



**Assinado
Digitalmente**

REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA INDÚSTRIA, COMÉRCIO EXTERIOR E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

CARTA PATENTE Nº PI 0312771-0

O INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL concede a presente PATENTE DE INVENÇÃO, que outorga ao seu titular a propriedade da invenção caracterizada neste título, em todo o território nacional, garantindo os direitos dela decorrentes, previstos na legislação em vigor.

(21) Número do Depósito: PI 0312771-0

(22) Data do Depósito: 10/07/2003

(43) Data da Publicação do Pedido: 29/01/2004

(51) Classificação Internacional: C12N 15/09; C12N 15/63; C12N 15/82; C12N 15/87; C12N 15/90

(30) Prioridade Unionista: US 60/396,665 de 18/07/2002

(54) Título: MÉTODOS PARA REDUZIR SILENCIAMENTO DE TRANSGENE E PARA DETECTAR POLINUCLEOTÍDEO EM PLANTAS, MOLÉCULA E CONSTRUTO DE DNA, BEM COMO KIT DE DETECÇÃO DE DNA

(73) Titular: MONSANTO TECHNOLOGY LLC, Sociedade Norte-Americana. Endereço: 800 North Lindbergh Boulevard, St. Louis, MO 63167, ESTADOS UNIDOS DA AMÉRICA(US)

(72) Inventor: STANISLAW FLASINSKI

Código de Controle: BB86314C31EB417B C5DCAD1E72219BF9

Prazo de Validade: 10 (dez) anos contados a partir de 02/10/2018, observadas as condições legais

Expedida em: 02/10/2018

Assinado digitalmente por:
Liane Elizabeth Caldeira Lage
Diretora de Patentes, Programas de Computador e Topografias de Circuitos Integrados

Relatório Descritivo da Patente de Invenção para **"MÉTODOS PARA REDUZIR SILENCIAMENTO DE TRANSGENE E PARA DETECTAR POLINUCLEOTÍDEO EM PLANTAS, MOLÉCULA E CONSTRUTO DE DNA, BEM COMO KIT DE DETECÇÃO DE DNA".**

5 FUNDAMENTOS DA INVENÇÃO

Campo da Invenção

A presente invenção refere-se ao engendramento genético vegetal. Mais particularmente, a um método para construir um polinucleotídeo artificial e aos métodos de uso para reduzir o silenciamento de transgene em plantas. A invenção também se refere às células vegetais contendo o polinucleotídeo artificial em que uma célula vegetal é transformada para expressar o polinucleotídeo artificial e a planta regenerada a partir deste.

Descrição da Técnica Relacionada

Os genes heterólogos podem ser isolados a partir de uma fonte outra que não a planta na qual eles serão transformados ou eles podem ser modificados ou planejados para ter qualidades diferentes ou melhoradas. Particularmente traços ou qualidades desejadas de interesse para o engendramento genético vegetal podem incluir, mas não são limitadas à resistência aos insetos, doenças fúngicas e outras pragas e agentes causadores de doença, tolerâncias aos herbicidas, estabilidade realçada ou vida de prateleira, rendimento, tolerâncias ao estresse ambiental e realces nutricionais.

Os métodos biológicos moleculares tradicionais para gerar novos genes e proteínas no geral envolveram a mutagênese aleatória ou direcionada. Um exemplo de mutagênese aleatória é uma técnica de recombinação conhecida como "embaralhamento de DNA" como divulgado nas Patentes US 5.605.793, 5.811.238, 5.830.721, 5.837.458 e Pedidos Internacionais WO 98/31837, WO 99/65927, a totalidade de todas as quais é aqui incorporada por referência. Um método alternativo de evolução molecular envolve um processo de extensão escalonado (StEP) para a mutagênese e recombinação *in vitro* de sequências de molécula de ácido nucléico, como divulgado na Patente US 5.965.408, aqui incorporada por referência. Um

exemplo de mutagênese direcionada é a introdução de uma mutação pontual em um sítio específico em um polipeptídeo.

Um método alternativo, útil quando o gene heterólogo é de uma fonte não-vegetal, é planejar um gene inseticida artificial que use o códon mais frequentemente usado na tabela de uso do códon na planta do milho (Koziel *et al.*, 1993, *Biotechnology* 11, 194 a 200). Fischhoff e Perlak (Patente US Nº 5.500.365, aqui incorporada por referência) relatam a expressão mais alta de proteína inseticida de *Bacillus thuringiensis* (Bt) comparada em plantas de safra quando a sequência de polinucleotídeo foi modificada para reduzir a ocorrência de seqüências desestabilizadoras. Foi necessário modificar a sequência de polinucleotídeo de Bt do tipo selvagem porque o polinucleotídeo de Bt de comprimento natural do tipo selvagem não expressou níveis suficientes de proteína inseticida em plantas para ser agronomicamente útil.

Os genes heterólogos são clonados em vetores adequados para a transformação de planta. As técnicas de transformação e regeneração úteis para incorporar genes heterólogos em um genoma da planta são bem conhecidos na técnica. O gene pode ser depois expressado na célula vegetal para exibir a característica ou traço adicionados. Entretanto, os genes heterólogos que normalmente expressam bem como transgenes podem experimentar o silenciamento de gene quando mais do que uma cópia dos mesmos genes é expressada na mesma planta. Isto pode ocorrer quando um primeiro gene heterólogo é muito similar a uma sequência de DNA de gene endógeno na planta. Outros exemplos incluem quando uma planta transgênica é subsequente cruzada com outras plantas transgênicas tendo os mesmos transgenes ou similares ou quando a planta transgênica é retransformada com um cassete de expressão vegetal que contém o mesmo gene ou similar. Similarmente, o silenciamento de gene pode ocorrer se a implantação de traços utilizam os mesmos elementos genéticos usados para direcionar a expressão do gene transgênico de interesse. De modo a implantar traços, linhagens transgênicas estáveis devem ser feitas com combinações diferentes de genes e elementos genéticos para evitar o silencia-

mento de gene.

A N-fosfonometilglicina, também conhecida como glifosato, é um herbicida bem conhecido que tem atividade em um amplo espectro de espécies vegetais. O glifosato é o ingrediente ativo do Roundup® (Monsanto Co.), um herbicida seguro tendo uma meia-vida desejavelmente curta no ambiente. Quando aplicado a uma superfície vegetal, o glifosato move-se sistemicamente através da planta. O glifosato é fitotóxico devido à sua inibição do caminho do ácido chiquímico, que fornece um precursor para a síntese de aminoácidos aromáticos. O glifosato inibe a enzima 5-enolpiruvil-3-fosfochiquimato sintase (EPSPS).

A tolerância ao glifosato também pode ser obtida pela expressão de variantes de EPSPS que têm afinidade mais baixa ao glifosato e portanto retêm a sua atividade catalítica na presença de glifosato (Patente U.S. Nº 5.633.435, aqui incorporada por referência). As enzimas que degradam glifosato em tecidos vegetais (Patente U.S. Nº 5.463.175) também são capazes de conferir tolerância celular ao glifosato. Tais genes são usados para a produção de safras transgênicas que são tolerantes ao glifosato, permitindo deste modo que o glifosato seja usado para o controle eficaz de erva-daninha com preocupação mínima de dano de safra. Por exemplo, a tolerância ao glifosato foi geneticamente engendrada no milho (Patente U.S. Nº 5.554.798), trigo (Pedido de Patente U.S. Nº 20020062503), soja (Pedido de Patente U.S. Nº 20020157139) e canola (WO 9204449), todas as quais são incorporadas por referência. Os transgenes para a tolerância ao glifosato e os transgenes para a tolerância a outros herbicidas, por exemplo, o gene *bar*, (Toki *et al.* Plant Physiol., 100: 1503 a 1507, 1992; Thompson *et al.* EMBO J. 6: 2519 a 2523, 1987, fosfinothricin acetyltransferase, BAR gene isolated from Streptomyces; DeBlock *et al.* EMBO J., 6:2513-2522, 1987, glufosinate herbicide) também são úteis como marcadores selecionáveis ou marcadores registráveis e podem fornecer um fenótipo útil para a seleção de plantas ligadas com outros traços agronomicamente úteis.

O que é necessário na técnica são métodos de planejar genes para a expressão em plantas para melhorar traços agronomicamente úteis

que evitem o silenciamento de gene quando cópias múltiplas são inseridas e a recombinação com genes vegetais endógenos.

BREVE DESCRIÇÃO DOS DESENHOS

Figura 1. Comparação acumulada das mudanças das seqüências de polinucleotídeo de duas versões da EPSPS de arroz artificial (OsEPSPS_AT, OsEPSPS_ZM) e uma EPSPS de arroz nativa (OsEPSPS_Nat) o polipeptídeo de cada uma modificado para ser resistente ao glifosato.

Figura 2. Comparação acumulada das seqüências de polinucleotídeo de uma EPSPS de milho nativa (ZmEPSPS_Nat) e uma artificial (ZmEPSPS_ZM) o polipeptídeo de cada uma modificado para ser resistente ao glifosato.

Figura 3. Comparação acumulada das seqüências de polinucleotídeo de uma EPSPS de soja nativa (GmEPSPS_Nat) e uma versão artificial (GmEPSPS_GM) o polipeptídeo de cada uma modificado para ser resistente ao glifosato.

Figura 4. Comparação acumulada das seqüências de polinucleotídeo de um gene BAR nativo (BAR1_Nat) e duas versões artificiais com códon de eleição de *Zea mays* (BAR1_ZM) e *Arabidopsis thaliana* (BAR1_AT).

Figura 5. Comparação acumulada das seqüências de polinucleotídeo de CTP2 e CP4EPSPS nativas (CTP2CP4_Nat) e versões artificiais (CTP2CP4_AT, CTP2CP4_ZM e CTP2CP4_GM).

Figura 6. Mapa plasmídico de pMON54949.

Figura 7. Mapa plasmídico de pMON54950.

Figura 8. Mapa plasmídico de pMON30151.

Figura 9. Mapa plasmídico de pMON59302.

Figura 10. Mapa plasmídico de pMON59307.

Figura 11. Mapa plasmídico de pMON42411.

Figura 12. Mapa plasmídico de pMON58400.

Figura 13. Mapa plasmídico de pMON58401.

Figura 14. Mapa plasmídico de pMON54964.

Figura 15. Mapa plasmídico de pMON25455.

Figura 16. Mapa plasmídico de pMON30152.

Figura 17. Mapa plasmídico de pMON54992.

Figura 18. Mapa plasmídico de pMON54985.

5 Figura 19. Mapa plasmídico de pMON20999.

Figura 20. Mapa plasmídico de pMON45313.

Figura 21. Mapa plasmídico de pMON59308.

Figura 22. Mapa plasmídico de pMON59309.

Figura 23. Mapa plasmídico de pMON59313.

10 Figura 24. Mapa plasmídico de pMON59396.

Figura 25. Mapa plasmídico de pMON25496.

BREVE DESCRIÇÃO DA LISTAGEM DE SEQUÊNCIA

SEQ ID NO: 1	OsEPSPS_TIPS	Uma sequência de proteína de EPSPS de arroz modificada para ser resistente ao glifosato, com peptídeo de trânsito de cloroplasto.
SEQ ID NO: 2	OsEPSPS_Nat	Sequência de polinucleotídeo de um polinucleotídeo de EPSPS nativa do arroz modificado para codificar uma proteína resistente ao glifosato.
SEQ ID NO: 3	OsEPSPS_AT	Sequência de polinucleotídeo de um polinucleotídeo de EPSPS de arroz artificial usando a tabela de uso do códon de <i>Arabidopsis</i> e os métodos da presente invenção e ainda modificado para codificar uma proteína resistente ao glifosato.

SEQ ID NO: 4	OsEPSPS_ZM	Seqüência de polinucleotídeo de um polinucleotídeo de EPSPS de arroz artificial usando a tabela de uso do códon de <i>Zea mays</i> e os métodos da presente invenção e ainda modificado para codificar uma proteína resistente ao glifosato.
SEQ ID NO: 5	GmEPSPS_IKS	Uma seqüência de proteína de EPSPS da soja modificada para ser resistente ao glifosato, com peptídeo de trânsito de cloroplasto.
SEQ ID NO: 6	GmEPSPS_Nat	Seqüência de polinucleotídeo de um polinucleotídeo de EPSPS nativa da soja modificado para codificar uma proteína resistente ao glifosato.
SEQ ID NO: 7	GmEPSPS_GM	Seqüência de polinucleotídeo de um polinucleotídeo de EPSPS da soja artificial usando a tabela de uso do códon de <i>Glicina max</i> e os métodos da presente invenção e ainda modificado para codificar uma proteína resistente ao glifosato.
SEQ ID NO: 8	ZmEPSPS_TIPS	Uma seqüência de proteína de EPSPS do milho modificada para ser resistente ao glifosato, com peptídeo de trânsito de cloroplasto.
SEQ ID NO: 9	ZmEPSPS_Nat	Seqüência de polinucleotídeo de um polinucleotídeo de EPSPS nativa do milho modificado para codificar uma proteína resistente ao glifosato.

SEQ ID NO: 10	ZmEPSPS_ZM	Seqüência de polinucleotídeo de um polinucleotídeo de EPSPS do milho artificial usando a tabela de uso do códon de <i>Zea mays</i> e os métodos da presente invenção e ainda modificado para codificar uma proteína resistente ao glifosato.
SEQ ID NO: 11	CTP2	Seqüência de proteína do peptídeo de trânsito de cloroplasto 2 do gene de EPSPS de <i>Arabidopsis</i> .
SEQ ID NO: 12	CTP2_Nat	Seqüência de polinucleotídeo do peptídeo de trânsito de cloroplasto de EPSPS de <i>Arabidopsis</i> .
SEQ ID NO: 13	CTP2_AT	Seqüência de polinucleotídeo de um polinucleotídeo artificial que codifica a CTP2 usando a tabela de uso do códon de <i>Arabidopsis</i> e os métodos da presente invenção.
SEQ ID NO: 14	CTP2_ZM	Seqüência de polinucleotídeo de um polinucleotídeo artificial que codifica a CTP2 usando a tabela de uso do códon de <i>Zea mays</i> e os métodos da presente invenção.
SEQ ID NO: 15	CP4EPSPS	A seqüência de proteína da proteína de EPSPS resistente ao glifosato da cepa CP4 de <i>Agrobacterium</i> .
SEQ ID NO: 16	CP4EPSPS_Nat	Seqüência de polinucleotídeo do polinucleotídeo nativo que codifica a proteína CP4EPSPS (Patente U.S. Nº 5.633.435).

SEQ ID NO: 17	CP4EPSPS_AT	Seqüência de polinucleotídeo de um polinucleotídeo artificial que codifica a proteína CP4EPSPS usando a tabela de uso do códon de <i>Arabidopsis</i> e os métodos da presente invenção.
SEQ ID NO: 18	CP4EPSPS_ZM	Seqüência de polinucleotídeo de um polinucleotídeo artificial que codifica a proteína CP4EPSPS usando a tabela de uso do códon de <i>Zea mays</i> e os métodos da presente invenção.
SEQ ID NO: 19	BAR1	A seqüência de proteína de uma fosfinotricina acetiltransferase.
SEQ ID NO: 20	BAR1_Nat	Seqüência de polinucleotídeo do polinucleotídeo nativo isolado de <i>Streptomyces</i> que codifica a fosfinotricina acetiltransferase.
SEQ ID NO: 21	BAR1_AT	Seqüência de polinucleotídeo de um polinucleotídeo artificial que codifica a fosfinotricina acetiltransferase usando a tabela de uso do códon de <i>Arabidopsis</i> e os métodos da presente invenção.
SEQ ID NO: 22	BAR1_ZM	Seqüência de polinucleotídeo de um polinucleotídeo artificial que codifica a fosfinotricina acetiltransferase usando a tabela de uso do códon de <i>Zea mays</i> e os métodos da presente invenção.
SEQ ID NO: 23	CP4EPSPS_Syn	Seqüência de polinucleotídeo de um polinucleotídeo artificial com códon de eleição de dicotiledônea.
SEQ ID NO: 24	CP4EPSPS_AT_p1	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_AT.

SEQ ID NO: 25	CP4EPSPS_AT_p2	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_AT.
SEQ ID NO: 26	CP4EPSPS_ZM_p1	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_ZM.
SEQ ID NO: 27	CP4EPSPS_ZM_p2	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_ZM.
SEQ ID NO: 28	CP4EPSPS_Nat_p1	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_Nat.
SEQ ID NO: 29	CP4EPSPS_Nat_p2	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_Nat.
SEQ ID NO: 30	CP4EPSPS_Syn_p1	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_Syn.
SEQ ID NO: 31	CP4EPSPS_Syn_p2	Diagnóstico de molécula iniciadora de DNA para o polinucleotídeo CP4EPSPS_Syn.
SEQ ID NO: 32	ZmAdh1 iniciador1	Diagnóstico de iniciador de controle 1 para o gene Adh1 do milho endógeno.
SEQ ID NO: 33	ZmAdh1 iniciador2	Diagnóstico de iniciador de controle 2 para o gene Adh1 do milho endógeno.
SEQ ID NO: 34	GNAGIAMKS	Motivo que fornece resistência ao glifosato a uma EPSPS vegetal.
SEQ ID NO: 35	CTPEPSPSCP4_GM	Seqüência de polinucleotídeo de um polinucleotídeo artificial que codifica a proteína CP4EPSPS usando a tabela de uso do códon de <i>Glicina max</i> .

SUMÁRIO DA INVENÇÃO

A presente invenção fornece métodos e composições para planejar uma seqüência de polinucleotídeo artificial que codifica uma proteína de interesse, em que o polinucleotídeo artificial é substancialmente diver-

gente de um polinucleotídeo que ocorre naturalmente em uma planta ou um polinucleotídeo que foi introduzido como um transgene em uma planta e o polinucleotídeo artificial e o polinucleotídeo codificam um polipeptídeo substancialmente idêntico.

5 Os polinucleotídeos artificiais da presente invenção que codificam proteínas que fornecem fenótipos agronomicamente úteis a uma planta transgênica contendo um constructo de DNA que compreende o polinucleotídeo artificial. Os fenótipos agronomicamente úteis incluem, mas não são limitados a: tolerância à seca, rendimento aumentado, tolerância ao frio, resistência às doenças, resistência a inseto e tolerância a herbicida.

Um outro aspecto da presente invenção são polinucleotídeos artificiais que codificam uma proteína de EPSPS resistente a herbicida, uma proteína de fosfinotricina acetiltransferase, uma proteína de peptídeo de trânsito de cloroplasto. Em modalidades preferidas da presente invenção, a
15 molécula de polinucleotídeo artificial é selecionada do grupo que consiste na SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 21, SEQ ID NO: 22 e SEQ ID NO: 35.

A presente invenção fornece constructos de DNA que compreendem: uma molécula promotora que funciona em plantas, operavelmente ligada a uma molécula de polinucleotídeo artificial da presente invenção, em que a molécula de polinucleotídeo artificial é selecionada do grupo que consiste na SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 21, SEQ ID NO: 22 e SEQ ID NO: 35, operavelmente ligada
20 a uma região de término de transcrição.

A presente invenção ainda fornece constructos de DNA que compreendem: uma molécula promotora que funciona em plantas, operavelmente ligada a uma molécula de polinucleotídeo artificial que codifica um peptídeo de trânsito de cloroplasto, operavelmente ligada a uma EPSPS resistente ao glifosato heteróloga, operavelmente ligada a uma região de sinal de término de transcrição, em que o polinucleotídeo artificial é substancial-

30

mente divergente na seqüência de polinucleotídeo dos polinucleotídeos conhecidos que codificam um peptídeo de trânsito de cloroplasto idêntico.

A presente invenção fornece constructos de DNA que compreendem pelo menos dois cassetes de expressão, o primeiro cassete de expressão compreendendo uma molécula promotora que funciona em plantas, operavelmente ligado a uma molécula de polinucleotídeo artificial da presente invenção, operavelmente ligada a uma região de sinal de término de transcrição e o segundo cassete de expressão que compreende uma molécula promotora que funciona em plantas, operavelmente ligado a uma molécula de polinucleotídeo que codifica um polipeptídeo substancialmente idêntico como o dito polinucleotídeo artificial e é menos do que oito-cinco por cento similar na seqüência de polinucleotídeo ao dito polinucleotídeo artificial, operavelmente ligada a uma região de sinal de término de transcrição.

A presente invenção fornece células vegetais, plantas ou sua progênie que compreendem um constructo de DNA da presente invenção. De interesse particular são as plantas de sua progênie selecionadas do grupo que consiste em trigo, milho, arroz, soja, algodão, batata, canola, grama-do, árvores de floresta, sorgo em grão, safras vegetais, plantas ornamentais, safras de forragem e safras de frutas.

Um método da presente invenção reduz o silenciamento de gene durante a geração das plantas transgênicas compreendendo as etapas de:

a) construir um polinucleotídeo artificial que é substancialmente divergente de polinucleotídeos conhecidos que codificam uma proteína substancialmente idêntica e

b) construir um constructo de DNA contendo a dita molécula de polinucleotídeo artificial; e

c) transformar o dito constructo de DNA em uma célula vegetal; e

d) regenerar a dita célula vegetal em uma planta transgênica; e

e) cruzar a dita planta transgênica com uma planta fértil, em que a dita planta fértil contém uma molécula de polinucleotídeo que codifica uma proteína substancialmente idêntica a uma proteína codificada pela dita molé-

cula de polinucleotídeo artificial e em que a dita molécula de polinucleotídeo artificial e a dita molécula de polinucleotídeo são substancialmente divergentes.

Um outro aspecto da invenção é uma célula vegetal transgênica que compreende dois polinucleotídeos, em que pelo menos um dos polinucleotídeos é um transgene e os dois polinucleotídeos codificam uma proteína substancialmente idêntica e são menos do que oito-cinco por cento similares na seqüência de polinucleotídeo.

Um outro aspecto da presente invenção em um método para reduzir o silenciamento de gene durante a produção de plantas transgênicas compreende as etapas de:

- a) construir um polinucleotídeo artificial que seja substancialmente divergente de polinucleotídeos conhecidos que codificam uma proteína substancialmente idêntica e
- b) construir um primeiro constructo de DNA contendo a dita molécula de polinucleotídeo artificial; e
- c) transformar o dito constructo de DNA em uma célula vegetal; e
- d) regenerar a dita célula vegetal em uma planta transgênica; e
- e) retransformar uma célula a partir da dita planta transgênica com um segundo constructo de DNA que compreende uma molécula de polinucleotídeo que codifica uma proteína substancialmente idêntica ao dito polinucleotídeo artificial e o dito polinucleotídeo e o polinucleotídeo artificial são substancialmente divergentes na seqüência de polinucleotídeo; e
- f) regenerar a dita célula da etapa d em uma planta transgênica que compreende tanto o dito polinucleotídeo artificial quanto o dito polinucleotídeo.

Ainda são fornecidos pela presente invenção métodos para a seleção de uma planta transformada com um constructo de DNA da invenção que compreende as etapas de:

- a) transformar uma célula vegetal com um constructo de DNA da presente invenção; e

b) cultivar a dita célula vegetal em um meio seletivo contendo um herbicida selecionado do grupo que consiste em: glifosato e glifosinato, para matar seletivamente as células que não foram transformadas com os ditos constructos de DNA; e

5 c) regenerar a dita célula vegetal em uma planta fértil.

Um outro aspecto da invenção é um método de detectar um polinucleotídeo artificial em uma célula vegetal, planta ou sua progênie transgênica que compreende as etapas:

a) contatar uma amostra de DNA isolada da dita célula vegetal, planta ou sua progênie com uma molécula de DNA, em que a dita molécula de DNA compreende pelo menos uma molécula de DNA de um par de moléculas de DNA que quando usado em uma reação de amplificação de ácido nucleico produz um amplicon que é diagnóstico para a dita molécula de polinucleotídeo artificial selecionada do grupo que consiste na: SEQ ID NO: 3, 10
SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 21, SEQ ID NO: 22 e SEQ ID NO: 35.

(a) realizar uma reação de amplificação de ácido nucleico, produzindo deste modo o amplicon; e

20 (b) detectar o amplicon.

Os reagentes fornecidos para realizar o método de detecção acima incluem, mas não são limitados a: moléculas de DNA que especificamente hibridizam com uma molécula de polinucleotídeo artificial selecionada do grupo que consiste em: SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, 25
SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 21 e SEQ ID NO: 22; e moléculas de DNA isoladas selecionadas do grupo que consiste em: SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 e SEQ ID NO: 27.

A presente invenção fornece plantas e progênie que compreende uma molécula de DNA selecionada do grupo que consiste em: SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO:

21, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 e SEQ ID NO: 27.

A presente invenção fornece pares de moléculas de DNA selecionados do grupo que compreendem: uma primeira molécula de DNA e uma
5 segunda molécula de DNA, em que a primeira molécula de DNA é a SEQ ID NO: 24 ou seu complemento e a segunda molécula de DNA é a SEQ ID NO: 25 ou seu complemento e o par de moléculas de DNA quando usado em um método de amplificação de DNA produz um amplicon e uma primeira molécula de DNA e uma segunda molécula de DNA, em que a primeira molécula
10 de DNA é a SEQ ID NO: 26 ou seu complemento e a segunda molécula de DNA é a SEQ ID NO: 27 ou seu complemento e o par de moléculas de DNA quando usados em um método de amplificação de DNA produz um amplicon, em que o amplicon é diagnóstico para a presença de um polinucleotídeo artificial da presente invenção no genoma de uma planta transgênica.

A presente invenção fornece uma planta e sua progênie identificada por um método de amplificação de DNA para conter no seu genoma uma molécula de DNA selecionada do grupo que consiste em: SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 21,
20 SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 e SEQ ID NO: 27.

A presente invenção fornece e considera kits de detecção de DNA que compreendem: pelo menos uma molécula de DNA de comprimento suficiente que seja especificamente homóloga ou complementar a um polinucleotídeo artificial selecionado do grupo que consiste na SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 21 e SEQ ID NO: 22, em que a dita molécula de DNA é útil como uma sonda de DNA ou iniciador de DNA; ou pelo menos uma molécula de DNA homóloga ou
30 complementar a uma molécula iniciadora de DNA selecionada do grupo que consiste em: SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26 e SEQ ID NO: 27.

A presente invenção ainda fornece um método de detectar a presença de um polinucleotídeo artificial que codifica uma EPSPS resistente ao glifosato em uma amostra de DNA, o método compreendendo:

(a) extrair uma amostra de DNA de uma planta; e

5 (b) contatar a amostra de DNA com uma molécula de DNA rotulada de comprimento suficiente que seja especificamente homóloga ou complementar a um polinucleotídeo artificial selecionado do grupo que consiste na: SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17 e SEQ ID NO: 18, em que a dita molécula de DNA rotulada é uma sonda de DNA; e

10 (c) submeter a amostra e a sonda de DNA às condições de hibridização severa; e

(d) detectar a sonda de DNA hibridizada à amostra de DNA.

A presente invenção fornece um polinucleotídeo isolado que codifica uma enzima de EPSPS, a enzima de EPSPS contém o motivo da SEQ ID NO: 34. A presente invenção fornece um constructo de DNA contendo um polinucleotídeo que codifica a enzima de EPSPS com o motivo da SEQ ID NO: 34. Uma célula vegetal, planta ou sua progênie que é tolerante ao glifosato como um resultado da expressão de uma enzima de EPSPS que contém o motivo da SEQ ID NO: 34 é um aspecto da invenção.

DESCRIÇÃO DETALHADA DA INVENÇÃO

As seguintes definições são fornecidas para melhor definir a presente invenção e para guiar aqueles versados na técnica na prática da presente invenção. A menos que de outro modo mencionado, os termos devem ser entendidos de acordo com o uso convencional por aqueles versados na técnica relevante. As definições de termos comuns na biologia molecular também podem ser encontradas em Rieger *et al.*, Glossary of Genetics: Classical e Molecular, 5ª edição, Springer-Verlag: Nova Iorque, (1991); e Lewin, Genes V, Oxford University Press: Nova Iorque, (1994). A nomenclatura para as bases de DNA como apresentada em 37 CFR § 1.822 é usada. A nomenclatura de uma e três letras padrão para os resíduos de aminoácido é usada.

