média - forte a forte foi encontrada em sementes médias, tardias e maduras. Não foi verificada funcionalidade parcial. Expressão média - forte a forte do promotor pLuPer foi encontrada em sementes prematuras, jovens, médias, tardias e maduras. Não foi verificada funcionalidade parcial.

Caracterização em linhaça: em embriões prematuros e jovens de linhaça, os promotores pAtPer e pLuPer não mostram expressão ou uma expressão fraca. A expressão média - forte destes promotores foi encontrada em sementes médias, tardias e maduras. Não foi verificada funcionalidade parcial.

EXEMPLO 7: Construção de vetor para experimentos de super-expressão e "abate" de genes

7.1 Super-expressão

Vetores usados para expressão de "genes candidatos" de comprimento completo de interesse em plantas (super-expressão) são projetados para super-expressar a proteína de interesse e são de dois tipos gerais, biolístico e binário, dependendo do método de transformação de planta a ser usado.

Para transformação biolística (vetores biolísticos), as

exigências são como a seguir:

1. uma estrutura dorsal com um marcador selecionável bacteriano (tipicamente, um gene de 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 de um promotor (por exemplo. ZmUBIint MOD), o gene de interesse (tipicamente, um cDNA de comprimento completo) e um terminador transcripcional (por exemplo, terminador nos de Agrobacterium tumefaciens);

b. um cassete de marcador selecionável de planta, consistindo de um promotor apropriado, gene marcador selecionável (por exemplo,, D-aminoácido oxidase; dao1) e terminador transcripcional (por exemplo, terminador nos).

Vetores projetados para transformação por Agrobacterium tumefaciens (*A. tumefaciens*; vetores binários) consistem de:

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 acima mencionados, mais o gene virG de *A. tumefaciens*;

2. uma porção específica de planta, como descrito para vetores biolísticos acima, exceto neste caso esta porção é flanqueada por seqüências de borda direita e esquerda de *A. tumefaciens* que mediam a transferência do DNA flanqueado por estas duas seqüências para a planta.

7.2 Vetores de silenciamento de genes

Vetores projetados para reduzir ou abolir a expressão de um gene único ou de uma família ou genes relacionados (vetores de silenciamento de genes) são também de dois tipos gerais correspondentes à metodologia usada para regular de modo negativo a expressão de genes: RNA

interferência anti-sentido ou filamento duplo (dsRNAi).

(a) Anti-sentido

Para vetores anti-sentido, um fragmento de gene de comprimento completo ou parcial (tipicamente, uma porção do cDNA) pode ser usado nos mesmos vetores descritos para expressão de comprimento completo, como parte do cassete de expressão do gene. Para a regulação de modo negativo mediada por anti-sentido da expressão de gene, a região de codificação do gene ou fragmento de gene estará na orientação oposta relativa ao promotor; assim, mRNA será feito a partir do filamento não de codificação (anti-sentido) *in planta.*

(b) dsRNAi

Para vetores dsRNAi, um fragmento de gene parcial (tipicamente, comprimento de 300 a 500 pares de bases) é usado no cassete de expressão do gene, e é expressado em ambas as orientações sentido e anti-sentido, separado por uma região de espaçador (tipicamente, um íntron de planta, por exemplo, o intron OsSH1 1, ou um marcador selecionável, por exemplo, conferindo resistência a canamicina). Vetores deste tipo são projetados para formar uma cepa de mRNA de filamento duplo, resultando da formação de pares de bases de dois fragmentos de genes complementares *in planta.*

Os vetores biolísticos ou binários projetados para super-expressão ou abate podem variar em vários diferentes modos, incluindo, por exemplo, os marcadores selecionáveis usados em plantas e bactérias, os terminadores transcripcionais usados nos cassetes de expressão de genes e de marcador selecionável de plantas, e as metodologias usadas para clonagem em gene ou fragmentos de genes de interesse (tipicamente, clonagem baseada em recombinase Gateway™ ou mediada por enzima de restrição convencional).

EXEMPLO 8: Análise do elemento promotor

A análise do motivo de promotor foi feita usando o software de Genomatix MatInspector versão profissional 7.3 (agosto de 2004). O promotor de perooxiredoxina (PER 1) de Linum usitatissimum (pLuPer; SEQ ID NO: 9; 1727bp) está compreendendo três regiões com densidade diferente de motivos de promotor (ver Fig. 2; região A, B, C). A seguinte lista está apresentando alguns dos motivos de promotor compreendidos. A posição está indicando a posição de nucleotídeo na seqüência descrita por SEQ ID NO: 9. O motivo é descrito de acordo como os seguintes princípios:

- atag**attaTTTA**gaa – Motivo de matriz (completo)
- **attaTTTA** – região conservada de motivo
- **TTA** – (negrito, letras maiúsculas) região de núcleo de motivo
- **att** – (negrito, letras minúsculas) parte de motivo de promotor conservada mas menos conservada do que a região de núcleo

8.1 Região A (1 – 794 bp; somente em SEQ ID NO: 9) está compreendendo numerosos motivos de promotor característicos para expressão específica de semente:

- Proteína de ligação de DNA de batata-doce que liga às seqüências theSP8a (ACTGTGTA) e SP8b (TACTATT) de genes esporamina e beta-amilase

Posição: 178-186 (+ filamento); Motivo: **agTACTctt**tga

- Elemento de choque térmico. Posição 223-235 (filamento+); Motivo: **aga**gga**gac**a**AGAA**a

- Caixa de Prolamina, conservada em promotores do gene de proteína de armazenamento de semente de cereal

Posição: 266-281 (filamento +); Motivo: **ga**agt**tgcAAAG**ttg**ag**

- Opaco-2. Posição: 316-330 (- filamento); Motivo: **CCATtccttgtagta**ct

- Storekeeper (STK), proteína de ligação de DNA específico

da planta importante para expressão do gene específico de tubérculo e indutível de sacarose.

Posição 621-628 (filamento -); Motivo: **TAAAtaatc**tat

- Caixa de legumina, elemento de seqüência altamente conservado, cerca de 100 bp a montante do TSS em genes de legumina

Posição: 645-669 (filamento -), 655-679 (filamento -). Motivo: t**ccattgCTATccttgttagaaacta**a

- Elementos de resposta ABA. Posição: 727-734 (filamento +). Motivo: ttttac**ACGTa**cc

8.2 Região B (795 – 1269bp; compreendida em SEQ ID NO: 9 e 10) está compreendendo uma densidade relativamente baixa de motivos de promotor específicos de semente, mas mais geral ou outros motivos.

- Determina a identidade de meristema floral e desenvolvimento de sépalas e pétalas

Posição: 857-872 (filamento +), 917-931 (filamento -); 1013-1027 (filamento+) e um motivo similar a 1134-1139 (filamento +); Motivo: acaaaacaAAAAtgttaatat

- Proteína semelhante a Myb de Petunia hybrida. Posição: 864-872 (filamento +), 801-809 (filamento -); 1178-1186 (filamento -); 1218-1226 (filamento +); 1268-1276 (filamento +); Motivo: acaaaaatGTTAata.

- SBF. Posição: 870-875 (filamento +), 898-903 (filamento -), 1179-1184 (filamento -). Motivo: aaaaatgTTAAtatttt

- Elementos semelhantes a OCS. Posição: 867-880 (filamento -). Motivo: atatgaaaatattaACATttt

- Agamous, requerido para desenvolvimento normal da flor, similaridade a proteínas SRF (humano) e MCM (levedura) Posição: 919-930 (filamento +); Motivo: aatTTCCatatttgtaggaaa

8.3 Região C (1270- 1727bp; compreendida pela SEQ IDS NO: 9, 10 e 11) (~ SEQ. ID NO 11) compreendendo agrupamentos de alguns

promotores de motivos específicos de semente:

- Motivos RY e Sph conservados em promotores específicos de semente Posição 1282-1287 (filamento -), 1522-1527 (filamento -) e 1615-1620 (filamento +); Motivo: cccaagta**CATGca**aaatttaacggtt

- Opaco-2. Posição: 1372-1386 (filamento -), 1608-1622 (-) e 1688-1696 (filamento +). Motivo: **TTATtataagtaaat**tg

- Embrião de soja fator 4. Posição: 1386-1390 (filamento -), 1713-1717 (filamento +); Motivo: atTTTTtatta

- Elementos semelhantes a OCS. Posição: 1446-159 (filamento +); Motivo: aaat**gtgg**ta**g**cgt**ACGT**gtg

- Elemento de choque térmico. Posição: 1571-1583 (filamento -); Motivo: t**gg**cggc**tc**c**AGAA**t

- Ativador 1- de transcrição de arroz (RITA), proteína zipper de leucina básica, altamente expressada durante desenvolvimento da semente. Posição: 1584-1587 (filamento -); Motivo: cttgcgt***ACGT***ggcggc

- Proteína 8 bZIP de Oryza sativa. Posição: 1602-1608 (filamento -), 1583-1589 (-); Motivo: tgaagaga**ACGTggc**cctgag

- Elementos de resposta ABA. Posição: 1598-1608 (filamento -), 1581-1589 (-), 1452-1460 (+); Motivo: tgcatgaa**g**aga**ACGTg**gccctgagtc

- Motivos RY e Sph conservados em promotores específicos de semente. Posição: -1615-1620 (filamento +); Motivo: gttctctt**CATGca**aggatagtagaac

- SEF3, embrião de soja fator 3. Posição: 1641-1650 (filamento +). Motivo: actc**cACCCa**c**c**tcc

- Opaco-2. Posição: 1688-1696 (filamento +); Motivo: ctttcc**ca**tc**CCATc**ca

- Embrião de soja fator 4. Posição:1713-1717 (filamento +); Motivo: ca**TTTTt**atca

REFERÊNCIAS

1. Abel et al., Science, 232:738 (1986).
2. Altschul et al., Nucleic Acids Res., 25:3389 (1997).
3. Altschul et al., J. Mol. Biol., 215:403 (1990).
4. An et al., EMBO J., 4:277 (1985).
5. Auch & Reth, Nucleic Acids Research, 18:6743 (1990).
6. Ausubel et al. (1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience
7. Ballas et al., Nucleic Acids Res., 17:7891 (1989).
8. Barkai-Golan et al., Arch. Microbiol., 116:119 (1978).
9. Batzer et al., Nucleic Acid Res., 19:5081 (1991).
10. Bäumlein et al. Mol Gen Genet 225:121-128 (1991)
11. Becker et al. (1994) Plant J., 5:299-307,
12. Bernal-Lugo e Leopold, Plant Physiol., 98:1207 (1992).
13. Bevan et al., Nature, 304:184 (1983).
14. Bevan et al., Nucl. Acids Res., 11:369 (1983).
15. Bevan, Nucl. Acids Res., 12:8711 (1984).
16. Blackman et al., Plant Physiol., 100:225 (1992).
17. Blochlinger & Diggelmann, Mol Cell Biol, 4:2929 (1984).
18. Bol et al., Ann. Rev. Phytopath., 28:113 (1990).
19. Bouchez et al., EMBO J., 8:4197 (1989).
20. Bourouis et al., EMBO J., 2:1099 (1983).
21. Bowler et al., Ann. Rev. Plant Physiol., 43:83 (1992).
22. Branson e Guss, Proc. North Central Branch Entomological Society of America (1972).
23. Broakgert et al., Science, 245:110 (1989).
24. Bustos MM et al. (1989) Plant Gell 1:839-853
25. Byrne et al. Plant Cell Tissue and Organ Culture, 8:3 (1987).

26. Callis et al., Genes and Develop., 1:1183 (1987).

27. Campbell e Gowri, Plant Physiol., 92:1 (1990).

28. Campbell, W. C., ed. Ivermectin and Abamectin, Springer-Verlag, New York, 1989.

29. Chee et al. Plant Physiol., 91:1212 (1989).

30. Chen e Winans (1991) J. Bacteriol. 173: 1139-1144

31. Christou et al. Proc. Natl. Acad. Sci USA, 86:7500 (1989).

32. Christou et al., Biotechnology, 9:957 (1991).

33. Christou et al., Plant Physiol., 87:671 (1988).

34. Chui et al. (1996) Curr Biol 6:325-330

35. Coe et al., In: Corn and Corn Improvement, Sprague et al. (eds.) pp. 81-258 (1988).

36. Corpet et al. Nucleic Acids Res., 16:10881 (1988).

37. Coxson et al., Biotropica, 24:121 (1992).

38. Crameri et al., Nature Biotech., 15:436 (1997).

39. Crameri et al., Nature, 391:288 (1998).

40. Crossway et al., BioTechniques, 4:320 (1986).

41. Cuozzo et al., Bio/Technology, 6:549 (1988).

42. Cutler et al., J. Plant Physiol., 135:351 (1989).

43. Czapla e Lang, J. Econ. Entomol., 83:2480 (1990).

44. Datta et al., Bio/Technology, 8:736 (1990).

45. Davies et al., Plant Physiol., 93:588 (1990).

46. Dayhoff et al., Atlas of Protein Sequence and Structure, Natl. Biomed. Res. Found., Washington, C. D. (1978).

47. De Blaere et al., Meth. Enzymol., 143:277 (1987).

48. De Block et al. Plant Physiol., 91:694 (1989).

49. De Block et al., EMBO Journal, 6:2513 (1987).

50. Deblaere et al. Nucl Acids Res 13:4777-4788 (1985)

51. Della-Cioppa et al. Bio/Technology 5:579-584 (1987)

52. Della-Cioppa et al., Plant Physiology, 84:965-968 (1987).

53. Dellaporta et al., in Chromosome Structure and Function, Plenum Press, 263-282 (1988).

54. Depicker et al., Plant Cell Reports, 7:63 (1988).

55. Dunn et al., Can. J. Plant Sci., 61:583 (1981).

56. Dure et al., Plant Mol. Biol., 12:475 (1989).

57. Ebinuma et al. Proc Natl Acad Sci USA 94:2117-2121 (2000a)

58. Ebinuma et al. Selection of Marker-free transgenic plants using the oncogenes (*ipt, rol* A, B, C) of *Agrobacterium* as selectable markers, In Molecular Biology of Woody Plants. Kluwer Academic Publishers (2000b)

59. Eichholtz et al. Somatic Cell and Molecular Genetics 13, 67-76 (1987)

60. Ellis et al., EMBO Journal, 6:3203 (1987).

61. Elroy-Stein et al., Proc. Natl. Acad. Sci. U.S.A., 86:6126 (1989).

62. English et al., Plant Cell, 8:179 (1996).

63. Erdmann et al., J. Gen. Microbiol., 138:363 (1992).

64. Erikson et al. Nat Biotechnol. 22(4):455-8 (2004)

65. Everett et al., Bio/Technology, 5:1201(1987).

66. Fedoroff NV & Smith DL Plant J 3:273- 289 (1993)

67. Fire A et al Nature 391:806-811 (1998)

68. Fitzpatrick, Gen. Engineering News, 22:7 (1993).

69. Fraley et al. Proc Natl Acad Sci USA 80:4803 (1983)

70. Fromm et al., Bio/Technology, 8:833 (1990).

71. Fromm et al., Nature (London), 319:791 (1986).

72. Galbiati et al. Funct. Integr Genozides 2000, 20 1:25-34

73. Gallie et al. Nucl Acids Res 15:8693-8711 (1987)

74. Gallie et al., Nucleic Acids Res., 15:3257 (1987).

75. Gallie et al., The Plant Cell, 1:301 (1989).

76. Gan et al., Science, 270:1986 (1995).

77. Gatehouse et al., J. Sci. Food Agric., 35:373 (1984).

78. Gelfand, eds., PCR Strategies Academic Press, New York (1995).

79. Gelvin et al., Plant Molecular Biology Manual, (1990).

80. Gleave et al. Plant Mol Biol. 40(2):223-35 (1999)

81. Gordon-Kamm et al., Plant Cell, 2:603 (1990).

82. Goring et al, PNAS, 88:1770 (1991).

83. Gruber, et al., Vectors for Plant Transformation, in: Methods in Plant Molecular Biology & Biotechnology" in Glich et al., (Eds. pp. 89-119, CRC Press, 1993).

84. Guerineau et al., Mol. Gen. Genet., 262:141 (1991).

85. Guerrero et al., Plant Mol. Biol., 15:11 (1990).

86. Gupta et al., PNAS, 90:1629 (1993).

87. Haines e Higgins (eds.), Nucleic Acid Hybridization, IRL Press, Oxford, U.K.

88. Hajdukiewicz et al. Plant Mol Biol 25:989-994 (1994)

89. Hammock et al., Nature, 344:458 (1990).

90. Hansen et al. Proc. Natl. Acad. Sci. USA 91:7603-7607 (1994)

91. Hayford et al. Plant Physiol. 86:1216 (1988)

92. Hemenway et al., EMBO Journal, 7:1273 (1988).

93. Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA, 89:10915 (1989).