"Substituições de aminoácido", "variantes de aminoácido", são preferivelmente substituições de resíduo de aminoácido único para um outro resíduo de aminoácido em qualquer posição dentro da proteína. As substituições, supressões, inserções ou qualquer combinação destas podem ser combinadas para chegar em um constructo final.

Um "polinucleotídeo artificial" como usado na presente invenção é uma seqüência de DNA planejada de acordo com os métodos da presente invenção e criada como uma molécula de DNA isolada para o uso em um constructo de DNA que forneça a expressão de uma proteína em células hospedeiras e para os propósitos de clonagem em constructos apropriadas ou outros usos conhecidos por aqueles versados na técnica. Programas de computador estão disponíveis para estes propósitos, incluindo mas não limitados aos programas "BestFit" ou "Gap" do Sequence Analysis Software Package, Genetics Computer Group (GCG), Inc., University of Wisconsin Biotechnology Center, Madison, WI 53711. O polinucleotídeo artificial pode ser criado por um ou mais métodos conhecidos na técnica, que incluem, mas não são limitados à: PCR de sobreposição. Um polinucleotídeo artificial da presente invenção é substancialmente divergente de outros polinucleotídeos que codificam a proteína idêntica ou quase idêntica.

O termo "quimérico" refere-se a uma seqüência de ácido nucléico ou de proteína de fusão. Uma seqüência codificadora de ácido nucléico quimérica é compreendida de duas ou mais seqüências unidas na matriz que codificam uma proteína quimérica. Um gene quimérico refere-se aos elementos genéticos múltiplos derivados de fontes heterólogas que compreendem um gene.

As frases "seqüência codificadora", "matriz de leitura aberta" e "seqüência estrutural" referem-se à região de tripletos de ácido nucléico seqüenciais contínuos que codificam uma proteína, polipeptídeo ou seqüência de peptídeo.

"Códon" refere-se a uma seqüência de três nucleotídeos que especificam um aminoácido particular.

"Uso de códon" ou "códon de eleição" referem-se à freqüência

de uso de códons que codificam aminoácidos nas seqüências codificadoras de organismos.

"Complementaridade" e "complemento" quando se referem às seqüências de ácido nucléico, referem-se à ligação específica de adenina à timina (uracila no RNA) e citosina à guanina nos filamentos opostos de DNA ou RNA.

"Constructo" refere-se aos elementos genéticos heterólogos operavelmente ligados entre si compondo uma molécula de DNA recombinante e pode compreender elementos que fornecem a expressão de uma molécula de polinucleotídeo de DNA em uma célula hospedeira e elementos que fornecem a manutenção do constructo.

"Região de terminal C" refere-se à região de uma cadeia de peptídeo, polipeptídeo ou proteína do seu centro até a extremidade que carrega o aminoácido tendo um grupo carboxila livre.

O termo "divergente", como aqui usado, refere-se à comparação das moléculas de polinucleotídeo que codificam a mesma ou quase a mesma proteína ou polipeptídeo. O código genético de quatro letras (A, G, C e T/U) compreende códons de três letras que direcionam as moléculas de t-RNA para montar aminoácidos em um polipeptídeo a partir de um padrão de mRNA. Tendo mais do que um códon que pode codificar o mesmo aminoácido é aludido como degenerado. Os códons degenerados são usados para construir moléculas de polinucleotídeo substancialmente divergentes que codificam o mesmo polipeptídeo onde estas moléculas têm uma seqüência de nucleotídeos do seu comprimento completo em que eles são menos do que 85 % idênticos e não existe nenhum comprimento de seqüência de polinucleotídeo maior do que 23 nucleotídeos que sejam idênticos.

O termo "que codifica DNA" refere-se ao DNA cromossômico, DNA plasmídico, cDNA ou polinucleotídeo de DNA artificial que codifica qualquer uma das proteínas aqui discutidas.

O termo "endógeno" refere-se aos materiais que se originam de dentro de um organismo ou célula.

"Endonuclease" refere-se a uma enzima que hidrolisa DNA de

filamento duplo nos locais internos.

"Exógeno" refere-se a materiais que se originam de fora de um organismo ou célula. Isto tipicamente se aplica às moléculas de ácido nucleico usadas na produção de células hospedeiras e plantas transformadas ou transgênicas.

"Éxon" refere-se à porção de um gene que é realmente traduzida em proteína, isto é, uma seqüência codificadora.

O termo "expressão" refere-se à transcrição ou tradução de um polinucleotídeo para produzir um produto de gene correspondente, um RNA ou proteína.

"Fragmentos". Um fragmento de um gene é uma porção de uma molécula de ácido polinucleico de comprimento total que é de pelo menos um comprimento mínimo capaz de transcrição em um RNA, tradução em um peptídeo ou útil como uma sonda ou iniciador em um método de detecção de DNA.

O termo "gene" refere-se ao DNA cromossômico, DNA plasmídico, cDNA, polinucleotídeo de DNA artificial ou outro DNA que codifica uma molécula de peptídeo, polipeptídeo, proteína ou RNA e os elementos genéticos que flanqueiam a seqüência codificadora que estão envolvidos na regulação da expressão.

O termo "genoma" como se aplica aos vírus abrange todas as seqüências de ácido nucleico contidas dentro do capsídeo do vírus. O termo "genoma" como se aplica às bactérias abrange tanto o cromossoma quanto os plasmídeos dentro de uma célula hospedeira bacteriana. Os ácidos nucleicos codificadores da presente invenção introduzidos em células hospedeiras bacterianas podem ser portanto cromossomicamente integrados ou localizados no plasmídeo. O termo "genoma" como se aplica às células vegetais abrange não apenas o DNA cromossômico encontrado dentro do núcleo, mas à organela de DNA encontrada dentro dos componentes subcelulares da célula. Os ácidos nucleicos da presente invenção introduzidos nas células vegetais podem estar portanto cromossomicamente integrados ou localizados na organela.

"Glifosato" refere-se à N-fosfonometilglicina e seus sais, o Glifosato é o ingrediente ativo do herbicida Roundup® (Monsanto Co.). Os tratamentos de plantas com "glifosato" referem-se aos tratamentos com as formulações de herbicida Roundup® ou Roundup Ultra®, a menos que de outro modo estabelecido. O Glifosato como N-fosfonometilglicina e seus sais (não formulado como o herbicida Roundup®) são componentes de meios de cultura sintéticos usados para a seleção de bactérias e tolerância da planta ao glifosato ou usados para determinar a resistência à enzima em ensaios bioquímicos *in vitro*.

Seqüência de "DNA heterólogo" refere-se a uma seqüência de polinucleotídeo que se origina de uma fonte ou espécies estranhas ou, se da mesma fonte, é modificada a partir da sua forma original.

"DNA homólogo" refere-se ao DNA da mesma fonte como aquela da célula receptora.

"Hibridização" refere-se à capacidade de um filamento de ácido nucléico para unir com um filamento complementar por intermédio do emparelhamento de base. A hibridização ocorre quando seqüências complementares nos dois filamentos de ácido nucléico se ligam entre si. As sondas e iniciadores de ácido nucléico da presente invenção hibridizam sob condições severas a uma seqüência de DNA alvo. Qualquer método de hibridização ou amplificação de ácido nucléico convencional pode ser usado para identificar a presença de DNA a partir de um evento transgênico em uma amostra. As moléculas de ácido nucléico ou seus fragmentos são capazes de hibridizar especificamente a outras moléculas de ácido nucléico sob certas circunstâncias. Como aqui usado, duas moléculas de ácido nucléico são ditas serem capazes de hibridizar especificamente entre si se as duas moléculas forem capazes de formar uma estrutura de ácido nucléico antiparalela, de filamento duplo. Uma molécula de ácido nucléico é dita ser o "complemento" de uma outra molécula de ácido nucléico se elas exibem complementaridade completa. Como aqui usado, as moléculas são ditas exibir "complementaridade completa" quando cada nucleotídeo de uma das moléculas é complementar a um nucleotídeo da outra. Duas moléculas são ditas serem "minimamente

complementar" se elas podem hibridizar entre si com estabilidade suficiente para permitir que elas permaneçam recozidas entre si pelo menos sob condições de "severidade baixa" convencionais. Similarmente, as moléculas são ditas serem "complementares" se elas podem hibridizar entre si com estabilidade suficiente para permitir que elas permaneçam recozidas entre si sob condições de "alta severidade" convencionais. As condições de severidade convencionais são descritas por Sambrook *et al.*, 1989 e por Haymes *et al.*, Em: Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, DC (1985), aqui incorporada por referência em sua totalidade. Afastamentos da complementaridade completa são portanto permissíveis, contanto que tais afastamentos não impeçam completamente a capacidade das moléculas para formarem uma estrutura de filamento duplo. De modo que uma molécula de ácido nucléico sirva como um iniciador ou sonda ela necessita ser apenas suficientemente complementar em seqüência para ser capaz de formar uma estrutura de filamento duplo estável sob o solvente e concentrações salinas particulares utilizados.

Como aqui usado, uma seqüência substancialmente homóloga é uma seqüência de ácido nucléico que especificamente hibridizará com o complemento da seqüência de ácido nucléico ao qual ela está sendo comparada sob condições de alta severidade. O termo "condições severas" é funcionalmente definido com respeito à hibridização de uma sonda de ácido nucléico a um ácido nucléico alvo (isto é, a uma seqüência de ácido nucléico particular de interesse) pelo procedimento de hibridização específico debatido em Sambrook *et al.*, 1989, em 9.52-9.55. Ver também, Sambrook *et al.*, 1989 em 9.47-9.52, 9.56-9.58 aqui incorporada por referência em sua totalidade; Kanehisa, (Nucl. Acids Res. 12: 203 a 213, 1984, aqui incorporada por referência em sua totalidade); e Wetmur e Davidson, (J. Mol. Biol. 31: 349 a 370, 1988, aqui incorporada por referência em sua totalidade). Conseqüentemente, as seqüências de nucleotídeo da invenção podem ser usadas quanto a sua capacidade para formar seletivamente moléculas duplex com trechos complementares de fragmentos de DNA. Dependendo da aplicação prevista, desejar-se-á utilizar condições variáveis de hibridização para se

obter graus variáveis de seletividade de sonda em relação à sequência alvo. Para aplicações que requerem seletividade alta, desejar-se-á tipicamente utilizar condições relativamente severas para formar os híbridos, por exemplo, selecionar-se-á condições relativamente de sal baixo e/ou temperatura

5 alta, tal como fornecidas por cerca de 0,02 M a cerca de 0,15 M de NaCl em temperaturas de cerca de 50° C a cerca de 70° C. Uma das condições severas, por exemplo, é lavar o filtro de hibridização pelo menos duas vezes com tampão de lavagem de alta severidade (0,2X SSC, 0,1 % de SDS, 65° C). As condições de severidade apropriadas que promovem a hibridização de DNA,

10 por exemplo, 6,0 x cloreto de sódio/citrato de sódio (SSC) a cerca de 45° C, seguida por uma lavagem de 2,0 x SSC a 50° C, são conhecidas por aqueles versados na técnica ou podem ser encontradas em Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Por exemplo, a concentração salina na etapa de lavagem pode ser selecionada de

15 uma severidade baixa de cerca de 2,0 x SSC a 50° C até uma severidade alta de cerca de 0,2 x SSC a 50° C. Além disso, a temperatura na etapa de lavagem pode ser aumentada de condições de severidade baixa na temperatura ambiente, cerca de 22° C, até condições de alta severidade a cerca de 65° C. Tanto a temperatura quanto o sal podem ser variados ou a tempe-

20 ratura ou a concentração de sal podem ser mantidos constantes enquanto a outra variável é mudada. Tais condições seletivas toleram pouco, se algum, desemparelhamento entre a sonda e o padrão ou filamento alvo. A detecção de seqüências de DNA por intermédio da hibridização é bem conhecida por aqueles versados na técnica e as divulgações das Patentes U.S. N^{os}

25 4.965.188 e 5.176.995 são exemplares dos métodos de análises de hibridização.

"Identidade" refere-se ao grau de similaridade entre duas seqüências de ácido polinucleico ou proteína. Um alinhamento das duas seqüências é realizado por um programa de computador adequado. Um pro-

30 grama de computador amplamente usado e aceito para realizar alinhamentos de seqüência é o CLUSTALW v1.6 (Thompson, *et al.* Nucl. Acids Res., 22: 4673 a 4680, 1994). O número de bases ou aminoácidos emparelhados

é dividido pelo número total de bases ou aminoácidos e multiplicado por 100 para se obter uma identidade percentual. Por exemplo, se duas seqüências de 580 pares de base tiveram 145 bases emparelhadas, elas podem ser 25 por cento idênticas. Se as duas seqüências comparadas são de comprimentos diferentes, o número de emparelhamentos é dividido pelo mais curto dos dois comprimentos. Por exemplo, se existem 100 aminoácidos emparelhados entre uma proteína de 200 e uma de 400 aminoácidos, elas são 50 por cento idênticas com respeito à seqüência mais curta. Se a seqüência mais curta é menor do que 150 bases ou 50 aminoácidos no comprimento, o número de emparelhamentos é dividido por 150 (para as bases de ácido nucleico) ou 50 (para os aminoácidos) e multiplicados por 100 para se obter uma identidade percentual.

Como aqui descrito uma proteína pode ser "substancialmente idêntica" às proteínas relacionadas. Estas proteínas com identidade substancial no geral compreendem pelo menos uma seqüência de polipeptídeo que tem pelo menos noventa e oito de identidade de seqüência percentual comparada com a sua outra seqüência de polipeptídeo relacionada. O programa Gap no WISCONSIN PACKAGE versão 10.0-UNIX da Genetics Computer Group, Inc. está fundamentado no método de Needleman e Wunsch (J. Mol. Biol. 48: 443 a 453, 1970) usando o conjunto de parâmetros de padrão para a comparação pelo modo de pares (para a comparação de seqüência de aminoácido: Penalidade de Criação de Intervalo = 8, Penalidade de Extensão de Intervalo = 20); ou usando o programa TBLASTN no conjunto de software BLAST 2.2.1 (Altschul *et al.*, Nucleic Acids Res. 25: 3389 a 3402), usando a matriz BLOSUM62 (Henikoff e Henikoff, Proc. Natl. Acad. Sci. U.S.A. 89: 10915 a 10919, 1992) e o conjunto de parâmetros padrão para a comparação pelo modo de pares (valor de criação de intervalo = 11, valor para a extensão de intervalo = 1.). No BLAST, o valor E ou o valor de probabilidade, representa o número de alinhamentos diferentes com pontuações equivalentes ou melhores do que as pontuações de alinhamento bruto, S, que são esperadas ocorrer em uma pesquisa de banco de dados por acaso. Quanto mais baixo o valor E, mais significativo o emparelhamento.

to. Porque o tamanho do banco de dados é um elemento nos cálculos do valor E, os valores E obtidos pela aplicação de "BLAST" contra banco de dados públicos, tais como o GenBank, tem no geral aumentado com o tempo para qualquer emparelhamento questão/entrada dado. A identidade percentual refere-se à porcentagem de resíduos de aminoácido identicamente emparelhados que existem ao longo do comprimento daquela porção das seqüências que é alinhada pelo algoritmo BLAST.

"Intron" refere-se a uma porção de um gene não traduzido em proteína, embora ela seja transcrita em RNA.

Uma seqüência de ácido nucléico "isolada" é substancialmente separada ou purificada de outras seqüências de ácido nucléico com as quais o ácido nucléico está normalmente associado na célula do organismo no qual o ácido nucléico naturalmente ocorre, isto é, outro DNA cromossômico ou extracromossômico. O termo abrange os ácido nucléicos que são bioquimicamente purificados de modo a remover substancialmente ácido nucléicos contaminantes e outros componentes celulares. O termo também abrange ácidos nucléicos recombinantes e ácido nucléicos quimicamente sintetizados.

Polipeptídeos "Isolados," "Purificados," "Homogêneos". Um polipeptídeo é "isolado" se ele foi separado dos componentes celulares (ácido nucléicos, lipídeos, carboidratos e outros polipeptídeos) que naturalmente o acompanham ou que é quimicamente sintetizado ou recombinante. Um polipeptídeo monomérico é isolado quando pelo menos 60 % em peso de uma amostra é composta do polipeptídeo, preferivelmente 90 % ou mais, mais preferivelmente 95 % ou mais e o mais preferivelmente mais do que 99 %. A pureza ou homogeneidade da proteína são indicadas, por exemplo, pela eletroforese em gel de poliacrilamida de uma amostra de proteína, seguido pela visualização de uma única faixa de polipeptídeo no tingimento do gel de poliacrilamida; a cromatografia líquida de alta pressão; ou outros métodos convencionais. As proteínas podem ser purificadas por qualquer um dos meios conhecidos na técnica, por exemplo como descrito no *Guide to Protein Purification*, ed. Deutscher, Meth. Enzymol. 185. Academic Press, San

Diego, 1990; e Scopes, *Protein Purification: Principles e Practice*, Springer Verlag, Nova Iorque, 1982.

"Rotulação" ou "rotulado". Existe uma variedade de métodos e reagentes convencionais para a rotulação de polinucleotídeos e polipeptídeos e fragmentos destes. Os rótulos típicos incluem isótopos, ligandos ou receptores de ligando radioativos, fluoróforos, agentes e enzimas quimioluminescentes. Os métodos para a rotulação e orientação na escolha de rótulos apropriados para os vários propósitos são debatidos, por exemplo, em Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989) e *Current Protocols in Molecular Biology*, ed. Ausubel *et al.*, Greene Publishing e Wiley-Interscience, Nova Iorque, (1992).

"Região codificadora de proteína madura", este termo refere-se à sequência de um produto de proteína processado, isto é, uma EPSP sintase madura que permanece depois que o peptídeo de trânsito de cloroplasto foi removido.

"Nativo", o termo "nativo" no geral refere-se a um ácido polinucleico ou polipeptídeo que ocorrem naturalmente ("tipo selvagem"). Entretanto, no contexto da presente invenção, alguma modificação de um polinucleotídeo e polipeptídeo isolados pode ter ocorrido para fornecer um polipeptídeo com um fenótipo particular, por exemplo, substituição de aminoácido na EPSPS sensível ao glifosato para fornecer uma EPSPS resistente ao glifosato. Para propósitos comparativos na presente invenção, o polinucleotídeo isolado que contém uns poucos nucleotídeos substituídos para fornecer a modificação de aminoácido para a tolerância a herbicida é aludido como o polinucleotídeo "nativo" quando comparado com o polinucleotídeo substancialmente divergente criado pelos métodos da presente invenção. Entretanto, o polinucleotídeo "nativo" modificado desta maneira é não nativo com respeito aos elementos genéticos normalmente encontrados ligados a um polinucleotídeo não-modificado que ocorre naturalmente.

"Região de terminal N" refere-se a uma região de uma cadeia de peptídeo, polipeptídeo ou proteína do aminoácido tendo um grupo amino livre no centro da cadeia.

"Ácido nucléico" refere-se ao ácido desoxirribonucléico (DNA) e ácido ribonucléico (RNA).

Códigos de ácido nucléico: A = adenosina; C = citosina; G = guanosina; T = timidina. Códigos usados para a síntese de oligonucleotí-
 5 deos: N = eqüimolar A, C, G e T; I = desoxiinosina; K = eqüimolar G e T; R = eqüimolar A e G; S = eqüimolar C e G; W = eqüimolar A e T; Y = eqüimolar C e T.

Um "segmento de ácido nucléico" ou uma "segmento de molé-
 cula de ácido nucléico" é uma molécula de ácido nucléico que foi isolada
 10 livre de DNA genômico total de uma espécie particular ou que foi sintetizado. Incluído com o termo "segmento de ácido nucléico" estão os segmentos de DNA, vetores recombinantes, plasmídeos, cosmídeos, fagomídeos, fago, vírus, *et cetera*.

"Variantes de Seqüência de Nucleotídeo", usando métodos bem
 15 conhecido, o versado na técnica pode facilmente produzir variantes de seqüência de nucleotídeo e aminoácido de genes e proteínas, respectivamente. Por exemplo, as moléculas de DNA "variantes" da presente invenção são moléculas de DNA contendo mudanças em uma seqüência de gene de EPSPS, isto é, mudanças que incluem um ou mais nucleotídeos da seqüên-
 20 cia de gene de EPSPS sendo suprimido, adicionado e/ou substituído, tal que o gene de EPSPS variante codifique uma proteína que retenha a atividade de EPSPS. As moléculas de DNA variantes podem ser produzidas, por exemplo, pelas técnicas de mutagênese de DNA padrão ou sintetizando-se quimicamente a molécula de DNA variante ou uma porção desta. Os méto-
 25 dos para a síntese química de ácidos nucléicos são debatidos, por exemplo, em Beaucage *et al.*, Tetra. Letts. 22: 1859 a 1862 (1981) e Matteucci *et al.*, J. Am. Chem. Soc. 103: 3185- (1981). A síntese química de ácido nucléicos pode ser realizada, por exemplo, em sintetizadores de oligonucleotídeo au-
 30 tomatizados. Tais variantes preferivelmente não mudam a matriz de leitura da região que codifica a proteína do ácido nucléico e preferivelmente codifi-
 cam uma proteína não tendo nenhuma mudança ou apenas uma redução menor.

"Matriz de leitura aberta (ORF)" refere-se a uma região de DNA ou RNA que codifica um peptídeo, polipeptídeo ou proteína.

5 "Operavelmente Ligada". Uma primeira seqüência de ácido nucléico está "operavelmente" ligada com uma segunda seqüência de ácido nucléico quando a primeira seqüência de ácido nucléico é colocada em uma relação funcional com a segunda seqüência de ácido nucléico. Por exemplo, um promotor está operavelmente ligado a uma seqüência codificadora de proteína se o promotor causa a transcrição ou expressão da seqüência codificadora. No geral, as seqüências de DNA operavelmente ligadas são contíguas e, onde necessário para unir duas regiões codificadoras de proteína,
10 na matriz de leitura.

"Superexpressão" refere-se à expressão de um RNA ou polipeptídeo ou proteína codificados por um DNA introduzido em uma célula hospedeira, em que o RNA ou polipeptídeo ou proteína não estão normalmente presentes na célula hospedeira ou em que o RNA ou polipeptídeo ou proteína estão presentes na dita célula hospedeira em um nível mais alto do que aquele normalmente expressado a partir do gene endógeno que codifica o RNA ou polipeptídeo ou proteína.
15

O termo "planta" abrange qualquer planta superior e sua progênie, incluindo monocotiledôneas (por exemplo, milho, arroz, trigo, cevada, etc.), dicotiledôneas (por exemplo, soja, algodão, canola, tomate, batata, *Arabidopsis*, tabaco, etc.), gimnospermas (pinheiros, abetos, cedros, etc.) e inclui partes de plantas, incluindo unidades reprodutivas de uma planta (por exemplo, sementes, bulbos, tubérculos, frutas, flores, etc.) ou outras partes
20 ou tecidos daquela planta podem ser reproduzidos.

"Cassete de expressão vegetal" refere-se a segmentos de DNA quimérico que compreendem os elementos reguladores que são operavelmente ligados para fornecer a expressão de um produto de transgene em plantas.

30 "Plasmídeo" refere-se a um pedaço circular, extracromossômico, auto-replicativo de DNA.

"Sinal de poliadenilação" ou "sinal poliA" refere-se a uma se-

qüência de ácido nucléico localizada a 3' em relação a uma região codificadora que cause a adição de nucleotídeos de adenilato à extremidade 3' do mRNA transcrito a partir da região codificadora.

"Reação de cadeia da polimerase (PCR)" refere-se a um método de amplificação de DNA que usa uma técnica enzimática para criar cópias múltiplas de uma seqüência de ácido nucléico (amplicon). As cópias de uma molécula de DNA são preparadas transportando-se um DNA polimerase entre dois amplímeros. A base deste método de amplificação são ciclos múltiplos de mudanças de temperatura para desnaturar, depois recozer os amplímeros (moléculas preparadoras de DNA), seguido pela extensão para sintetizar novos filamentos de DNA na região localizada entre os amplímeros flaqueadores. A amplificação de ácido nucléico pode ser realizada por qualquer uma dos vários métodos de amplificação de ácido nucléico conhecido na técnica, incluindo a reação de cadeia da polimerase (PCR). Uma variedade de métodos de amplificação são conhecidos na técnica e são descritos, *inter alia*, nas Patentes U.S. Nºs 4.683.195 e 4.683.202 e no PCR Protocols: A Guide to Method e Applications, ed. Innis *et al.*, Academic Press, San Diego, 1990. Os métodos de amplificação pela PCR foram desenvolvidos para amplificar até 22 kb de DNA genômico e até 42 kb de DNA de bacteriófago (Cheng *et al.*, Proc. Natl. Acad. Sci. USA 91: 5695 a 5699, 1994). Estes métodos assim como outros métodos conhecidos na técnica de amplificação de DNA podem ser usados na prática da presente invenção.

Polinucleotídeo refere-se a uma extensão de moléculas de ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA) maiores do que dois, que são conectados para formar uma molécula maior.

Fragmentos de polipeptídeo. A presente invenção também abrange fragmentos de uma proteína que carece de pelo menos um resíduo de uma proteína de comprimento natural nativa, mas que substancialmente mantém a atividade da proteína.

O termo "promotor" ou "região promotora" refere-se a uma molécula de ácido polinucleico que funciona como um elemento regulador, usualmente encontrado a montante (5') em relação a uma seqüência codificado-

ra, que controla a expressão da seqüência codificadora controlando-se a produção de RNA mensageiro (mRNA) pelo qual fornece o sítio de reconhecimento para a RNA polimerase e/ou outros fatores necessários para o início da transcrição no sítio correto. Como aqui considerado, um promotor ou região promotora inclui variações de promotores derivados por meio de ligação a várias seqüências reguladoras, mutagênese aleatória ou controlada e a adição ou duplicação de seqüências realçadoras. A região promotora aqui divulgada e seus equivalentes biologicamente funcionais, são responsáveis por conduzir a transcrição de seqüências codificadoras sob o seu controle quando introduzido em um hospedeiro como parte de um vetor recombinante adequado, como demonstrado pela sua capacidade para produzir mRNA.

"Recombinante". Um ácido nucléico "recombinante" é feito por uma combinação de dois segmentos de outro modo separados de seqüência, por exemplo, pela síntese química ou pela manipulação de segmentos isolados de ácidos nucléicos pelas técnicas de engendramento genético.

O termo "constructo de DNA recombinante" ou "vetor recombinante" refere-se a qualquer agente tal como um plasmídeo, cosmídeo, vírus, seqüência que replica autonomamente, fago ou seqüência de nucleotídeo de DNA ou RNA de filamento único ou de filamento duplo linear ou circular, derivada de qualquer fonte, capaz de integração genômica ou replicação autônoma, compreendendo uma molécula de DNA que uma ou mais seqüências de DNA foram ligadas em uma maneira funcionalmente operativa. Tais constructos de DNA recombinantes ou vetores são capazes de introduzir uma seqüência reguladora 5' ou região promotora e uma seqüência de DNA para um produto de gene selecionado em uma célula em uma maneira tal que a seqüência de DNA é transcrita em um mRNA funcional que é traduzido e portanto expressado. Os constructos de DNA recombinantes ou vetores recombinantes podem ser construídos para serem capazes de expressar RNAs de anti-sentido, de modo a inibir a tradução de um RNA específico de interesse.

"Regeneração" refere-se ao processo de cultivar uma planta a partir de uma célula vegetal (por exemplo, protoplasto vegetal ou explante).

"Repórter" refere-se a um gene e produto de gene correspondente que quando expressados em organismos transgênicos produzem um produto detectável pelos métodos químicos ou moleculares ou produzem um fenótipo observável.