94. Hiei et al. Plant J 6: 271-282 (1994)

95. Higgins et al., Gene, 73:237 (1988).

96. Higo et al. (1999) Nucl Acids Res 27(1): 297-300

97. Hilder et al., Nature, 330:160 (1987).

98. Hille et al. Plant Mol. Biol. 7:171 (1986)

99. Hinchee et al. Bio/Technology 6:915 (1988).

100. Hoekema *et al.* (1983) Nature 303:179-181

101. Hoekema, In: The Binary Plant Vector System. Offset-drukkerij Kanters B.V.; Alblasserdam (1985).

102. Hood et al. J Bacteriol 168:1291-1301 (1986)

103. Huang et al., CABIOS, 8:155 (1992).

104. Ikeda et al., J. Bacteriol., 169:5612 (1987).

105. Ikuta et al., Biotech., 8:241 (1990).

106. Ingelbrecht et al., Plant Cell, 1:671 (1989).

107. Innis and Gelfand, eds., PCR Methods Manual (Academic Press, New York) (1999).

108. Innis et al., eds., PCR Protocols: A Guide to Methods and Applications (Academic Press, New York (1995).

109. Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., San Diego, Calif. (1990).

110. Ishida Y et al. Nature Biotech 745-750 (1996)

111. Jefferson et al. EMBO J 6:3901-3907 (1987)

112. Jefferson et al. Plant Mol Biol Rep 5:387-405 (1987)

113. Jenes B 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)

114. Jobling et al., Nature, 325:622 (1987).

115. Johnson et al., PNAS USA, 86:9871 (1989)

116. Jones et al. Mol. Gen. Genet., 210:86 (1987)

117. Joshi et al., Nucleic Acid Res., 15:9627 (1987).

118. Kaasen et al., J. Bacteriol., 174:889 (1992).

119. Karlin and Altschul, Proc. Natl. Acad Sci. USA, 87:2264 (1990).

120. Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 90:5873 (1993).

121. Karsten et al., Botanica Marina, 35:11 (1992).

122. Katz et al., J. Gen. Microbiol., 129:2703 (1983).

123. Keller et al., EMBO Journal, 8:1309 (1989).

124. Keller et al., Genes Dev., 3:1639 (1989). (Eine der beiden « Keller » Referenzen ist zuviel)

125. Klapwijk et al. J. Bacteriol., 141,128-136 (1980)

126. Klein et al., Bio/Technoloy, 6:559 (1988).

127. Klein et al., Plant Physiol., 91:440 (1988).

128. Klein et al., Proc. Natl. Acad. Sci. USA, 85:4305 (1988).

129. 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.

130. Koncz & Schell Mol Gen Genet 204:383-396 (1986)

131. Koprek T et al. Plant J 19(6): 719-726 (1999)

132. Koster e Leopold, Plant Physiol., 88:829 (1988).

133. Koziel et al., Biotechnology, 11:194 (1993).

134. Kunkel et al., Methods in Enzymol., 154:367 (1987).

135. Kunkel, Proc. Natl. Acad. Sci. USA, 82:488 (1985).

136. Lam E und Chua NH, J Biol Chem; 266(26):17131-17135 (1991)

137. Laufs et al., PNAS, 87:7752 (1990).

138. Lawton et al., Mol. Cell Biol., 7:335 (1987).

139. Lee and Saier, J. Bacteriol., 153 (1982).

140. Leffel et al. Biotechniques 23(5):912-8 (1997)

141. Lescot et al. Nucleic Acids Res 30(1):325-7 (2002)

142. Levings, Science, 250:942 (1990).

143. Li et al. Plant Mol Biol 20:1037-1048 (1992)

144. Lindsey et al., Transgenic Research, 2:3347 (1993).

145. Liu et al.,. Plant J. 8, 457-463 (1995)

146. Lommel et al., Virology, 181:382 (1991).

147. Loomis et al., J. Expt. Zool., 252:9 (1989).

148. Lorz et al., Mol. Gen. Genet., 199:178 (1985).

149. Ma et al., Nature, 334:631 (1988).

150. Macejak et al., Nature, 353:90 (1991).

151. Maki et al., Methods in Plant Molecular Biology & Biotechnology, Glich et al., 67-88 CRC Press, (1993).

152. Maniatis T, Fritsch EF, and Sambrook J Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY), (1989)

153. Mariani et al, Nature, 347:737 (1990).

154. Matzke et al. (2000) Plant Mol Biol 43:401-415;

155. McBride et al., PNAS USA, 91:7301 (1994).

156. McCabe et al., Bio/Technology, 6:923 (1988).

157. Meinkoth e Wahl, Anal. Biochem., 138:267 (1984).

158. Messing and Vierra, Gene, 19:259 (1982).

159. Michael et al., J. Mol. Biol., 26:585 (1990). (im Text steht: Michael et al. 1994)

160. Millar et al. Plant Mol Biol Rep 10:324-414 (1992)

161. Mogen et al., Plant Cell, 2:1261 (1990).

162. Moore et al., J. Mol. Biol., 272:336 (1997).

163. Mozo & Hooykaas Plant Mol. Biol. 16:917-918

(1991)

164. Mundy and Chua, EMBO J., 7:2279 (1988).

165. Munroe et al., Gene, 91:151 (1990).

166. Murakami et al., Mol. Gen. Genet., 205:42 (1986).

167. Murata et al., FEBS Lett., 296:187 (1992).

168. Murdock et al., Phytochemistry, 29:85 (1990).

169. Murray et al., Nucleic Acids Res., 17:477 (1989).

170. Myers and Miller, CABIOS, 4:11 (1988).

171. Naested H Plant J 18:571-576 (1999)

172. Napoli et al., Plant Cell, 2:279 (1990).

173. Needleman and Wunsch, J. Mol. Biol., 48:443-453 (1970).

174. Nehra et al. Plant J. 5:285-297 (1994)

175. Niedz et al., Plant Cell Reports, 14:403 (1995).

176. Odell et al., Mol. Gen. Genet., 113:369 (1990).

177. Odell et al., Nature, 313:810 (1985).

178. Ohtsuka et al., J. Biol. Chem., 260:2605 (1985)

179. Olhoft et al. Plant Cell Rep 20: 706–711 (2001)

180. Ow et al., Science, 234:856 (1986).

181. Pacciotti et al., Bio/Technology, 3:241 (1985).

182. Park et al., J. Plant Biol., 38:365 (1985).

183. Paszkowski et al., EMBO J., 3:2717 (1984).

184. Pearson and Lipman, Proc. Natl. Acad. Sci., 85:2444 (1988).

185. Pearson et al., Meth. Mol. Biol., 24:307 (1994).

186. Perera RJ et al. Plant Mol. Biol 23(4): 793-799 (1993)

187. Perlak et al., Proc. Natl. Acad. Sci. USA, 88:3324 (1991).

188. Phillips et al., In Corn & Corn Improvement, 3rd Edition 10 Sprague et al. (Eds. pp. 345-387)(1988).

189. Phi-Van et al., Mol. Cell. Biol., 10:2302 (1990).

190. Piatkowski et al., Plant Physiol., 94:1682 (1990).

191. Potrykus et al., Mol. Gen. Genet., 199:183 (1985).

192. Potrykus, Trends Biotech., 7:269 (1989).

193. Prasher et al., Biochem. Biophys. Res. Comm., 126:1259 (1985).

194. Proudfoot, Cell, 64:671 (1991).

195. Reed et al., J. Gen. Microbiol., 130:1 (1984).

196. Riggs et al., Proc. Natl. Acad. Sci. USA, 83:5602 (1986).

197. Rossolini et al., Mol. Cell. Probes, 8:91 (1994).

198. Ruiz, Plant Cell, 10:937 (1998).

199. Sambrook et al., Molecular Cloning: A Laboratory Manual (2d ed., Cold Spring Harbor Laboratory Press, Plainview, N.Y.) (1989).

200. Sanfacon et al., Genes Dev., 5:141 (1991).

201. Sanford et al., Particulate Science and Technology, 5:27 (1987).

202. Scheeren-Groot et al. J. Bacteriol 176: 6418-6426 (1994)

203. Schenborn e Groskreutz Mol Biotechnol 13(1): 29-44 (1999)

204. Schlaman e Hooykaas Plant J 11:1377-1385 (1997)

205. Schoffl F et al. (1989) Mol Gen Genetics 217(2-3):246-53

206. Shagan et al., Plant Physiol., 101:1397 (1993).

207. Shah et al. Science 233: 478 (1986)

208. Shapiro, Mobile Genetic Elements, Academic Press, N.Y. (1983).

209. Shimamoto et al., Nature, 338:274 (1989).

210. Silhavy TJ, Berman ML, and Enquist LW Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), (1984)

211. Skuzeski et al., Plant Molec. Biol. 15: 65-79 (1990).

212. Smith et al., Adv. Appl. Math., 2:482.(1981).

213. Smith et al., Mol. Gen. Genet., 224:447 (1990).

214. Spencer et al., Theor. Appl. Genet, 79:625 (1990). Spencer 1992 Referenz fehlt

215. Stalker et al., Science, 242:419 (1988).

216. Staub et al., EMBO J., 12:601 (1993).

217. Staub et al., Plant Cell, 4:39 (1992).

218. Steifel et al., The Plant Cell, 2:785 (1990).

219. Stemmer, Nature, 370:389 (1994).

220. Stemmer, Proc. Natl. Acad. Sci. USA, 91:10747 (1994).

221. Stief et al., Nature, 341:343 (1989).

222. Stougaard Plant J 3:755-761 (1993)

223. Sukhapinda et al., Plant Mol. Biol., 8:209 (1987).

224. Sundaresan et al. Gene Develop 9: 1797-1810 (1995)

225. Sutcliffe, PNAS USA, 75:3737 (1978).

226. Svab et al., Plant Mol. Biol. 14:197 (1990)

227. Svab et al., Proc. Natl. Acad. Sci. USA, 87:8526 (1990).

228. Svab et al., Proc. Natl. Acad. Sci. USA, 90:913 (1993).

229. Tarczynski et al., PNAS USA, 89:2600 (1992).

230. Thillet et al., J. Biol. Chem., 263:12500 (1988).

231. Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, Elsevier, N.Y. (1993).

232. Tomes et al., Plant Cell, Tissue and Organ Culture: Fundamental Methods, Springer Verlag, Berlin (1995).

233. Tomic et al., NAR, 12:1656 (1990).

234. Thompson JD et al., NAR 22(22):4673-4680 (1994).

235. Turner et al., Molecular Biotechnology, 3:225 (1995).

236. Twell et al., Plant Physiol., 91:1270 (1989).

237. Ugaki et al., Nucl. Acids Res., 19:371 (1991).

238. Ulmasov et al., Plant Mol. Biol., 35:417 (1997).

239. Upender et al., Biotechniques, 18:29 (1995).

240. van der Krol et al., Plant Cell, 2:291 (1990).

241. Vanden Elzen et al. Plant Mol Biol. 5:299 (1985)

242. Vasil et al. Bio/Technology, 10:667-674 (1992)

243. Vasil et al. Bio/Technology, 11:1153-1158 (1993)

244. Vasil et al., Mol. Microbiol., 3:371 (1989).

245. Vasil et al., Plant Physiol., 91:1575 (1989).

246. Vernon e Bohnert, EMBO J., 11:2077 (1992).

247. Walker e Gaastra, eds., Techniques in Molecular Biology, MacMillan Publishing Company, New York (1983).

248. Wan & Lemaux (1994) Plant Physiol., 104:3748

249. Wang et al., Mol. Cell. Biol., 12:3399 (1992).

250. Waterman, M. S. Introduction to Computational Biology: Maps, sequences and genomes. Chapman & Hall. London (1995).

251. Watrud et al., in Engineered Organisms and the Environment (1985).

252. Watson et al. J. Bacteriol 123, 255-264 (1975)

253. Watson et al., Corn: Chemistry and Technology (1987).

254. Weeks et al. Plant Physiol 102:1077-1084 (1993)

255. Weissinger et al., Annual Rev. Genet., 22:421 (1988).

256. White et al, Nucl Acids Res, 18, 1062 (1990).

257. Wingender E et al. Nucleic Acids Res 29(1):281-3 (2001)

258. Wolter et al., EMBO Journal, 11:4685 (1992).

259. Wyn-Jones and Storey, Physiology and Biochemistry of Drought Resistance in Plants, Paleg et al. (eds.), pp. 171-204 (1981).

260. Yamaguchi-Shinozaki et al., Plant Cell Physiol., 33:217 (1992).

261. Zhang et al., Proc. Natl. Acad. Sci. USA, 94:4504 (1997).

262. Zukowsky et al., PNAS USA, 80:1101 (1983).

Todas as publicações, patentes e pedidos de patente são aqui incorporados por referência. Apesar do relatório acima desta invenção ter sido descrito 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 arte que a invenção é susceptível de formas de realização adicionais e que alguns dos detalhes aqui descritos podem ser variados consideravelmente sem sair dos princípios básicos da invenção.

LISTAGEM DE SEQÜÊNCIA

<110> BASF Plant Science GmbH
<120> Cassetes de expressão para expressão preferencial de semente em plantas
<130> AE20050148 / PF56382-AT
<160> 127
<170> PatentIn versão 3.3
<210> 1
<211> 1683
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> promotor
<222> (1)..(1683)
<223> seqüência nucleotídica reguladora da transcrição de gene de Arabidopsis
thaliana At1g48130 (peroxiredoxina) compreendendo
região não traduzida5'

<220>
<221> misc_aspecto
<222> (1503)..(1508)
<223> motivos RY e Sph conservados em promotores específicos de semente (
Filamento +) ctcttttgCATGcacctcaatttgaac

<220>
<221> sinal TATA
<222> (1586)..(1593)
<223> caixa TATA putativa

<220>
<221> 5'UTR
<222> (1620)..(1683)
<223> região não traduzida 5' putativa

<400> 1
gccacatcat gtttagactt atctccataa agaaaaccac tcatcaaagc taatacaaaa 60
gctctagtgt gacacttcaa cgtcacttca tcaggatcag caggtaaatt ccgaaaattc 120
tcacgcagcc acgctaagga cacatgagaa ccatgaagat ccttgggacc tggcctatga 180
ccaagcaaat cctcacataa atcagcccag ttgtattttg tactaccagt tactgcaggt 240
ccatcgacac gtagacccaa caaaatattc acatcttgta aagtcacagt gatctctcca 300
gcaggaagat gaaaagtatg cgtttcgggt ctccatctct ccaccaaagc tgttatcaga 360
gcataatcaa gttgtataaa ggcaaccttg taaactccat atagaccaaa ctctatcaac 420
ttttgacaca cgagaggatc cagaggccaa tctcgcatcc ccaataactt gtgccgacat 480
gtcagttcac gaggaggaac ctgaatgtga agtataacgg taaaaaggaa ataattaaaa 540
caacggaagc aaaacaagaa acaagatgaa atgagaaact agtaacacac ctcatcttcc 600
catatagcag ctgatctatg ctcatgttgc cacaccaata tagattgatc aactggacca 660
ggatccaaat caaagtttaa tagactttgc acctccatct atataatata tcacaggaca 720
ataaacacaa tgatcagtga ttatacaaca tcaaagaaaa cttgtaattc tgggaatata 780
actgagaaat gagaattaaa gattcataat ttgaacacaa gaaatcctaa actggtacga 840
aagaaaaatt gctcaacaaa aaaatcaagc taattactcg tatacaaaga cacgaagaac 900
taatacaaga aacaagaaac aacaaaccac aaagagattg aaccaatcca aattcgacaa 960
cataaaccaa gtgtgtgggt gattggaatc agaggacgta ccaaagaaaa gcgtccgtga 1020
tcaacaaaaa ccaaaaagag acgtttgaaa taaaccagag gaagacgaag aataattaag 1080
caaagagaag cgttaagcgg gagcgagaaa ggaaacgaga gaaagagaga gcttccagat 1140
ccgacagaag ttttcggctt cttctttttc gtttaagaac ttctgatcct cctaggtctg 1200
tccgaagaac taatcttttt gaggtaacga cgccgttttt ctcaaaacat gggcccatta 1260
accatagtct cggcccaaac gaaacttaat acgacaatgt ttgggtgtaa acgcaaagat 1320
tttgtcgatt atcacaagta aaaaaataaa tacaaacact tgagtctctc tagacatcgt 1380
gcgtcgcctt agctttaagt tttttctcga aacaaaagag ttattttatt tgaactttga 1440

agattatacg aagacacgtg gcgtgaaccc aattcataac aacgccacgc tatactcttt 1500
tgcatgcacc tcaatttgaa catcatcaag tctctctctc tctttttctg actttgatcc 1560
acgaacctaa ccagcttgcg atctctattt aatcggtcct cgacgcaact tcaacttcta 1620
ctacatccat tcacatcaaa tcaatacaga aagtttttc tatatataaa tataaaaggt 1680
aaa 1683