- 5 "Resistência" refere-se a uma enzima que é capaz de funcionar na presença de uma toxina, por exemplo, EPSP sintases de classe II resistentes ao glifosato. Uma enzima que tem resistência a uma toxina pode ter a função de destoxificar a toxina, por exemplo, a fosfinotricina acetiltransferase, a glifosato oxidoreductase ou pode ser uma enzima mutante tendo atividade catalítica que não é afetada por um herbicida que rompe a mesma atividade na enzima do tipo selvagem, por exemplo, a acetolactato sintase, EPSP sintases de classe I mutantes.

- 15 "Enzima de restrição" refere-se a uma enzima que reconhece uma sequência palindrômica específica de nucleotídeos no DNA de filamento duplo e cliva ambos os filamentos; também chamada de uma endonuclease de restrição. A clivagem tipicamente ocorre dentro do sítio de restrição.

- 20 "Marcador selecionável" refere-se a uma molécula de ácido polinucleico que codifica uma proteína, que confere um fenótipo que facilita a identificação de células contendo a molécula de ácido polinucleico. Marcadores selecionáveis incluem aqueles genes que conferem resistência aos antibióticos (por exemplo, ampicilina, canamicina), complementam uma deficiência nutricional (por exemplo, uracila, histidina, leucina) ou comunica uma característica que se distinga visualmente (por exemplo, mudanças de cor ou fluorescência). Os genes selecionável marcadores dominantes úteis incluem genes que codificam os genes de resistência a antibiótico (por exemplo, a neomicina fosfotransferase, *aad*); e os genes de resistência a herbicida (por exemplo, a fosfinotricina acetiltransferase, EPSP sintase de classe II, EPSP sintase de classe I modificada). Uma estratégia útil para a seleção de transformantes para a resistência a herbicida é descrita, por exemplo, em Vasil, *Cell Culture e Somatic Cell Genetics of Plants*, Vols. I-III, Laboratory Procedures e Their Applications Academic Press, Nova Iorque (1984).
- 25
- 30

O termo "específica para (uma seqüência alvo)" indica que uma sonda de DNA ou iniciador de DNA hibridizam sob condições de hibridização dadas apenas com a seqüência alvo em um amostra que compreende a seqüência alvo.

5 O termo "substancialmente purificado", como aqui usado, refere-se a uma molécula separada de outras moléculas normalmente associadas com ela no seu estado nativo. Mais preferivelmente, uma molécula substancialmente purificada é a espécie predominante presente em uma preparação. Uma molécula substancialmente purificada pode ser maior do que 60 %
10 livre, preferivelmente 75 % livre, mais preferivelmente 90 % livre das outras moléculas (exclusivo de solvente) presentes na mistura natural. O termo "substancialmente purificado" não é intencionado a abranger moléculas presentes no seu estado nativo.

"Tolerante" ou "tolerância" refere-se a um efeito reduzido de um
15 agente biótico ou abiótico sobre o cultivo e desenvolvimento de organismos e plantas, por exemplo, uma praga ou um herbicida.

"Transcrição" refere-se ao processo de produzir uma cópia de RNA a partir de um DNA padrão.

"Transformação" refere-se a um processo de introduzir uma
20 molécula de ácido polinucleico exógeno (por exemplo, um constructo de DNA, uma molécula de ácido polinucleico recombinante) em uma célula ou protoplasto e que a molécula de ácido polinucleico exógena seja incorporada em um cromossoma ou seja capaz de replicação autônoma.

"Transformado" ou "transgênico" referem-se a uma célula, teci-
25 do, órgão ou organismo nos quais um ácido polinucleico estranho, tal como um vetor de DNA ou molécula de ácido polinucleico recombinante. Uma célula ou organismo "transgênicos" ou "transformados" também incluem a progênie da célula ou organismo e a progênie produzida a partir de um programa de procriação que utiliza uma tal planta "transgênica" como um precursor
30 em um cruzamento e que exhibe um fenótipo aumentado resultante da presença da molécula de ácido polinucleico estranha.

O termo "transgene" refere-se a qualquer molécula de ácido po-

linucleico não nativa para uma célula ou organismo transformados na célula ou organismo. "Transgene" também abrange as partes componentes de um gene vegetal nativo modificado pela inserção de uma molécula de ácido polinucleico não nativa pela recombinação direcionada ou mutação específica ao sítio.

Moléculas de "peptídeo de trânsito" ou "peptídeo alvejador", estes termos no geral referem-se às moléculas de peptídeo que quando ligadas a uma proteína de interesse direcionam a proteína a um tecido, célula, localização subcelular ou organela celular particulares. Os exemplos incluem, mas não são limitados a, peptídeos de trânsito de cloroplasto, sinais de alvejamento nuclear e sinais vacuolares. O peptídeo de trânsito de cloroplasto é de utilidade particular na presente invenção para direcionar a expressão da enzima de EPSPS ao cloroplasto.

O termo "tradução" refere-se à produção do produto de gene correspondente, isto é, um peptídeo, polipeptídeo ou proteína a partir de um mRNA.

"Vetor" refere-se a um plasmídeo, cosmídeo, bacteriófago ou vírus que carrega DNA estranho em um organismo hospedeiro.

Polinucleotídeos

Os métodos da presente invenção incluem planejar genes que conferem um traço de interesse à planta na qual eles são introduzidos. Os transgenes de interesse agrônômico que fornecem traços agrônômicos benéficos para as plantas de safra, por exemplo, incluindo, mas não limitadas aos elementos genéticos que compreendem resistência à herbicida (Patente US Nº 5.633.435; Patente US Nº 5.463.175), rendimento aumentado (Patente US Nº 5.716.837), controle de inseto (Patente US Nº 6.063.597; Patente US Nº 6.063.756; Patente US Nº 6.093.695; Patente US Nº 5.942.664; Patente US Nº 6.110.464), resistência fúngica às doenças (Patente US Nº 5.516.671; Patente US Nº 5.773.696; Patente US Nº 6.121.436; e Patente US Nº 6.316.407 e Patente US Nº 6.506.962), resistência a vírus (Patente US Nº 5.304.730 e Patente US Nº 6.013.864), resistência à nematóide (Patente US Nº 6.228.992), resistência bacteriana às doenças (Patente US Nº

5.516.671), produção de amido (Patente US Nº 5.750.876 e Patente US Nº 6.476.295), produção de óleo modificado (Patente US Nº 6.444.876), produção de óleo alta (Patente US Nº 5.608.149 e Patente US Nº 6.476.295), teor de ácido graxo modificado (Patente US Nº 6.537.750), produção de proteína alta (Patente US Nº 6.380.466), maturação de fruta (Patente US Nº 5.512.466), nutrição animal e humana realçada (Patente US Nº 5.985.605 e Patente US Nº 6.171.640), biopolímeros (Patente US Nº 5.958.745 e Pedido de Patente US Nº US20030028917), 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) rafinose baixa (Patente US Nº 6.166.292), produção de enzima industrial (Patente US Nº 5.543.576), sabor melhorado (Patente US Nº 6.011.199), fixação de nitrogênio (Patente US Nº 5.229.114), produção de semente híbrida (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.

Os herbicidas para os quais a tolerância da planta transgênica foi demonstrada e o método da presente invenção pode ser aplicado, incluem mas não são limitados aos: herbicidas de glifosato, glifosinato, sulfoniluréias, imidazolinonas, bromoxinila, delapon, cicloezanodiona, inibidores da protoporfirionogênio oxidase e isoxasflutol. As moléculas de polinucleotídeo que codificam proteínas envolvidas na tolerância a herbicida são conhecidas na técnica e incluem, mas não são limitadas a uma molécula de polinucleotídeo que codifica a 5-enolpiruvilshiquimato-3-fosfato sintase (EPSPS, descrita nas Patentes U.S. Nºs 5.627.061, 5.633.435, 6.040.497; Padgett *et al.* Herbicide Resistant Crops, Lewis Publishers, 53 a 85, 1996; e Penaloza-Vazquez, *et al.* Plant Cell Reports 14: 482 a 487, 1995; e *aroA* (Patente U.S. Nº 5.094.945) para a tolerância a glifosato; bromoxinil nitrilase (*Bxn*) para a tolerância a Bromoxinil (Patente U.S. Nº 4.810.648); fitoeno dessaturase (*crtl*, Misawa *et al.*, (1993) Plant J. 4: 833 a 840 e (1994) Plant J. 6: 481 a 489); para a tolerância ao norflurazon, acetoidroxiácido sintase (AHAS, *aka*

ALS, Sathasiivan *et al.* Nucl. Acids Res. 18: 2188 a 2193. 1990); e o gene *bar* para a tolerância ao glufosinato e bialafos (DeBlock, *et al.* EMBO J. 6: 2513 a 2519, 1987).

A tolerância a herbicida é um fenótipo desejável para plantas de safra. N-fosfonometilglicina, também conhecido como glifosato, é um herbicida bem conhecido que tem atividade em um amplo espectro de espécies vegetais. O glifosato é o ingrediente ativo de Roundup® (Monsanto Co.), um herbicida seguro tendo uma meia vida desejavelmente curta no ambiente. Quando aplicado a uma superfície da planta, o glifosato move-se sistemicamente através da planta. O glifosato é tóxico às plantas pela inibição do caminho do ácido chiquímico, que fornece um precursor para a síntese de aminoácidos aromáticos. Especificamente, o glifosato afeta a conversão de fosfoenolpiruvato e ácido 3-fosfochiquímico em ácido 5-enolpiruvil-3-fosfochiquímico pela inibição da enzima 5-enolpiruvil-3-fosfochiquimato sintase (daqui em diante aludido como EPSP sintase ou EPSPS). Para os propósitos da presente invenção, o termo "glifosato" deve ser considerado incluir qualquer forma herbicidamente eficaz de N-fosfonometilglicina (incluindo qualquer sal desta) e outras formas que resultam na produção do ânion glifosato na planta.

Através dos métodos de engendramento genético vegetal, é possível produzir plantas tolerantes ao glifosato inserindo-se no genoma da planta uma molécula de DNA que cause a produção de níveis mais altos de EPSPS do tipo selvagem (Shah *et al.*, Science 233: 478 a 481, 1986). A tolerância ao glifosato também pode ser obtida pela expressão de variantes de EPSPS que tenham afinidade mais baixa ao glifosato e portanto retêm a sua atividade catalítica na presença de glifosato (Patente U.S. Nº 5.633.435). As enzimas que degradam glifosato nos tecidos da planta (Patente U.S. Nº 5.463.175) também são capazes de conferir tolerância celular ao glifosato. Tais genes, portanto, permitem a produção de safras transgênicas que são tolerantes ao glifosato, permitindo deste modo que o glifosato seja usado para o controle eficaz de ervas daninhas com preocupações mínimas de dano à safra. Por exemplo, a tolerância a glifosato foi geneticamente engen-

drada no milho (Patente U.S. Nº 5.554.798, 6.040.497), trigo (Zhou *et al.* Plant Cell Rep. 15: 159 a 163, 1995), soja (WO 9200377) e canola (WO 9204449).

As variantes da enzima de EPSPS do tipo selvagem foram isoladas que são resistentes ao glifosato como um resultado de alterações na sequência codificadora de aminoácido EPSPS (Kishore *et al.*, Annu. Rev. Biochem. 57: 627 a 663, 1988; Schulz *et al.*, Arch. Microbiol. 137: 121 a 123. 1984; Sost *et al.*, FEBS Lett. 173: 238 a 241, 1984; Kishore *et al.*, Em "Biotechnology for Crop Protections" ACS Symposium Series Nº 379. eds. Hedlin *et al.*, 37 a 48, 1988). Estas variantes tipicamente têm um K_i mais alto para o glifosato do que a enzima de EPSPS do tipo selvagem que confere o fenótipo tolerante a glifosato, mas estas variantes também são caracterizadas por um alto K_m para PEP que torna a enzima cineticamente menos eficiente. Por exemplo, a K_m aparente para PEP e o K_i aparente para glifosato para a EPSPS nativa da *E. coli* são 10 μ M e 0,5 μ M enquanto para um isolado resistente ao glifosato tendo uma única substituição de aminoácido de uma alanina no lugar da glicina na posição 96 estes valores são 220 μ M e 4,0 mM, respectivamente. A Patente US Nº 6.040.497 relata que a mutação conhecida como a mutação TIPS (uma substituição de isoleucina no lugar da treonina na posição de aminoácido 102 e uma substituição de serina no lugar de prolina na posição de aminoácido 106) compreende duas mutações que quando introduzidas na sequência de polipeptídeo da EPSPS de *Zea mays* confere a resistência ao glifosato à enzima. As plantas transgênicas contendo esta enzima mutante são tolerantes ao glifosato. As mutações idênticas podem ser feitas nas enzimas EPSPS sensíveis ao glifosato de outras fontes para criar enzimas resistente ao glifosato.

Uma variedade de enzimas de EPSPS nativas e variantes foi expressada em plantas transgênicas de modo a conferir a tolerância a glifosato (Singh, *et al.*, Em "Biosynthesis e Molecular Regulation of Amino Acids in Plants", Amer Soc Plant Phys. Pubs., 1992). Os exemplos de algumas destas EPSPSs incluem aqueles descritos e/ou isolados de acordo com a Patente U.S. Nº 4.940.835. Patente U.S. Nº 4.971.908, Patente U.S. Nº

5.145.783, Patente U.S. Nº 5.188.642, Patente U.S. Nº 5.310.667 e Patente U.S. Nº 5.312.910. Elas também podem ser derivadas de uma classe estruturalmente distinta de genes de EPSPS não homólogos, tais como os genes de EPSPS de classe II isolado de *Agrobacterium* sp. cepa CP4 como descrito na Patente U.S. Nº 5.633.435 e Patente U.S. Nº 5.627.061.

Peptídeos de trânsito de cloroplasto (CTPs) são engendrados para serem fundidos ao término N da EPSPS bacteriana para direcionar as enzimas resistentes ao glifosato no cloroplasto da planta. Na EPSPS vegetal nativa, as regiões de peptídeo de trânsito de cloroplasto estão contidas na sequência codificadora nativa (por exemplo, CTP2, Klee *et al.*, Mol. Gen. Genet. 210: 47 a 442, 1987, aqui incorporada por referência em sua totalidade). O CTP nativo pode ser substituído com um CTP heterólogo durante a construção de um cassete de expressão vegetal de transgene. Muitas proteínas localizadas no cloroplasto, incluindo a EPSPS, são expressadas a partir de genes nucleares como precursores e são alvejadas ao cloroplasto por um peptídeo de trânsito de cloroplasto (CTP) que é removido durante as etapas de importação. Os exemplos de outras de tais proteínas de cloroplasto incluem a subunidade pequena (SSU) de Ribulose-1,5-bisfosfato carboxilase, Ferredoxina, Ferredoxina oxidorreductase, a proteína I e a proteína II do complexo de colheita de luz e Tiorredoxina F. Foi demonstrado *in vivo* e *in vitro* que as proteínas que não de cloroplasto podem ser alvejadas ao cloroplasto pelo uso de fusões de proteína com um CTP e que a sequência de CTP é suficiente para alvejar uma proteína ao cloroplasto. A incorporação de um peptídeo de trânsito de cloroplasto adequado, tal como, o CTP de EPSPS da *Arabidopsis thaliana* (Klee *et al.*, Mol. Gen. Genet. 210: 437 a 442 (1987) e a CTP de EPSPS da *Petunia hybrida* (della-Cioppa *et al.*, Proc. Natl. Acad. Sci. USA 83: 6873 a 6877 (1986) foi mostrada alvejar as seqüências de proteína de EPSPS heterólogas aos cloroplastos nas plantas transgênicas. A produção de plantas tolerantes ao glifosato pela expressão de uma proteína de fusão que compreende um CTP de terminal amino com uma enzima de EPSPS resistente ao glifosato é bem conhecida por aqueles versados na técnica, (Patente U.S. Nº 5.627.061, Patente U.S. Nº 5.633.435.

Patente U.S. Nº 5.312.910, EP 0218571, EP 189707, EP 508909 e EP 924299). Aqueles versados na técnica reconhecerão que vários constructos quiméricos podem ser feitas que utilizam a funcionalidade de um CTP particular para importar enzimas de EPSPS resistentes ao glifosato no cloroplasto de célula vegetal.

Modificação e mudanças podem ser feitas na estrutura dos polinucleotídeos da invenção e ainda obter-se uma molécula que codifique uma proteína ou peptídeo funcionais com características desejáveis. O seguinte é um método com base na substituição do(s) códon(s) de um primeiro polinucleotídeo para criar um polinucleotídeo artificial equivalente ou ainda um melhorado, de geração secundária, onde este novo polinucleotídeo artificial é útil nos métodos de implantação de gene transgênico e expressão realçada. É considerado que as substituições de códon no polinucleotídeo de geração secundária pode em certos casos resultar em pelo menos um aminoácido diferente daquele do primeiro polinucleotídeo. A substituição de aminoácido pode fornecer uma característica melhorada à proteína, por exemplo, uma EPSP sintase resistente ao glifosato ou ela pode ser uma mudança conservada que não afeta substancialmente as características da proteína. O método fornece um polinucleotídeo artificial criado pela retro tradução de uma seqüência de polipeptídeo em um polinucleotídeo usando uma tabela do uso de códon, seguido pelas etapas de realçar as características do polipeptídeo artificial que o torna particularmente útil em plantas transgênicas.

Em formas de modalidades particulares da invenção, os polipeptídeos modificados que codificam proteínas resistentes a herbicida são considerados serem úteis para pelo menos um dos seguintes: conferir tolerância a herbicida em uma planta transformada ou transgênica, para melhorar a expressão de genes de resistência a herbicida em plantas, para o uso como marcadores selecionáveis para a introdução de outros traços de interesse em uma planta e para impedir a recombinação com um gene vegetal endógeno similar ou transgene existente permitindo ainda a implantação de gene sem o silenciamento de gene.

É conhecido que o código genético é degenerado. Os aminoáci-

dos e seus códon(s) de RNA são listados abaixo na Tabela 1.

TABELA 1. Aminoácidos e os códons de RNA que os codificam.

Aminoácido Códons

Nome completo; código de 3 letras; código de 1 letra

5	Alanina; Ala; A	GCA GCC GCG GCU
	Cisteína; Cys ; C	UGC UGU
	Ácido aspártico; Asp; D	GAC GAU
	Ácido glutâmico; Glu; E	GAA GAG
	Fenilalanina; Phe; F	UUC UUU
10	Glicina; Gly; G	GGA GGC GGG GGU
	Histidina; His; H	CAC CAU
	Isoleucina; Ile; I	AUA AUC AUU
	Lisina; Lys; K	AAA AAG
	Leucina ; Leu ; L	UUA UUG CUA CUC CUG CUU
15	Metionina; Met; M	AUG
	Asparagina; Asn; N	AAC AAU
	Prolina; Pro; P	CCA CCC CCG CCU
	Glutamina; Gln; Q	CAA CAG
	Arginina; Arg; R	AGA AGG CGA CGC CGG CGU
20	Serina; Ser; S	AGC AGU UCA UCC UCG UCU
	Treonina; Thr; T	ACA ACC ACG ACU
	Valina; Val; V	GUA GUC GUG GUU
	Triptofano; Trp; W	UGG
	Tirosina; Tyr; Y	UAC UAU

25 Os códons são descritos em termos de bases de RNA, por exemplo, adenina, uracila, guanina e citosina, é o mRNA que é diretamente traduzido em polipeptídeos. É entendido que quando do planejamento de um polinucleotídeo de DNA para o uso em um constructo, as bases de DNA podem ser substituídas, por exemplo, timina ao invés de uracila.

30 É desejável fornecer plantas transgênicas que tenham fenótipos agronomicamente melhorados múltiplos. Frequentemente a tolerância a herbicida é usada como um marcador selecionável para ajudar na produção de

plantas transgênicas que podem possuir genes adicionais de importância agronômica. A implantação dos transgenes pelos métodos de cruzamento tradicionais ou pela retransformação de uma primeira planta transgênica com um cassete de expressão vegetal adicional pode incluir a introdução de genes ou elementos genéticos que têm seqüência de polinucleotídeo idêntica ou quase idêntica. A progênie contendo estes genes implantados pode ser suscetível à perda da expressão de gene devido ao silenciamento de gene. O método da presente invenção fornece uma molécula de polinucleotídeo modificada que codifica uma proteína resistente a herbicida. As moléculas de polinucleotídeo são planejadas para serem suficientemente divergentes em seqüência de polinucleotídeo de outras moléculas de polinucleotídeo que codificam a mesma proteína de resistência a herbicida. Estas moléculas podem então coexistir na mesma célula vegetal sem a preocupação do silenciamento de gene.

A seqüência de polinucleotídeo divergente é criada usando-se uma tabela de uso de códon construída a partir de seqüências codificadoras conhecidas de várias espécies vegetais. Por exemplo, as tabelas de uso de códon para *Arabidopsis thaliana*, *Zea mays* e *Glicina max* podem ser usadas no método para planejar os polinucleotídeos da presente invenção. Outras tabelas de uso de códon de outras plantas também podem ser usadas por aqueles versados na técnica.

A primeira etapa no método para planejar uma nova molécula de polinucleotídeo artificial que codifica uma proteína de tolerância a herbicida é o uso de uma tabela de uso de códon para determinar o uso de códon percentual em uma espécie vegetal para cada aminoácido da proteína de tolerância a herbicida, seguida pela substituição de pelo menos um de cada oito códons contíguos com um códon diferente selecionado da tabela de uso de códon e ajustando o uso de códon percentual para cada aminoácido codificado pelo polinucleotídeo até substancialmente o mesmo uso de códon percentual encontrado na tabela de uso de códon. As etapas adicionais podem incluir a introdução de um códon de parada translacional na segunda e terceira matriz de leitura aberta da nova seqüência de polinucleotídeo; elimi-

- nando alguns códons de partida translacionais na segunda e na terceira matrizes de leitura aberta; ajustando a relação GC:AT local para cerca de 2:1 em uma faixa de cerca de 50 nucleotídeos; rompendo sinais de poliadenilação potenciais ou sítios de junção de intron potenciais; removendo pelo menos um sítio de enzima de restrição de seis nucleotídeos contíguos ou mais; e comparando a identidade de seqüência do novo polinucleotídeo artificial com um polinucleotídeo existente que codifica a mesma ou proteína similar de modo que a identidade de seqüência entre os dois polinucleotídeos não seja mais do que 85 por cento.
- 10 Uma retrotradução de uma seqüência de proteína para uma seqüência de nucleotídeo pode ser realizada usando uma tabela de uso de códon, tal como aquela encontrada no Genetics Computer Group (GCG) SeqLab ou outros programas de análise de DNA conhecidos por aqueles versados na técnica de análise de DNA ou como fornecido nas Tabelas 2, 3 e 4
- 15 da presente invenção. A tabela de uso de códon para *Arabidopsis thaliana* (Tabela 2), *Zea mays* (Tabela 3) e *Glicina max* (Tabela 4) são exemplos de tabelas que podem ser construídas para espécies vegetais, as tabelas de uso de códon também podem ser construídas que representam o uso de códon de monocotiledônea ou dicotiledônea.
- 20 Tabela 2. Tabela de uso de códon de *Arabidopsis thaliana*.

Aminoácido	Codon	Número	/1000	Fração
Gly	GGG	188335,00	10,18	0,16
Gly	GGA	443469,00	23,98	0,37
Gly	GGT	409478,00	22,14	0,34
Gly	GGC	167099,00	9,03	0,14
Glu	GAG	596506,00	32,25	0,48
Glu	GAA	639579,00	34,58	0,52
Asp	GAT	683652,00	36,96	0,68
Asp	GAC	318211,00	17,20	0,32
Val	GTG	320636,00	17,34	0,26

Aminoácido	Codon	Número	/1000	Fração
Val	GTA	185889,00	10,05	0,15
Val	GTT	505487,00	27,33	0,41
Val	GTC	235004,00	12,71	0,19
Ala	GCG	162272,00	8,77	0,14
Ala	GCA	323871,00	17,51	0,27
Ala	GCT	521181,00	28,18	0,44
Ala	GCC	189049,00	10,22	0,16
Arg	AGG	202204,00	10,93	0,20
Arg	AGA	348508,00	18,84	0,35
Ser	AGT	260896,00	14,11	0,16
Ser	AGC	206774,00	11,18	0,13
Lys	AAG	605882,00	32,76	0,51
Lys	AAA	573121,00	30,99	0,49
Asn	AAT	418805,00	22,64	0,52
Asn	AAC	385650,00	20,85	0,48
Met	ATG	452482,00	24,46	1,00
Ile	ATA	235528,00	12,73	0,24
Ile	ATT	404070,00	21,85	0,41
Ile	ATC	341584,00	18,47	0,35
Thr	ACG	140880,00	7,62	0,15
Thr	ACA	291436,00	15,76	0,31
Thr	ACT	326366,00	17,65	0,34
Thr	ACC	190135,00	10,28	0,20
Trp	TGG	231618,00	12,52	1,00
End	TGA	19037,00	1,03	0,43
Cys	TGT	196601,00	10,63	0,60
Cys	TGC	131390,00	7,10	0,40

Aminoácido	Codon	Número	/1000	Fração
End	TAG	9034,00	0,49	0,20
End	TAA	16317,00	0,88	0,37
Tyr	TAT	276714,00	14,96	0,52
Tyr	TAC	254890,00	13,78	0,48
Leu	TTG	389368,00	21,05	0,22
Leu	TTA	237547,00	12,84	0,14
Phe	TTT	410976,00	22,22	0,52
Phe	TTC	380505,00	20,57	0,48
Ser	TCG	167804,00	9,07	0,10
Ser	TCA	334881,00	18,11	0,20
Ser	TCT	461774,00	24,97	0,28
Ser	TCC	203174,00	10,99	0,12
Arg	CGG	88712,00	4,80	0,09
Arg	CGA	115857,00	6,26	0,12
Arg	CGT	165276,00	8,94	0,17
Arg	CGC	69006,00	3,73	0,07
Gln	CAG	280077,00	15,14	0,44
Gln	CAA	359922,00	19,46	0,56
His	CAT	256758,00	13,88	0,62
His	CAC	160485,00	8,68	0,38
Leu	CTG	183128,00	9,90	0,11
Leu	CTA	184587,00	9,98	0,11
Leu	CTT	447606,00	24,20	0,26
Leu	CTC	294275,00	15,91	0,17
Pro	CCG	155222,00	8,39	0,17
Pro	CCA	298880,00	16,16	0,33

Aminoácido	Codon	Número	/1000	Fração
Pro	CCT	342406,00	18,51	0,38
Pro	CCC	97639,00	5,28	0,11

Tabela 3. Tabela de uso do códon de *Zea mays*

Aminoácido	Codon	Número	/1000	Fração
Gly	GGG	8069,00	15,19	0,21
Gly	GGA	7100,00	13,37	0,18
Gly	GGT	7871,00	14,82	0,20
Gly	GGC	15904,00	29,94	0,41
Glu	GAG	22129,00	41,67	0,68
Glu	GAA	10298,00	19,39	0,32
Asp	GAT	11996,00	22,59	0,41
Asp	GAC	17045,00	32,09	0,59
Val	GTG	13873,00	26,12	0,38
Val	GTA	3230,00	6,08	0,09
Val	GTT	8261,00	15,55	0,23
Val	GTC	11330,00	21,33	0,31
Ala	GCG	11778,00	22,18	0,24
Ala	GCA	8640,00	16,27	0,18
Ala	GCT	11940,00	22,48	0,24
Ala	GCC	16768,00	31,57	0,34
Arg	AGG	7937,00	14,94	0,27
Arg	AGA	4356,00	8,20	0,15
Ser	AGT	3877,00	7,30	0,10
Ser	AGC	8653,00	16,29	0,23
Lys	AAG	22367,00	42,11	0,74
Lys	AAA	7708,00	14,51	0,26
Asn	AAT	6997,00	13,17	0,36