<210> 2
<211> 1619
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> promotor
<222> (1)..(1619)
<223> seqüência nucleotídica reguladora da transcrição de gene de Arabidopsis
thaliana At1g48130 (peroxiredoxina)

<220>
<221> sinal TATA
<222> (1586)..(1593)
<223> caixa TATA putativa

<400> 2
gccacatcat gtttagactt atctccataa agaaaaccac tcatcaaagc taatacaaaa 60
gctctagtgt gacacttcaa cgtcacttca tcaggatcag caggtaaatt ccgaaaattc 120
tcacgcagcc acgctaagga cacatgagaa ccatgaagat ccttgggacc tggcctatga 180
ccaagcaaat cctcacataa atcagcccag ttgtattttg tactaccagt tactgcaggt 240
ccatcgacac gtagacccaa caaaatattc acatcttgta aagtcacagt gatctctcca 300
gcaggaagat gaaaagtatg cgtttcgggt ctccatctct ccaccaaagc tgttatcaga 360
gcataatcaa gttgtataaa ggcaaccttg taaactccat atagaccaaa ctctatcaac 420
ttttgacaca cgagaggatc cagaggccaa tctcgcatcc ccaataactt gtgccgacat 480
gtcagttcac gaggaggaac ctgaatgtga agtataacgg taaaaaggaa ataattaaaa 540
caacggaagc aaaacaagaa acaagatgaa atgagaaact agtaacacac ctcatcttcc 600
catatagcag ctgatctatg ctcatgttgc cacaccaata tagattgatc aactggacca 660
ggatccaaat caaagtttaa tagactttgc acctccatct atataatata tcacaggaca 720
ataaacacaa tgatcagtga ttatacaaca tcaaagaaaa cttgtaattc tgggaatata 780
actgagaaat gagaattaaa gattcataat ttgaacacaa gaaatcctaa actggtacga 840
aagaaaaatt gctcaacaaa aaaatcaagc taattactcg tatacaaaga cacgaagaac 900
taatacaaga aacaagaaac aacaaaccac aaagagattg aaccaatcca aattcgacaa 960
cataaaccaa gtgtgtgggt gattggaatc agaggacgta ccaaagaaaa gcgtccgtga 1020
tcaacaaaaa ccaaaaagag acgtttgaaa taaaccagag gaagacgaag aataattaag 1080
caaagagaag cgttaagcgg gagcgagaaa ggaaacgaga gaaagagaga gcttccagat 1140
ccgacagaag ttttcggctt cttctttttc gtttaagaac ttctgatcct cctaggtctg 1200
tccgaagaac taatcttttt gaggtaacga cgccgttttt ctcaaaacat gggcccatta 1260
accatagtct cggcccaaac gaaacttaat acgacaatgt ttgggtgtaa acgcaaagat 1320
tttgtcgatt atcacaagta aaaaaataaa tacaaacact tgagtctctc tagacatcgt 1380
gcgtcgcctt agctttaagt tttttctcga aacaaaagag ttattttatt tgaactttga 1440
agattatacg aagacacgtg gcgtgaaccc aattcataac aacgccacgc tatactcttt 1500
tgcatgcacc tcaatttgaa catcatcaag tctctctctc tctttttctg actttgatcc 1560
acgaacctaa ccagcttgcg atctctattt aatcggtcct cgacgcaact tcaacttct 1619

<210> 3
<211> 985
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> promotor
<222> (1)..(985)
<223> seqüência nucleotídica reguladora da transcrição (região de promotor de núcleo putativo) de gene de Arabidopsis thaliana At1g48130

(peroxiredoxina) compreendendo região não traduzida 5'

<220>
<221> sinal TATA
<222> (888)..(895)
<223> caixa TATA putativa

<220>
<221> 5'UTR
<222> (922)..(985)
<223> região não traduzida 5' putativa

<400> 3
ctatataata tatcacagga caataaacac aatgatcagt gattatacaa catcaaagaa 60
aacttgtaat tctgggaata taactgagaa atgagaatta aagattcata atttgaacac 120
aagaaatcct aaactggtac gaaagaaaaa ttgctcaaca aaaaaatcaa gctaattact 180
cgtatacaaa gacacgaaga actaatacaa gaaacaagaa acaacaaacc acaaagagat 240
tgaaccaatc caaattcgac aacataaacc aagtgtgtgg gtgattggaa tcagaggacg 300
taccaaagaa aagcgtccgt gatcaacaaa aaccaaaaag agacgtttga aataaaccag 360
aggaagacga agaataatta agcaaagaga agcgttaagc gggagcgaga aaggaaacga 420
gagaaagaga gagcttccag atccgacaga agttttcggc ttcttctttt tcgtttaaga 480
acttctgatc ctcctaggtc tgtccgaaga actaatcttt ttgaggtaac gacgccgttt 540
ttctcaaaac atgggcccat taaccatagt ctcggcccaa acgaaactta atacgacaat 600
gtttgggtgt aaacgcaaag attttgtcga ttatcacaag taaaaaaata aatacaaaca 660
cttgagtctc tctagacatc gtgcgtcgcc ttagctttaa gtttttttctc gaaacaaaag 720
agttatttta tttgaacttt gaagattata cgaagacacg tggcgtgaac ccaattcata 780
acaacgccac gctatactct tttgcatgca cctcaatttg aacatcatca agtctctctc 840
tctcttttc tgactttgat ccacgaacct aaccagcttg cgatctctat ttaatcggtc 900
ctcgacgcaa cttcaacttc tactacatcc attcacatca aatcaataca gaaagttttt 960
tctatatata aatataaaag gtaaa 985

<210> 4
<211> 921
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> promotor
<222> (1)..(921)
<223> seqüência nucleotídica reguladora da transcrição (região de promotor de núcleo putativo) de gene de Arabidopsis thaliana At1g48130
(peroxiredoxina)

<220>
<221> sinal TATA
<222> (888)..(895)
<223> caixa TATA putativa

<400> 4
ctatataata tatcacagga caataaacac aatgatcagt gattatacaa catcaaagaa 60
aacttgtaat tctgggaata taactgagaa atgagaatta aagattcata atttgaacac 120
aagaaatcct aaactggtac gaaagaaaaa ttgctcaaca aaaaaatcaa gctaattact 180
cgtatacaaa gacacgaaga actaatacaa gaaacaagaa acaacaaacc acaaagagat 240
tgaaccaatc caaattcgac aacataaacc aagtgtgtgg gtgattggaa tcagaggacg 300
taccaaagaa aagcgtccgt gatcaacaaa aaccaaaaag agacgtttga aataaaccag 360
aggaagacga agaataatta agcaaagaga agcgttaagc gggagcgaga aaggaaacga 420
gagaaagaga gagcttccag atccgacaga agttttcggc ttcttctttt tcgtttaaga 480
acttctgatc ctcctaggtc tgtccgaaga actaatcttt ttgaggtaac gacgccgttt 540
ttctcaaaac atgggcccat taaccatagt ctcggcccaa acgaaactta atacgacaat 600
gtttgggtgt aaacgcaaag attttgtcga ttatcacaag taaaaaaata aatacaaaca 660
cttgagtctc tctagacatc gtgcgtcgcc ttagctttaa gtttttttctc gaaacaaaag 720
agttatttta tttgaacttt gaagattata cgaagacacg tggcgtgaac ccaattcata 780

acaacgccac gctatactct tttgcatgca cctcaatttg aacatcatca agtctctctc 840
tctctttttc tgactttgat ccacgaacct aaccagcttg cgatctctat ttaatcggtc 900
ctcgacgcaa cttcaacttc t 921

<210> 5
<211> 1696
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> promotor
<222> (1)..(1696)
<223> seqüência nucleotídica reguladora da transcrição (derivado) de gene de Arabidopsis thaliana gene At1g48130 (peroxiredoxina) compreendendo região não traduzida 5'

<220>
<221> sinal TATA
<222> (1595)..(1602)
<223> caixa TATA putativa

<220>
<221> 5'UTR
<222> (1633)..(1696)
<223> região não traduzida 5' putativa

<400> 5
gccacatcat gtttagactt atctccataa agaaaaccac tcatcaaagc taatacaaaa 60
gctctagtgt gacacttcaa cgtcacttca tcaggatcag caggtaaatt ccgaaaattc 120
tcacgcagcc acgctaagga cacatgagaa ccatgaagat ccttgggacc tggcctatga 180
ccaagcaaat cctcacataa atcagcccag ttgtattttg tactaccagt tactgcaggt 240
ccatcgacac gtagacccaa caaaatattc acatcttgta aagtcacagt gatctctcca 300
gcaggaagat gaaaagtatg cgtttcgggt ctccatctct ccaccaaagc tgttatcaga 360
gcataatcaa gttgtataaa ggcaaccttg taaactccat atagaccaaa ctctatcaac 420
ttttgacaca cgagaggatc cagaggccaa tctcgcatcc ccaataactt gtgccgacat 480
gtcagttcac gaggaggaac ctgaatgtga agtataacgg taaaaaggaa atatatatta 540
aaacaacgga agcaaaacaa gaaacaagat gaaatgagaa actagtaaca cacctcataa 600
tcttcccata tagcagctga tctatgctca tgttgccaca ccaatataga ttgatcaacc 660
attggaccag gatccaaatc aaagtttaat agactttgca cctccatcta tataatatat 720
cacaggacaa taaacacaat gatcagtgat tatacaacat caaagaaaac ttgtaattct 780
gggaatataa ctgagaaatg agaattaaag attcataatt tgaacacaag aaatcctaaa 840
ctggtacgaa agaaaaattg ctcaacaaaa aaatcaagct aattactcgt atacaaagac 900
acgaagaact aatacaagaa acaagaaaca acaaaccaca aagagattga accaatccaa 960
attcgacaac ataaaccaag tgtgtgggtg attggaatca gaggacgtac caaagaaaag 1020
cgtccgtgat caacaaaaac caaaaagaga cgtttgaaat aaaccagagg aagacgaaga 1080
ataattaagc aaagagaagc gttaagcggg agcgagaaag gaaacgagag aaagagagag 1140
cttccagatc cgacagaagt tttcggcttc ttctttttcg tttaagaact tctgatcctc 1200
ctaggtctgt ccgaagaact aatctttttg aggtaacgac gccgtttttc tcaaaacatg 1260
ggcccattaa ccatagtctc ggcccaaacg aaacttaata cgacaatgtt tgggtgtaaa 1320
cgcaaagatt ttgtcgatta tcacaagtaa aaaaataaat acaaacactt gagtctctct 1380
agacatcgtg cgtcgcctta gctttaagtt ttttctcgaa acaaaagagt tattttattt 1440
gaactttgaa gattatacga agacacgtgg cgtgaaccca attcataaca acgccacgct 1500
atactctttt gcatgcacct caatttgaac atcatcaagt ctctctctct ctttttctga 1560
ctttgatcca cgaacctaac cagcttgcga tctctattta atcggtcctc gacgcaactt 1620
catataactt ctactacatc cattcacatc aaatcaatac agaaagtttt ttctatatat 1680
aaatataaaa ggtaaa 1696

<210> 6
<211> 1632
<212> DNA
<213> Arabidopsis thaliana

<220>

<221> promotor
<222> (1)..(1632)
<223> seqüência nucleotídica reguladora da transcrição (derivado) de gene de Arabidopsis thaliana At1g48130 (peroxiredoxina)

<220>
<221> sinal TATA
<222> (1595)..(1602)
<223> caixa TATA putativa

<400> 6
gccacatcat gtttagactt atctccataa agaaaaccac tcatcaaagc taatacaaaa 60
gctctagtgt gacacttcaa cgtcacttca tcaggatcag caggtaaatt ccgaaaattc 120
tcacgcagcc acgctaagga cacatgagaa ccatgaagat ccttgggacc tggcctatga 180
ccaagcaaat cctcacataa atcagcccag ttgtattttg tactaccagt tactgcaggt 240
ccatcgacac gtagacccaa caaaatattc acatcttgta aagtcacagt gatctctcca 300
gcaggaagat gaaaagtatg cgtttcgggt ctccatctct ccaccaaagc tgttatcaga 360
gcataatcaa gttgtataaa ggcaaccttg taaactccat atagaccaaa ctctatcaac 420
ttttgacaca cgagaggatc cagaggccaa tctcgcatcc ccaataactt gtgccgacat 480
gtcagttcac gaggaggaac ctgaatgtga agtataacgg taaaaaggaa atatatatta 540
aaacaacgga agcaaaacaa gaaacaagat gaaatgagaa actagtaaca cacctcataa 600
tcttcccata tagcagctga tctatgctca tgttgccaca ccaatataga ttgatcaacc 660
attggaccag gatccaaatc aaagtttaat agactttgca cctccatcta tataatatat 720
cacaggacaa taaacacaat gatcagtgat tatacaacat caaagaaaac ttgtaattct 780
gggaatataa ctgagaaatg agaattaaag attcataatt tgaacacaag aaatcctaaa 840
ctggtacgaa agaaaaattg ctcaacaaaa aaatcaagct aattactcgt atacaaagac 900
acgaagaact aatacaagaa acaagaaaca acaaaccaca aagagattga accaatccaa 960
attcgacaac ataaaccaag tgtgtgggtg attggaatca gaggacgtac caaagaaaag 1020
cgtccgtgat caacaaaaac caaaaagaga cgtttgaaat aaaccagagg aagacgaaga 1080
ataattaagc aaagagaagc gttaagcggg agcgagaaag gaaacgagag aaagagagag 1140
cttccagatc cgacagaagt tttcggcttc ttctttttcg tttaagaact tctgatcctc 1200
ctaggtctgt ccgaagaact aatctttttg aggtaacgac gccgtttttc tcaaaacatg 1260
ggcccattaa ccatagtctc ggcccaaacg aaacttaata cgacaatgtt tgggtgtaaa 1320
cgcaaagatt ttgtcgatta tcacaagtaa aaaaataaat acaaacactt gagtctctct 1380
agacatcgtg cgtcgcctta gctttaagtt ttttctcgaa acaaaagagt tattttattt 1440
gaactttgaa gattatacga agacacgtgg cgtgaaccca attcataaca acgccacgct 1500
atactctttt gcatgcacct caatttgaac atcatcaagt ctctctctct ctttttctga 1560
ctttgatcca cgaacctaac cagcttgcga tctctattta atcggtcctc gacgcaactt 1620
catataactt ct 1632

<210> 7
<211> 989
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> promotor
<222> (1)..(989)
<223> seqüência nucleotídica reguladora da transcrição (derivado; região de promotor de núcleo putativo) de gene de Arabidopsis thaliana At1g48130 (peroxiredoxina) compreendendo região não traduzida 5'

<220>
<221> sinal TATA
<222> (888)..(895)
<223> caixa TATA putativa

<220>
<221> 5'UTR
<222> (926)..(989)
<223> região não traduzida 5' putativa

<400> 7
ctatataata tatcacagga caataaacac aatgatcagt gattatacaa catcaaagaa 60
aacttgtaat tctgggaata taactgagaa atgagaatta aagattcata atttgaacac 120
aagaaatcct aaactggtac gaaagaaaaa ttgctcaaca aaaaaatcaa gctaattact 180
cgtatacaaa gacacgaaga actaatacaa gaaacaagaa acaacaaacc acaaagagat 240
tgaaccaatc caaattcgac aacataaacc aagtgtgtgg gtgattggaa tcagaggacg 300
taccaaagaa aagcgtccgt gatcaacaaa aaccaaaaag agacgtttga aataaaccag 360
aggaagacga agaataatta agcaaagaga agcgttaagc gggagcgaga aaggaaacga 420
gagaaagaga gagcttccag atccgacaga agttttcggc ttcttctttt tcgtttaaga 480
acttctgatc ctcctaggtc tgtccgaaga actaatcttt ttgaggtaac gacgccgttt 540
ttctcaaaac atgggcccat taaccatagt ctcggcccaa acgaaactta atacgacaat 600
gtttgggtgt aaacgcaaag attttgtcga ttatcacaag taaaaaaata aatacaaaca 660
cttgagtctc tctagacatc gtgcgtcgcc ttagctttaa gttttttctc gaaacaaaag 720
agttatttta tttgaacttt gaagattata cgaagacacg tggcgtgaac ccaattcata 780
acaacgccac gctatactct tttgcatgca cctcaatttg aacatcatca agtctctctc 840
tctcttttc tgactttgat ccacgaacct aaccagcttg cgatctctat ttaatcggtc 900
ctcgacgcaa cttcatataa cttctactac atccattcac atcaaatcaa tacagaaagt 960
tttttctata tataaatata aaaggtaaa 989

<210> 8
<211> 925
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> promotor
<222> (1)..(925)
<223> seqüência nucleotídica reguladora da transcrição (derivado; (região de promotor de núcleo putativo) de gene de Arabidopsis thaliana
At1g48130 (peroxiredoxina)

<220>
<221> sinal TATA
<222> (888)..(895)
<223> caixa TATA putativa

<400> 8
ctatataata tatcacagga caataaacac aatgatcagt gattatacaa catcaaagaa 60
aacttgtaat tctgggaata taactgagaa atgagaatta aagattcata atttgaacac 120
aagaaatcct aaactggtac gaaagaaaaa ttgctcaaca aaaaaatcaa gctaattact 180
cgtatacaaa gacacgaaga actaatacaa gaaacaagaa acaacaaacc acaaagagat 240
tgaaccaatc caaattcgac aacataaacc aagtgtgtgg gtgattggaa tcagaggacg 300
taccaaagaa aagcgtccgt gatcaacaaa aaccaaaaag agacgtttga aataaaccag 360
aggaagacga agaataatta agcaaagaga agcgttaagc gggagcgaga aaggaaacga 420
gagaaagaga gagcttccag atccgacaga agttttcggc ttcttctttt tcgtttaaga 480
acttctgatc ctcctaggtc tgtccgaaga actaatcttt ttgaggtaac gacgccgttt 540
ttctcaaaac atgggcccat taaccatagt ctcggcccaa acgaaactta atacgacaat 600
gtttgggtgt aaacgcaaag attttgtcga ttatcacaag taaaaaaata aatacaaaca 660
cttgagtctc tctagacatc gtgcgtcgcc ttagctttaa gttttttctc gaaacaaaag 720
agttatttta tttgaacttt gaagattata cgaagacacg tggcgtgaac ccaattcata 780
acaacgccac gctatactct tttgcatgca cctcaatttg aacatcatca agtctctctc 840
tctcttttc tgactttgat ccacgaacct aaccagcttg cgatctctat ttaatcggtc 900
ctcgacgcaa cttcatataa cttct 925

<210> 9
<211> 1727
<212> DNA
<213> Linum usitatissimum

<220>
<221> promotor
<222> (1)..(1727)

<223> seqüência nucleotídica reguladora da transcrição (derivado) de gene ortólogo de Linum usitartissimum para gene de Arabidopsis thaliana
At1g48130 (peroxiredoxina) compreendendo região não traduzida 5'

<220>
<221> misc_aspecto
<222> (1282)..(1287)
<223> motivos RY e Sph conservados em promotores específicos de semente (filamento-)
cccaagtaCATGcaaaatttaacggtt

<220>
<221> misc_aspecto
<222> (1615)..(1620)
<223> RY e Sph motivos conservados em promotores específicos de semente (+
filamento) gttctcttCATGcaaggatagtagaac

<220>
<221> sinal TATA
<222> (1652)..(1656)
<223> caixa TATA putativa

<400> 9
cacgggcagg acatagggac tactacaagc atagtatgct tcagacaaag agctaggaaa 60
gaactcttga tggaggttaa gagaaaaaag tgctagaggg gcatagtaat caaacttgtc 120
aaaaccgtca tcatgatgag ggatgacata atataaaaag ttgactaagg tcttggtagt 180
actctttgat tagtattata tattggtgag aacatgagtc aagaggagac aagaaaccga 240
ggaaccatag tttagcaaca agatggaagt tgcaaagttg agctagccgc tcgattagtt 300
acatctccta agcagtacta caaggaatgg tctctatact ttcatgttta gcacatggta 360
gtgcggattg acaagttaga aacagtgctt aggagacaaa gagtcagtaa aggtattgaa 420
agagtgaagt tgatgctcga caggtcagga gaagtccctc cgccagatgg tgactaccaa 480
ggggttggta tcagctgaga cccaaataag attcttcggt tgaaccagtg gttcgaccga 540
gactcttagg gtgggatttc actgtaagat ttgtgcattt tgttgaatat aaattgacaa 600
ttttttttat ttaattatag attatttaga atgaattaca tatttagttt ctaacaagga 660
tagcaatgga tgggtatggg tacaggttaa acatatctat tacccaccca tctagtcgtc 720
gggttttaca cgtacccacc cgtttacata aaccagaccg gaattttaaa ccgtacccgt 780
ccgttagcgg gtttcagatt tacccgttta atcgggtaaa acctgattac taaatatata 840
ttttttattt gataaacaaa acaaaaatgt taatattttc atattggatg caattttaag 900
aaacacatat tcataaattt ccatatttgt aggaaaataa aaagaaaaat atattcaaga 960
acacaaattt caccgacatg acttttatta cagagttgga attagatcta acaattgaaa 1020
aattaaaatt aagatagaat atgttgagga acatgacata gtataatgct gggttacccg 1080
tcgggtaggt atcgaggcgg atactactaa atccatccca ctcgctatcc gataatcact 1140
ggtttcgggt atacccattc ccgtcaacag gccttttaa ccggataatt tcaacttata 1200
gtgaatgaat tttgaataaa tagttagaat accaaaatcc tggattgcat ttgcaatcaa 1260
attttgtgaa ccgttaaatt ttgcatgtac ttgggataga tataatagaa ccgaattttc 1320
attagtttaa tttataactt actttgttca aagaaaaaaa atatctatcc aatttactta 1380
taataaaaaa taatctatcc aagttactta ttataatcaa cttgtaaaaa ggtaagaata 1440
caaatgtggt agcgtacgtg tgattatatg tgacgaaatg ttatatctaa caaaagtcca 1500
aattcccatg gtaaaaaaaa tcaaaatgca tggcaggctg tttgtaacct tggaataaga 1560
tgttggccaa ttctggagcc gccacgtacg caagactcag ggccacgttc tcttcatgca 1620
aggatagtag aacaccactc cacccacctc ctatattaga cctttgccca accctcccca 1680
actttcccat cccatccaca aagaaaccga catttttatc ataaatc 1727

<210> 10
<211> 933
<212> DNA
<213> iniciador de oligonucleotídeo

<220>
<221> promotor
<222> (1)..(933)

<223> seqüência nucleotídica reguladora da transcrição (derivado; região de promotor de núcleo putativo) de gene ortólogo de Linum usitartissimum para gene de Arabidopsis thaliana At1g48130 (peroxiredoxina) compreendendo região não traduzida 5'

<220>
<221> sinal TATA
<222> (858)..(862)
<223> caixa TATA putativa

<400> 10
cagatttacc cgtttaatcg ggtaaaacct gattactaaa tatatatttt ttatttgata 60
aacaaaacaa aaatgttaat attttcatat tggatgcaat tttaagaaac acatattcat 120
aaatttccat atttgtagga aaataaaaag aaaaatatat tcaagaacac aaatttcacc 180
gacatgactt ttattacaga gttggaatta gatctaacaa ttgaaaaatt aaaattaaga 240
tagaatatgt tgaggaacat gacatagtat aatgctgggt tacccgtcgg gtaggtatcg 300
aggcggatac tactaaatcc atcccactcg ctatccgata atcactggtt tcgggtatac 360
ccattcccgt caacaggcct ttttaaccgg ataatttcaa cttatagtga atgaattttg 420
aataaatagt tagaatacca aaatcctgga ttgcatttgc aatcaaattt tgtgaaccgt 480
taaattttgc atgtacttgg gatagatata atagaaccga attttcatta gtttaattta 540
taacttactt tgttcaaaga aaaaaaatat ctatccaatt tacttataat aaaaaataat 600
ctatccaagt tacttattat aatcaacttg taaaaaggta agaatacaaa tgtggtagcg 660
tacgtgtgat tatatgtgac gaaatgttat atctaacaaa agtccaaatt cccatggtaa 720
aaaaaatcaa aatgcatggc aggctgtttg taaccttgga ataagatgtt ggccaattct 780
ggagccgcca cgtacgcaag actcagggcc acgttctctt catgcaagga tagtagaaca 840
ccactccacc cacctcctat attagacctt tgcccaaccc tccccaactt tcccatccca 900
tccacaaaga aaccgacatt tttatcataa atc 933

<210> 11
<211> 458
<212> DNA
<213> iniciador de oligonucleotídeo

<220>
<221> promotor
<222> (1)..(458)
<223> seqüência nucleotídica reguladora da transcrição (derivado; região de promotor de núcleo putativo) de gene ortólogo de Linum usitartissimum para gene de Arabidopsis thaliana At1g48130 (peroxiredoxina) compreendendo região não traduzida 5'

<220>
<221> sinal TATA
<222> (383)..(387)
<223> caixa TATA putativa

<400> 11
accgttaaat tttgcatgta cttgggatag atataataga accgaatttt cattagttta 60
atttataact tactttgttc aaagaaaaaa aatatctatc caatttactt ataataaaaa 120
ataatctatc caagttactt attataatca acttgtaaaa aggtaagaat acaaatgtgg 180
tagcgtacgt gtgattatat gtgacgaaat gttatatcta acaaaagtcc aaattcccat 240
ggtaaaaaaa atcaaaatgc atggcaggct gtttgtaacc ttggaataag atgttggcca 300
attctggagc cgccacgtac gcaagactca gggccacgtt ctcttcatgc aaggatagta 360
gaacaccact ccacccacct cctatattag acctttgccc aaccctcccc aactttccca 420
tcccatccac aaagaaaccg acatttttat cataaatc 458

<210> 12
<211> 973
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS

<222> (109)..(759)
<223> codificando para produto de gene de Arabidopsis thaliana At1g48130 (peroxiredoxina)

<400> 12
tggcgatctc tatttaatcg gtcctcgacg caacttcaac ttctactaca tccattcaca 60
tcaaatcaat acagaaagtt ttttctatat ataaatataa aaggtaaa atg cca ggg 117
Met Pro Gly
1
atc aca cta gga gac acg gtg ccg aac cta gaa gtg gag acg aca cat 165
Ile Thr Leu Gly Asp Thr Val Pro Asn Leu Glu Val Glu Thr Thr His
5 10 15
gac aag ttc aaa ctt cat gac tac ttc gcc aat tct tgg acc gtc ctc 213
Asp Lys Phe Lys Leu His Asp Tyr Phe Ala Asn Ser Trp Thr Val Leu
20 25 30 35
ttc tct cat ccc ggg gat ttc aca ccc gtg tgc acc acg gag ctt ggt 261
Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr Glu Leu Gly
40 45 50
gcg atg gcc aaa tac gct cat gag ttc gat aag aga ggc gtg aag ctc 309
Ala Met Ala Lys Tyr Ala His Glu Phe Asp Lys Arg Gly Val Lys Leu
55 60 65
ctt ggt ctg tct tgt gac gat gta cag tca cac aaa gac tgg atc aaa 357
Leu Gly Leu Ser Cys Asp Asp Val Gln Ser His Lys Asp Trp Ile Lys
70 75 80
gat att gaa gcc ttt aat cac gga agc aag gtg aat tac ccg ata atc 405
Asp Ile Glu Ala Phe Asn His Gly Ser Lys Val Asn Tyr Pro Ile Ile
85 90 95
gct gac ccg aac aaa gag atc att cct cag ctt aac atg att gat cca 453
Ala Asp Pro Asn Lys Glu Ile Ile Pro Gln Leu Asn Met Ile Asp Pro
100 105 110 115
att gag aac gga ccg tct cgt gcc ctt cat att gtt ggt cct gac agt 501
Ile Glu Asn Gly Pro Ser Arg Ala Leu His Ile Val Gly Pro Asp Ser
120 125 130
aag ata aaa ttg agc ttc ttg tat ccg tcg acc acg gga cgt aac atg 549
Lys Ile Lys Leu Ser Phe Leu Tyr Pro Ser Thr Thr Gly Arg Asn Met
135 140 145
gac gaa gta ttg agg gct tta gac tcg ttg ttg atg gcg tcc aag cac 597
Asp Glu Val Leu Arg Ala Leu Asp Ser Leu Leu Met Ala Ser Lys His
150 155 160
aat aac aag atc gcc act ccg gtt aac tgg aag cca gat caa ccg gtt 645
Asn Asn Lys Ile Ala Thr Pro Val Asn Trp Lys Pro Asp Gln Pro Val
165 170 175
gtg att tcg cct gct gtg tcg gac gag gaa gcc aaa aag atg ttc cca 693
Val Ile Ser Pro Ala Val Ser Asp Glu Glu Ala Lys Lys Met Phe Pro
180 185 190 195
cag ggt ttc aag acc gcc gat ctt ccg tca aag aaa ggc tat ctg cgt 741
Gln Gly Phe Lys Thr Ala Asp Leu Pro Ser Lys Lys Gly Tyr Leu Arg
200 205 210
cgc aca gag gtc tct tga aaagataata acaaaagcct actatataac 789
Arg Thr Glu Val Ser
215
gtacatgcaa gtattgtatg atattaatgt ttttacgtac gtgtaaacaa aaataattac 849
gtttgtaacg tatggtgatg atgtggtgca ctaggtgtag gccttgtatt aataaaaaga 909
agtttgttct atatagagtg gtttagtacg acgatttatt tactaaacaa aaaaaaaaaa 969
aaaa 973

<210> 13
<211> 216
<212> PRT
<213> Arabidopsis thaliana

<400> 13
Met Pro Gly Ile Thr Leu Gly Asp Thr Val Pro Asn Leu Glu Val Glu

```
1               5                  10                  15

Thr Thr His Asp Lys Phe Lys Leu His Asp Tyr Phe Ala Asn Ser Trp
            20                  25                  30

Thr Val Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
        35                  40                  45

Glu Leu Gly Ala Met Ala Lys Tyr Ala His Glu Phe Asp Lys Arg Gly
    50                  55                  60

Val Lys Leu Leu Gly Leu Ser Cys Asp Asp Val Gln Ser His Lys Asp
65                  70                  75                  80

Trp Ile Lys Asp Ile Glu Ala Phe Asn His Gly Ser Lys Val Asn Tyr
                85                  90                  95

Pro Ile Ile Ala Asp Pro Asn Lys Glu Ile Ile Pro Gln Leu Asn Met
            100                 105                 110

Ile Asp Pro Ile Glu Asn Gly Pro Ser Arg Ala Leu His Ile Val Gly
        115                 120                 125

Pro Asp Ser Lys Ile Lys Leu Ser Phe Leu Tyr Pro Ser Thr Thr Gly
    130                 135                 140

Arg Asn Met Asp Glu Val Leu Arg Ala Leu Asp Ser Leu Leu Met Ala
145                 150                 155                 160

Ser Lys His Asn Asn Lys Ile Ala Thr Pro Val Asn Trp Lys Pro Asp
                165                 170                 175

Gln Pro Val Val Ile Ser Pro Ala Val Ser Asp Glu Glu Ala Lys Lys
            180                 185                 190

Met Phe Pro Gln Gly Phe Lys Thr Ala Asp Leu Pro Ser Lys Lys Gly
        195                 200                 205

Tyr Leu Arg Arg Thr Glu Val Ser
    210                 215
```

<210> 14
<211> 862
<212> DNA
<213> Brassica napus

<220>
<221> CDS
<222> (37)..(687)
<223> codificando para ortólogo de Brassica napus de produto de gene de Arabidopsis thaliana
At1g48130 (peroxiredoxina)