Aminoácido	Codon	Número	/1000	Fração
Asn	AAC	12236,00	23,04	0,64
Met	ATG	12841,00	24,18	1,00
Ile	ATA	3997,00	7,53	0,16
Ile	ATT	7457,00	14,04	0,31
Ile	ATC	12925,00	24,34	0,53
Thr	ACG	5665,00	10,67	0,22
Thr	ACA	5408,00	10,18	0,21
Thr	ACT	5774,00	10,87	0,22
Thr	ACC	9256,00	17,43	0,35
Trp	TGG	6695,00	12,61	1,00
End	TGA	591,00	1,11	0,45
Cys	TGT	2762,00	5,20	0,30
Cys	TGC	6378,00	12,01	0,70
End	TAG	411,00	0,77	0,32
End	TAA	299,00	0,56	0,23
Tyr	TAT	4822,00	9,08	0,31
Tyr	TAC	10546,00	19,86	0,69
Leu	TTG	6677,00	12,57	0,14
Leu	TTA	2784,00	5,24	0,06
Phe	TTT	6316,00	11,89	0,32
Phe	TTC	13362,00	25,16	0,68
Ser	TCG	5556,00	10,46	0,14
Ser	TCA	5569,00	10,49	0,15
Ser	TCT	6149,00	11,58	0,16
Ser	TCC	8589,00	16,17	0,22
Arg	CGG	4746,00	8,94	0,16

Aminoácido	Codon	Número	/1000	Fração
Arg	CGA	2195,00	4,13	0,07
Arg	CGT	3113,00	5,86	0,10
Arg	CGC	7374,00	13,88	0,25
Gln	CAG	13284,00	25,01	0,64
Gln	CAA	7632,00	14,37	0,36
His	CAT	5003,00	9,42	0,39
His	CAC	7669,00	14,44	0,61
Leu	CTG	13327,00	25,09	0,28
Leu	CTA	3785,00	7,13	0,08
Leu	CTT	8238,00	15,51	0,17
Leu	CTC	12942,00	24,37	0,27
Pro	CCG	8274,00	15,58	0,27
Pro	CCA	7845,00	14,77	0,26
Pro	CCT	7129,00	13,42	0,23
Pro	CCC	7364,00	13,87	0,24

Tabela 4, Tabela de uso do códon de *Glicina max*

Aminoácido	Codon	Número	/1000	Fração
Gly	GGG	3097,00	12,82	0,18
Gly	GGA	5434,00	22,49	0,32
Gly	GGT	5248,00	21,72	0,31
Gly	GGC	3339,00	13,82	0,20
Glu	GAG	8296,00	34,33	0,50
Glu	GAA	8194,00	33,91	0,50
Asp	GAT	7955,00	32,92	0,62
Asp	GAC	4931,00	20,40	0,38
Val	GTG	5342,00	22,11	0,32
Val	GTA	1768,00	7,32	0,11

Aminoácido	Codon	Número	/1000	Fração
Val	GTT	6455,00	26,71	0,39
Val	GTC	2971,00	12,29	0,18
Ala	GCG	1470,00	6,08	0,08
Ala	GCA	5421,00	22,43	0,31
Ala	GCT	6796,00	28,12	0,38
Ala	GCC	4042,00	16,73	0,23
Arg	AGG	3218,00	13,32	0,28
Arg	AGA	3459,00	14,31	0,30
Ser	AGT	2935,00	12,15	0,17
Ser	AGC	2640,00	10,92	0,15
Lys	AAG	9052,00	37,46	0,59
Lys	AAA	6370,00	26,36	0,41
Asn	AAT	5132,00	21,24	0,48
Asn	AAC	5524,00	22,86	0,52
Met	ATG	5404,00	22,36	1,00
Ile	ATA	3086,00	12,77	0,23
Ile	ATT	6275,00	25,97	0,47
Ile	ATC	3981,00	16,47	0,30
Thr	ACG	1006,00	4,16	0,08
Thr	ACA	3601,00	14,90	0,29
Thr	ACT	4231,00	17,51	0,34
Thr	ACC	3562,00	14,74	0,29
Trp	TGG	2866,00	11,86	1,00
End	TGA	221,00	0,91	0,36
Cys	TGT	1748,00	7,23	0,49
Cys	TGC	1821,00	7,54	0,51

Aminoácido	Codon	Número	/1000	Fração
End	TAG	143,00	0,59	0,23
End	TAA	256,00	1,06	0,41
Tyr	TAT	3808,00	15,76	0,51
Tyr	TAC	3667,00	15,17	0,49
Leu	TTG	5343,00	22,11	0,24
Leu	TTA	2030,00	8,40	0,09
Phe	TTT	4964,00	20,54	0,49
Phe	TTC	5067,00	20,97	0,51
Ser	TCG	1107,00	4,58	0,06
Ser	TCA	3590,00	14,86	0,21
Ser	TCT	4238,00	17,54	0,24
Ser	TCC	2949,00	12,20	0,17
Arg	CGG	683,00	2,83	0,06
Arg	CGA	964,00	3,99	0,08
Arg	CGT	1697,00	7,02	0,15
Arg	CGC	1538,00	6,36	0,13
Gln	CAG	4147,00	17,16	0,46
Gln	CAA	4964,00	20,54	0,54
His	CAT	3254,00	13,47	0,55
His	CAC	2630,00	10,88	0,45
Leu	CTG	2900,00	12,00	0,13
Leu	CTA	1962,00	8,12	0,09
Leu	CTT	5676,00	23,49	0,26
Leu	CTC	4053,00	16,77	0,18
Pro	CCG	1022,00	4,23	0,08
Pro	CCA	4875,00	20,17	0,37
Pro	CCT	4794,00	19,84	0,36

Aminoácido	Codon	Número	/1000	Fração
Pro	CCC	2445,00	10,12	0,19

As tabelas de uso de códon são bem conhecidas na técnica e podem ser encontradas nos bancos de dados de gene por exemplo, banco de dados do Genbank. O Banco de Dados do Uso de Códon é uma versão WWW extendida de CUTG (Uso de Códon Tabulado do Genbank). A frequência de uso de códon em cada organismo é feito pesquisável através deste local da World Wide Web (Nakamura *et al.* Nucleic Acids Res. 28: 292, 2000).

Em várias formas de realização da invenção, as etapas podem ser realizadas em qualquer ordem ou simultaneamente. Qualquer ou todas as etapas podem ser realizadas no planejamento de um polinucleotídeo artificial da invenção. Cada etapa é descrita em detalhes abaixo.

Os códons diferentes para um aminoácido particular deve ser distribuído por todo o polinucleotídeo com base no uso de códon percentual aproximado para espécies particulares a partir de uma tabela de uso de códon. Cachos locais de códons idênticos devem ser evitados. Pelo menos um códon é substituído para cada oito códons contíguos para fornecer divergência suficiente de seqüências de polinucleotídeo que codificam proteínas idênticas ou similares. Exceto onde especificamente desejado, por exemplo, para fornecer uma enzima tolerante a herbicida, a proteína codificada permanece inalterada pela substituição de um códon no lugar de um outro códon que é traduzido ao mesmo aminoácido como listado na Tabela 1.

Em formas de realização da presente invenção, correções são feitas para a relação GC:AT local de um polinucleotídeo ajustando-se a relação GC:AT local para ser aproximadamente a mesma relação como o polinucleotídeo de tamanho natural, mas não maior do que 2X em uma faixa de cerca de 50 nucleotídeos contíguos da molécula de polinucleotídeo. A faixa de relações GC:AT de um polinucleotídeo usando as tabelas de uso de códon de plantas dicotiledôneas deve ser de cerca de 0,9 a cerca de 1,3 e para as plantas monocotiledôneas de cerca de 1,2 a cerca de 1,7. A relação

GC:AT local pode ser importante na manutenção da estrutura secundária apropriada de RNA. As regiões que compreendem muitas bases A+T ou bases G+C consecutivas são prognosticadas ter uma probabilidade maior para formar estruturas de grampo de cabelo devido à autocomplementaridade.

5 Portanto, a substituição com um códon diferente pode reduzir a probabilidade de formação de estrutura secundária autocomplementar, que é conhecida reduzir a transcrição e/ou tradução em alguns organismos. Na maioria dos casos, os efeitos adversos podem ser minimizados usando-se moléculas de polinucleotídeo que não contenham mais do que cinco A+T ou G+C consecutivos.

10 O comprimento máximo da trilha GC local (sem qualquer nucleotídeo AT) não deve ser maior do que 10 nucleotídeos. Portanto os códons que codificam proteínas ricas em Gly, Ala, Arg, Ser e Pro podem ser substituídos para impedir cachos longos de nucleotídeos GC. Os códons ricos em GC listados podem ser usados em combinação com os códons ricos em AT para

15 os aminoácidos Lys, Asn, Ile, Tyr, Leu, Phe e vice versa para corrigir a relação GC:AT local.

Uma checagem da identidade de sequência usando ferramentas de alinhamento de sequência de nucleotídeo tais como o programa GAP (GCG, Madison, WI) pode ser feita imediatamente depois da retrotradução

20 para garantir que a sequência gerada tenha grau apropriado de diversidade de sequência. A sequência de polinucleotídeo contígua maior do que 23 nucleotídeos que têm cem por cento de identidade de sequência deve ser eliminado realizando-se substituições de códon nestes comprimentos de sequência.

25 Os códons de partida translacionais (ATG do DNA, AUG no mRNA) presentes na segunda matriz de leitura (matriz "b"), a terceira matriz de leitura (matriz "c") e as matrizes de leitura reversa (matriz "d", "e", "f"). O segundo e terceiro códons de início de matriz podem iniciar a tradução, entretanto muito menos eficientemente do que o primeiro. Portanto, se um ou

30 dois AUG são encontrados próximos à extremidade 5' de uma molécula de mRNA reside na matriz "b" ou "c" pode ser benéfico eliminá-los em uma região de polinucleotídeo que contenha pelo menos os primeiros três códons

Met na matriz "a". Também, se a sequência de proteína não tem mais do que um Met na matriz "a", então eliminar tanto quanto possível das matrizes avançadas "b" ou "c". Para realizar isto, por exemplo, os códons para aminoácidos, Asp, Asn, Tyr, His na proteína de interesse seguido por qualquer um dos aminoácidos: Gly, Glu, Asp, Val ou Ala, pode ser substituído para eliminar um códon de partida na segunda matriz. A sequência GATGGG codifica os aminoácidos Asp-Gly e forma um ATG na matriz de leitura "b". Quando a sequência é modificada para GACGGG, o início ATG é eliminado e a sequência ainda codifica Asp-Gly. Uma estratégia similar é usada para eliminar códons de partida na matriz de leitura "c". A combinação de um aminoácido selecionado do grupo de Gly, Glu, Val, Ala, Arg, Lys, Ile, Thr, Cys, Tyr, Leu, Ser, His ou Pro seguido por Trp pode resultar na formação no ATG na terceira matriz de leitura do gene. Nesta situação, o primeiro códon pode ser mudado para ter um nucleotídeo outro que não A na terceira posição.

A eliminação do códon de ATG no filamento de DNA complementar do gene nas matrizes alternadas ("d", "e", e/ou "f") sem mudar a sequência de aminoácido da proteína pode ser realizada em uma maneira similar. Esta modificação reduz a probabilidade de tradução mesmo se o transgene é integrado em um genoma vegetal em uma orientação que permite a transcrição do mRNA de complemento reverso de um promotor vegetal nativo. A tradução de qualquer matriz de leitura reversa pode ser minimizada pela introdução de um códon de parada em todas as três matrizes de leitura reversa como descrito abaixo.

A criação de códons de parada para todas as três matrizes do DNA complementar pode ser realizada como segue. Os códons Leu (TTA e CTA) e Ser (TCA) produzem três códons de parada diferentes no filamento de complemento reverso. Se aqueles aminoácidos podem ser encontrados no término C da proteína de interesse, seus códons podem ser usados para gerar paradas no filamento complementar na matriz de leitura "d". Para gerar um códon de parada na matriz de leitura "e" de filamento complementar, encontrar os aminoácidos Ala, Arg, Asn, Asp, Cys, Gly, His, Ile, Leu Phe, Pro,

Ser, Thr, Tyr ou Val seguidos pelos aminoácidos Gln, His ou Tyr da proteína de interesse. Por exemplo, a seqüência de polinucleotídeo de GCCCAC que codifica os aminoácidos Ala-His pode ser modificada para GCTCAC. A seqüência complementar, GTGAGC, terá agora o códon de parada TGA mostrado em *itálicos*. Quando a proteína de interesse tem um Ala, Ile, Leu, Phe, Pro, Ser, Thr ou Val seguido por um Arg, Asn, Ile, Lys, Met ou Ser a matriz de leitura no filamento complementar pode ser modificada para ter um códon de parada na matriz de leitura "f" do filamento complementar. A seqüência de polinucleotídeo ATATCT para Ile e Ser pode ser modificada para ATCAGT para gerar códon de parada no filamento complementar como mostrado em *itálicos*, ACTGAT. A combinação de códons para Phe seguido por qualquer um dos códons para os aminoácidos Asn, Ile, Lys, Met ou Thr sempre gerará códon de parada na matriz do filamento complementar "e" ou "f".

Para criar um códon de parada na matriz de leitura avançada "b", a matriz de leitura a deve terminar nos nucleotídeos TA ou TG. A pesquisa da proteína de interesse quanto aos aminoácidos Ile, Leu, Met ou Val em combinação com qualquer um dos seguintes aminoácidos: Ala, Arg, Asn Asp, Glu, Gly, Ile, Lys, Met, Ser, Thr ou Val. Por exemplo, se a seqüência de polinucleotídeo que codifica os aminoácidos Met-Ser é ATGTCT, ela pode ser modificada para ATGAGT para produzir um códon de parada TGA na segunda matriz de leitura.

Para ser capaz de criar um códon de parada para a matriz de leitura "c", a matriz de leitura "a" deve ter o nucleotídeo T na terceira posição e o códon seguinte deve começar de AA, AG ou GA. Para encontrar códons adequados para modificar, procurar a proteína de interesse para qualquer um dos aminoácidos: Ala, Asn, Asp, Arg, Cys, Gly, His, Ile, Leu Phe, Pro, Ser, Thr, Tyr ou Val seguidos por qualquer um dos seguintes aminoácidos: Arg, Asn, Asp, Glu, Lys ou Ser. Por exemplo, se a seqüência de nucleotídeo para os aminoácidos Gly-Glu é GGAGAG, a seqüência pode ser modificada to GGTGAG para criar um códon de parada TGA na terceira matriz de leitura.

Uma outra modificação útil nos métodos de planejamento de polinucleotídeo artificial da presente invenção é eliminar sítios de restrição não desejados e outros padrões de sequência específicos. Os sítios de restrição podem interferir com a clonagem e as manipulações de gene futuras. Por exemplo, alguns sítios de restrição habitualmente usados na clonagem de gene incluem, mas não são limitados, às enzimas de restrição do Tipo II com 6 ou mais bases que não de N listadas na Tabela 5 abaixo que é um excerto do banco de dados de endonuclease de restrição da New England Biolabs, Inc. (Beverly, MA, USA). A procura quanto aos sítios de reconhecimento de enzima de restrição pode ser feita usando a aplicação da função Map encontrada no GCG SeqLab ou uma aplicação similar contida em outros programas de análise de DNA conhecidos por aqueles versados na técnica de análise de DNA. As enzimas de restrição também podem ser adicionadas à sequência para facilitar a clonagem. Por exemplo, o sítio de restrição ClaI é colocado em CP4EPSPS versão AT (SEQ ID NO: 17) e ZM (SEQ ID NO: 18) para gerar seqüências recombinantes pela mudança de fragmento e para facilitar a síntese de gene usando fragmentos de nucleotídeo que podem ser montados no gene inteiro. A seqüência de polinucleotídeo CTP2 de peptídeo de trânsito (SEQ ID NO: 12) é conectado com CP4EPSPS pelo sítio de restrição SphI para facilitar a substituição de CTP2 com versões de nucleotídeo diferentes de CTP2 (SEQ ID NO: 13, SEQ ID NO: 14) ou polinucleotídeos que codificam peptídeos de trânsito de cloroplasto diferentes. Por exemplo, na EPSPS de arroz, o sítio de restrição NgaMIV é preservado em torno da posição de nucleotídeo 205 em todas as versões artificiais para facilitar a mudança da região codificadora do peptídeo de trânsito de cloroplasto. Também, para a EPSPS da soja a seqüência de polinucleotídeo para o peptídeo de trânsito de cloroplasto é separada do peptídeo maduro pelo sítio de restrição para SacII endonuclease.

É entendido que a modificação dos sítios de endonuclease de restrição não é requerida, mas é útil para outras manipulações das moléculas de DNA. A Tabela 5 fornece uma lista de endonucleases de restrição, aquelas de interesse particular para a presente invenção são marcadas com

um asterisco. Outros sítios de endonuclease de restrição desejáveis para a eliminação ou adição a um polinucleotídeo artificial da presente invenção estarão evidentes àqueles versados na técnica e não são limitadas por aquelas listadas na Tabela 5.

5 Tabela 5. Sequências de reconhecimento de enzimas de restrição

Enzimas	Sequência de reconhecimento
AatII	G_ACGT^C
*AccI	GT^MK_AC
Acc65I	G^GTAC_C
AceII	G_CTAG^C -
AcII	AA^CG_TT
AcyI	GR^CG_YC
AfeI	AGC^GCT
AflII	C^TTAA_G
*AflIII	A^CRYG_T
AgeI	A^CCGG_T
AhaIII	TTT^AAA
ApaI	G_GGCC^C
ApaLI	G^TGCA_C
ApoI	R^AATT_Y
AscI	GG^CGCG_CC
Asel	AT^TA_AT
AsiSI	GCG_AT^CGC
AsuII	TT^CG_AA
AvaI	C^YCGR_G
AvaIII	ATGCAT
AvrII	C^CTAG_G
BalI	TGG^CCA
*BamHI	G^GATC_C
BanI	G^GYRC_C
BanII	G_RGCY^C
*BbeI	G_GCGC^C
BbvCI	CC^TCA_GC
*BclI	T^GATC_A
BetI	W^CCGG_W
BfrBI	ATG^CAT

Enzimas	Seqüência de reconhecimento
*BglII	A^GATC_T
BloHII	CTGCA^G
BlpI	GC^TNA_GC
Bme1580I	G_KGCM^C
BmgI	GKGCCC
Bpu10I	CC^TNA_GC
BsaI	GGTCTCN^NNNN_
BsaAI	YAC^GTR
BsaHI	GR^CG_YC
BsaWI	W^CCGG_W
Bsbl	CAACAC
BsePI	G^CGCG_C
BseSI	G_KGCM^C
Bsil	C^ACGA_G
BsiEI	CG_RY^CG
BsiWI	C^GTAC_G
BsmI	GAATG_CN^
Bsp1286I	G_DGCH^C
Bsp1407I	T^GTAC_A
BspEI	T^CCGG_A
BspGI	CTGGAC
BspHI	T^CATG_A
BspLU11I	A^CATG_T
BspMII	T^CCGG_A
BsrBI	CCG^CTC
BsrDI	GCAATG_NN^
BsrFI	R^CCGG_Y
BsrGI	T^GTAC_A
BssHII	G^CGCG_C
BssSI	C^ACGA_G
BstAPI	GCAN_NNN^NTGC
BstBI	TT^CG_AA
BstEII	G^GTNAC_C
BstXI	CCAN_NNNN^NTGG
BstYI	R^GATC_Y

Enzimas	Seqüência de reconhecimento
BstZ17I	GTA^TAC
Bsu36I	CC^TNA_GG
BtgI	C^CRYG_G
BtrI	CAC^GTC
BtsI	GCAGTG_NN^
CfrI	Y^GGCC_R
Cfr10I	R^CCGG_Y
*ClaI	AT^CG_AT
DraI	TTT^AAA
DraII	RG^GNC_CY
DrdII	GAACCA
DsaI	C^CRYG_G
EaeI	Y^GGCC_R
EagI	C^GGCC_G
Ecl136II	GAG^CTC
Eco47III	AGC^GCT
EcoNI	CCTNN^N_NNAGG
EcoO109I	RG^GNC_CY
*EcoRI	G^AATT_C
*EcoRV	GAT^ATC
EspI	GC^TNA_GC
*FseI	GG_CCGG^CC
*FspI	TGC^GCA
FspAI	RTGC^GCAY
GdIII	C^GGCC_R
HaeI	WGG^CCW
HaeII	R_GCGC^Y
HgiAI	G_WGCW^C
HgiCI	G^GYRC_C
HgiJII	G_RGCY^C
*HincII	GTY^RAC
HindII	GTY^RAC
*HindIII	A^AGCT_T
*HpaI	GTT^AAC
KasI	G^GCGC_C

Enzimas	Seqüência de reconhecimento
*KpnI	G_GTAC^C
LpnI	RGC^GCY
McrI	CG_RY^CG
MfeI	C^AATT_G
*MluI	A^CGCG_T
MscI	TGG^CCA
MspA1I	CMG^CKG
MstI	TGC^GCA
NaeI	GCC^GGC
NarI	GG^CG_CC
*NcoI	C^CATG_G
*NdeI	CA^TA_TG
*NgoMIV	G^CCGG_C
*NheI	G^CTAG_C
Nli3877I	C_YCGR^G
*NotI	GC^GGCC_GC
*NruI	TCG^CGA
*NsiI	A_TGCA^T
NspI	R_CATG^Y
NspBII	CMG^CKG
*PacI	TTA_AT^TAA
*PciI	A^CATG_T
Pfi1108I	TCGTAG
*PfiMI	CCAN_NNN^NTGG
PmaCI	CAC^GTG
PmeI	GTTT^AAAC
PmlI	CAC^GTG
Ppu10I	A^TGCA_T
*PpuMI	RG^GWC_CY
PshAI	GACNN^NNGTC
PsiI	TTA^TAA
*PspOMI	G^GGCC_C
PssI	RG_GNC^CY
*PstI	C_TGCA^G
*PvuI	CG_AT^CG

Enzimas	Seqüência de reconhecimento
*PvuII	CAG [^] CTG
RsrII	CG [^] GWC_CG
*SacI	G_AGCT [^] C
*SacII	CC_GC [^] GG
*Sall	G [^] TCGA_C
SanDI	GG [^] GWC_CC
SapI	GCTCTTCN [^] NNN_
SauI	CC [^] TNA_GG
SbfI	CC_TGCA [^] GG
*ScaI	AGT [^] ACT
Scil	CTC [^] GAG
SduI	G_DGCH [^] C
SexAI	A [^] CCWGG_T
SfcI	C [^] TRYA_G
SfeI	C [^] TRYA_G
SfiI	GGCCN_NNN [^] NGGCC
SfoI	GGC [^] GCC
SgfI	GCG_AT [^] CGC
SgrAI	CR [^] CCGG_YG
*SmaI	CCC [^] GGG
SmlI	C [^] TYRA_G
SnaI	GTATAC
*SnaBI	TAC [^] GTA
*SpeI	A [^] CTAG_T
*SphI	G_CATG [^] C
SpII	C [^] GTAC_G
SrfI	GCCC [^] GGGC
Sse232I	CG [^] CCGG_CG
Sse8387I	CC_TGCA [^] GG
Sse8647I	AG [^] GWC_CT
*SspI	AAT [^] ATT
*StuI	AGG [^] CCT
*StyI	C [^] CWWG_G
*Swal	ATTT [^] AAAT
TatI	W [^] GTAC_W

Enzimas	Seqüência de reconhecimento
UbaMI	TCCNGGA
UbaPI	CGAACG
*VspI	AT [^] TA_AT
*XbaI	T [^] CTAG_A
*XhoI	C [^] TCGA_G
XhoII	R [^] GATC_Y
*XmaI	C [^] CCGG_G
XmaIII	C [^] GGCC_G
XmnI	GAANN [^] NNTTC
ZraI	GAC [^] GTC

Uma pesquisa padrão pode ser realizada para encontrar seqüências desestabilizadoras e sítios de poliadenilação potenciais e depois romper ou eliminá-las como descrito na Patente US Nº 5.500.365. Certos trechos longos de regiões ricas em AT, por exemplo, o motivo de seqüência

5 ATTTA (ou AUUUA, como aparece no RNA) foi implicado como uma seqüência desestabilizadora em mRNA de célula de mamífero (Shaw e Kamen, Cell 46: 659 a 667, 1986). Muitos mRNAs de vida curta têm regiões não traduzidas 3' ricas em A+T e estas regiões freqüentemente têm a seqüência ATTTA, algumas vezes presente em cópias múltiplas ou como mul-

10 tímetros (por exemplo, ATTTATTTA . . .). Shaw e Kamen mostraram que a transferência da extremidade 3' de um mRNA instável para um RNA estável (globina ou VA1) diminuiu a meia vida do RNA estável dramaticamente. Eles mostraram ainda que um pentâmero de ATTTA teve um efeito desestabilizador profundo em uma mensagem estável e que este sinal pode exercer o

15 seu efeito se ele estiver localizado na extremidade 3' ou dentro da seqüência codificadora. Entretanto, o número de seqüências ATTTA e/ou o contexto de seqüência em que elas ocorrem também parece ser importante na determinação se elas funcionam como seqüências desestabilizadoras. Eles também mostraram que um trîmero de ATTTA teve muito menos efeito do que um

20 pentâmero sobre a estabilidade de mRNA e um dímero ou um monômero não teve nenhum efeito sobre a estabilidade. Note que os multímeros de ATTTA tais como um pentâmero automaticamente cria uma região rica em

A+T. Em outros mRNAs instáveis, a sequência ATTTA pode estar presente em apenas uma única cópia, mas está frequentemente contida em uma região rica em A+T. Uma repetição de 11 pentâmeros de AUUUA mostraram alvejar transcritos repórter para a degradação rápida em plantas (Ohmetakagi *et al.*, Proc. Nat. Acad. Sci. USA 90, 11811 a 11815, 1993). A sequência ATTTA pode ser formada pela combinação de códons para o aminoácido Ile (ATT) e Tyr (TAT) como mostrado ATTTAT. Um outro exemplo pode ser os códons que terminam em AT como em Asn, Asp, His ou Tyr, seguido pelo códon TTA para Leu (por exemplo, AATTTA). Também o códon para Phe (UUU) quando colocado entre os códons que termina em A e começa em A formará o motivo ATTTA. Para eliminar este motivo usualmente a mudança de nucleotídeo único é suficiente como no exemplo mostrado: GCATTTAGC muda para GCATTCAGC ou GCCTTTAGC. Todos os três polinucleotídeos mostrados codificam Ala-Phe-Arg.

Mais sequências que atuam em cis que alvejam o transcrito para a rotação rápida em plantas e em outro sistema foram identificadas (Abler e Green, Plant Mol. Biol. 32: 63 a 78, 1997). Estas incluem o elemento DST que consiste em três subdomínios altamente conservados separados pelas duas regiões variáveis encontradas a jusante do códon de parada dos transcritos SAUR (Newmaan *et al.*, Vegetal Cell 5: 701 a 714, 1993). A sequência conservada de DST consiste em GGAG(N₅)CATAGATTG(N₇)CATTTTGTAT, onde os resíduos altamente conservados são mostrados em tipo itálico. O segundo e terceiro subdomínios de elementos DST contêm resíduos que são invariáveis entre os elementos DST e são chamados de ATAGAT e GTA, respectivamente. Ambos destes subdomínios são necessários para a função DST. Novas sequências de polinucleotídeo artificiais são triadas quanto a presença de motivos conservados de elementos DST GGAG, ATAGATT, CATTT e CATTTTGTAT. Estas sequências são eliminadas pelas substituições de base de códons que preservam a sequência de proteína codificada pelo polinucleotídeo. Os motivos de sequência DST GGAG, ATAGAT, CATTT e GTAT que apareceram em cachos ou padrões similares à sequência de DST conservada também são eliminados pelas substituições

de base.