<400> 14

```
ccaatacata cattttcatt tcgataacaa gaaaaa atg cca ggg atc aca cta    54
                       Met Pro Gly Ile Thr Leu
                       1               5
gga gac acg gtg cca aac cta gag gtg gag acg aca cac aag aac ttc   102
Gly Asp Thr Val Pro Asn Leu Glu Val Glu Thr Thr His Lys Asn Phe
            10                  15                  20
aag ctc cat gac tac ttc gcc gat tct tgg acc gtc ctc ttc tct cac   150
Lys Leu His Asp Tyr Phe Ala Asp Ser Trp Thr Val Leu Phe Ser His
        25                  30                  35
```

```
cct ggg gat ttc aca ccc gta tgc aca acc gag ctt ggt gcg atg gga   198
Pro Gly Asp Phe Thr Pro Val Cys Thr Thr Glu Leu Gly Ala Met Gly
  40         45         50
aaa tac gct cat gag ttc gag cag aga ggt gtg aag ctc ctt ggt ttg   246
Lys Tyr Ala His Glu Phe Glu Gln Arg Gly Val Lys Leu Leu Gly Leu
55         60         65         70
tct tgt gac gac ata cag tcg cac aaa gat tgg atc ccc gac att gag   294
Ser Cys Asp Asp Ile Gln Ser His Lys Asp Trp Ile Pro Asp Ile Glu
        75         80         85
gca ttt act cct gga agc aaa gtg aca tac cca ata atc gct gat ccg   342
Ala Phe Thr Pro Gly Ser Lys Val Thr Tyr Pro Ile Ile Ala Asp Pro
      90         95         100
aac aaa gag atc atc cct cag ctt aac atg att gat cct att gag aac   390
Asn Lys Glu Ile Ile Pro Gln Leu Asn Met Ile Asp Pro Ile Glu Asn
    105         110         115
ggt ccc tct cgt gcc ctt cat gta gtt ggt cct gat tgc aag ata aag   438
Gly Pro Ser Arg Ala Leu His Val Val Gly Pro Asp Cys Lys Ile Lys
  120         125         130
ttg agt ttc ctg tat ccg tcc acc acg gga cgt aac atg gac gaa gtg   486
Leu Ser Phe Leu Tyr Pro Ser Thr Thr Gly Arg Asn Met Asp Glu Val
135         140         145         150
ctg agg gct tta gac tcg ttg ttg atg gcg gcc aaa cat aag aac aag   534
Leu Arg Ala Leu Asp Ser Leu Leu Met Ala Ala Lys His Lys Asn Lys
        155         160         165
att gcc act ccg gtt aac tgg aaa cca gat gaa ccc gtt gtt atc tca   582
Ile Ala Thr Pro Val Asn Trp Lys Pro Asp Glu Pro Val Val Ile Ser
      170         175         180
cca gct gtg tca gac gaa gag gct aag aag ttg ttt cct cag ggt ttc   630
Pro Ala Val Ser Asp Glu Glu Ala Lys Lys Leu Phe Pro Gln Gly Phe
    185         190         195
aag act gcc aaa ctt ccg tcc aag aaa ggc tac cta cgt gtc gcc gat   678
Lys Thr Ala Lys Leu Pro Ser Lys Lys Gly Tyr Leu Arg Val Ala Asp
  200         205         210
gtc tct tga aagaccgaaa agaaagaaac tagttatgtt gtctgagtgt       727
Val Ser
215
attgtataat aatgctcaca atgggttgat aaaaatatga aaaactacgt ttgtaacgta  787
tcgtgtgtgt gctctggagg atatagcctt atattaatag aaataatgtg ttttataaaa  847
aaaaaaaaaa aaaaa                                              862
```

<210> 15
<211> 216
<212> PRT
<213> Brassica napus

<400> 15

```
Met Pro Gly Ile Thr Leu Gly Asp Thr Val Pro Asn Leu Glu Val Glu
1         5         10         15

Thr Thr His Lys Asn Phe Lys Leu His Asp Tyr Phe Ala Asp Ser Trp
      20         25         30

Thr Val Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
    35         40         45

Glu Leu Gly Ala Met Gly Lys Tyr Ala His Glu Phe Glu Gln Arg Gly
  50         55         60

Val Lys Leu Leu Gly Leu Ser Cys Asp Asp Ile Gln Ser His Lys Asp
65         70         75         80

Trp Ile Pro Asp Ile Glu Ala Phe Thr Pro Gly Ser Lys Val Thr Tyr
        85         90         95
```

```
Pro Ile Ile Ala Asp Pro Asn Lys Glu Ile Ile Pro Gln Leu Asn Met
      100         105         110

Ile Asp Pro Ile Glu Asn Gly Pro Ser Arg Ala Leu His Val Val Gly
    115         120         125

Pro Asp Cys Lys Ile Lys Leu Ser Phe Leu Tyr Pro Ser Thr Thr Gly
  130         135         140

Arg Asn Met Asp Glu Val Leu Arg Ala Leu Asp Ser Leu Leu Met Ala
145         150         155         160

Ala Lys His Lys Asn Lys Ile Ala Thr Pro Val Asn Trp Lys Pro Asp
        165         170         175

Glu Pro Val Val Ile Ser Pro Ala Val Ser Asp Glu Glu Ala Lys Lys
      180         185         190

Leu Phe Pro Gln Gly Phe Lys Thr Ala Lys Leu Pro Ser Lys Lys Gly
    195         200         205

Tyr Leu Arg Val Ala Asp Val Ser
  210         215
```

<210> 16
<211> 816
<212> DNA
<213> Bromus secalinus

<220>
<221> CDS
<222> (2)..(610)
<223> codificando para ortólogo de Bromus secalinus de produto de gene de Arabidopsis thaliana At1g48130 (peroxiredoxina)

<400> 16

```
c tcc acc cat ggc aag atc cgc atc cac gac tac gtc gcc aac ggc tac   49
 Ser Thr His Gly Lys Ile Arg Ile His Asp Tyr Val Ala Asn Gly Tyr
 1        5          10         15
gtc atc ctc ttc tcg cac ccc ggc gat ttc acc ccg gtg tgc acg acg    97
Val Ile Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
      20         25         30
gag ctg gcg gcg atg gcc aac tac gcc aag gag ttc gag aag agg ggc   145
Glu Leu Ala Ala Met Ala Asn Tyr Ala Lys Glu Phe Glu Lys Arg Gly
    35         40         45
gtc aag ctg ctg ggc atc tcc tgc gac gac gtg cag tcc cac aag gag   193
Val Lys Leu Leu Gly Ile Ser Cys Asp Asp Val Gln Ser His Lys Glu
  50         55         60
tgg acc aag gac atc gag gcc tac aag cct ggg agc aag gtg acg tac   241
Trp Thr Lys Asp Ile Glu Ala Tyr Lys Pro Gly Ser Lys Val Thr Tyr
65         70         75         80
ccg atc atg gcg gac ccg gac cgg tcg gcc atc aag cag ctc aac atg   289
Pro Ile Met Ala Asp Pro Asp Arg Ser Ala Ile Lys Gln Leu Asn Met
        85         90         95
gtg gat ccg gac gag aag gac gcg gag ggg cag ctg ccg tcg cgc acg   337
Val Asp Pro Asp Glu Lys Asp Ala Glu Gly Gln Leu Pro Ser Arg Thr
      100         105         110
ctg cac atc gtg ggg ccg gac aag aag gtg aag ctg agc ttc ctg tac   385
Leu His Ile Val Gly Pro Asp Lys Lys Val Lys Leu Ser Phe Leu Tyr
    115         120         125
ccg tcg tgc acg ggg agg aac atg gac gag gtt gtg cgc gcg gtg gac   433
Pro Ser Cys Thr Gly Arg Asn Met Asp Glu Val Val Arg Ala Val Asp
```

```
  130         135         140
tcg ctg ctg acg gcg gcc aag cac aag gtg gcc acc ccg gcc aac tgg   481
Ser Leu Leu Thr Ala Ala Lys His Lys Val Ala Thr Pro Ala Asn Trp
145         150         155         160
aag ccc ggg gag tgc gtg gtg atc gcg ccg ggc gtg tcc gac gag gag   529
Lys Pro Gly Glu Cys Val Val Ile Ala Pro Gly Val Ser Asp Glu Glu
        165         170         175
gcc aag aag ttg ttc ccg cag ggg ttc gag acc aag gac ctg ccc tcc   577
Ala Lys Lys Leu Phe Pro Gln Gly Phe Glu Thr Lys Asp Leu Pro Ser
      180         185         190
aag aag gga tac ctc cgc ttc acc aag gtc tag gcgtgcccgt cgccgtgggc  630
Lys Lys Gly Tyr Leu Arg Phe Thr Lys Val
    195         200
tagctcgtcg gcgctttaag agttcacctg tggtacaagt attgtgtcgt cgttgtggta  690
cttttggggt acgtagtagc atggcgtttg tttctgttgt aatccagtac tcgactgctg  750
gtggcgtcga gctaatatat gcgtgtactg tgtggactgt tggccccttt gaaaaaaaaa  810
aaaaaa                                    816

<210> 17
<211> 202
<212> PRT
<213> Bromus secalinus

<400> 17
Ser Thr His Gly Lys Ile Arg Ile His Asp Tyr Val Ala Asn Gly Tyr
1        5          10         15

Val Ile Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
      20         25         30

Glu Leu Ala Ala Met Ala Asn Tyr Ala Lys Glu Phe Glu Lys Arg Gly
    35         40         45

Val Lys Leu Leu Gly Ile Ser Cys Asp Asp Val Gln Ser His Lys Glu
  50         55         60

Trp Thr Lys Asp Ile Glu Ala Tyr Lys Pro Gly Ser Lys Val Thr Tyr
65         70         75         80

Pro Ile Met Ala Asp Pro Asp Arg Ser Ala Ile Lys Gln Leu Asn Met
        85         90         95

Val Asp Pro Asp Glu Lys Asp Ala Glu Gly Gln Leu Pro Ser Arg Thr
      100         105         110

Leu His Ile Val Gly Pro Asp Lys Lys Val Lys Leu Ser Phe Leu Tyr
    115         120         125

Pro Ser Cys Thr Gly Arg Asn Met Asp Glu Val Val Arg Ala Val Asp
  130         135         140

Ser Leu Leu Thr Ala Ala Lys His Lys Val Ala Thr Pro Ala Asn Trp
145         150         155         160

Lys Pro Gly Glu Cys Val Val Ile Ala Pro Gly Val Ser Asp Glu Glu
        165         170         175

Ala Lys Lys Leu Phe Pro Gln Gly Phe Glu Thr Lys Asp Leu Pro Ser
      180         185         190

Lys Lys Gly Tyr Leu Arg Phe Thr Lys Val
    195         200
```

<210> 18
<211> 864
<212> DNA
<213> Hordeum vulgare

<220>
<221> CDS
<222> (51)..(707)
<223> codificando para ortólogo de Hordeum vulgare de produto de gene de Arabidopsis thaliana At1g48130 (peroxiredoxina)

<400> 18
gttcgcagca gcaaagagca cgcagcaact agctttcagt tgtttcaacc atg ccg 56
Met Pro
1

ggg ctc acc att ggc gac acc gtc ccc aac ctg gag ctg gac tcc acc 104
Gly Leu Thr Ile Gly Asp Thr Val Pro Asn Leu Glu Leu Asp Ser Thr
5 10 15

cat ggc aag atc cgc atc cac gac tac gtc ggc aac ggc tac gtc atc 152
His Gly Lys Ile Arg Ile His Asp Tyr Val Gly Asn Gly Tyr Val Ile
20 25 30

ctc ttc tcg cac ccc ggt gat ttc acc ccg gtg tgc acg acg gag ctg 200
Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr Glu Leu
35 40 45 50

gcg gcc atg gcc aac tac gcc aag gag ttc gag aag cgg ggc gtg aag 248
Ala Ala Met Ala Asn Tyr Ala Lys Glu Phe Glu Lys Arg Gly Val Lys
55 60 65

ctg ctc ggc atc tcc tgc gac gac gtg cag tcc cac aag gag tgg acc 296
Leu Leu Gly Ile Ser Cys Asp Asp Val Gln Ser His Lys Glu Trp Thr
70 75 80

aag gac atc gag gcc tac aag cct ggg agc aag gtg acg tac ccg atc 344
Lys Asp Ile Glu Ala Tyr Lys Pro Gly Ser Lys Val Thr Tyr Pro Ile
85 90 95

atg gcg gac ccg gac cgg tcg gcg atc aag cag ctc aac atg gtg gac 392
Met Ala Asp Pro Asp Arg Ser Ala Ile Lys Gln Leu Asn Met Val Asp
100 105 110

ccg gac gag aag gac gcg cag ggg cag ctg ccg tcc cgc acc ctg cac 440
Pro Asp Glu Lys Asp Ala Gln Gly Gln Leu Pro Ser Arg Thr Leu His
115 120 125 130

atc gtg ggg ccg gac aag gtg gtg aag ctg agc ttc ctg tac ccg tcg 488
Ile Val Gly Pro Asp Lys Val Val Lys Leu Ser Phe Leu Tyr Pro Ser
135 140 145

tgc acg ggg cgg aac atg gac gag gtg gtg cgc gcc gtg gac tcg ctg 536
Cys Thr Gly Arg Asn Met Asp Glu Val Val Arg Ala Val Asp Ser Leu
150 155 160

ctg acg gcg gca aag cac aag gtg gca acc ccg gcc aac tgg aag ccc 584
Leu Thr Ala Ala Lys His Lys Val Ala Thr Pro Ala Asn Trp Lys Pro
165 170 175

ggc gag tgc gtt gtg atc gcg ccc ggc gtc tcc gac gag gag gcc aag 632
Gly Glu Cys Val Val Ile Ala Pro Gly Val Ser Asp Glu Glu Ala Lys
180 185 190

aag atg ttc ccg cag ggg ttc gag acc gcc gac ctg ccc tcc aag aag 680
Lys Met Phe Pro Gln Gly Phe Glu Thr Ala Asp Leu Pro Ser Lys Lys
195 200 205 210

ggc tac ctc cgc ttc acc aag gtc tag gcgtccgtcc gtgctagctc 727
Gly Tyr Leu Arg Phe Thr Lys Val
215

gtcggcgctt gctcggttac ttatggtggt ggtatgtgtc gttgtggtgc gtgcgtactt 787
ttgtggtagc atgtgtcgtt tcttctgtaa ttctactcgc gtcgccgagc tatctctact 847
gtgacttcgc ctctttc 864

<210> 19
<211> 218

<212> PRT
<213> Hordeum vulgare

<400> 19
Met Pro Gly Leu Thr Ile Gly Asp Thr Val Pro Asn Leu Glu Leu Asp
1 5 10 15

Ser Thr His Gly Lys Ile Arg Ile His Asp Tyr Val Gly Asn Gly Tyr
20 25 30

Val Ile Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
35 40 45

Glu Leu Ala Ala Met Ala Asn Tyr Ala Lys Glu Phe Glu Lys Arg Gly
50 55 60

Val Lys Leu Leu Gly Ile Ser Cys Asp Asp Val Gln Ser His Lys Glu
65 70 75 80

Trp Thr Lys Asp Ile Glu Ala Tyr Lys Pro Gly Ser Lys Val Thr Tyr
85 90 95

Pro Ile Met Ala Asp Pro Asp Arg Ser Ala Ile Lys Gln Leu Asn Met
100 105 110

Val Asp Pro Asp Glu Lys Asp Ala Gln Gly Gln Leu Pro Ser Arg Thr
115 120 125

Leu His Ile Val Gly Pro Asp Lys Val Val Lys Leu Ser Phe Leu Tyr
130 135 140

Pro Ser Cys Thr Gly Arg Asn Met Asp Glu Val Val Arg Ala Val Asp
145 150 155 160

Ser Leu Leu Thr Ala Ala Lys His Lys Val Ala Thr Pro Ala Asn Trp
165 170 175

Lys Pro Gly Glu Cys Val Val Ile Ala Pro Gly Val Ser Asp Glu Glu
180 185 190

Ala Lys Lys Met Phe Pro Gln Gly Phe Glu Thr Ala Asp Leu Pro Ser
195 200 205

Lys Lys Gly Tyr Leu Arg Phe Thr Lys Val
210 215

<210> 20
<211> 273
<212> DNA
<213> Linum usitatissimum

<220>
<221> CDS
<222> (1)..(273)
<223> codificando para ortólogo de Linum usitatissimum de produto de gene de Arabidopsis thaliana At1g48130 (peroxiredoxina)

<400> 20
atg cct gga ctc acc atc ggg gac acc gtc cca aat ctt gaa gtg gag 48
Met Pro Gly Leu Thr Ile Gly Asp Thr Val Pro Asn Leu Glu Val Glu
1 5 10 15
agc act cat gga gtc atc aat ctc cac gac ttc ttc acc aat tcc tgg 96
Ser Thr His Gly Val Ile Asn Leu His Asp Phe Phe Thr Asn Ser Trp

```
      20         25         30
acc atc ctc ttc tct cac ccc ggt gat ttc acg ccg gtg tgc acc acg   144
Thr Ile Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
    35         40         45
gaa cta ggg aag atg gca gct tat gcg ccg gag ttt gag aag aga ggg   192
Glu Leu Gly Lys Met Ala Ala Tyr Ala Pro Glu Phe Glu Lys Arg Gly
  50         55         60
gtc aaa ctc ctg ggc ttg tca tgc gac gac gtt ttg tct cat gtc gag   240
Val Lys Leu Leu Gly Leu Ser Cys Asp Asp Val Leu Ser His Val Glu
65         70         75         80
tgg atc aag gac gta gaa gcc ttc acc gtg agc             273
Trp Ile Lys Asp Val Glu Ala Phe Thr Val Ser
        85         90
```

<210> 21
<211> 91
<212> PRT
<213> Linum usitatissimum

<400> 21

```
Met Pro Gly Leu Thr Ile Gly Asp Thr Val Pro Asn Leu Glu Val Glu
1        5          10         15

Ser Thr His Gly Val Ile Asn Leu His Asp Phe Phe Thr Asn Ser Trp
      20         25         30

Thr Ile Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
    35         40         45

Glu Leu Gly Lys Met Ala Ala Tyr Ala Pro Glu Phe Glu Lys Arg Gly
  50         55         60

Val Lys Leu Leu Gly Leu Ser Cys Asp Asp Val Leu Ser His Val Glu
65         70         75         80

Trp Ile Lys Asp Val Glu Ala Phe Thr Val Ser
        85         90
```

<210> 22
<211> 827
<212> DNA
<213> Triticum turgidum

<220>
<221> CDS
<222> (1)..(657)
<223> codificando para ortólogo de Triticum turgidum de produto de gene de Arabidopsis thaliana At1g48130 (peroxiredoxina)

<400> 22

```
atg ccg ggg ctg acg ata ggc gac acc gtc ccc aac ctg gag ctg gac    48
Met Pro Gly Leu Thr Ile Gly Asp Thr Val Pro Asn Leu Glu Leu Asp
1        5          10         15
tcc acc cat ggc aag atc cgg atc cac gac tac gtc ggc aac ggc tac    96
Ser Thr His Gly Lys Ile Arg Ile His Asp Tyr Val Gly Asn Gly Tyr
      20         25         30
gtc atc ctc ttc tca cac ccc ggt gat ttc acc ccg gtg tgc acg acg   144
Val Ile Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
    35         40         45
gag tta gcg gcg atg gcc aac tac gcc aag gag ttc gag aag cgg ggc   192
Glu Leu Ala Ala Met Ala Asn Tyr Ala Lys Glu Phe Glu Lys Arg Gly
  50         55         60
gtg aag ctg ctg ggc atc tcc tgc gac gac gtg cag tcc cac aag gaa   240
```

```
Val Lys Leu Leu Gly Ile Ser Cys Asp Asp Val Gln Ser His Lys Glu
65         70         75         80
tgg acc aag gac atc gag gcc tac aag cct ggg agc aag gtg acg tac   288
Trp Thr Lys Asp Ile Glu Ala Tyr Lys Pro Gly Ser Lys Val Thr Tyr
        85         90         95
ccg atc atg gcg gac ccg gac cgg tcg gcc atc aag caa ctg aac atg   336
Pro Ile Met Ala Asp Pro Asp Arg Ser Ala Ile Lys Gln Leu Asn Met
      100         105         110
gtc gac ccc gac gag aag gac gcg gag ggg cag ctg ccg tcc cgc acc   384
Val Asp Pro Asp Glu Lys Asp Ala Glu Gly Gln Leu Pro Ser Arg Thr
    115         120         125
ttg cac atc gtg ggg ccg gac aag aag gtg aag ctg agc ttc ctg tac   432
Leu His Ile Val Gly Pro Asp Lys Lys Val Lys Leu Ser Phe Leu Tyr
  130         135         140
ccg tcg tgc acg ggg cgg aac atg gac gag gtg gtg cgc gcc gtg gac   480
Pro Ser Cys Thr Gly Arg Asn Met Asp Glu Val Val Arg Ala Val Asp
145         150         155         160
tcg ctg ctg acg gcg gcc aag cac aag gtg gcc acc ccg gcc aac tgg   528
Ser Leu Leu Thr Ala Ala Lys His Lys Val Ala Thr Pro Ala Asn Trp
        165         170         175
aat ccc ggg gag tgc gtg gtg atc gcg ccc ggc gtc tcc gac gac ggg   576
Asn Pro Gly Glu Cys Val Val Ile Ala Pro Gly Val Ser Asp Asp Gly
      180         185         190
gcc aag aag atg ttc ccg cag ggg ttc gag act gcc gac ctg ccc tcc   624
Ala Lys Lys Met Phe Pro Gln Gly Phe Glu Thr Ala Asp Leu Pro Ser
    195         200         205
aag aag ggg tac ctc cgc ttc acc aag gtc tag gcgtgcgtcc gtgctagctc  677
Lys Lys Gly Tyr Leu Arg Phe Thr Lys Val
  210         215
gtcggcgcct gctcggttac tcgtggcggt ctgtgtcgtt gtggtccgtg cgtacttttg  737
tggtggtacg tgtcgtttct tctgtaaatc taccagcgtc tccgagctat gtatgtgtac  797
tgtgactttg tctcaaaaaa aaaaaaaaaa                    827

<210> 23
<211> 218
<212> PRT
<213> Triticum turgidum

<400> 23
Met Pro Gly Leu Thr Ile Gly Asp Thr Val Pro Asn Leu Glu Leu Asp
1         5          10         15

Ser Thr His Gly Lys Ile Arg Ile His Asp Tyr Val Gly Asn Gly Tyr
      20         25         30

Val Ile Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr
    35         40         45

Glu Leu Ala Ala Met Ala Asn Tyr Ala Lys Glu Phe Glu Lys Arg Gly
  50         55         60

Val Lys Leu Leu Gly Ile Ser Cys Asp Asp Val Gln Ser His Lys Glu
65         70         75         80

Trp Thr Lys Asp Ile Glu Ala Tyr Lys Pro Gly Ser Lys Val Thr Tyr
        85         90         95

Pro Ile Met Ala Asp Pro Asp Arg Ser Ala Ile Lys Gln Leu Asn Met
      100         105         110

Val Asp Pro Asp Glu Lys Asp Ala Glu Gly Gln Leu Pro Ser Arg Thr
    115         120         125
```

Leu His Ile Val Gly Pro Asp Lys Lys Val Lys Leu Ser Phe Leu Tyr
130 135 140

Pro Ser Cys Thr Gly Arg Asn Met Asp Glu Val Val Arg Ala Val Asp
145 150 155 160

Ser Leu Leu Thr Ala Ala Lys His Lys Val Ala Thr Pro Ala Asn Trp
165 170 175

Asn Pro Gly Glu Cys Val Val Ile Ala Pro Gly Val Ser Asp Asp Gly
180 185 190

Ala Lys Lys Met Phe Pro Gln Gly Phe Glu Thr Ala Asp Leu Pro Ser
195 200 205

Lys Lys Gly Tyr Leu Arg Phe Thr Lys Val
210 215

<210> 24
<211> 39
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 24
ttttttgtcg acgccacatc atgtttagac ttatctcca 39

<210> 25
<211> 44
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 25
ttaattcata tgtttacctt ttatatttat atatagaaaa aact 44

<210> 26
<211> 32
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 26
ttaattaagc ttagaagttg aagttgcgtc ga 32

<210> 27
<211> 34
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 27
ttttttacta gtctatataa tatatcacag gaca 34

<210> 28

<211> 33
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 28
ttaattcata tgtttacctt ttatatttat ata 33

<210> 29
<211> 32
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 29
ttaattcata tgagaagttg aagttgcgtc ga 32

<210> 30
<211> 36
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 30
ttaattaagc ttagaagtta tatgaagttg cgtcga 36

<210> 31
<211> 33
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 31
ttaattggta cctttacctt ttatatttat ata 33

<210> 32
<211> 33
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 32
ttaattcata tgagaagtta tatgaagttg cgt 33

<210> 33
<211> 38
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 33

tttttttgtcg accacgggca ggacataggg actactac 38

<210> 34
<211> 44
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 34
ttaattacta gtgatttatg ataaaaatgt cggtttcttt gtgg 44

<210> 35
<211> 36
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 35
tttttttgtcg accagattta cccgtttaat cgggta 36

<210> 36
<211> 34
<212> DNA
<213> Artificial

<220>
<223> iniciador de oligonucleotídeo

<400> 36
tttttttgtcg acaccgttaa attttgcatg tact 34

<210> 37
<211> 7
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 37
Gly Asp Thr Val Pro Asn Leu
1 5

<210> 38
<211> 14
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<220>
<221> VARIANTE
<222> (4)..(4)
<223> (I/L) variação

<220>

<221> VARIANTE
<222> (6)..(6)
<223> variação(L/I)

<400> 38
Met Pro Gly Ile Thr Leu Gly Asp Thr Val Pro Asn Leu Glu
1 5 10

<210> 39
<211> 7
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 39
Gly Asp Phe Thr Pro Val Cys
1 5

<210> 40
<211> 16
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 40
Leu Phe Ser His Pro Gly Asp Phe Thr Pro Val Cys Thr Thr Glu Leu
1 5 10 15

<210> 41
<211> 7
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 41
Val Lys Leu Leu Gly Leu Ser
1 5

<210> 42
<211> 12
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene Arabidopsis thaliana At1g48130)

<400> 42
Arg Gly Val Lys Leu Leu Gly Leu Ser Cys Asp Asp
1 5 10

<210> 43
<211> 7
<212> PRT

<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<220>
<221> VARIANTE
<222> (4)..(4)
<223> variação(I/M)

<400> 43
Tyr Pro Ile Ile Ala Asp Pro
1 5

<210> 44
<211> 11
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<220>
<221> VARIANTE
<222> (4)..(4)
<223> variação(T/N)

<220>
<221> VARIANTE
<222> (8)..(8)
<223> variação(I/M)

<400> 44
Ser Lys Val Thr Tyr Pro Ile Ile Ala Asp Pro
1 5 10

<210> 45
<211> 7
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 45
Lys Leu Ser Phe Leu Tyr Pro
1 5

<210> 46
<211> 10
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<220>
<221> VARIANTE
<222> (1)..(1)

<223> variação(K/V)

<220>
<221> VARIANTE
<222> (2)..(2)
<223> variação(I/V)

<400> 46
Lys Ile Lys Leu Ser Phe Leu Tyr Pro Ser
1 5 10

<210> 47
<211> 7
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 47
Gly Arg Asn Met Asp Glu Val
1 5

<210> 48
<211> 11
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 48
Thr Gly Arg Asn Met Asp Glu Val Leu Arg Ala
1 5 10

<210> 49
<211> 7
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<220>
<221> VARIANTE
<222> (1)..(1)
<223> (I/V) variação

<220>
<221> VARIANTE
<222> (5)..(5)
<223> (V/A) variação

<400> 49
Ile Ala Thr Pro Val Asn Trp
1 5

<210> 50
<211> 10
<212> PRT

<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<220>
<221> VARIANTE
<222> (2)..(2)
<223> (I/V) variação

<220>
<221> VARIANTE
<222> (6)..(6)
<223> variação(V/A)

<400> 50
Lys Ile Ala Thr Pro Val Asn Trp Lys Pro
1 5 10

<210> 51
<211> 7
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 51
Pro Ser Lys Lys Gly Tyr Leu
1 5

<210> 52
<211> 9
<212> PRT
<213> Artificial

<220>
<223> motivo de aminoácido para peroxiredoxina (ortólogos de produto de gene de Arabidopsis thaliana At1g48130)

<400> 52
Leu Pro Ser Lys Lys Gly Tyr Leu Arg
1 5

<210> 53
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 53
tact 4

<210> 54
<211> 4
<212> DNA
<213> Artificial

<220>

<223> motivo de promotor

<400> 54
agaa 4

<210> 55
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 55
aaag 4

<210> 56
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 56
ccat 4

<210> 57
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 57
taaa 4

<210> 58
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 58
ctat 4

<210> 59
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 59
acgt 4

<210> 60
<211> 4
<212> DNA

<213> Artificial

<220>
<223> motivo de promotor

<400> 60
aaaa 4

<210> 61
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 61
gtta 4

<210> 62
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 62
ttaa 4

<210> 63
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 63
acat 4

<210> 64
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (2)..(2)
<223> n é a, c, g, ou t

<400> 64
tncc 4

<210> 65
<211> 4
<212> DNA
<213> Artificial

<220>

<223> motivo de promotor

<400> 65
catg 4

<210> 66
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 66
ttat 4

<210> 67
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 67
tttt 4

<210> 68
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 68
acgt 4

<210> 69
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 69
agaa 4

<210> 70
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 70
acgt 4

<210> 71
<211> 4
<212> DNA

<213> Artificial

<220>
<223> motivo de promotor

<400> 71
acgt 4

<210> 72
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 72
acgt 4

<210> 73
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 73
catg 4

<210> 74
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 74
accc 4

<210> 75
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 75
ccat 4

<210> 76
<211> 4
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 76
tttt 4

<210> 77
<211> 9
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (2)..(2)
<223> n é a, c, g, ou t

<400> 77
antactctt 9

<210> 78
<211> 13
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (3)..(6)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (9)..(9)
<223> n é a, c, g, ou t

<400> 78
gannnnacna gaa 13

<210> 79
<211> 16
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (3)..(5)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (13)..(15)
<223> n é a, c, g, ou t

<400> 79
gannntgcaa agnnna 16

<210> 80
<211> 15
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 80
ccattccttg tagta 15

<210> 81
<211> 9
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (5)..(5)
<223> n é a, c, g, ou t

<400> 81
taaanaatc 9

<210> 82
<211> 25
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (14)..(14)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (16)..(16)
<223> n é a, c, g, ou t

<400> 82
ccattgctat cctngntaga aacta 25

<210> 83
<211> 9
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (2)..(4)
<223> n é a, c, g, ou t

<400> 83
tnnnacgta 9

<210> 84
<211> 15
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (2)..(5)
<223> n é a, c, g, ou t

<400> 84
annnnaaaaa tgtta 15

<210> 85
<211> 9
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (2)..(3)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (5)..(5)
<223> n é a, c, g, ou t

<400> 85
annangtta 9

<210> 86
<211> 6
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (5)..(5)
<223> n é a, c, g, ou t

<400> 86
ttaana 6

<210> 87
<211> 14
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (5)..(6)
<223> n é a, c, g, ou t

<220>

<221> misc_aspecto
<222> (8)..(10)
<223> n é a, c, g, ou t

<400> 87
gaaannannn acat 14

<210> 88
<211> 12
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (2)..(2)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (5)..(6)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (8)..(10)
<223> n é a, c, g, ou t

<400> 88
tnccnnannn gt 12

<210> 89
<211> 6
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 89
catgca 6

<210> 90
<211> 15
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 90
ttattataag taaat 15

<210> 91
<211> 5
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 91
ttttt 5

<210> 92
<211> 14
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (5)..(6)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (8)..(10)
<223> n é a, c, g, ou t

<400> 92
gtggnngnnn acgt 14

<210> 93
<211> 13
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (3)..(6)
<223> n é a, c, g, ou t

<220>
<221> misc_aspecto
<222> (9)..(9)
<223> n é a, c, g, ou t

<400> 93
ggnnnntcna gaa 13

<210> 94
<211> 10
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 94
cgtacgtggc 10

<210> 95
<211> 7
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 95
acgtggc 7

<210> 96
<211> 9
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (2)..(4)
<223> n é a, c, g, ou t

<400> 96
gnnnacgtg 9

<210> 97
<211> 6
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 97
catgca 6

<210> 98
<211> 8
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (7)..(7)
<223> n é a, c, g, ou t

<400> 98
cacccanc 8

<210> 99
<211> 9
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<220>
<221> misc_aspecto
<222> (3)..(4)
<223> n é a, c, g, ou t

<400> 99
cannccatc 9

<210> 100
<211> 5
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 100
ttttt 5

<210> 101
<211> 12
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 101
agtactcttt ga 12

<210> 102
<211> 15
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 102
agaggagaca agaaa 15