As seqüências de polinucleotídeo que possivelmente podem funcionar como sítios de poliadenilação são eliminados no novo planejamento de polinucleotídeo (Patente U.S. Nº 5.500.365). Estes sinais de poliadenilação podem não atuar como sítios poliA apropriados, mas ao invés funcionam como sítios aberrantes que dão origem aos mRNAs instáveis.

A adição de um cordão poliadenilado à extremidade 3' de um mRNA é comum à maior parte dos mRNAs eucarióticos. Contidos dentro estes transcritos de mRNA são sinais para a poliadenilação e a formação da extremidade 3' apropriada. Este processamento na extremidade 3' envolve a clivagem do mRNA e a adição de poliA à extremidade 3' madura. Pesquisando-se quanto as seqüências de consenso próximas ao trato poliA nos mRNAs tanto vegetal quanto animal, foi possível identificar seqüências de consenso que aparentemente estão envolvidas na adição de poliA e na clivagem da extremidade 3'. As mesmas seqüências de consenso parecem ser importantes a ambos destes processos. Estes sinais são tipicamente uma variação na seqüência AATAAA. Em células de animal, algumas variantes desta seqüência que são funcionais foram identificadas; em células vegetais parece haver uma faixa ampliada de seqüências funcionais (Dean *et al.*, Nucl Acid Res., 14: 2229 a 2240, 1986; Hunt, Annu Rev. Plant Physiol. Plant Mol. Biol 45: 47 a 60, 1994; Rothine, Plant Mol. Biol. 32: 43 a 61, 1996). Todas estas seqüências de consenso são variações em AATAAA, portanto, todas elas são seqüências ricas em A+T.

Tipicamente, para se obter expressão suficiente de transgenes modificados em plantas, a seqüência codificadora de polinucleotídeo estrutural existente ("gene estrutural") que codifica a proteína de interesse é modificada pela remoção das seqüências ATTTA e dos sinais de poliadenilação putativos pela mutagênese direcionada a sítio do DNA que compreende o gene estrutural. Substancialmente todos os sinais de poliadenilação conhecidos e as seqüências ATTTA são removidos no polinucleotídeo modificado, embora os níveis de expressão realçados são freqüentemente observados apenas com a remoção de algumas das seqüências de sinal de poliadenila-

ção identificadas acima. Alternativamente, se um polinucleotídeo artificial é preparado codificando a proteína objeto, os códons são selecionados para se evitar a seqüência ATTTA e os sinais de poliadenilação putativos. Para os propósitos da presente invenção os sinais de poliadenilação putativos incluem, mas não são necessariamente limitados a, AATAAA, AATAAT, AAC-
5 CAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA e CATAAA.

A seqüência de DNA selecionada é escaneada para identificar regiões com mais do que quatro nucleotídeos de adenina (A) ou timina (T)
10 consecutivos. As regiões A+T são escaneadas quanto aos sinais de poliadenilação vegetal potenciais. Embora a ausência de cinco ou mais nucleotídeos A ou T consecutivos elimina a maior parte dos sinais de poliadenilação vegetal, se existe mais do que um dos sinais de poliadenilação menores identificados dentro de dez nucleotídeos entre si, então a seqüência de nu-
15 cleotídeo desta região é alterada para remover estes sinais enquanto mantém a seqüência de aminoácido codificada original.

A etapa seguinte é considerar os cerca de 15 a cerca de 30 ou mais resíduos de nucleotídeo que circundam a região rica em A+T. Se o teor de A+T da região circundante for menor do que 80 %, a região deve ser
20 examinada quanto a sinais de poliadenilação. A alteração da região com base em sinais de poliadenilação é dependente (1) do número de sinais de poliadenilação presentes e (2) da presença de um sinal de poliadenilação vegetal maior. Os sinais de poliadenilação são removidos pela substituição de base da seqüência de DNA no contexto da substituição de códon.

Dois padrões adicionais não identificados na Patente US Nº 5.500.365. são procurados e eliminados nas formas de realização da presente invenção. As seqüências AGGTAA e GCAGGT são seqüências de consenso para os sítios de junção 5' e 3' de intron, respectivamente, em plantas monocotiledôneas e plantas dicotiledôneas. Apenas GT do sítio de
25 junção 5' e o AG no sítio de junção 3' são requeridos para ser um emparelhamento exato. Entretanto, quando da condução de uma pesquisa quanto a estas seqüências de consenso, nenhum desemparelhamento é permitido
30

para cada base.

Depois de cada etapa o mapeamento de seqüência é feito usando o programa MAP da GCG para determinar a localização das matrizes de leitura aberta e identificar padrões de seqüência que ainda precisam ser modificados. A etapa final seria realizar a análise da identidade de seqüência usando por exemplo o programa GAP do pacote da GCG para determinar o grau de divergência de seqüência e a identidade percentual.

Polipeptídeos

No geral, a proteína traduzida do polinucleotídeo artificial terá a mesma seqüência de aminoácido como a proteína traduzida da região codificadora não-modificada. Entretanto, a substituição de códons que codificam aminoácidos que fornecem um homólogo funcional da proteína é um aspecto da invenção. Por exemplo, certos aminoácidos podem ser substituídos no lugar de outros aminoácidos em uma estrutura de proteína sem perda apreciável da capacidade de ligação interativa com as estruturas tais como, por exemplo, regiões de ligação de antígeno de anticorpos ou sítios de ligação nas moléculas do substrato. Visto que é a capacidade e a natureza interativas de uma proteína que define essa atividade funcional biológica da proteína, certas substituições de seqüência de aminoácido podem ser feitas em uma seqüência de proteína, e, naturalmente, a sua seqüência codificadora de DNA subjacente, e não obstante obter uma proteína com propriedades semelhantes. É assim considerado pelos inventores que várias mudanças podem ser feitas nas seqüências de peptídeo das composições divulgadas realizando-se mudanças nas seqüências de DNA correspondentes que codificam os peptídeos em que os peptídeos não mostraram nenhuma perda apreciável de sua utilidade ou atividade biológicas.

Um outro aspecto da invenção compreende homólogos funcionais, que diferem em um ou mais aminoácidos daqueles de um polipeptídeo aqui fornecido como o resultado de uma ou mais substituições de aminoácido conservativas. É bem conhecido na técnica que um ou mais aminoácidos em uma seqüência nativa podem ser substituídos com pelo menos um outro aminoácido, a carga e a polaridade do qual são similares àquelas do amino-

ácido nativo, resultando em uma mudança silenciosa. Por exemplo, valina é um substituto conservativo para alanina e treonina é um substituto conservativo para serina. As substituições conservativas para um aminoácido dentro da seqüência de polipeptídeo nativa podem ser selecionadas a partir de outros membros da classe a qual o aminoácido que ocorre naturalmente pertence. Os aminoácidos podem ser divididos nos seguintes quatro grupos: (1) aminoácidos ácidos, (2) aminoácidos básicos, (3) aminoácidos polares neutros e (4) aminoácidos não-polares neutros. Os aminoácidos representativos dentro destes vários grupos incluem, mas não são limitados a: (1) aminoácidos ácidos (negativamente carregados) tais como ácido aspártico e ácido glutâmico; (2) aminoácidos básicos (positivamente carregados) tais como arginina, histidina e lisina; (3) aminoácidos polares neutros tais como glicina, serina, treonina, cisteína, tirosina, asparagina e glutamina; e (4) aminoácidos não-polares neutros (hidrofóbicos) tais como alanina, leucina, isoleucina, valina, prolina, fenilalanina, triptofano e metionina. Os substitutos conservados para um aminoácido dentro de uma seqüência de aminoácido nativa pode ser selecionada de outros membros do grupo ao qual o aminoácido que ocorre naturalmente pertence. Por exemplo, um grupo de aminoácidos tendo cadeias laterais alifáticas é glicina, alanina, valina, leucina e isoleucina; um grupo de aminoácidos tendo cadeias laterais hidroxílicas alifáticas é serina e treonina; um grupo de aminoácidos tendo cadeias laterais contendo amida é asparagina e glutamina; um grupo de aminoácidos tendo cadeias laterais aromáticas é fenilalanina, tirosina e triptofano; um grupo de aminoácidos tendo cadeias laterais básicas é lisina, arginina e histidina; e um grupo de aminoácidos tendo cadeias laterais contendo enxofre é cisteína e metionina. Os grupos de substituição de aminoácidos naturalmente conservativos são: valina-leucina, valina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, ácido aspártico-ácido glutâmico e asparagina-glutamina.

Constructos de DNA

O Material genético exógeno pode ser transferido em uma planta pelo uso de um constructo de DNA projetada para tal propósito pelos métodos que utilizam *Agrobacterium*, bombardeamento de partícula ou outros

métodos conhecidos por aqueles versados na técnica. O projeto de tal constructo de DNA está em geral dentro do conhecimento da técnica (Plant Molecular Biology: A Laboratory Manual, eds. Clark, Springer, Nova Iorque (1997). Os exemplos de tais plantas nas quais o material genético exógeno

5 pode ser transferido, incluem, sem limitação, alfalfa, *Arabidopsis*, cevada, *Brassica*, brócoli, repolho, *Citrus*, algodão, alho, aveia, colza, cebola, canola, linho, milho, uma planta anual ornamental e perene ornamental, ervilha, amendoim, pimentão, batata, arroz, centeio, sorgo, soja, morango, cana de açúcar, beterraba, tomate, trigo, álamo, pinheiro, abeto, eucalipto, maçã,

10 alface, lentilhas, uva, banana, chá, gramas, girassol, dendezeiro, *Phaseolus*, árvores, arbustos, videiras, etc. É bem conhecido que as plantas agronomicamente importantes compreendem genótipos, variedades e cultivares e que os métodos e composições da presente invenção podem ser testados nestas plantas por aqueles versados comumente na técnica da biologia molecular

15 vegetal e do cruzamento vegetal.

Um grande número de moléculas promotoras de DNA isoladas que são ativas como um elemento genético de um transgene em células vegetais foi descrito. Estas incluem o promotor da nopalina sintase (P-nos) (Ebert *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 84: 5745 a 5749, 1987), a totali-

20 dade da qual é aqui incorporada por referência), o promotor da octopina sintase (P-ocs), que são transportadas em plasmídeos indutores de tumor de *Agrobacterium tumefaciens*, os promotores de *caulimovirus*, tais como o promotor 19S do vírus mosaico da couve-flor (CaMV) (Lawton *et al.*, Plant Mol. Biol. 9: 315 a 324, 1987), a totalidade da qual é aqui incorporada por

25 referência) e o promotor 35S de CaMV (Odell *et al.*, Nature 313: 810 a 812, 1985), a totalidade da qual é aqui incorporada por referência), o promotor 35S do vírus mosaico de escrofulária (Patente U.S. Nº 6.018.100, a totalidade da qual é aqui incorporada por referência), o promotor indutível por luz da subunidade pequena de ribulose-1,5-bis-fosfato carboxilase (ssRUBISCO), o

30 promotor Adh (Walker *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 84: 6624 a 6628, 1987), a totalidade da qual é aqui incorporada por referência), o promotor da sacarose sintase (Yang *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 87: 4144 a

4148, 1990), a totalidade da qual é aqui incorporada por referência), o promotor do complexo do gene R (Chandler *et al.*, Plant Cell 1: 1175 a 1183, 1989, a totalidade da qual é aqui incorporada por referência) e o promotor do gene da proteína de ligação de clorofila a/b, etc.

5 Uma variedade de promotores especificamente ativos em tecidos vegetativos, tais como folhas, hastes, raízes e tubérculos, podem ser usados para expressar as moléculas de ácido nucléico da presente invenção. Os exemplos de promotores específicos de tubérculos incluem, mas não são limitados aos promotores de patatina classe I e II (Bevan *et al.*, EMBO J. 8: 1899 a 1906, 1986); Koster-Topfer *et al.*, Mol Gen Genet. 219: 390 a 396, 1989); Mignery *et al.*, Gene 62: 27 a 44, 1988); Jefferson *et al.*, Plant Mol. Biol. 14: 995 a 1006, 1990), aqui incorporada por referência nas suas totalidades), o promotor para os genes ADPGPP do tubérculos, tanto a subunidade grande quanto a pequena; o promotor da sacarose sintase (Salanoubat e Belliard, Gene 60: 47 a 56, 1987), Salanoubat e Belliard, Gene 84: 181 a 185, 1989), aqui incorporada por referência em sua totalidade); e o promotor para as proteínas de tubérculos principais incluindo os complexos de proteína de 22 kd e inibidores da proteinase (Hannapel, Plant Physiol. 101: 703 a 704, 1993), aqui incorporada por referência em sua totalidade).

10 Os exemplos de promotores específicos de folha incluem mas não são limitados aos promotores da ribulose bifosfato carboxilase (RbcS ou RuBISCO) (ver, por exemplo, Matsuoka *et al.*, Plant J. 6: 311 a 319, 1994), aqui incorporada por referência em sua totalidade); o promotor do gene da proteína de ligação da clorofila de coleta de luz a/b (ver, por exemplo, Shiina *et al.*, Plant Physiol. 115: 477 a 483. 1997; Casal *et al.*, Plant Physiol. 116: 1533 a 1538, 1998, aqui incorporada por referência em suas totalidades); e o promotor do gene relacionado com myb da *Arabidopsis thaliana* (Atmyb5) (Li *et al.*, FEBS Lett. 379: 117 a 121, 1996, aqui incorporada por referência em sua totalidade). Os exemplos de promotor específico de raiz incluem mas não são limitados ao promotor para o gene da quitinase ácida (Samac *et al.*, Plant Mol. Biol. 25: 587 a 596, 1994), aqui incorporada por referência em sua totalidade); os subdomínios específicos de raiz do promotor CaMV35S que foram

20

25

30

identificados (Lam *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 86: 7890 a 7894. 1989, aqui incorporada por referência em sua totalidade); o promotor da ORF13 de *Agrobacterium rhizogenes* que exibe alta atividade nas raízes (Hansen *et al.*, Mol. Gen. Genet. 254: 337 a 343. 1997), aqui incorporada por
5 referência em sua totalidade); o promotor para o gene específico da raiz do tabaco RB7 (Patente US 5.750.386; Yamamoto *et al.*, Plant Cell 3: 371 a 382, 1991, aqui incorporada por referência em sua totalidade); e os promotores específicos da célula de raiz reportado por Conkling *et al.* (Conkling *et al.*, Plant Physiol. 93: 1203 a 1211, 1990, aqui incorporada por referência em
10 sua totalidade) e o promotor de POX1 (Pox1, pox1) (Hertig, *et al.* Plant Mol. Biol. 16: 171, 1991).

Uma outra classe de promotores específicos de tecido vegetativo útil são os promotores meristemáticos (ponta de raiz e ápice de broto). Por exemplo, os promotores "SHOOTMERISTEMLESS" e "SCARECROW",
15 que são ativos no desenvolvimento de meristemas apicais de broto ou raiz (Di Laurenzio *et al.*, Cell 86: 423 a 433, 1996; Long, Nature 379: 66 a 69, 1996); aqui incorporada por referência em suas totalidades), podem ser usados. Um outro exemplo de um promotor útil é aquele que controla a expressão do gene da 3-hidróxi-3- metilglutaril coenzima A reductase HMG2, cuja
20 expressão é restrita aos tecidos meristemáticos e florais (zona secretora do estigma, grãos de pólen maduros, tecido vascular de gineceu e óvulos fertilizados) (ver, por exemplo, Enjuto *et al.*, Plant Cell, 7: 517 a 527, 1995, aqui incorporada por referência em sua totalidade). Também um outro exemplo de um promotor útil é aquele que controla a expressão de genes relaciona-
25 dos com kn1 de milho e outras espécies que apresentam a expressão específica de meristema (ver, por exemplo, Granger *et al.*, Plant Mol. Biol. 31: 373 a 378, 1996; Kerstetter *et al.*, Plant Cell 6: 1877 a 1887, 1994; Hake *et al.*, Philos. Trans. R. Soc. Lond. B. Biol. Sci. 350: 45 a 51, 1995, aqui incorporada por referência em suas totalidades). Um outro exemplo de um promotor
30 meristemático é o promotor KNAT1 *Arabidopsis thaliana*. No ápice de broto, o transcrito KNAT1 está localizado primariamente no meristema apical de broto; a expressão de KNAT1 no meristema de broto diminui durante a tran-

sição floral e é restrita ao córtex da haste de inflorescência (ver, por exemplo, Lincoln *et al.*, Plant Cell 6: 1859 a 1876, 1994, aqui incorporada por referência em sua totalidade).

Os promotores específicos de semente e realçados de semente adequados podem ser derivados dos seguintes genes: MAC1 do milho (Sheridan *et al.*, Genetics 142: 1009 a 1020, 1996, aqui incorporada por referência em sua totalidade); Cat3 do milho (Genbank N^o L05934. Abler *et al.*, Plant Mol. Biol. 22: 10131 a 1038, 1993. aqui incorporada por referência em sua totalidade); viviparous-1 de *Arabidopsis* (Genbank N^o U93215); Atimyc1 de *Arabidopsis* (Urao *et al.*, Plant Mol. Biol. 32: 571-57, 1996; Conceicao *et al.*, Plant 5: 493 a 505, 1994, aqui incorporada por referência em suas totalidades); napA de *Brassica napus* (Genbank N^o J02798); a família do gene napin de *Brassica napus* (Sjodahl *et al.*, Planta 197: 264 a 271, 1995, aqui incorporada por referência em sua totalidade).

O promotor específico de óvulo para o gene BEL1 (Reiser *et al.* Cell 83: 735 a 742, 1995, Genbank N^o U39944; Ray *et al.*, Proc. Natl. Acad. Sci. USA 91: 5761 a 5765, 1994, todas as quais são aqui incorporada por referência em suas totalidades) também podem ser usados. Os promotores MEA (FIS1) e FIS2 específicos de ovo e célula central também são promotores específicos de tecido reprodutivo úteis (Luo *et al.*, Proc. Natl. Acad. Sci. USA, 97: 10637 a 10642, 2000; Vielle-Calzada, *et al.*, Genes Dev. 13: 2971 a 2982, 1999; aqui incorporada por referência em suas totalidades).

Um promotor específico de pólen do milho foi identificado no milho (Guerrero *et al.*, Mol. Gen. Genet. 224: 161 a 168, 1990, aqui incorporada por referência em sua totalidade). Outros genes especificamente expressados no pólen foram descritos (ver, por exemplo, Wakeley *et al.*, Plant Mol. Biol. 37: 187 a 192, 1998; Ficker *et al.*, Mol. Gen. Genet. 257: 132 a 142, 1998; Kulikauskas *et al.*, Plant Mol. Biol. 34: 809 a 814, 1997; Treacy *et al.*, Plant Mol. Biol. 34: 603 a 611, 1997; todas as quais são aqui incorporadas por referência em suas totalidades).

Os promotores derivados de genes que codificam proteínas de armazenagem embriônica, que inclui o gene que codifica a proteínas de ar-

mazenagem 2S de *Brassica napus* (Dasgupta *et al*, Gene 133: 301 a 302, 1993, aqui incorporada por referência em sua totalidade); a família de gene de proteínas de armazenagem em semente 2s de *Arabidopsis*; o gene que codifica a oleosina 20kD de *Brassica napus* (GenBank Nº M63985); os genes que codificam a oleosina A (Genbank Nº U09118) e a oleosina B (GenBank Nº U09119) da soja; o gene que codifica a oleosina de *Arabidopsis* (GenBank Nº Z17657); o gene que codifica a oleosina 18kD do milho (GenBank Nº J05212, Lee, Plant Mol. Biol. 26: 1981 a 1987, 1994), aqui incorporada por referência em sua totalidade); e o gene que codifica a proteína rica em enxofre de peso molecular baixo da soja (Choi *et al.*, Mol. Gen. Genet. 246: 266 a 268, 1995, aqui incorporada por referência em sua totalidade), também podem ser usados.

Os promotores derivados de zeína que codifica genes (incluindo os genes de 15 kD, 16 kD, 19 kD, 22 kD, 27 kD e gama; Pedersen *et al.*, Cell 29: 1015 a 1026, 1982, aqui incorporada por referência em sua totalidade) também podem ser usados. As zeínas são um grupo de proteínas de armazenagem encontradas no endosperma do milho.

Outros promotores conhecidos funcionar, por exemplo, no milho, incluem os promotores para os seguintes genes: *waxy*, *Brittle*, *Shrunken 2*, Enzimas de ramificação I e II, amido sintases, enzimas de desramificação, oleosinas, glutelinas e sacarose sintases. Um promotor particularmente preferido para a expressão do endosperma do milho é o promotor para o gene da glutelina do arroz, mais particularmente o promotor Osgt-1 (Zheng *et al.*, Mol. Cell Biol. 13: 5829 a 5842, 1993, aqui incorporada por referência em sua totalidade). Os exemplos de promotores adequados para a expressão no trigo incluem aqueles promotores para as subunidades de ADPGlicose pirofosforilase (ADPGPP), a ligação de grânulo e outras amido sintases, as enzimas de ramificação e desramificação, as proteínas abundantes em embriogênese, as gliadinas e as gluteninas. Os exemplos de tais promotores no arroz incluem aqueles promotores para as subunidades de ADPGPP, a ligação de grânulo e outras amido sintases, as enzimas de ramificação, as enzimas de desramificação, as sacarose sintases e as glutelinas. Um promotor

particularmente preferido é o promotor para a glutelina do arroz, Osgt-1. Os exemplos de tais promotores para a cevada incluem aqueles para as subunidades de ADPGPP, a ligação de grânulo e outras amido sintases, as enzimas de ramificação, as enzimas de desramificação, as sacarose sintases, as hordeínas, as globulinas de embrião e as proteínas específicas de aleurona.

Um promotor do tomate ativo durante o amadurecimento da fruta, senescência e abscisão das folhas e, em um grau menor, das flores pode ser usado (Blume *et al.*, Plant J. 12: 731 a 746, 1997, aqui incorporada por referência em sua totalidade). Outros promotores exemplares incluem o promotor específico do gene SK2 de *pistol* na batata (*Solanum tuberosum* L.), que codifica uma endoquitinase básica específica de pistilo (Ficker *et al.*, Plant Mol. Biol. 35: 425 a 431, 1997, aqui incorporada por referência em sua totalidade); o gene Blec4 da ervilha (*Pisum sativum* cv. Alaska), ativo no tecido epidérmico de ápices de broto vegetativo e floral de alfalfa transgênica. Isto o torna uma ferramenta útil para alvejar a expressão de genes estranhos à camada epidérmica de brotos que crescem ativamente. O promotor E8 específico de tecido do tomate também é útil para direcionar a expressão de gene em frutas (Deikman, *et al.*, Plant Physiology 100: 2013 a 2017, 1992).

É ainda reconhecido que, visto que na maioria dos casos os limites exatos de seqüências reguladoras não foram completamente definidos, fragmentos de DNA de comprimentos diferentes podem ter atividade promotora idêntica.

Os promotores que são conhecidos ou são descoberto causar a transcrição de DNA em células vegetais podem ser usados na presente invenção. Tais promotores podem ser obtidos a partir de uma variedade de fontes tais como plantas e vírus de planta. Além dos promotores que são conhecidos por causar a transcrição de DNA em células vegetais, outras moléculas promotoras podem ser identificadas para o uso na corrente invenção triando-se uma biblioteca de cDNA vegetal para os genes que são seletiva ou preferivelmente expressados nos tecidos ou células alvo e isolando a região genômica 5' dos cDNAs identificados.

Os constructos ou vetores também podem incluir com a região codificadora de interesse um ácido polinucleico que atua, no todo ou em parte, para terminar a transcrição daquela região. Por exemplo, tais seqüências foram isoladas incluindo a seqüência Tr7 3' e a seqüência *nos* 3' (Ingelbrecht *et al.*, The Plant Cell 1: 671 a 680, 1989, a totalidade da qual é aqui incorporada por referência; Bevan *et al.*, Nucleic Acids Res. 11: 369 a 385, 1983, a totalidade da qual é aqui incorporada por referência).

Um vetor ou constructo também podem incluir elementos reguladores. Os exemplos de tais incluem o intron Adh 1 (Callis *et al.*, Genes e Develop. 1: 1183 a 1200, 1987, a totalidade da qual é aqui incorporada por referência), o intron da sacarose sintase (Vasil *et al.*, Plant Physiol. 91: 1575 a 1579, 1989, a totalidade da qual é aqui incorporada por referência) e o elemento TMV ômega (Gallie *et al.*, Plant Cell 1: 301 a 311, 1989, a totalidade da qual é aqui incorporada por referência). Estes e outros elementos reguladores podem ser incluídos quando apropriado.

Um vetor ou constructo também podem incluir um marcador selecionável. Os marcadores selecionáveis também podem ser usados para selecionar plantas ou células vegetais que contêm o material genético exógeno. Os exemplos de tais incluem, mas não são limitados a, um gene *neo* (Potrykus *et al.*, Mol. Gen. Genet. 199: 183 a 188, 1985, a totalidade da qual é aqui incorporada por referência) que codifica a resistência à canamicina e pode ser selecionada para usar a canamicina, G418, etc.; um gene *bar* que fornece a resistência ao bialafos; um gene da EPSP sintase mutante (Hinchey *et al.*, Bio/Technology 6: 915 a 922 (1988), a totalidade da qual é aqui incorporada por referência) que fornece a resistência ao glifosato; um gene da nitrilase que fornece resistência à bromoxinila (Stalker *et al.*, J. Biol. Chem. 263: 6310 a 6314 (1988), a totalidade da qual é aqui incorporada por referência); um gene da acetolactato sintase mutante (ALS) que confere imidazolinona ou sulfoniluréia; e um gene DHFR resistente ao metotrexato (Thillet *et al.*, J. Biol. Chem. 263: 12500 a 12508, 1988, a totalidade da qual é aqui incorporada por referência).

Um vetor ou constructo também pode incluir um marcador triá-

vel. Os marcadores triáveis podem ser usados para monitorar a expressão. Os marcadores triáveis exemplares incluem um gene da β -glicuronidase ou *uidA* (GUS) que codifica uma enzima para a qual vários substratos cromogênicos são conhecidos (Jefferson, Plant Mol. Biol. Rep. 5: 387 a 405 (1987), a totalidade da qual é aqui incorporada por referência; Jefferson *et al.*, EMBO J. 6: 3901 a 3907 (1987), a totalidade da qual é aqui incorporada por referência); um gene do local R, que codifica um produto que regula a produção de pigmentos de antocianina (cor vermelha) em tecidos vegetais (Dellaporta *et al.*, Stadler Symposium 11: 263 a 282 (1988), a totalidade da qual é aqui incorporada por referência); um gene da β -lactamase (Sutcliffe *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 75: 3737 a 3741 (1978), a totalidade da qual é aqui incorporada por referência), um gene que codifica uma enzima para a qual vários substratos cromogênicos são conhecidos (por exemplo, PADAC, uma cefalosporina cromogênica); um gene da luciferase (Ow *et al.*, Science 234: 856 a 859 (1986), a totalidade da qual é aqui incorporada por referência); um gene *xylE* (Zukowsky *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 80: 1101 a 1105 (1983), a totalidade da qual é aqui incorporada por referência) que codifica uma catecol dioxigenase que pode converter catecóis cromogênicos; um gene da α -amilase (Ikata *et al.*, Bio/Technol. 8: 241 a 242, 1990, a totalidade da qual é aqui incorporada por referência); um gene da tirosinase (Katz *et al.*, J. Gen. Microbiol. 129: 2703 a 2714, 1983, a totalidade da qual é aqui incorporada por referência) que codifica uma enzima capaz de oxidar a tirosina a DOPA e dopaquinona que por sua vez condensa até melanina; e uma α -galactosidase.

25 Introdução de Polinucleotídeos em Plantas

Existem muitos métodos para introduzir moléculas de ácido nucléico transformadoras em células vegetais. Acredita-se que os métodos adequados incluam virtualmente qualquer método pelo qual as moléculas de ácido nucléico possam ser introduzidas em uma célula, tal como pela infecção com *Agrobacterium* ou liberação direta de moléculas de ácido nucléico tais como, por exemplo, pela transformação mediada por PEG, pela eletroporação ou pela aceleração de partículas revestidas com DNA, etc. (Po-

trykus, Ann. Rev. Plant Physiol. Plant Mol. Biol. 42: 205 a 225 (1991), a totalidade da qual é aqui incorporada por referência; Vasil, Plant Mol. Biol. 25: 925 a 937 (1994), a totalidade da qual é aqui incorporada por referência). Por exemplo, a eletroporação foi usada para transformar protoplastos de *Zea mays* (Fromm *et al.*, Nature 312: 791 a 793, 1986, a totalidade da qual é aqui incorporada por referência).

Outros sistemas de vetor adequados para introduzir DNA transformador em uma célula hospedeira vegetal incluem mas não são limitados aos vetores de cromossoma artificial binário (BIBAC) (Hamilton *et al.*, Gene 200: 107 a 116, 1997, a totalidade da qual é aqui incorporada por referência) e a transfecção com vetores de RNA viral (Della-Cioppa *et al.*, Ann. N. Y. Acad. Sci. (1996), 792 pp Engineering Plants for Commercial Products e Applications, pp 57 a 61, a totalidade da qual é aqui incorporada por referência).

A tecnologia para a introdução de DNA em células é bem conhecida por aqueles versados na técnica. Quatro métodos gerais para a liberação de um gene em células foram descritos: (1) métodos químicos (Graham e van der Eb, Virology 54: 536 a 539, 1973, a totalidade da qual é aqui incorporada por referência); (2) métodos físicos tais como microinjeção (Capecchi, Cell 22: 479 a 488 (1980), a totalidade da qual é aqui incorporada por referência), eletroporação (Wong e Neumann, Biochem. Biophys. Res. Commun. 107: 584 a 587 (1982); Fromm *et al.*, Proc. Natl. Acad. Sci. (U.S.A.) 82: 5824 a 5828 (1985); Patente U.S. Nº 5.384.253, todas as quais são aqui incorporadas em sua totalidade); e a pistola de gene (Johnston e Tang, Methods Cell Biol. 43: 353 a 365 (1994), a totalidade da qual é aqui incorporada por referência); (3) vetores virais (Clapp, Clin. Perinatol. 20: 155 a 168, 1993; Lu *et al.*, J. Exp. Med. 178: 2089 a 2096, 1993; Eglitis e Anderson, Biotechniques 6: 608 a 614, 1988, todas as quais são aqui incorporadas em sua totalidade); e (4) mecanismos mediados por receptor (Curiel *et al.*, Hum. Gen. Ther. 3: 147 a 154, 1992, Wagner *et al.*, Proc. Natl. Acad. Sci. USA 89: 6099 a 6103, 1992, todas as quais são incorporadas por referência em sua totalidade).

Os métodos de aceleração que podem ser usados incluem, por exemplo, o bombardeamento de microprojétil e outros. Um exemplo de um método para a liberação de moléculas de ácido nucléico transformadoras às células vegetais é o bombardeamento de microprojétil. Este método foi revisado por Yang e Christou, eds., *Particle Bombardment Technology for Gene Transfer*, Oxford Press, Oxford, Inglaterra (1994), a totalidade da qual é aqui incorporada por referência. As partículas não-biológicas (microprojéteis) que podem ser revestidas com ácidos nucléicos e liberados nas células por uma força de propulsão. As partículas exemplares incluem aquelas compreendidas de tungstênio, ouro, platina e outros.

A transferência mediada por *Agrobacterium* é um sistema amplamente aplicável para introduzir genes em células vegetais porque o DNA pode ser introduzido nos tecidos vegetais integrais, desviando deste modo a necessidade para a regeneração de uma planta intacta a partir de um protoplasto. O uso de vetores de integração vegetal medidos por *Agrobacterium* para introduzir DNA em células vegetais é bem conhecido na técnica. Ver, por exemplo os métodos descritos por Fraley *et al.*, *Bio/Technology* 3: 629 a 635 (1985) e Rogers *et al.*, *Methods Enzymol.* 153: 253 a 277 (1987), ambas as quais são aqui incorporadas por referência em sua totalidade. Além disso, a integração do Ti-DNA é um processo relativamente preciso que resulta em poucos rearranjos. A região de DNA a ser transferida é definida pelas seqüências da extremidade e o DNA intervector é usualmente inserido no genoma da planta como descrito (Spielmann *et al.*, *Mol. Gen. Genet.* 205: 34 (1986), a totalidade da qual é aqui incorporada por referência).

Uma planta transgênica que resulta de métodos de transformação com *Agrobacterium* freqüentemente contém um único gene em um cromossoma. Tais plantas transgênicas podem ser aludidas como sendo homozigotas em relação ao gene adicionado. Mais preferido é uma planta transgênica que seja homozigota em relação ao gene estrutural adicionado; isto é, uma planta transgênica que contenha dois genes adicionados, um gene no mesmo local em cada cromossoma de um par de cromossomas. Uma

planta transgênica homozigota pode ser obtida acasalando-se sexualmente (autopolinização) uma planta transgênica segregante independente que contenha um único gene adicionado, germinando algumas das sementes produzidas e analisando as plantas resultantes produzidas para o gene de interesse.

Também deve ser entendido que duas plantas transgênicas diferentes também podem ser acasaladas para produzir progênie que contenha dois genes exógeno independentemente segregados adicionados. A autopolinização de progênie apropriada pode produzir plantas que sejam homozigotas em relação a ambos os genes exógenos adicionados que codificam um polipeptídeo de interesse. O retro-cruzamento com uma planta precursora e o cruzamento com uma planta não -transgênica também são considerados, como o é a propagação vegetativa.

A regeneração, desenvolvimento e cultivo de plantas a partir de um único transformante de protoplasto vegetal ou a partir de vários explantes transformados são bem conhecidos na técnica (Weissbach e Weissbach, Em: Methods for Plant Molecular Biology, (eds.), Academic Press, Inc. San Diego, CA, (1988), a totalidade da qual é aqui incorporada por referência). Este processo de regeneração e cultivo tipicamente inclui as etapas de seleção de células transformadas, cultivando-se estas células individualizadas através dos estágios usuais de desenvolvimento embriônico até o estágio de muda enraizada. Os embriões e as sementes transgênicas são similarmente regenerados. Os brotos enraizados transgênicos resultantes são depois disso plantados em um meio de cultivo de planta apropriado tal como o solo.

O desenvolvimento ou regeneração de plantas contendo o gene estranho, exógeno que codifica uma proteína de interesse é bem conhecido na técnica. Preferivelmente, as plantas regeneradas são autopolinizadas para fornecer plantas transgênicas homozigotas. De outro modo, o pólen obtido a partir das plantas regeneradas é cruzado com plantas desenvolvidas com semente de linhagens agronomicamente importantes. Ao contrário, o pólen de plantas destas linhagens importantes é usado para polinizar plantas regeneradas. Uma planta transgênica da presente invenção conten-

do um polipeptídeo desejado é cultivada usando métodos bem conhecidos por uma pessoa versada na técnica.

A presente invenção também fornece partes das plantas da presente invenção. As partes de planta, sem limitação, incluem semente, endosperma, óvulo e pólen. Em uma forma de realização particularmente preferida da presente invenção, a parte de planta é uma semente.

Os métodos para transformar dicotiledôneas, primariamente pelo uso de *Agrobacterium tumefaciens* e a obtenção de plantas transgênicas foi publicado, por exemplo, algodão (Patente U.S. Nº 5.004.863. Patente U.S. Nº 5.159.135. Patente U.S. Nº 5.518.908, todas as quais são aqui incorporada por referência em sua totalidade), soja (Patente U.S. Nº 5.569.834. a totalidade da qual é aqui incorporada por referência) e Brassica (Patente U.S. Nº 5.463.174. a totalidade da qual é aqui incorporada por referência).

A transformação de monocotiledôneas usando a eletroporação, bombardeamento de partícula e *Agrobacterium* também foi relatada. Por exemplo, a transformação e a regeneração de planta foi obtida no aspargo, cevada, *Zea mays* (Fromm *et al.*, Bio/Technology 8: 833 (1990), Armstrong *et al.*, Crop Science 35: 550 a 557 (1995), todas as quais são aqui incorporada por referência em sua totalidade); aveia; arroz, cana de açúcar; festuca alta e trigo (Patente U.S. Nº 5.631.152, a totalidade da qual é aqui incorporada por referência).

Além dos procedimentos debatidos acima, os técnicos estão familiarizados com os materiais de recurso padrão que descrevem condições e procedimentos específicos para a construção, manipulação e isolamento de macromoléculas (por exemplo, moléculas de DNA, plasmídeos, etc.), a geração de organismos recombinantes e a triagem e isolamento de clones, (ver por exemplo, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press (1989); Mailga *et al.*, Methods in Plant Molecular Biology, Cold Spring Harbor Press (1995), a totalidade da qual é aqui incorporada por referência; Birren *et al.*, Genome Analysis: Detecting Genes, 1, Cold Spring Harbor, Nova Iorque (1998), a totalidade da qual é aqui incorporada por referência; Birren *et al.*, Genome Analysis: Analyzing DNA, 2, Cold

Spring Harbor, Nova Iorque (1998), a totalidade da qual é aqui incorporada por referência; Plant Molecular Biology: A Laboratory Manual, eds. Clark, Springer, Nova Iorque (1997), a totalidade da qual é aqui incorporada por referência).

- 5 Tendo agora no geral descrito a invenção, a mesma será mais facilmente entendida através da referência aos seguintes exemplos, que são fornecidos por via de ilustração e não são intencionados a serem limitantes da presente invenção, a menos que especificado.

EXEMPLOS

10 EXEMPLO 1

- Quando um polinucleotídeo vegetal nativo isolado que compreende uma sequência codificadora é reconstruído como um transgene, depois introduzido na planta pelos métodos de transformação vegetal existe um risco de que a expressão do gene vegetal homólogo endógeno interagirá negativamente com o transgene. Para evitar estas interações negativas pode ser necessário fornecer um polinucleotídeo de transgene substancialmente divergente em sequência do gene vegetal nativo. Uma molécula de polinucleotídeo artificial pode ser produzida pelo método da presente invenção e usada para reduzir a ocorrência de silenciamento de transgene.

- 20 Este exemplo serve para ilustrar os métodos da presente invenção que resultam na produção de um polinucleotídeo que codifica uma EPSP sintase vegetal modificada. A enzima de EPSPS e o peptídeo de trânsito de cloroplasto do arroz nativo (*Oryza sativa*) são usados para construir uma molécula de polinucleotídeo artificial que também inclui códons que codificam aminoácidos substituídos que não ocorrem naturalmente na enzima de EPSPS do arroz. Estes aminoácidos substituídos fornecem uma enzima de EPSPS do arroz resistente ao glifosato (OsEPSPS_TIPS, SEQ ID NO: 1).

- 25 As etapas descritas na Tabela 6 são usadas para construir uma tal sequência de polinucleotídeo artificial (OsEPSPS_AT, SEQ ID NO: 3)
- 30 usando uma tabela de uso do códon de *Arabidopsis* e os parâmetros para a construção de uma molécula de polinucleotídeo substancialmente divergente, que quando expressada em plantas codifica uma enzima de EPSPS do

arroz modificada resistente ao herbicida de glifosato. A comparação da sequência de gene de EPSPS do arroz nativo aludida como OsEPSPS_Nat (SEQ ID NO: 2) que foi previamente modificada para codificar uma enzima resistente ao glifosato com a molécula de polinucleotídeo modificada quanto ao uso de códon de *Arabidopsis*, OsEPSPS_AT (SEQ ID NO: 3) e com a sequência modificada quanto ao uso de códon de *Zea mays*, OsEPSPS_ZM (SEQ ID NO: 4) por este método é mostrada na Figura 1. A Figura 1 mostra bases de nucleotídeo mudadas nos polinucleotídeos modificados comparados à OsEPSPS_Nat, SEQ ID NO: 2.

10 Tabela 6. Planejamento de polinucleotídeo para uma EPSP sintase de arroz modificada (OsEPSPS_AT)

1. Substituir os aminoácidos nas posições 173 e 177 para fornecer uma enzima de EPSPS do arroz modificada resistente ao herbicida de glifosato mostrada na SEQ ID NO: 1.
- 15 2. Retro traduzir a SEQ ID NO: 1 para gerar uma sequência de polinucleotídeo artificial usando a tabela de uso de códon de *Arabidopsis thaliana* (Tabela 2).
3. Realizar o alinhamento de sequência com a sequência de polinucleotídeo OsEPSPS nativa (SEQ ID NO: 2) e a sequência de polinucleotídeo artificial para determinar o grau de identidade de sequência, mapear as matrizes de leitura aberta, selecionar padrões para procurar e identificar sequências de reconhecimento de enzimas de restrição.
- 20 4. Fazer correções nos códons usados na sequência de polinucleotídeo artificial para se obter a porcentagem desejada de identidade de sequência e evitar a formação de cacho de códons idênticos. Isto é especialmente importante para os aminoácidos que ocorrem em alta frequência, isto é, alanina, glicina, histidina, leucina, serina e valina. Distribuição aproximada de uso de códon na sequência de polinucleotídeo de acordo com o uso de códon de *Arabidopsis*, Tabela 2.
- 25 5. A sequência de polinucleotídeo é inspecionada quanto a regiões locais que tenham uma relação GC:AT mais alta do que cerca de 2 em uma faixa de cerca de 50 nucleotídeos contíguos. A sequência de polinucleotídeo

é ajustada como necessário, substituindo-se os códons nestas regiões tal que a relação GC:AT local seja menor do que cerca de 2 e a composição de polinucleotídeo inteira esteja na faixa de 0,9 a 1,3.

6. Introduzir códons de parada às matrizes de tradução "b", "c", "d", "e" e "f". Os códons de parada de tradução são criados nas matrizes de tradução "b", "c", "d", "e" e "f" substituindo-se um ou mais códons dentro de cerca de 130 pares de bases (pares de base) das extremidades do polinucleotídeo artificial que cria um códon de parada sem mudar a seqüência codificadora de aminoácido da matriz um.
7. Eliminar os códons de ATG das matrizes de leitura aberta avançada (matrizes "b" e "c") e reversa (matrizes "d", "e", "f"). As matrizes de leitura avançada e reversa são inspecionadas quanto a presença de códons ATG. Qualquer códon ATG nas matrizes "b" e "c" encontrado na seqüência de polinucleotídeo antes do terceiro Met na matriz "a" do polinucleotídeo é eliminado pela substituição de um ou mais códons que envolve o ATG mudando um dos nucleotídeos sem mudar a seqüência codificadora de aminoácido da matriz "a". Nas matrizes reversas, a substituição de ATG ou a introdução do códon de parada podem ser feitos para interromper matrizes de leitura potenciais.
8. Eliminar os padrões de reconhecimento das enzimas de restrição não desejadas e outros padrões específicos (poliadenilação, junção de RNA, padrões de instabilidade de seqüência). A seqüência de polinucleotídeo é inspecionada quanto a presença de qualquer padrão de polinucleotídeo não desejado e os padrões são rompidos pela substituição de códons nestas regiões.
9. Checar a identidade de seqüência entre um primeiro polinucleotídeo e o polinucleotídeo artificial criado pelo método da presente invenção. Eliminar a identidade de seqüência em um polinucleotídeo contíguo que seja maior do que 23 pares de base. É desejável eliminar a identidade de seqüência maior do que cerca de 15 pares de base. É útil selecionar a partir de aminoácidos tais como, serina, arginina e leucina que têm 6 códons ou a partir de aminoácidos com 4 códons para eliminar a identidade de seqüên-

cia.

10. Revisar a seqüência de polinucleotídeo artificial que resulta de qualquer uma das etapas de 1 a 9 quanto a qualquer uma das características de seqüência identificadas nas etapas de 4 a 9 e se a seqüência não estiver de acordo com as condições realizar substituições de códon adicionais à seqüência até que as condições das etapas de 4 a 9 sejam atingidas.

11. Construir a molécula de polinucleotídeo artificial pelos métodos conhecidos na técnica, por exemplo, usando PCR com uma mistura de iniciadores envolvidos. Os iniciadores nas extremidades do gene podem conter sítios de restrição convenientes para permitir a clonagem fácil do gene no vetor selecionado. Na extremidade 5' usualmente AlflII, BspHI, NcoI, NdeI, PciI ou SphI são mais convenientes visto que a sua seqüência contém um códon de partida ATG, entretanto outras enzimas também podem ser usadas se um polinucleotídeo modificado é planejado para criar uma fusão com um outro segmento de polinucleotídeo, por exemplo, peptídeo de trânsito de cloroplasto e seqüência codificadora de EPSPS.

12. Realizar uma análise de seqüência de DNA do polinucleotídeo artificial para confirmar a construção sintética resultante na molécula de polinucleotídeo desejada. Se erros são encontrados, então eliminar os mesmos pela mutagênese direcionada ao sítio para a qual muitos métodos são conhecidos por aqueles versados na técnica da mutagênese de DNA.

Uma versão do uso de códon de *Zea mays* (*Zea mays*, Tabela 3) da seqüência de enzima de EPSPS do arroz resistente ao glifosato (enzima de EPSPS do *Oryzae sativa* com mutações TIPS, SEQ ID NO: 1) é fabricada. O polinucleotídeo que codifica esta enzima inclui códons que codificam aminoácidos substituídos que não ocorrem naturalmente na enzima de EPSPS do arroz nativa. Estes aminoácidos substituídos fornecem uma enzima de EPSPS do arroz resistente ao glifosato. As etapas descritas na Tabela 7 são usadas para construir uma seqüência de polinucleotídeo artificial modificada (OsEPSPS_ZM, SEQ ID NO: 4) com base em uma tabela de uso do códon de *Zea mays* que codifica uma enzima de EPSPS do arroz modificada resistente ao herbicida de glifosato. A comparação da seqüência de

polinucleotídeo OsEPSPS_Nat (SEQ ID NO: 2) com a seqüência de polinucleotídeo artificial OsEPSPS_ZM (SEQ ID NO: 4) usando o uso de códon de *Zea mays* é mostrada na Figura 2.

Tabela 7. Construção de polinucleotídeo para a EPSP sintase de arroz modificada (OsEPSPS_ZM)

1. Retro traduzir a SEQ ID NO: 1 para gerar uma seqüência de polinucleotídeo artificial usando a tabela de uso do códon de *Zea mays* (Tabela 3).
2. Realizar o alinhamento de seqüência com a seqüência de polinucleotídeo OsEPSPS nativa (SEQ ID NO: 2) e a seqüência de polinucleotídeo artificial para determinar o grau de identidade de seqüência, mapear as matrizes de leitura aberta, selecionar padrões para procurar e identificar seqüências de reconhecimento de enzimas de restrição.
3. Fazer correções nos códons usados na seqüência de polinucleotídeo artificial para se obter a porcentagem desejada de identidade de seqüência e evitar a formação de cachos de códons idênticos. Isto é especialmente importante para os aminoácidos que ocorrem em alta freqüência, isto é, alanina, glicina, histidina, leucina, serina e valina. Distribuição aproximada de uso de códon na seqüência de polinucleotídeo de acordo com o uso de códon de *Zea mays*, Tabela 3.
4. A seqüência de polinucleotídeo é inspecionada quanto às regiões locais que têm uma relação GC:AT maior do que cerca de 2 em uma faixa de cerca de 50 nucleotídeos contíguos. A seqüência de polinucleotídeo é ajustada como necessário, pela substituição de códons nestas regiões tal que a relação GC:AT local seja menor do que cerca de 2 e a composição de polinucleotídeo inteira esteja na faixa de 1,2 a 1,7.
5. Seguir as etapas de 6 a 12 da Tabela 6.

Tabela 8. Identidade de seqüência percentual entre polinucleotídeos OsEPSPS.

	OsEPSPS_ZM	OsEPSPS_AT	OsEPSPS_Nat
OsEPSPS_ZM	100,00	73,51	71,58
OsEPSPS_AT		100,00	74,03
OsEPSPS_Nat			100,00

Tabela 9. A composição de nucleotídeo e a relação GC:AT das seqüências de polinucleotídeo modificadas para a seqüência de gene de EPSPS do arroz.

	A	C	G	T	GC:AT
OsEPSPS_AT	377	336	444	391	1.02
OsEPSPS_ZM	365	381	470	332	1.22

As duas seqüências de polinucleotídeo artificial de EPSPS do arroz (SEQ ID NO: 3 e SEQ ID NO: 4) são modificadas tal que a identidade percentual esteja abaixo de 75 por cento comparada com a SEQ ID NO: 2 ou em relação uma com a outra (Tabela 8). A composição de nucleotídeo e a relação GC:AT das seqüências de polinucleotídeo para a seqüência de gene de EPSPS do arroz são mostradas na Tabela 9. Estes polinucleotídeos podem ser selecionados para o uso em constructos de expressão vegetal junto com elementos reguladores diferentes ou eles podem ser combinados em uma única planta pela retransformação com um constructo de DNA ou pelos métodos de cruzamento vegetal. Problemas com o silenciamento e a recombinação de gene são reduzidos quando os constructos de DNA têm níveis reduzidos de DNA homólogo.

EXEMPLO 2

O Milho (*Zea mays*) foi geneticamente modificado para ter resistência ao herbicida de glifosato (Patente US Nº 6.040.497). Estas plantas de milho contêm um transgene com uma EPSP sintase do milho modificado quanto a tolerância ao glifosato. Os métodos da presente invenção podem ser usados para construir um novo polinucleotídeo artificial que codifica uma EPSP sintase do milho que é substancialmente diferente na identidade percentual ao gene da EPSP sintase do milho endógeno. O polinucleotídeo arti-

ficial da EPSP sintase do milho recém construído pode ser usado como um marcador selecionável durante a seleção de linhagens de planta transgênica que podem conter traços agrônômicos transgênicos adicionais. Durante a produção de semente de milho híbrido, é útil ter ambos os precursores tolerantes ao glifosato usando transgenes que não interfiram.

Tabela 10. Constructo de polinucleotídeo para a EPSP sintase do milho modificada (ZmEPSPS_ZM, SEQ ID NO: 10)

1. Retro traduzir a SEQ ID NO: 8 para gerar uma seqüência de polinucleotídeo usando a tabela de uso do códon de *Zea mays* (Tabela 3).
- 10 2. Realizar o alinhamento de seqüência com a seqüência de polinucleotídeo ZmEPSPS_Nat (SEQ ID NO: 9) e a seqüência de polinucleotídeo artificial para determinar o grau de identidade de seqüência, mapear as matrizes de leitura aberta, selecionar padrões para procurar e identificar seqüências de reconhecimento de enzima de restrição.
- 15 3. Fazer correções nos códons usados na seqüência de polinucleotídeo artificial para se obter a porcentagem desejada de identidade de seqüência e para evitar a formação de cachos de códons idênticos. Isto é especialmente importante para os aminoácidos que ocorrem em alta frequência, isto é, alanina, glicina, histidina, leucina, serina e valina. Distribuição
- 20 aproximada de uso de códon na seqüência de polinucleotídeo de acordo com o uso de códon de *Zea mays*, Tabela 3.
4. A seqüência de polinucleotídeo artificial é inspecionada quanto às regiões locais que têm uma relação GC:AT maior do que cerca de 2 em uma faixa de cerca de 50 nucleotídeos contíguos. A seqüência de polinucleotídeo é ajustada como necessário, pela substituição de códons nestas regiões tal que a relação GC:AT local seja menor do que cerca de 2 e a composição de polinucleotídeo inteira esteja na faixa de 1,2 a 1,7.
- 25 5. Seguir as etapas de 6 a 12 da Tabela 6.

Tabela 11. Identidade de sequência percentual entre polinucleotídeos ZmEPSPS.

	ZmEPSPS_ZM	ZmEPSPS_Nat
ZmEPSPS_ZM	100.00	74.81
ZmEPSPS_Nat		100.00

A sequência de nucleotídeo do gene de EPSPS do milho também é modificada para reduzir identidade entre o gene sintético e nativo e manter a relação GC:AT global típica para as monocotiledôneas. A relação GC:AT para a sequência de ZmEPSPS_ZM é de 1,38. A identidade de sequência é reduzida a cerca de 75 % entre a nativa (ZmEPSPS_Nat, SEQ ID NO: 9) e a sintética (ZmEPSPS_ZM, SEQ ID NO: 10).

A comparação de polinucleotídeos nativos que codificam a EPSPS indica que o peptídeo de trânsito de cloroplasto é o fragmento mais divergente do gene. A similaridade na sequência de nucleotídeo de peptídeos maduros é maior do que 88 % para as enzimas do milho e do arroz e algumas regiões conservadas têm identidade de sequência tão longas quanto 50 pares de base. O silenciamento de gene pós-transcricional foi observado para seqüências tão pequenas quanto 60 polinucleotídeos (Sijen *et al.*, Plant Cell, 8: 2277 a 2294. 1996; Mains, Plant Mol. Biol. 43: 261 a 273, 2000).

EXEMPLO 3

A soja (*Glicina max*) foi geneticamente modificada para ser tolerante ao glifosato pela expressão de uma EPSPS de classe II isolada de *Agrobacterium* (Padgett *et al.* Crop Sci. 35: 1451 a 1461, 1995). Uma sequência de gene de EPSPS da soja nativa foi identificada e uma sequência de polinucleotídeo artificial planejada usando o método da presente invenção. O polinucleotídeo artificial codifica uma sequência de proteína que é modificada para produzir uma enzima de EPSPS resistente ao glifosato (GmEPSPS_IKS, SEQ ID NO: 5) pela substituição dos aminoácidos T para I, R para K e P para S dentro do motivo GNAGTAMRP, resultando em uma enzima de EPSPS de soja modificada com o motivo GNAGIAMKS (SEQ ID NO: 34), também aludido como o mutante IKS. A expressão de uma enzima

de EPSPS modificada nas células de uma planta pela transformação com um cassete de expressão vegetal transgênico, que contém um polinucleotídeo que codifica a EPSPS modificada com o motivo GNAGIAMKS conferirá a tolerância ao glifosato às plantas. As substituições de aminoácido adicionais para a arginina (R) no motivo também pode incluir asparagina (N).

Tabela 12. Constructo de polinucleotídeo para o gene da EPSP sintase de soja modificado (GmEPSPS_GM, SEQ ID NO: 7).

1. Retro traduzir a SEQ ID NO: 5 para gerar uma seqüência de polinucleotídeo artificial usando a tabela de uso do códon de *Glicina max* (Tabela 4).
2. Realizar o alinhamento de seqüência com a seqüência de polinucleotídeo GmEPSPS_Nat (SEQ ID NO: 6) e a seqüência de polinucleotídeo artificial para determinar o grau de identidade de seqüência, mapear as matrizes de leitura aberta, selecionar os padrões para procurar e identificar as seqüências de reconhecimento de enzima de restrição.
3. Fazer as correções para os códons usados na seqüência de polinucleotídeo artificial para se obter a porcentagem desejada de identidade de seqüência e para evitar a formação de cachos de códons idênticos. Isto é especialmente importante para os aminoácidos que ocorrem em alta frequência, isto é, alanina, glicina, histidina, leucina, serina e valina. Distribuição aproximada de uso de códon na seqüência de polinucleotídeo de acordo com o uso de códon de *Glicina max*, Tabela 4.
4. A seqüência de polinucleotídeo é inspecionada quanto às regiões locais que têm uma relação GC:AT maior do que cerca de 2 em uma faixa de cerca de 50 nucleotídeos contíguos. A seqüência de polinucleotídeo é ajustada como necessário, pela substituição de códons nestas regiões tal que a relação GC:AT local seja menor do que cerca de 2 e a composição de polinucleotídeo inteira esteja na faixa de 0,9 a 1,3.
5. Seguir as etapas de 6 a 12 da Tabela 6.