<210> 103
<211> 17
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 103
gaagttgcaa agttgag 17

<210> 104
<211> 17
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 104
ccattccttg tagtact 17

<210> 105
<211> 12
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 105
taaataatct at 12

<210> 106
<211> 27
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 106
tccattgcta tccttgttag aaactaa 27

<210> 107
<211> 13
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 107
ttttacacgt acc 13

<210> 108
<211> 21
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 108
acaaaacaaa aatgttaata t 21

<210> 109
<211> 15
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 109
acaaaaatgt taata 15

<210> 110
<211> 17
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 110
aaaaatgtta atatttt 17

<210> 111
<211> 21
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 111
atatgaaaat attaacattt t 21

<210> 112
<211> 21
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 112
aatttccata tttgtaggaa a 21

<210> 113
<211> 27
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 113
cccaagtaca tgcaaaattt aacggtt 27

<210> 114
<211> 17
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 114
ttattataag taaattg 17

<210> 115
<211> 11
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 115
atttttatt a 11

<210> 116
<211> 21
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 116
aaatgtggta gcgtacgtgt g 21

<210> 117
<211> 15

<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 117
tggcggctcc agaat 15

<210> 118
<211> 17
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 118
cttgcgtacg tggcggc 17

<210> 119
<211> 21
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 119
tgaagagaac gtggccctga g 21

<210> 120
<211> 27
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 120
tgcatgaaga gaacgtggcc ctgagtc 27

<210> 121
<211> 27
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 121
gttctcttca tgcaaggata gtagaac 27

<210> 122
<211> 15
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 122
actccaccca cctcc 15

<210> 123
<211> 17
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 123
ctttcccatc ccatcca 17

<210> 124
<211> 11
<212> DNA
<213> Artificial

<220>
<223> motivo de promotor

<400> 124
cattttatc a 11

<210> 125
<211> 876
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> terminador
<222> (1)..(876)
<223> seqüência seguindo CDS de gene At1g48130 de Arabidopsis thaliana peroxiredoxina (PER1) compreendendo 3'-UTR usada como seqüência de terminação transcripcional

<400> 125
aaagataata acaaaagcct actatataac gtacatgcaa gtattgtatg atattaatgt 60
ttttacgtac gtgtaaacaa aaataattac gtttgtaacg tatggtgatg atgtggtgca 120
ctaggtgtag gccttgtatt aataaaaaga agtttgttct atatagagtg gtttagtacg 180
acgatttatt tactagtcgg attggaatag agaaccgaat tcttcaatcc ttgcttttga 240
tcaagaattg aaaccgaatc aaatgtaaaa gttgatatat ttgaaaaacg tattgagctt 300
atgaaaatgc taatactctc atctgtatgg aaaagtgact ttaaaaccga acttaaaagt 360
gacaaaaggg gaatatcgca tcaaaccgaa tgaaaccgat atgaaaccgc atcaaaccga 420
tatagcgtga accaaaatgt ggattaatga catttagacc ggaggaacat aagtggggac 480
tatacggcat cgttttagtg tataagtgaa ttcgcatatt cacgagttcc ccgtattcac 540
cggttgttat atcgtcgatt catcgtcgtc gacggcgagt acttgtgacc tccacaccca 600
gattaatctt ctataactga ttttatcgct tagacgatca cgatctcatc gattaactgt 660
gagtataatc gattcctcct ctttatttttg tgtgttgtat cgatttacga taatttttg 720
gaggaacatc tcctttcgga tcggttttgt tgggaattta tatttctttg ccatggtttt 780
atcttgattt ggtgatttac cttgtgtttg ttctggtgta ctatttagct aatttagtat 840
aagctttcag atgattctag gaaacttttg tgctac 876

<210> 126
<211> 400
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> terminador
<222> (1)..(400)
<223> seqüência seguindo CDS de gene At1g48130 de Arabidopsis thaliana peroxiredoxina (PER1) compreendendo 3'-UTR (derivado) usada como seqüência de terminação transcripcional

<400> 126
aaagataata acaaaagcct actatataac gtacatgcaa gtattgtatg atattaatgt 60
ttttacgtac gtgtaaacaa aaataattac gtttgtaacg tatggtgatg atgtggtgca 120
ctaggtgtag gccttgtatt aataaaaaga agtttgttct atatagagtg gtttagtacg 180
acgatttatt tactagtcgg attggaatag agaaccgaat tcttcaatcc ttgcttttga 240
tcaagaattg aaaccgaatc aaatgtaaaa gttgatatat ttgaaaaacg tattgagctt 300
atgaaaatgc taatactctc atctgtatgg aaaagtgact ttaaaaccga acttaaaagt 360
gacaaaaggg gaatatcgca tcaaaccgaa tgaaaccgat 400

<210> 127
<211> 227
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> 3'UTR
<222> (1)..(227)
<223> 3' região não traduzida de transcrito At1g48130 de Arabidopsis thaliana
peroxiredoxina (PER1)

<400> 127
aaagataata acaaaagcct actatataac gtacatgcaa gtattgtatg atattaatgt 60
ttttacgtac gtgtaaacaa aaataattac gtttgtaacg tatggtgatg atgtggtgca 120
ctaggtgtag gccttgtatt aataaaaaga agtttgttct atatagagtg gtttagtacg 180
acgatttatt tactagtcgg attggaatag agaaccgaat tcttcaa 227

PF 56382
2

PF 56382

REIVINDICAÇÕES

1. Cassetes de expressão para regular expressão específica de semente ou preferencial de semente em plantas, caracterizados de compreender

a) pelo menos uma seqüência nucleotídica reguladora da transcrição de um gene codificando uma proteína peroxiredoxina (PER1), e funcionalmente ligado à mesma,

b) pelo menos uma seqüência de ácido nucleico que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

2. Cassetes de expressão para regular expressão específica de semente ou preferencial de semente em plantas, caracterizados de compreender

a) pelo menos uma seqüência nucleotídica reguladora da transcrição de um gene de planta, referido gene de planta selecionado dentre o grupo consistindo de

i) o gene descrito pelo *loci* do genoma de *Arabidopsis thaliana* do GenBank At1g48130, e

ii) genes ortólogos do gene descrito pelo *loci* do genoma de *Arabidopsis thaliana* do GenBank At1g48130,

e funcionalmente ligados ao mesmo,

b) pelo menos uma seqüência de ácido nucleico que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

3. Cassete de expressão de acordo com qualquer uma das reivindicações 1 ou 2, caracterizado pelo fato de que a seqüência nucleotídica reguladora da transcrição é obtenível a partir de DNA genômico de planta de um gene codificando um polipeptídeo que

a) compreende pelo menos um motivo de seqüência

selecionado dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL	(SEQ ID NO: 37),
ii) GDFTPVC	(SEQ ID NO: 39),
iii) VKLLGLS	(SEQ ID NO: 41),
iv) (YPI(I/M)ADP	(SEQ ID NO: 43),
v) KLSFLYP	(SEQ ID NO: 45),
(vi) GRNMDEV	(SEQ ID NO: 47),
vii) (I/V)ATP(V/A)NW	(SEQ ID NO: 49), e
viii) PSKKGYL	(SEQ ID NO: 51),

ou

b) tem pelo menos 50% identidade de seqüência de aminoácido para um polipeptídeo selecionado dentre o grupo descrito por SEQ ID NO: 13, 15, 17, 19, 21, e 23.

4. Cassete de expressão de acordo com qualquer uma das reivindicações 1 a 3, caracterizado pelo fato de que a seqüência nucleotídica reguladora da transcrição é obtenível a partir de DNA genômico de planta de um gene codificando um polipeptídeo que

a) compreende pelo menos um motivo de seqüência selecionado dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL	(SEQ ID NO: 37),
ii) GDFTPVC	(SEQ ID NO: 39),
iii) VKLLGLS	(SEQ ID NO: 41),
iv) YPI(I/M)ADP	(SEQ ID NO: 43),
v) KLSFLYP	(SEQ ID NO: 45),
vi) GRNMDEV	(SEQ ID NO: 47),
vii) (I/V)ATP(V/A)NW	(SEQ ID NO: 49), e
viii) PSKKGYL	(SEQ ID NO: 51),

e

b) tem pelo menos 50% identidade de seqüência de

aminoácido para um polipeptídeo selecionado dentre o grupo descrito por SEQ ID NO: 13, 15, 17, 19, 21, e 23.

5. Cassete de expressão de acordo com qualquer uma das reivindicações 1 a 4, caracterizado pelo fato de que a seqüência nucleotídica reguladora da transcrição compreende pelo menos uma região de seqüência selecionada dentre o grupo consistindo de

a) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 53, 54, 55, 56, 57, 58, ou 59,

b) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 60, 61, 62, 63 ou 64,

c) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ou 76.

6. Cassete de expressão de acordo com a reivindicação 5, caracterizado pelo fato de que a seqüência nucleotídica reguladora da transcrição compreende pelo menos uma região de seqüência selecionada dentre o grupo consistindo de

a) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 77, 78, 79, 80, 81, 82, ou 83,

b) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 84, 85, 86, 87, ou 88,

c) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, ou 100.

7. Cassete de expressão de acordo com a reivindicação 5 ou 6, caracterizado pelo fato de que a seqüência nucleotídica reguladora da transcrição compreende pelo menos uma região de seqüência selecionada dentre o grupo consistindo de

a) uma região compreendendo pelo menos dois dos motivos de

promotor descritos por SEQ ID NO: 101, 102, 103, 104, 105, 106, ou 107,

b) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 108, 109, 110, 111, ou 112,

c) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, ou 124.

8. Cassete de expressão de acordo com qualquer uma das reivindicações 1 a 7, caracterizado pelo fato de que a seqüência nucleotídica reguladora da transcrição é selecionada dentre o grupo de seqüências consistindo de

i) a seqüência descrita por SEQ ID NOs: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11,

ii) um fragmento de pelo menos 50 bases consecutivas de uma seqüência sob i),

iii) uma seqüência nucleotídica tendo similaridade substancial com uma identidade de seqüência de pelo menos 50% para uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11;

iv) uma seqüência nucleotídica capaz de hibridizar sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C para uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11, ou o seu complemento;

v) uma seqüência nucleotídica capaz de hibridizar sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo 50 a 200 ou mais nucleotídeos consecutivos de uma seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11, ou o seu complemento;

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

9. Cassete de expressão de acordo com a reivindicação 8, caracterizado pelo fato de que as seqüências especificadas sob ii), iii), iv) v) e vi) de acordo com a reivindicação 8 são capazes de modificar transcrição em uma célula de planta ou organismo.

10. Cassete de expressão de acordo com a reivindicação 8, caracterizado pelo fato de que as seqüências especificadas sob ii), iii), iv) v) e vi) de acordo com a reivindicação 8 tem substancialmente a mesma atividade reguladora de transcrição como a seqüência nucleotídica reguladora da transcrição descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11.

11. Cassete de expressão de acordo com a reivindicação 8, caracterizado pelo fato de que as seqüências especificadas sob iii) de acordo com a reivindicação 8 tem uma identidade de seqüência de pelo menos 80% para uma seqüência descrita por SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, ou 11.

12. Cassete de expressão de acordo com a reivindicação 8, caracterizado pelo fato de que as seqüências especificadas sob iv) ou v) de acordo com a reivindicação 8 são hibridizadas sob condições estringentes com a seqüência de marcação especificada.

13. Cassete de expressão de acordo com qualquer uma das reivindicações 1 a 12, caracterizado pelo fato de que a expressão da seqüência de ácido nucleico resulta em expressão de uma proteína, ou expressão de um RNA anti-sentido, RNA sentido ou filamento duplo.

14. Cassete de expressão de acordo com qualquer uma das reivindicações 1 a 13, caracterizado pelo fato de que a expressão da seqüência de ácido nucleico confere à planta um traço agronomicamente valioso.

15. Cassete de expressão de acordo com qualquer

reivindicação 1 a 14, caracterizado pelo fato de que o cassete de expressão ainda compreende funcionalmente associado à seqüência nucleotídica a ser expressada pelo menos uma seqüência selecionado dentre o grupo consistindo de:

a) as seqüências descritas por SEQ ID NO: 125, 126, e 127, e

b) uma seqüência compreendendo um fragmento de pelo menos 25 nucleotídeos consecutivos de uma seqüência descrita por SEQ ID NO: 125, 126, ou 127,

c) uma seqüência tendo uma identidade de pelo menos 50% para uma seqüência descrita por SEQ ID NO: 125, 126, ou 127

d) uma seqüência capaz de hibridizar sob condições equivalentes a hibridização em 7% dodecil sulfato de sódio (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo 50 a 200 ou mais (preferivelmente todos) nucleotídeos consecutivos de como uma seqüência descrita por SEQ ID NO: 125, 126, ou 127, ou o seu complemento,

em que referida seqüência é capaz de mediar estabilização de RNA, estabilização de expressão de RNA, terminação de transcrição de RNA e/ou poliadenilação de RNA.

16. Seqüência nucleotídica isolada, caracterizada pelo fato de ser selecionada dentre o grupo de seqüências consistindo de:

a) a seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e seqüências tendo uma identidade de pelo menos 99% para uma seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e seqüências compreendendo pelo menos 600 nucleotídeos consecutivos de uma seqüência como descrita por SEQ ID NO: 1, 2, 3, ou 4, e

b) a seqüência como descrita por SEQ ID NO: 9, 10, ou 11, e seqüências tendo uma identidade de pelo menos 40% para uma seqüência como descrita por SEQ ID NO: 5 ou 26, e seqüências compreendendo pelo

menos 15 nucleotídeos consecutivos de uma seqüência como descrita por SEQ ID NO: 9, 10, ou 11,

17. Vetor, caracterizado pelo fato de compreender uma seqüência isolada de ácido nucleico como definida na reivindicação 16 ou um cassete de expressão como definido em qualquer uma das reivindicações 1 a 15.

18. Célula hospedeira transgênica ou organismo não humano, caracterizada pelo fato de compreender um cassete de expressão como definido em qualquer uma das reivindicações 1 a 15, ou um vetor como definido na reivindicação 17.

19. Célula de planta ou planta transgênica, caracterizada pelo fato de compreender o cassete de expressão como definido em qualquer uma das reivindicações 1 a 15, ou um vetor como definido na reivindicação 17.

20. Célula de planta ou planta transgênica de acordo com a reivindicação 19, caracterizada pelo fato de que referida planta ou célula de planta é de uma planta usada para produção de óleo.

21. Método para identificar e/ou isolar uma seqüência nucleotídica reguladora da transcrição, caracterizado pelo fato de que referida identificação e/ou isolamento utiliza uma seqüência de ácido nucleico codificando um polipeptídeo como descrito por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de pelo menos 15 nucleotídeos consecutivos de referida seqüência de ácido nucleico.

22. Método de acordo com a reivindicação 21, caracterizado pelo fato de que as seqüências de ácido nucleico são descritas por SEQ ID NO: 12, 14, 16, 18, 20, ou 22 ou uma parte de pelo menos 15 nucleotídeos consecutivos das mesmas.

23. Método de acordo com a reivindicação 21 ou 22, caracterizado pelo fato de que referida identificação e/ou isolamento é realizada por um método selecionado dentre reação em cadeia da polimerase,

hibridização, e triagem de base de dados.

24. Método para prover ou produzir um cassete de expressão transgênica para expressão heteróloga em plantas, caracterizado pelo fato de compreender as etapas de:

I. isolar de uma seqüência nucleotídica reguladora da transcrição de um gene de planta utilizando pelo menos uma seqüência de ácido nucleico codificando um polipeptídeo descrito por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de pelo menos 15 nucleotídeos consecutivos de referida seqüência de ácido nucleico, e

II. funcionalmente ligar referida seqüência nucleotídica reguladora da transcrição a outra seqüência nucleotídica de interesse, que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição.

25. Método para prover um cassete de expressão transgênica para expressão específica de semente ou preferencial de semente, caracterizado pelo fato de compreender as etapas de:

I. isolar de uma seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente utilizando pelo menos uma seqüência de ácido nucleico ou uma parte da mesma, em que referida seqüência está codificando um polipeptídeo descrito por SEQ ID NO: 13, 15, 17, 19, 21, ou 23, ou uma parte de pelo menos 15 nucleotídeos consecutivos do mesmo, e

II. funcionalmente ligar referida seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente a outra seqüência nucleotídica de interesse, que é heteróloga com relação à referida seqüência nucleotídica reguladora da transcrição preferencial de semente ou específica de semente.