Tabela 13. Comparação da identidade de seqüência percentual do GmEPSPS modificado ao nível da seqüência de polinucleotídeo.

	GmEPSPS_GM	GmEPSPS_Nat
GmEPSPS_GM	100,00	72,43
GmEPSPS_Nat		100,00

O gene EPSPS da soja nativa é modificado usando uma tabela de códon de soja (Tabela 4) e as condições do método da presente invenção. A relação relativa de GC:AT não é mudada no gene modificado, entretanto a identidade de seqüência entre as duas é reduzida em 72 %.

EXEMPLO 4

O gene do polinucleotídeo *aroA* nativo isolado de *Agrobacterium* cepa CP4 (Patente U.S. Nº 5.633.435, aqui incorporada por referência em sua totalidade) que codifica uma EPSP sintase resistente ao glifosato (SEQ ID NO: 15) pode ser modificada pelo método da presente invenção para fornecer um polinucleotídeo que tenha o uso de códon de *Arabidopsis*, *Zea mays* ou *Glicina max*. Para a expressão apropriada de CP4EPSPS para conferir a tolerância ao glifosato em plantas, um peptídeo de trânsito de cloroplasto é necessariamente fundido à seqüência codificadora de CP4EPSPS para alvejar o acúmulo da enzima nos cloroplastos. O peptídeo de trânsito de cloroplasto CTP2 é habitualmente usado para a expressão deste gene em plantas transgênicas (Nida *et al.*, J. Agric. Food Chem. 44: 1960 a 1966, 1996). A seqüência de CP4EPSPS junta com o polinucleotídeo CTP2 (SEQ ID NO: 11) foram modificados pelo método da presente invenção. Outros peptídeos de trânsito de cloroplasto conhecidos na técnica podem ser fundidos a CP4EPSPS para direcionar a enzima aos cloroplastos.

Tabela 14. Construção de polinucleotídeo para a seqüência codificadora da EPSP sintase de *aroA*:CP4 (CP4EPSPS_AT, CP4EPSPS_ZM ou CP4EPSPS_GM)

1. Colocar a seqüência de peptídeo de trânsito de CTP2 (SEQ ID NO: 11) em frente de CP4EPSPS (SEQ ID NO: 15) como um polipeptídeo de fusão. Retro traduzir o polipeptídeo de fusão para produzir uma seqüência de polinucleotídeo artificial usando a tabela de uso de códon de *Arabidopsis*

thaliana (Tabela 2), ou a tabela de uso do códon de *Zea mays* (Tabela 3) ou a tabela de uso do códon de *Glicina max* (Tabela 4).

2. Realizar o alinhamento de sequência com as sequências de polinucleotídeo CTP2 nativo (SEQ ID NO: 12) e CP4EPSPS nativo (SEQ ID NO: 16) e a sequência de polinucleotídeo artificial para determinar grau de identidade de sequência, mapear as matrizes de leitura aberta, selecionar padrões para procurar e identificar sequências de reconhecimento de enzima de restrição.

3. Fazer correções nos códons usados na sequência de polinucleotídeo artificial para se obter a porcentagem desejada de identidade de sequência e para evitar a formação de cachos de códons idênticos. Isto é especialmente importante para os aminoácidos que ocorrem em alta frequência, isto é, alanina, glicina, histidina, leucina, serina e valina. Distribuição aproximada do uso de códon na sequência de polinucleotídeo de acordo com o uso de códon de *Arabidopsis thaliana*, Tabela 2 ou a tabela de uso do códon de *Zea mays* (Tabela 3) dependendo da tabela em uso.

4. A sequência de polinucleotídeo artificial é inspecionada quanto às regiões locais que têm uma relação GC:AT maior do que cerca de 2 em uma faixa de cerca de 50 nucleotídeos contíguos. A sequência de polinucleotídeo é ajustada como necessário, pela substituição de códons nestas regiões tal que a relação GC:AT local seja menor do que cerca de 2 e a composição de polinucleotídeo inteira estando na faixa de 0,9 a 1,3 é a Tabela 2 que é usada e de 1,2 a 1,7 se a Tabela 3 é usada.

5. Seguir as etapas de 6 a 12 da Tabela 6.

25 Tabela 15. Comparação da identidade de sequência percentual dos polinucleotídeos CP4EPSPS artificiais.

	CTP2CP _GM	CTP2CP4 _AT	CTP2CP4 _ZM	CTP2CP4 _Syn	CTP2CP4 _NAT
CTP2CP4_GM	100,00	75,66	74,12	75,15	74,37
CTP2CP4_AT		100,00	76,13	74,56*	72,93
CTP2CP4_ZM			100,00	77,76*	82,58
CTP2CP4_Syn				100,00	82,70
CTP2CP4_NAT					100,00

*A porcentagem de identidade diz respeito à CP4EPSPS e não incluem o peptídeo de trânsito.

Tabela 16. A composição de nucleotídeo e a relação GC:AT das seqüências de polinucleotídeo artificiais para a seqüência de gene de CP4EPSPS.

	A	C	G	T	GC:AT
CTP2CP4_GM	382	375	442	397	1.05
CTP2CP4_AT	369	408	469	350	1.22
CTP2CP4_ZM	312	487	577	290	1.65

5 A seqüência de polinucleotídeo CTP2_Nat (SEQ ID NO: 12) mais CP4EPSPS_Nat (SEQ ID NO: 16) designada como CTP2CP4_Nat é comparada na Tabela 15 com as seqüências de polinucleotídeo artificiais designadas como CTP2CP4_AT (CTP2_AT, SEQ ID NO: 13 fundida a CP4EPSPS_AT, SEQ ID NO: 17) e CTP2CP4_ZM (CTP2_AT, SEQ ID NO: 10 14 fundida a CP4EPSPS_ZM, SEQ ID NO: 18) produzidas pelo método da presente invenção. A seqüência de polinucleotídeo que é a mais divergente das seqüências nativas CTP2CP4_NAT e CTP2CP4EPSPS_Syn é a CTP2CP4_AT tendo cerca de 73 % e 75 % de identidade de seqüência, respectivamente. A seqüência de polinucleotídeo CTP2CP4_ZM comparada 15 com a CTP2CP4_Nat e CP4EPSPS_Syn tem cerca de 83 % e 78 % de identidade com estas duas seqüências, respectivamente.

Um critério primário para a seleção de transgenes para combinar em uma planta é a identidade percentual. A Tabela 15 pode ser usada para selecionar uma molécula de polinucleotídeo de CP4EPSPS para a construção do cassete de expressão vegetal quando é conhecido que a planta receptora conterà mais do que um polinucleotídeo CP4EPSPS. A relação GC:AT na CP4EPSPS nativa é de cerca de 1,7. A versão artificial com o códon de eleição de *Zea mays* é produzida para ter uma relação GC:AT muito similar. Na versão do códon de *Arabidopsis*, a relação GC:AT é diminuída 20 em cerca de 1,2.

A expressão de gene também é um critério para a seleção de transgenes a serem expressados. A expressão de um transgene pode variar em plantas de safra diferentes, portanto ter várias seqüências codificadoras

de polinucleotídeo artificiais disponíveis para testar em plantas de safra e genótipos, variedades ou cultivares diferentes é uma vantagem e um aspecto da invenção.

EXEMPLO 5

5 A seqüência de polinucleotídeo *bar* (SEQ ID NO: 20) que codifica uma proteína de fosfinotricina acetil transferase (SEQ ID NO: 19) foi usada para modificar geneticamente plantas quanto a resistência ao herbicida de glufosinato. Duas seqüências de polinucleotídeo *bar* novas foram planejadas usando o método da presente invenção. O alinhamento de BAR1_Nat
10 com os dois polinucleotídeos BAR1 novos artificiais é mostrado na Figura 4.

Tabela 17. Construção de gene de polinucleotídeo para BAR1_AT (SEQ ID NO: 21) e BAR1_ZM (SEQ ID NO: 22)

1. Retro traduzir a SEQ ID NO: 19 para gerar uma seqüência de polinucleotídeo usando a tabela de uso de códon de *Arabidopsis thaliana*
15 (Tabela 2) ou a tabela de uso do códon de *Zea mays* (Tabela 3)

2. Realizar o alinhamento de seqüência com a seqüência de polinucleotídeo BAR1_Nat nativo (SEQ ID NO: 20) e a seqüência de polinucleotídeo artificial para determinar o grau de identidade de seqüência, mapear as estruturas de leitura aberta, selecionar os padrões para procurar e identificar
20 as seqüências de reconhecimento de enzima de restrição.

3. Fazer correções para os códons usados na seqüência de polinucleotídeo artificial para se obter a porcentagem desejada de identidade de seqüência e para evitar a formação de cachos de códons idênticos. Isto é especialmente importante para os aminoácidos que ocorrem em alta frequência, isto é, alanina, glicina, histidina, leucina, serina e valina. Distribuição aproximada de uso de códon na seqüência de polinucleotídeo de acordo com o uso de códon de *Arabidopsis thaliana*, Tabela 2 ou a tabela de uso do códon de *Zea mays* (Tabela 3) dependendo da tabela em uso.
25

4. A seqüência de polinucleotídeo artificial é inspecionada quanto
30 às regiões locais que têm uma relação GC:AT maior do que cerca de 2 em uma faixa de cerca de 50 nucleotídeos contíguos. A seqüência de polinucleotídeo é ajustada como necessário, pela substituição de códons nestas regi-

ões tal que a relação GC:AT local seja menor do que cerca de 2 e a composição de polinucleotídeo inteira esteja na faixa de 0,9 a 1,3 se a Tabela 2 é usada e 1,2 a 1,7 se a Tabela 3 é usada.

5. Seguir as etapas de 6 a 12 da Tabela 6.

- 5 A identidade de seqüência de polinucleotídeos BAR artificiais está na faixa de 73 a 77 % (Tabela 18). O polinucleotídeo nativo é altamente rico em GC. A versão artificial (BAR1_ZM) com o códon de eleição de *Zea mays* reduziu a relação GC:AT a cerca de 1,3 e a versão artificial (BAR1_AT) com códon de eleição de *Arabidopsis* a relação é de cerca de 1,0 (Tabela 19).

Tabela 18. Identidade de seqüência percentual entre os genes *bar* ao nível de seqüência de polinucleotídeo.

	BAR1_ZM	BAR1_AT	BAR1_Nat
BAR1_ZM	100.00	77.35	76.99
BAR1_AT		100.00	73.73
BAR1_Nat			100.00

Tabela 19. Composição de nucleotídeo e a relação GC:AT das seqüências de polinucleotídeo artificiais para a seqüência de gene *bar*.

BAR_AT	139	130	144	139	1.01
BAR_ZM	122	156	154	120	1.28

15 EXEMPLO 6

Este exemplo serve para ilustrar os constructos de DNA para a expressão dos polinucleotídeos artificiais da presente invenção em plantas. Um cassete de expressão vegetal de DNA transgênico compreende elementos reguladores que controlam a transcrição de um mRNA do cassete.

- 20 Um cassete de expressão vegetal é construído para incluir um promotor que funciona em plantas que é operavelmente ligado a uma região líder 5' que é operavelmente ligada a uma seqüência de DNA de interesse operavelmente ligada a uma região de término 3'. Estes cassetes são construídos em vetores plasmídicos, que podem ser depois transferidos em plantas pelos métodos de transformação mediados por *Agrobacterium* ou outros métodos conhecidos por aqueles versados na técnica da transformação de planta. Os

seguintes constructos de vetor plasmídico são ilustradas para fornecer exemplos de plasmídeos contendo cassetes de expressão vegetal que compreendem as moléculas de polinucleotídeo artificiais da presente invenção e não são limitadas a estes exemplos.

5 As moléculas de polinucleotídeo artificiais da presente invenção, por exemplo, CP4EPSPS_AT e CP4EPSPS_ZM são sintetizadas usando iniciadores envolvidos. O produto de tamanho natural é depois amplificado com iniciadores específicos de gene contendo projeções com SphI (iniciador avançado) e EcoRI (iniciador reverso). Os genes são clonados no vetor
10 pCRII-TOPO (Invitrogen, CA). Os plasmídeos pMON54949 (CP4EPSPS_AT, Figura 6) e pMON54950 (CP4EPSPS_ZM, Figura 7) resultantes contêm o polinucleotídeo artificial e estes polinucleotídeos são seqüenciados usando métodos de seqüenciamento de DNA para confirmar que as modificações projetadas pelo método da presente invenção estão contidas nos polinucleo-
15 tídeos artificiais. Na etapa seguinte, o polinucleotídeo artificial que codifica o peptídeo de trânsito de cloroplasto CTP2 é ligado à extremidade 5' dos polinucleotídeos CP4EPSPS. O promotor CaMV 35S com um realçador duplicado (P-CaMV35S) e um intron de actina 1 de arroz (I-OsAct1) derivado de pMON30151 (Figura 8) pela digestão com SphI e HindIII ligados aos polinucleotídeos CTP2CP4EPSPS para criar os plasmídeos pMON59302
20 (CTP2CP4EPSPS_AT, Figura 9) e pMON59307 (CTP2CP4EPSPS_ZM, Figura 10).

Para a expressão dos novos polinucleotídeos artificiais em plantas monocotiledôneas, os genes são colocados nos cassetes de expressão vegetal contendo na extremidade 5' do polinucleotídeo, um promotor e
25 um intron, uma região não-traduzida 5' e na extremidade 3' do polinucleotídeo um sinal de término de transcrição. Para este propósito, pMON42411 (Figura 11) contendo P-CaMV35S:en, I-HSP70, CTP2CP4_Nat e NOS 3' são digeridos com enzimas de restrição NcoI e EcoRI. O pMON59302 (Figura 9) e pMON59307 (Figura 10) são digeridos com as mesmas enzimas de restrição.
30 Os fragmentos são purificados em gel usando o kit de purificação em gel da Qiagen e ligados para formar pMON58400 (CP4EPSPS_AT, Figura

12) e pMON58401 (CP4EPSPS_ZM, Figura 13). O vetor pMON54964 adicional (Figura 14), contendo P-OsAct1/ I-OsAct1 é fabricado pela substituição de P-e35S/I-Hsp70 a partir de pMON58400 (Figura 12) usando o fragmento HindIII /NcoI de pMON25455 (Figura 15). Para criar um vetor de expressão de monocotiledônea contendo o promotor P-FMV, pMON30152 (Figura 16) é digerido com NheI, as extremidades são feitas abruptas com T4DNA polimerase na presença de 4 dNTP-s (200 µM) e NcoI. Os fragmentos de DNA CPT2CP4_AT ou CTP2CP4_ZM são isolados de pMON59302 (Figura 9) ou pMON59307 (Figura 10), respectivamente digerindo-se com EcoRI, fazendo-se abrupto com T4 DNA polimerase e digestão com NcoI. Os fragmentos de DNA purificados em gel são ligados e novos plasmídeos pMON54992 (CTP2CP4_AT, Figura 17) e pMON54985 (CTP2CP4_ZM, Figura 18) são criados. Em cada caso a construção plasmídica bem sucedida é confirmada pela digestão com endonuclease de restrição, usando entre outras ClaI (introduzida em ambos os polinucleotídeos artificiais) e Pst I (introduzida no CP4EPSPS_ZM). O CP4EPSPS_Nat presente nos vetores precursores tem tanto ClaI quanto dois sítios de restrição PstI na região codificadora em localização diferente daquela nos polinucleotídeos artificiais.

Para a expressão de polinucleotídeos CP4EPSPS artificiais em plantas dicotiledôneas, dois vetores precursores são usados: pMON20999 (P-FMV/CTP2CP4_Syn/3'E-9, Figura 19) e pMON45313 (P-e35S/CTP2CP4_Syn/3'E9, Figura 20). Em cada plasmídeo, um fragmento de DNA contendo o polinucleotídeo CTP2CP4_Syn é substituído com CTP2CP4_AT ou CTP2CP4_ZM. Para criar pMON59308 (P-CaMVe35S/CTP2CP4_AT, Figura 21) ou pMON59309 (P-CaMVe35S/CTP2CP4_ZM, Figura 22), pMON45313 é digerido com NcoI e EcoRI e os fragmentos de restrição de DNA derivados de digestões com NcoI/EcoRI de pMON59302 (CTP2CP4_AT, Figura 9) ou pMON59307 (CTP2CP4_ZM, Figura 10) são ligados, respectivamente. Para criar pMON59313 (P-FMV/CTP2CP4_AT/3'E9, Figura 23) e pMON59396 (P-FMV/CTP2CP4_ZM/3'E9, Figura 24) o plasmídeo precursor pMON20999 é digerido com NcoI e BamHI para remover CTP2CP4_Syn e os fragmentos

de restrição NcoI/BamHI derivados de pMON59308 (CTP2CP4_AT, Figura 21) ou pMON59309 (CTP2CP4_ZM, Figura 22) são ligados, respectivamente.

EXEMPLO 7

5 Os polinucleotídeos artificiais são testados para determinar a eficácia para conferir tolerância ao glifosato às plantas transgênicas. Cinco cassetes de expressão diferentes (Tabela 20) com os novos polinucleotídeos CP4EPSPS artificiais são transformados no milho e as plantas de milho transgênico resultante comparadas com o padrão comercial (Roundup Ready® Corn 603, Monsanto Co.). O plasmídeo pMON25496 (Figura 25) contido
10 no padrão comercial tem duas cópias do polinucleotídeo CP4EPSPS_Nat, a expressão dirigida pelos promotores P-CaMVe35S (P-CaMVe35S) e P-OsAct1, respectivamente. Os plasmídeos contendo os novos polinucleotídeos de CP4EPSPS artificiais contêm apenas uma única cópia do polinucleotídeo a ser testado. A expressão destes polinucleotídeos são dirigidas pelo
15 promotor P-CaMVe35S com o intron da proteína de choque térmico I-Hsp70 ou o promotor P-FMV com um intron da sacarose sintase do arroz (I-OsSS). O plasmídeo pMON54964 contém o promotor da actina 1 do arroz com o primeiro intron nativo (Patente U.S. Nº 5.641.876, aqui incorporada por referência em sua totalidade).
20

Estes plasmídeos são transformados em células de milho por um método mediado por *Agrobacterium* e as linhagens de milho transgênico regeneradas em seleção com glifosato. As plantas de milho transgênico podem ser produzidas por um método de transformação mediado por *Agrobacterium*. Uma *Agrobacterium* cepa C58 desarmada (ABI) que abriga um
25 constructo binário da presente invenção é usada. Esta é transferida em *Agrobacterium* por um método de acasalamento triparental (Ditta *et al.*, Proc. Natl. Acad. Sci. 77: 7347 a 7351). Culturas líquidas de *Agrobacterium* são iniciadas a partir de estoques de glicerol ou a partir de uma placa recém riscada e cultivada durante a noite entre 26° C e 28° C com agitação (aproximadamente 150 rpm) até a fase de cultivo de hemi-log em meio LB líquido,
30 pH 7,0 contendo os antibióticos apropriados. As células de *Agrobacterium*

são ressuspensos no meio de inoculação (CM4C líquido) e a densidade é ajustada até OD₆₆₀ de 1. Os embriões de milho HillxLH198 e Hill Tipo II imaturos recém isolados são inoculados com um constructo contendo *Agrobacterium* e co-cultivados vários dias no escuro a 23° C. Os embriões são

5 depois transferidos para meio de demora e incubados a 28° C durante vários ou mais dias. Todas as culturas subsequentes são mantidas nesta temperatura. Os embriões são transferidos para um primeiro meio de seleção contendo carbenicilina 500/0,5 mM de glifosato). Duas semanas mais tarde, o tecido sobrevivente é transferido a um segundo meio de seleção contendo

10 carbenicilina 500/1,0 mM de glifosato). Subcultivar o calo sobrevivente a cada 2 semanas até que os eventos possam ser identificados. Isto pode levar cerca de 3 subculturas em glifosato 1,0 mM. Uma vez que os eventos são identificados, avolumar o tecido a regenerar. As mudas são transferidas ao meio MSOD em vaso de cultura e mantido durante duas semanas. De-

15 pois as plantas com raízes são transferidas para o solo. Aqueles versados na técnica de transformação de milho podem modificar este método para fornecer plantas de milho transgênico substancialmente idênticas contendo as composições de DNA da presente invenção.

Cerca de 30 linhagens de milho transgênico para cada constructo de plasmídeo são testadas e a eficiência de transformação e os níveis de expressão da enzima CP4EPSPS são mostradas na Tabela 20. As linhagens de milho transgênico são tratadas com glifosato em uma taxa de 64 oz/acre (1,8 kg/acre) como plantas jovens, as plantas sobreviventes são en-

20 saçadas pelo CP4EPSPS ELISA (Padgett *et al.* Crop Sci. 35: 1451 a 1461, 1995) para determinar os níveis de expressão de proteína de EPSPS CP4 (% exp CP4) mostrada na Tabela 20 e o nível de expressão é comparado com a planta do milho tolerante ao glifosato padrão comercialmente disponíveis (Roundup Ready® Corn 603, Monsanto Co., St. Louis, MO) como uma

25 porcentagem da quantidade de expressão de proteína determinada no padrão comercial. No geral, mais do que 50 % das linhagens de milho sobrevivem a pulverização com 1,8 kg/acre (64 oz/acre) de glifosato. As plantas sobreviventes são mostradas ter alto nível de expressão de CP4EPSPS que

30

variam de cerca de 75 a 86 % de padrão comercial 603.

Tabela 20. Eficiência de transformação (TE), expressão de CP4 (% média) derivada da transformação de constructos CP4-alt diferentes.

pMON	Promotor/Intron #	TE (%)	CP4 exp (%)*
58400 (CP4_AT)	P-CaMVe35S/IHsp70	5.4	75.5
58401 (CP4_ZM)	P-CaMVe35S/IHsp70	7.2	84.7
54964 (CP4_AT)	P-OsAct1	8.2	78.1
54985 (CP4_ZM)	P-FMV/IOsSS	11.5	85.7
54992 (CP4_AT)	P-FMV/IOsSS	11.5	78.2
nk603 (control)	P-OsAct1/P-e35S:	-	100

* A expressão de CP4EPSPS é calculada como a porcentagem de controle (603) feita em plantas que sobreviveram à pulverização com glifosato (1,8 kg/acre (64 oz/acre)).

EXEMPLO 8

Três constructos plasmídicos são avaliados em plantas de algodão transgênico (Tabela 21). O constructo de controle (pMON20999) contém P-FMV/CP4EPSPS_Syn este cassete de expressão está contido na linhagem 1445 de algodão tolerante ao glifosato comercialmente disponível (Roundup Ready® cotton, Monsanto Co., St. Louis, MO). Os constructos plasmídicos, pMON59313 e pMON59396 contendo os polinucleotídeos CP4EPSPS_AT e CP4EPSPS_ZM, respectivamente, são ensaiados quanto a eficiência de transformação e os níveis de enzima CP4EPSPS em relação ao cassete de expressão tolerante ao glifosato comercial. Cerca de cinquenta linhagens de algodão transgênico são avaliadas para cada constructo. O polinucleotídeo CP4EPSPS_AT artificial dirigido pelo promotor P-FMV dá uma porcentagem mais alta de plantas com um único inserto e um aumento no nível de expressão da enzima CP4EPSPS em relação ao cassete de expressão de pMON20999 como medido pelo ELISA.

Tabela 21. Eficiência de transformação (TE), expressão de CP4EPSPS média em linhagens de algodão R0 derivada da transformação de constructos de CP4EPSPS diferentes.

pMON	Promotor	TE (%)	CP4 Exp (%)
20999 (CP4EPSPS_Syn)	P-FMV	15,0	100,0
59313 (CP4EPSPS_AT)	P-FMV	15,0	116,4
59396 (CP4EPSPS_ZM)	P-FMV	16,1	52,0

EXEMPLO 9

5 Os constructos contendo os polinucleotídeos CP4EPSPS artificiais, CP4EPSPS_AT e CP4EPSPS_ZM são avaliados na soja (Tabela 22). Os constructos plasmídicas contêm todas o promotor P-FMV para direcionar a expressão dos novos polinucleotídeos CP4EPSPS e são comparadas ao P-FMV/CP4EPSPS_Syn contido em pMON20999. Cerca de 25 a 30 plantas

10 de soja transgênica são produzidas para cada construção. A eficiência de transformação e os níveis de enzima CP4EPSPS são medidos. Um nível de expressão surpreendentemente alto de proteína CP4EPSPS é medido em plantas de soja contendo a seqüência codificadora de CP4EPSPS_ZM (Tabela 22).

15 **Tabela 22.** Eficiência de transformação (TE), expressão CP4 média derivada da transformação de constructos de CP4EPSPS diferentes.

pMON	Promotor	TE (%)	CP4Exp (%)
20999 (CP4_Syn)	P-FMV	0,55	100,0
59313 (CP4_AT)	P-FMV	0,40	66,6
54996 (CP4_ZM)	P-FMV	0,29	242,5

EXEMPLO 10

20 As células de tabaco são transformadas com três constructos plasmídicos contendo seqüências de polinucleotídeo de CP4EPSPS diferentes e regeneradas em plantas. Cerca de vinte linhagens transgênicas são avaliadas de cada constructo. A expressão de cada um dos polinucleotídeos CP4EPSPS é direcionada pelo promotor realçador duplicado P-CaMVe35S (Tabela 23). A eficiência de transformação e o nível de expressão da enzima CP4EPSPS é medido. Os constructos de polinucleotídeo de CP4EPSPS

diferentes são mostradas realizar aproximadamente a mesma em tabaco transgênico quanto a eficiência de transformação e expressão.

Tabela 23. Eficiência de transformação (TE), expressão de CP4 média em linhagens de tabaco R0 derivadas da transformação de constructos de EPSPS de CP4 diferentes.

pMON	Promotor	TE (%)	CP4 exp. (%)
59308 CP4EPSPS_AT	P-CaMVe35S	35	100.0
59309 CP4EPSPS_ZM	P-CaMVe35S	35	91.0
54313 CP4EPSPS_Syn	P-CaMVe35S	35	100.0

EXEMPLO 11

A *Arabidopsis thaliana* é transformada com quatro constructos plasmídicos pela infiltração a vácuo (Bechtold N, *et al.*, CR Acad Sci Paris Sciences di la vie/life sciences 316: 1194 a 1199, (1993) e a progênie V1 avaliada para comparar a eficácia das seqüências de polinucleotídeo CP4EPSPS diferentes e promotores diferentes para o uso na seleção de plantas em glifosato (Tabela 24). Cerca de 30 plantas V1 transgênicas (+) são produzidas para cada constructo. Os constructos direcionadas pelo P-CaMVe35S com o realçador duplicado (pMON45313, pMON59308 e pMON59309) não mostram nenhuma diferença substancial no nível de expressão em folhas como determinado pelo ELISA. As plantas são transformadas com pMON26140 que contém CP4EPSPS_Syn dirigido pelo promotor P-FMV, estas plantas mostram o nível de expressão mais alto, os níveis de expressão detectados a partir das plantas dos constructos de teste são comparados a pMON26140.

Tabela 24. Avaliação de cassetes de expressão CP4 diferentes em *Arabidopsis*

pMON	Promotor/	Plantas produzidas	CP4 exp.(%)
45313(CPEPSPS4_Syn)	P-CaMVe35S	+	82,1
59308(CP4EPSPS_AT)	P-CaMVe35S	+	79,3
59309(CP4EPSPS_ZM)	P-CaMVe35S	+	77,3
26140(CP4EPSPS_Syn)	P-FMV	+	100,0

EXEMPLO 12

Plantas de trigo transformadas com os novos polinucleotídeos CP4EPSPS são comparadas quanto a eficiência de transformação e a expressão da enzima CP4EPSPS determinada pelo ELISA (Tabela 25). O CP4EPSPS_ZM fornece a expressão da enzima CPEPSPS pelo menos sete vezes mais alta do que o CP4EPSPS_AT. A expressão média de CP4EPSPS nas folhas de plantas de trigo contendo o polinucleotídeo CP4EPSPS_ZM é de cerca de 64 % daquele encontrado no trigo resistente ao glifosato que contém um constructo de cassete duplo, pMON30139: P-e35S/I-Hsp70/CP4EPSPS_Nat e P-OsAct1/I-OsAct1/CP4EPSPS_Nat (WO/0022704).