26. Métodos de acordo com qualquer uma das reivindicações 21 a 25, caracterizados pelo fato de que a seqüência nucleotídica utilizada

para isolamento de referida seqüência nucleotídica reguladora da transcrição está codificando um polipeptídeo compreendendo pelo menos um motivo de seqüência selecionado dentre o grupo consistindo das seqüências de aminoácido

i) GDTVPNL (SEQ ID NO: 37),
ii) GDFTPVC (SEQ ID NO: 39),
iii) VKLLGLS (SEQ ID NO: 41),
iv) YPI(I/M)ADP (SEQ ID NO: 43),
v) KLSFLYP (SEQ ID NO: 45),
vi) GRNMDEV (SEQ ID NO: 47),
vii) (I/V)ATP(V/A)NW (SEQ ID NO: 49), e
viii) PSKKGYL (SEQ ID NO: 51),

27. Seqüência reguladora de transcrição sintética, caracterizada pelo fato de compreender pelo menos uma região de seqüência selecionada dentre o grupo consistindo de

a) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 53, 54, 55, 56, 57, 58, ou 59,

b) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 60, 61, 62, 63 ou 64,

c) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ou 76.

28. Seqüência reguladora de transcrição sintética de acordo com a reivindicação 27, caracterizada pelo fato de que referida seqüência está compreendendo pelo menos uma região de seqüência selecionada dentre o grupo consistindo de

a) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 77, 78, 79, 80, 81, 82, ou 83,

b) uma região compreendendo pelo menos dois dos motivos de

promotor descritos por SEQ ID NO: 84, 85, 86, 87, ou 88,

c) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, ou 100.

29. Seqüência reguladora de transcrição sintética de acordo com a reivindicação 27 ou 28, caracterizada pelo fato de que a referida seqüência está compreendendo pelo menos uma região de seqüência selecionada dentre o grupo consistindo de

a) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 101, 102, 103, 104, 105, 106, ou 107,

b) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 108, 109, 110, 111, ou 112,

c) uma região compreendendo pelo menos dois dos motivos de promotor descritos por SEQ ID NO: 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, ou 124.

30. Método para prover uma seqüência nucleotídica reguladora de transcrição sintética, caracterizado pelo fato de que os motivos de promotores isolados ou agrupamentos de motivos de promotores são combinados, referidos motivos ou agrupamentos de motivos compreendendo pelo menos 5 dos motivos descritos por qualquer uma dentre SEQ ID NO: 53, 54, 55, 56, 57, 58, 59,: 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, ou 76.

31. Método de acordo com a reivindicação 30, caracterizado pelo fato de que os motivos isolados de promotores ou agrupamentos de motivos de promotores são combinados, referidos motivos ou agrupamentos de motivos compreendendo pelo menos 5 dos motivos descritos por qualquer uma dentre SEQ ID NO: 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, ou 100.

32. Método de acordo com a reivindicação 30 ou 31,

caracterizado pelo fato de que os motivos de promotores isolados ou agrupamentos de motivos de promotores são combinados, referidos motivos ou agrupamentos de motivos compreendendo pelo menos 5 dos motivos descritos por qualquer uma dentre SEQ ID NO: 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112,: 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, ou 124.

33. Cassete de expressão, caracterizado pelo fato de compreender

i) pelo menos uma seqüência nucleotídica reguladora da transcrição capaz de mediar a transcrição em plantas e funcionalmente ligado à mesma,

ii) pelo menos uma seqüência de ácido nucleico, e funcionalmente ligada à mesma

iii) pelo menos uma seqüência selecionada dentre o grupo consistindo de:

a) as seqüências descritas por SEQ ID NO: 125, 126, e 127, e

b) uma seqüência compreendendo um fragmento de pelo menos 25 nucleotídeos consecutivos de uma seqüência descrita por SEQ ID NO: 125, 126, ou 127,

c) uma seqüência tendo uma identidade de pelo menos 50% para uma seqüência descrita por SEQ ID NO: 125, 126, ou 127

d) uma seqüência capaz de hibridizar sob condições equivalentes à hibridização em 7% dodecil sulfato de sódio (SDS), 0,5 M $NaPO_4$, 1 mM EDTA a 50°C com lavagem em 2 X SSC, 0,1% SDS a 50°C para um ácido nucleico compreendendo 50 a 200 ou mais (preferivelmente todos) nucleotídeos consecutivos de uma seqüência descrita por SEQ ID NO: 125, 126, ou 127, ou o seu complemento,

em que referida seqüência iii) é capaz de mediar estabilização de RNA, estabilização de expressão de RNA, terminação de transcrição de

RNA e/ou poliadenilação de RNA; e

em que ou referida seqüência reguladora de transcrição i) ou referida seqüência iii) ou ambas i e iii) são heterólogas com relação à referida seqüência nucleotídica ii).

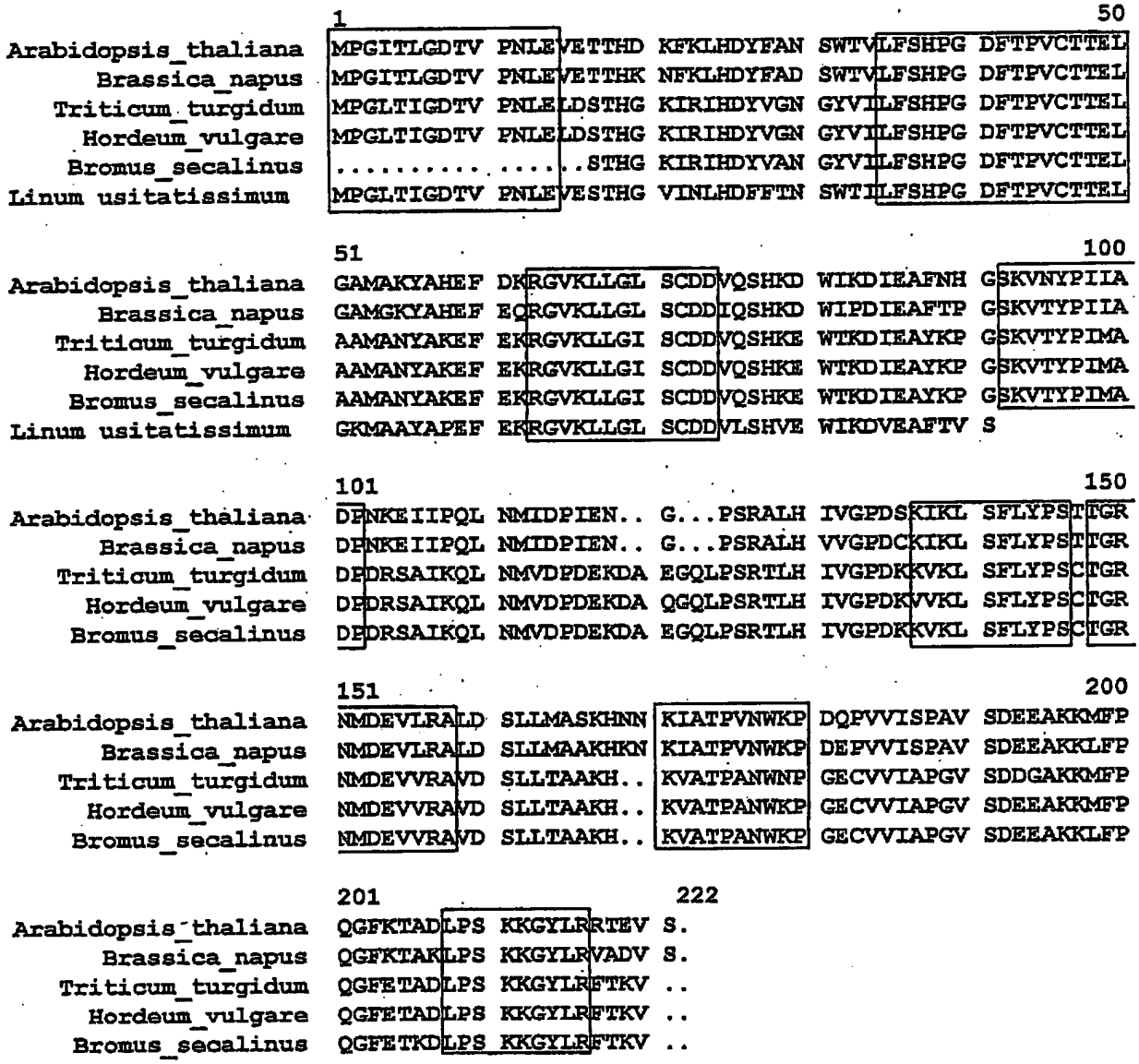

1 50
Arabidopsis_thaliana MPGITLGDTV PNLEVETTHD KFKLHDYFAN SWTVLFSHPG DFTPVCTTEL
Brassica_napus MPGITLGDTV PNLEVETTHK NFKLHDYFAD SWTVLFSHPG DFTPVCTTEL
Triticum_turgidum MPGLTIGDTV PNLELDSTHG KIRIHDYVGN GYVILFSHPG DFTPVCTTEL
Hordeum_vulgare MPGLTIGDTV PNLELDSTHG KIRIHDYVGN GYVILFSHPG DFTPVCTTEL
Bromus_secalinus .....STHG KIRIHDYVAN GYVILFSHPG DFTPVCTTEL
Linum usitatissimum MPGLTIGDTV PNLEVESTHG VINLHDFFTN SWTILFSHPG DFTPVCTTEL
51 100
Arabidopsis_thaliana GAMAKYAHEF DKRGVKLLGL SCDDVQSHKD WIKDIEAFNH GSKVNYPIIA
Brassica_napus GAMGKYAHEF EQRGVKLLGL SCDDIQSHKD WIPDIEAFTP GSKVTYPIIA
Triticum_turgidum AAMANYAKEF EKRGVKLLGI SCDDVQSHKE WTKDIEAYKP GSKVTYPIMA
Hordeum_vulgare AAMANYAKEF EKRGVKLLGI SCDDVQSHKE WTKDIEAYKP GSKVTYPIMA
Bromus_secalinus AAMANYAKEF EKRGVKLLGI SCDDVQSHKE WTKDIEAYKP GSKVTYPIMA
Linum usitatissimum GKMAAYAPEF EKRGVKLLGL SCDDVLSHVE WIKDVEAFTV S
101 150
Arabidopsis_thaliana DFNKEIIPQL NMIDPIEN.. G...PSRALH IVGPDSKIKL SFLYPSTTGR
Brassica_napus DFNKEIIPQL NMIDPIEN.. G...PSRALH VVGPDCKIKL SFLYPSTTGR
Triticum_turgidum DFDRSAIKQL NMVDPDEKDA EGQLPSRTLH IVGPDKKVKL SFLYPSCTGR
Hordeum_vulgare DFDRSAIKQL NMVDPDEKDA QGQLPSRTLH IVGPDKVVKL SFLYPSCTGR
Bromus_secalinus DFDRSAIKQL NMVDPDEKDA EGQLPSRTLH IVGPDKKVKL SFLYPSCTGR
151 200
Arabidopsis_thaliana NMDEVLRALD SLLMASKHNN KIATPVNWKP DQPVVISPAV SDEEAKKMFP
Brassica_napus NMDEVLRALD SLLMAAKHKN KIATPVNWKP DEPVVISPAV SDEEAKKLFP
Triticum_turgidum NMDEVVRAVD SLLTAAKH.. KVATPANWNP GECVVIAPGV SDDGAKKMFP
Hordeum_vulgare NMDEVVRAVD SLLTAAKH.. KVATPANWKP GECVVIAPGV SDEEAKKMFP
Bromus_secalinus NMDEVVRAVD SLLTAAKH.. KVATPANWKP GECVVIAPGV SDEEAKKLFP
201 222
Arabidopsis_thaliana QGFKTADLPS KKGYLRRTEV S.
Brassica_napus QGFKTAKLPS KKGYLRVADV S.
Triticum_turgidum QGFETADLPS KKGYLRFTKV ..
Hordeum_vulgare QGFETADLPS KKGYLRFTKV ..
Bromus_secalinus QGFETKDLPS KKGYLRFTKV ..

Fig. 1

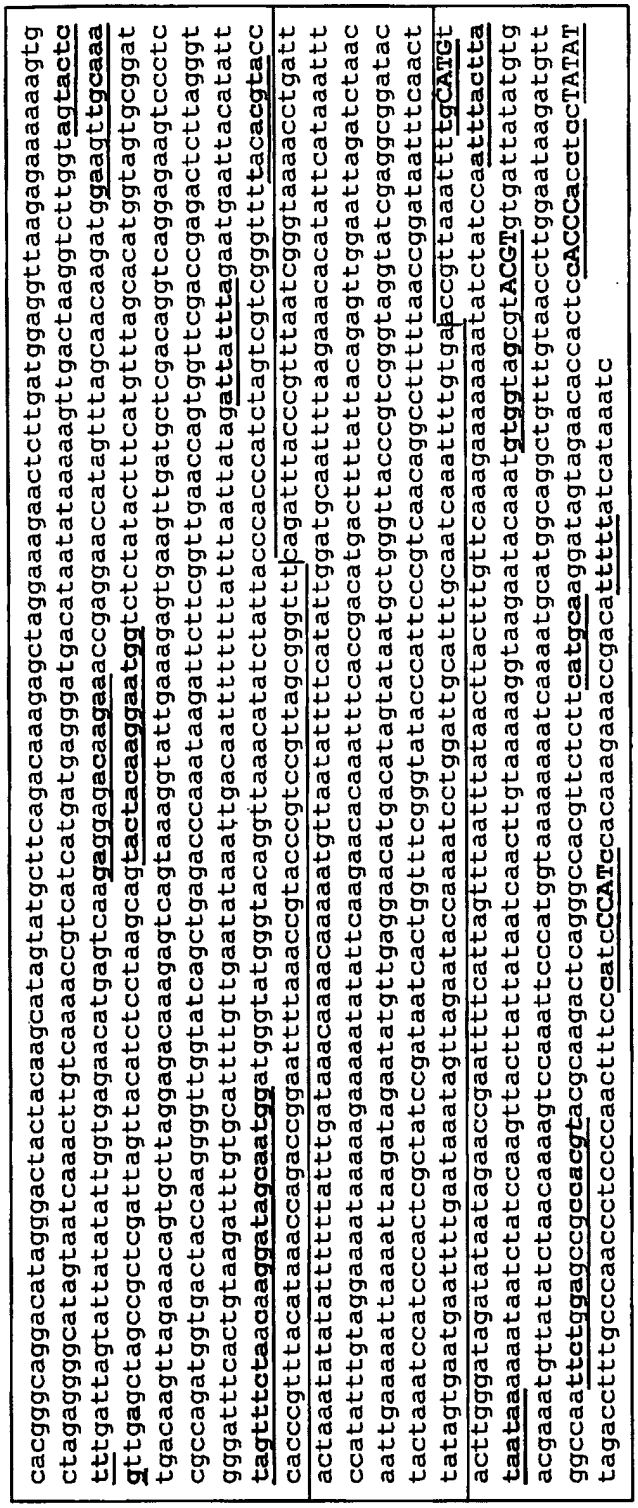


Fig. 2

RESUMO

"CASSETES DE EXPRESSÃO, SEQÜÊNCIA NUCLEOTÍDICA ISOLADA, VETOR, CÉLULA HOSPEDEIRA TRANSGÊNICA OU ORGANISMO NÃO HUMANO, CÉLULA DE PLANTA OU PLANTA TRANSGÊNICA, MÉTODOS PARA IDENTIFICAR E/OU ISOLAR UMA SEQÜÊNCIA NUCLEOTÍDICA REGULADORA DA TRANSCRIÇÃO, PARA PROVER OU PRODUZIR UM CASSETE DE EXPRESSÃO TRANSGÊNICA, E PARA PROVER UMA SEQÜÊNCIA NUCLEOTÍDICA REGULADORA DE TRANSCRIÇÃO SINTÉTICA, E, SEQÜÊNCIA REGULADORA DE TRANSCRIÇÃO SINTÉTICA"

A presente invenção refere-se a cassetes de expressão compreendendo seqüências nucleotídicas reguladoras de transcrição com perfis de expressão preferenciais de semente ou específicos de semente em plantas obteníveis a partir do gene de Arabidopsis thaliana descrito pelo *locus* Atlg48130 ou seus ortólogos de outras espécies de plantas (como Linum usitatissimum). As seqüências nucleotídicas reguladoras de transcrição preferivelmente demonstram uma forte atividade de expressão especialmente em sementes e tecidos de semente.