Tabela 25. Desempenho de polinucleotídeos CP4EPSPS diferentes no trigo

pMON	Promotor/Intron	TE (%)	CP4 Exp. (%)
58400 CP4EPSPS_AT	P-e35S/I-Hsp70	0.25	9,2
58401 CP4EPSPS_ZM	P-e35S/I-Hsp70	0.35	64,0
30139 CP4EPSPS_Nat	P-e35S:P-OsAct1	-	100,0

EXEMPLO 13

Este exemplo serve para ilustrar a detecção de polinucleotídeos artificiais diferentes em plantas transgênicas, especificamente CP4EPSPS_AT e CP4EPSPS_ZM. Os outros polinucleotídeos artificiais, OsEPSPS_AT, OsEPSPS_ZM, GmEPSPS_GM, ZmEPSPS_ZM, CTP2_AT, CTP2_ZM, Bar1_AT e Bar1_ZM podem ser todos especificamente detectados em plantas transgênicas pelos métodos que fornecem um amplicon de DNA ou pela hibridização de uma sonda de DNA a uma amostra de planta. Aqueles versados na técnica da detecção de DNA podem facilmente planejar moléculas preparadoras a partir das seqüências de polinucleotídeo artificiais fornecidas na presente invenção para permitir um método que especificamente detectará o polinucleotídeo artificial em uma amostra de planta. O uso de um método ou de um kit que forneça iniciadores de DNA ou sondas homólogas ou complementares aos polinucleotídeos artificiais aqui descritos é um aspecto da presente invenção.

Um método de detecção de DNA (reação de cadeia da polime-

rase, PCR) é planejado para detectar os polinucleotídeos CP4EPSPS artificiais em plantas transgênicas. Os conjuntos únicos de iniciadores de DNA mostrados na Tabela 26 são planejados para amplificar um polinucleotídeo de CP4EPSPS específico e para fornecer amplicons distintamente dimensionados. Os amplicons diferem suficientemente no comprimento de polinucleotídeo entre os vários polinucleotídeos CP4EPSPS para tornar fácil a separação dos amplicons pela eletroforese em gel de agarose padrão. A presença de mais do que um dos polinucleotídeos artificiais pode ser detectada em uma planta usando-se um método de PCR multiplex.

10 **Tabela 26.** Seqüência de iniciadores usados para a detecção de genes CP4 diferentes em plantas transgênicas.

Par de iniciador	Especificidade de gene	produto de PCR (bps)
SEQ ID NOs: 24 e 25	CP4EPSPS_AT	938 (940)
SEQ ID NOs: 26 e 27	CP4EPSPS_ZM	595 (600)
SEQ ID NOs: 28 e 29	CP4EPSPS_Nat	712 (710)
SEQ ID NOs: 30 e 31	CP4EPSPS_Syn	443 (440)

Os pares iniciadores de DNA (Tabela 26) são usados para produzir um diagnóstico de amplicon para um polinucleotídeo CP4EPSPS específico em uma planta transgênica. Estes pares iniciadores incluem, mas não são limitados à SEQ ID NO: 24 e SEQ ID NO: 25 para o polinucleotídeo CP4EPSPS_AT; SEQ ID NO: 26 e SEQ ID NO: 27 para o CP4EPSPS_ZM; SEQ ID NO: 28 e SEQ ID NO: 29 para o CP4EPSPS_Nat e SEQ ID NO: 30 e SEQ ID NO: 31 para a molécula de polinucleotídeo CP4EPSPS_Syn. Além destes pares preparadores, qualquer par preparador derivado da SEQ ID NO: 17 ou SEQ ID NO: 18 que quando usados em uma reação de amplificação de DNA produz um diagnóstico de amplicon de DNA para o respectivo polinucleotídeo CP4EPSPS é um aspecto da presente invenção.

As condições de amplificação para esta análise é ilustrada na Tabela 27 e Tabela 28, entretanto, qualquer modificação destas condições incluindo o uso de fragmentos das moléculas de DNA da presente invenção ou seus complementos como moléculas iniciadoras, produzindo um diagnóstico de molécula de DNA de amplicon para os polinucleotídeos artificiais

aqui descritos está dentro da versatibilidade da técnica. As moléculas de DNA da presente invenção incluem pelo menos a SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 7, SEQ ID NO: 10, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 21, SEQ ID NO: 22 e SEQ ID NO: 35.

5 As moléculas de DNA que funcionam como moléculas iniciadores em um método de amplificação de DNA para detectar a presença dos polinucleotídeos artificiais incluem, mas não são limitadas à SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30 e SEQ ID NO: 31.

10 Em um método para determinar a presença de polinucleotídeos da presente invenção, a análise de amostra de extrato de DNA de tecido vegetal deve incluir um controle positivo conhecido conter o polinucleotídeo artificial e um controle de extrato de DNA negativo de uma planta que não é transgênica ou não contém o polinucleotídeo artificial e um controle negativo
15 que não contenha nenhum padrão no extrato de DNA.

As moléculas iniciadoras de DNA adicionais de comprimento suficiente pode ser selecionada da SEQ ID NO: 17 e SEQ ID NO: 18 e condições otimizadas para a produção de um amplicon que pode diferir dos métodos mostrados na Tabela 27 e Tabela 28, mas resultam em um diagnóstico de amplicon para os polinucleotídeos artificiais. O uso destas se-
20 quências preparadoras de DNA homólogas ou complementares à SEQ ID NO: 17 e SEQ ID NO: 18 usadas com ou sem as modificações para os métodos das Tabelas 27 e 28 estão dentro do escopo da invenção. O ensaio para os amplicons CP4EPSPS_AT e CP4EPSPS_ZM pode ser realizado
25 pelo uso de um termociclador Stratagene Robocycler, MJ Engine, Perkin-Elmer 9700 ou Eppendorf Mastercycler Gradient como mostrado na Tabela 28 ou pelos métodos e aparelho conhecido por aqueles versados na técnica.

Tabela 27. Procedimento de amplificação de DNA e mistura de reação para a confirmação de polinucleotídeo de EPSPS artificial CP4EPSPS_AT em plantas de milho.

Etapa	Reagente	Quantidade	Comentários
1	Água isenta de Nuclease	adicionar até volume final de 20 μ l	-
2	tampão de reação 10X (com $MgCl_2$)	2,0 μ l	Concentração final de 1X de tampão, concentração final de 1,5 mM de $MgCl_2$
3	10 mM de solução de dATP, dCTP, dGTP e dTTP	0,4 μ l	Concentração final de 200 μ M de cada dNTP
4	Iniciador (SEQ ID NO: 24) (ressuspenso em tampão 1X TE ou água isenta de nuclease a uma concentração de 10 μ M)	0,4 μ l	Concentração final de 0,2 μ M
5	Iniciador (SEQ ID NO: 25) (ressuspenso em tampão 1X TE ou água isenta de nuclease a uma concentração de 10 μ M)	0,4 μ l	Concentração final de 0,2 μ M
6	Iniciador de controle (SEQ ID NO: 32) (ressuspenso em tampão 1X TE ou água isenta de nuclease a uma concentração de 10 μ M)	0,2 μ l	Concentração final de 0,1 μ M

Etapa	Reagente	Quantidade	Comentários
7	Iniciador de controle (SEQ ID NO: 33) (ressuspenso em tampão 1X TE ou água isenta de nuclease a uma concentração de 10 μ M)	0,2 μ l	Concentração final de 0,1 μ M
8	RNase, DNase livre (500 ng/ μ l)	0,1 μ l	50 ng/reação
9	REDTaq DNA polimerase (1 unidade/ μ l)	1,0 μ l (recomendado sacudir as pipetas antes da etapa seguinte)	1 unidade/reação
10	DNA extraído (padrão): • Amostras a serem analisadas * folhas individuais * folhas agrupadas (máximo de 50 folhas/ agrupamento) • Controle negativo • Controle negativo • Controle positivo	• 10 a 200 ng de DNA genômico • 200 ng de DNA genômico • 50 ng de DNA genômico de planta não transgênica • nenhum DNA padrão • 5 ng DNA plasmídico	-

Tabela 28. Parâmetros de PCR Sugeridos para termocicladores diferentes

Mistura suave e, se necessário (nenhum topo quente no termociclador), adicionar 1 a 2 gotas de óleo mineral no topo de cada reação. Processar com a PCR em um termociclador Stratagene Robocycler, MJ Engine, Perkin-Elmer 9700 ou Eppendorf Mastercycler Gradient usando os seguintes parâmetros de ciclicização.

Nota: Os termocicladores MJ Engine ou Eppendorf Mastercycler Gradient devem ser conduzidos no modo calculado. Conduzir o termociclador Perkin-Elmer 9700 com a velocidade de elevação ajustada no máximo.

Nº de Ciclos	Ajustes: Stratagene Robocycler	
1	94° C	3 minutos
38	94° C	1 minuto
	60° C	1 minuto
	72° C	1 minuto e 30 segundos
1	72° C	10 minutos

Nº de Ciclos	Ajustes: MJ Engine ou Perkin-Elmer 9700	
1	94° C	3 minutos
38	94° C	10 segundos
	60° C	30 segundos
	72° C	1 minuto
1	72° C	10 minutos

Nº de Ciclos	Ajustes: Eppendorf Mastercycler Gradient	
1	94° C	3 minutos
38	94° C	15 segundos
	60° C	15 segundos
	72° C	1 minuto
1	72° C	10 minutos

Todas as composições e métodos descritos e reivindicados aqui podem ser feitas e executadas considerando-se a presente descrição. Embora as composições e métodos desta invenção tenham sido descritas, estará evidente por aqueles de habilidade na técnica que as variações podem ser aplicadas às composições e métodos e nas etapas ou na seqüência de etapas dos métodos aqui descritos sem divergir do conceito, espírito e escopo da invenção. Mais especificamente, estará evidente que certos agentes que são tanto química e fisiologicamente relacionados podem ser substituídos no lugar dos agentes aqui descritos enquanto os mesmos resultados ou similares podem ser obtidos. Todos de tais substitutos e modificações similares evidentes àqueles versados na técnica são considerados estar dentro do espírito, escopo e conceito da invenção como definido pelas reivindicações anexas.

Todas as publicações e pedidos de patente aqui citados são incorporados por referência em sua totalidade no mesmo grau como se cada

publicação ou pedido de patente individual fosse específico e individualmente indicado como sendo incorporado por referência.

REIVINDICAÇÕES

1. Método para reduzir o silenciamento de transgene em plantas transgênicas, caracterizado pelo fato de que compreende as etapas de:

5 a) obter um construto de DNA compreendendo um polinucleotídeo artificial que codifica um peptídeo de trânsito de cloroplasto operavelmente ligado à SEQ ID NO: 18;

b) transformar o referido construto de DNA dentro de uma célula vegetal que compreende SEQ ID NO: 16; e

10 c) regenerar a referida célula vegetal em uma planta transgênica fértil.

2. Método de acordo com a reivindicação 1, caracterizado pelo fato de que o referido polinucleotídeo artificial é expresso na referida planta transgênica fértil.

15 3. Método de acordo com a reivindicação 2, caracterizado pelo fato de que o referido polinucleotídeo artificial fornece um fenótipo agronomicamente útil selecionado do grupo que consiste em: tolerância a herbicida e resistência às doenças.

4. Molécula de polinucleotídeo artificial, caracterizada pelo fato de que compreende SEQ ID NO: 18.

20 5. Construto de DNA, caracterizado pelo fato de que compreende: uma molécula promotora que funciona em plantas, operavelmente ligada à referida molécula de polinucleotídeo artificial como definida na reivindicação 4.

25 6. Método de detectar um polinucleotídeo artificial em célula vegetal, planta ou sua progênie que compreende uma molécula de polinucleotídeo artificial, como definida na reivindicação 4, caracterizado pelo fato de que compreende as etapas:

30 (a) contatar uma amostra de DNA isolado da referida célula vegetal, planta ou sua progênie com uma molécula de DNA, em que a referida molécula de DNA compreende pelo menos uma molécula de DNA de um par de moléculas de DNA que quando usado em uma reação de amplificação de ácido nucléico produz um amplicon que é diagnóstico para a referida molé-

cula de polinucleotídeo artificial, como definida na reivindicação 4, em que a molécula de DNA compreende um fragmento de SEQ ID NO: 18;

(b) realizar uma reação de amplificação de ácido nucléico, produzindo deste modo o amplicon; e

5 (c) detectar o amplicon.

7. Kit de detecção de DNA, caracterizado pelo fato de que compreende pelo menos uma molécula de DNA de comprimento suficiente para ser especificamente homóloga ou complementar a um polinucleotídeo artificial, como definido na reivindicação 4, em que a molécula de DNA compre-

10 ende um fragmento de SEQ ID NO:18 e a molécula de DNA não é completamente complementar a SEQ ID NO:16, em que a referida molécula de DNA é útil como uma sonda de DNA ou iniciador de DNA.

Figura 1A
Figura 1B
Figura 1C
Figura 1D
Figura 1E
Figura 1F
Figura 1G
Figura 1H

Fig 1

Figura 3A
Figura 3B
Figura 3C
Figura 3D
Figura 3E
Figura 3F
Figura 3G

Fig 3

Figura 5A
Figura 5B
Figura 5C
Figura 5D
Figura 5E
Figura 5F
Figura 5G

Fig 5

Figura 2A
Figura 2B
Figura 2C
Figura 2D
Figura 2E
Figura 2F

Fig 2

Figura 4A
Figura 4B
Figura 4C

Fig 4

1	OsEPSPS_Nat	ATGGCGGCGA	CCATGGCGTC	CAACGCCGCG	GCTGCGGCGG	CGGTGTCCCT	50
	OsEPSPS_AT	T.A.	TAG	A...	C.T.	..C..T...T.	
	OsEPSPS_ZM	C..T.	TAG	G..A	..C..T....	..T..T..G...	
51	OsEPSPS_Nat	GGACCAGGCC	GTGGCGGCGT	CGGCGGCGGT	CTCGTCGCGG	AAGCAGCTGC	100
	OsEPSPS_AT	A....A..A	..A..A...A	GC..T..A..	...A..A..T	A..A..A.	
	OsEPSPS_ZM	A..T.....G	..T..T..CA	GC..T..C..	...CAGTA..	..A...T.A..	
101	OsEPSPS_Nat	GGCTGCCCGC	CGCGGCGCGC	GGGGGGATGC	GGGTGCGGGT	GCGGGCGCGG	150
	OsEPSPS_AT	A..A..	A..C..TA.A	..C..C...A	..A..TA.A..	..A...TA.A	
	OsEPSPS_ZM	..T..C..G..	G..T..C..A	..T..A...A	..A....C...	..T..A..CA.A	
151	OsEPSPS_Nat	GGGCGGCGGG	AGGCGGTGGT	GGTGCGGTCC	GCGTCGTCGT	CGTCGGTGCG	200
	OsEPSPS_AT	..TA.....	T..A..	C..A..C...	..T..TAGCA	GT.....	
	OsEPSPS_ZM	..CA.A..T.	C.....	A..C..C..G	..CAGC....	..CAGT..T..	

Fig 1A

201	OsEPSPS_Nat	AGCGCCGGCG	GCGAAGGCGG	AGGAGATCGT	GCTCCAGCCC	ATCAGGGAGA	250
	OsEPSPS_AT	T.....T	..T....A.	T..	TT.A..A..T	..T....A.	
	OsEPSPS_ZM	T.....A	..T....C.	.A..A..T..	T..G..A..G	..AC....A.	
251	OsEPSPS_Nat	TCTCCGGGGC	GGTTCAGCTG	CCAGGGTCCA	AGTCGCTCTC	CAACAGGATC	300
	OsEPSPS_AT	.A..G.....	C..A..AT.A	..T..AAG..	..AGC..T..	T	
	OsEPSPS_ZM	.T..G..C..	...G.....C	..T..C..G.	C..G..	G..TC.T...	
301	OsEPSPS_Nat	CTCCTCCTCT	CCGCCCTCTC	CGAGGGCACA	ACAGTGGTGG	ACAACTTGCT	350
	OsEPSPS_AT	..GT.G..T.	.A..T.....	G.....A...	T....	..T..TC..T.	
	OsEPSPS_ZM	..T..G..A.	:G..G..G..	...A..G..C	..G..T..A.	..T..TC..T.	
351	OsEPSPS_Nat	GAACAGTGAG	GATGTTCACT	ACATGCTTGA	GGCCCTGAAA	GCCCTCGGGC	400
	OsEPSPS_AT	...T.....	G....	..T....A..	...T..C..G	..T..A....	
	OsEPSPS_ZM	...TCC..A	..C..G....	C..	A..G..A...	..GT.G..CT	

Fig 1B

401	OsEPSPS_Nat	TCTCTGTGGA	AGCAGATAAA	GTTGCAAAAA	GAGCTGTAGT	CGTTGGCTGT	450
	OsEPSPS_AT	.T....A..	...G.....	..A.....C	.C..A..G..	T..A..T...	
	OsEPSPS_ZM	.GAGC..T..	G..T..C...	..G..C..GC	...A..C..	A..G..A..C	
451	OsEPSPS_Nat	GGTGGCAAGT	TTCCTGTTGA	GAAGGATGCG	AAAGAGGAAG	TGCAACTCTT	500
	OsEPSPS_AT	...G....	.C..A.....	A.....T	..G..A....	.A..G.....	
	OsEPSPS_ZM	..C..T....	.C..A..G..	...A..C..C	A..G.	.T...T.G..	
501	OsEPSPS_Nat	CTTGGGGAAC	GCTGGAATCG	CAATGCGATC	CTTGACAGCA	GCCGTGACTG	550
	OsEPSPS_AT	TC.C.....	T..C..G....	.C.....GAG	T.....T..T	..G..C..A.	
	OsEPSPS_ZM	A...	C....	A..AG	.C....T..C	.G..C.....	
551	OsEPSPS_Nat	CTGCTGGTGG	AAATGCAACT	TATGTGCTTG	ATGGAGTGCC	ACGAATGAGG	600
	OsEPSPS_AT	.C.....A..	C..C.....A	.C..C..A.	...G.....	GA.....C.T	
	OsEPSPS_ZM	.C..A..G..	T..C.....	..C..C..G.	.C..T..T..	...G...C.C	

Fig 1C

601						650
OsEPSPS_Nat	GAGAGACCGA	TTGGTGACTT	GGTTGTCGGG	TTGAAACAAC	TTGGTGCGGA	
OsEPSPS_AT	...C.T..T.	TC.	T..C..A..T	C.C..G....	.C..C..T..	
OsEPSPS_ZM	..AC....C.	.C.....TC.	C..G..A..A	C.T.....T	.G..A..C..	
651						700
OsEPSPS_Nat	TGTCGACTGT	TTCCTTGGCA	CTGAATGCCC	ACCTGTTTCGT	GTC AAGGGAA	
OsEPSPS_AT	C..A..T...	G..T.	G..T..	G..A..CA.A	..T..A....	
OsEPSPS_ZM	..A..T...	...T.G..G.	A..G.....	T..A..G..G	..A.....C.	
701						750
OsEPSPS_Nat	TTGGAGGACT	TCCTGGTGGC	AAGTTAAGC	TCTCTGGTTC	CATCAGCAGT	
OsEPSPS_AT	.C..T..G..	G..G..C..A	C....	.G..G..CAG	T..TTCG...	
OsEPSPS_ZM	.C..G..T..	C..A..C..A	G...T	.A..A..GAG	T...TC.TCG	
751						800
OsEPSPS_Nat	CAGTACTTGA	GTGCCCTTGCT	GATGGCTGCT	CCTTTGGCCC	TTGGGGATGT	
OsEPSPS_AT	TC.TT	C...TC.C..	G	..A..A..TT	.G..A.....	
OsEPSPS_ZM	C.CT	CC..GC..T.	A.:A	..AC....GT	.G..C..C..	

Fig 1D

```

801                               850
OsEPSPS_Nat GGAGATCGAA ATCATTGACA AACTAATCTC CATTCCTTAC GTTGAAATGA
OsEPSPS_AT T.....G .....T. ....T.A.. T..C..G..T ..C..G.....
OsEPSPS_ZM T..A..T... ..T..C..... .G..C..A.. ..C..G..T ..G..G.....

851                               900
OsEPSPS_Nat CATTGAGATT GATGGAGCGT TTTGGTGTGA AGGCAGAGCA TTCTGATAGT
OsEPSPS_AT .T..A...C. T.....A..G .....G..T. ....C..... .AGC..C...
OsEPSPS_ZM .C..C.GC. ....AA.A .....A.... .A..T..A.. .AG...CTC.

901                               950
OsEPSPS_Nat TGGGACAGAT TCTATATTAA GGGAGGGCAG AAGTACAAAT CTCCTGGAAA
OsEPSPS_AT ....TC.T. ....C..A.. .....C... .....T..G. ....G...
OsEPSPS_ZM ....TC.G. ....C..C.. ...C..A... ..A..T..GA G...A..C...

951                               1000
OsEPSPS_Nat TGCCTATGTT GAAGGTGATG CCTCAAGCGC GAGCTATTTC TTGGCTGGTG
OsEPSPS_AT ...T.....A .....G.... .T...TCT.. .TCT..C... C.T..G..A.
OsEPSPS_ZM ...G..C..G ..G..C..C. ....TTC... C.....C... ..A..C.

```

Fig 1E

1001
 OsEPSPS_Nat CTGCAATCAC TGGAGGCACT GTGACAGTTC AAGTTGTGG TACGACCAGT 1050
 OsEPSPS_AT .G..T..A.. ..A..A..C.....G..C..C.. ..A..A..
 OsEPSPS_ZM .A..T..T..C.....T... ..T.....G. G..... A..A..G..C

1051
 OsEPSPS_Nat TTGCAGGGTG ATGTCAAATT TGCTGAGGTA CTTGAGATGA TGGGAGCAAA 1100
 OsEPSPS_AT A.... .C..G..G.. ..C..... ..A.... T..C..
 OsEPSPS_ZM C....A..C. .C..A..G.. C..A..A..C A.... C..T..

1101
 OsEPSPS_Nat GGTTACATGG ACTGACACCA GTGTAACCGT AACTGGTCCA CCACGTGAGC 1150
 OsEPSPS_AT A..A..G... ..A....AT CG..G..A.. T.....T A..A..
 OsEPSPS_ZM ...G..T... ..A..T..GT CA..C..... C.....G..T ..T..G..A.

1151
 OsEPSPS_Nat CTTATGGGAA GAAACACCTG AAAGCTGTTG ATGTCAACAT GAACAAAATG 1200
 OsEPSPS_AT C..C.. A..G..T..T ..G..C..G. T..T.. T..G..
 OsEPSPS_ZM .C..C..T.. T... ..G..A..G. .C..G..... T..G...

Fig 1F

	1201	1250
OsEPSPS_Nat	CCTGATGTTG	CCATGACCCT TGCCGTTGTT GCACTCTTCG CTGATGGTCC
OsEPSPS_AT	C....	T.....A.....C..T..T..A..C..C..
OsEPSPS_ZM	..A..C.....	..G.....C..A..A..C..T..G.....A..C..A..

	1251	1300
OsEPSPS_Nat	AACTGCTATC	AGTAAAGGAA
OsEPSPS_AT	...G..G..A	C..C.....
OsEPSPS_ZM	T..C.....	G..C.....

	1301	1350
OsEPSPS_Nat	TGGTTGCAAT	TCGGACCGAG
OsEPSPS_AT	...G..T..	...A..T..A
OsEPSPS_ZM	...C..T..	C..C.....A
		CTAACAAAGC
		TGGAGCATC
		GGTTGAAGAA

	1351		1400
OsEPSPS_Nat	GGTCCTGACT	ACTGCATCAT	CACCCACCG
OsEPSPS_AT	.C.....T.	T....	T.A..C..T
OsEPSPS_ZM	A....	A..G..T..C	A.....C
			ACATCACGGC

Fig 1G

	1401	1450
OsEPSPS_Nat	AATCGACACC	TACGATGATC ACAGGATGGC CATGGCCTTC TCCCTCGCTG
OsEPSPS_AT	T..T..T..A	C....T..A.....T....T...A..G..C..
OsEPSPS_ZM	T..A.....T	C.....A.....G.....G.....A..

	1451	1500
OsEPSPS_Nat	CCTGCGCCGA	CGTGCCCGTG ACGATCAGGG ACCCTGGTTG CACCCGCAAG
OsEPSPS_AT	.T..T..A..	T..T..T..C..A.....A..T....C..T..TA.A...
OsEPSPS_ZM	.G.....G..	...T..G..T.....T..A..T.....G....T..T..A

	1501	1548
OseSPS_Nat	ACCTTCCCCGA ACTACTTCGA CGTTCTAAGC ACTTTCGTCA GGA	ACTGTA
OseSPS_AT	. . G G T TTCA . . A C . C . . T	T
OseSPS_ZM	. . T T T . . G TCA . . C . . T . . AC . C	C

Fig 1H

1	ZmEPSPS_Nat	ATGGCGGGCCA	TGGCGACCAA	GGCCGCCGCG	GGCACCGTGT	CGCTGGACCT	50
	ZmEPSPS_ZM	T.	...C..G..	...A..G..C	..T..A..AA	GC..C..TT.	
51	ZmEPSPS_Nat	CGCCGCGCCG	TCGCGCCGCC	ACCACCGCCC	GAGCTCGGCG	CGCCCGCCCT	100
	ZmEPSPS_ZM	G..G..C..C	..C..TA.G.	G..	A...AGT...	A.G..A..G.	
101	ZmEPSPS_Nat	TCCGCCCCGC	CGTCCGCGGG	CTGCGGGCGC	CTGGGCGCCG	CGTGATCGCC	150
	ZmEPSPS_ZM	..A.G..A..	A..T.....T	..TA.A....	TA.AA.	G..T.....A	
151	ZmEPSPS_Nat	GCGCCGCCGG	CGCGGCAGC	GGCGGCGGCG	GTGCAGGCGG	GTGCCGAGGA	200
	ZmEPSPS_ZM	A....	T..C..	T.....A...	C.	.C..G..A..	
201	ZmEPSPS_Nat	GATCGTGCTG	CAGCCCATCA	AGGAGATCTC	CGGCACCGTC	AAGCTGCCGG	250
	ZmEPSPS_ZM	C..A	T.....	A.....	T..T.:G..A	...T.A..A..	

251	ZmEPSPS_Nat	GGTCCAAGTC	GCTTTCCAAC	CGGATCCTCC	TACTCGCCGC	CCTGTCCGAG	300
	ZmEPSPS_ZM	.CAG...AAG	T...AG....	.A.....G.	.GT.G..G..	A..C..T..A	
301	ZmEPSPS_Nat	GGGACAACAG	TGGTTGATAA	CCTGCTGAAC	AGTGAGGATG	TCCACTACAT	350
	ZmEPSPS_ZM	C..G.	.C..A....;	T.....C...	..C..A..C.	.G.....T..	
351	ZmEPSPS_Nat	GCTCGGGGCC	TTGAGGACTC	TTGGTCTCTC	TGTCGAAGCG	GACAAAGCTG	400
	ZmEPSPS_ZM	.T.G..T...	C.....G.	.A.....G..	A..G.....C	..T..G...C.	
401	ZmEPSPS_Nat	CCAAAAGAGC	TGTAGTTGTT	GGCTGTGGTG	GAAAGTTCCC	AGTTGAGGAT	450
	ZmEPSPS_ZM	...GC.C..	...C.C...	C..C.	.T.....	C..G.....C	
451	ZmEPSPS_Nat	GCTAAAGAGG	AAGTGCAGCT	CTTCTTGGGG	AATGCTGGAA	TCGCAATGCG	500
	ZmEPSPS_ZM	..G.....A.	.G.:.....T.	A..TC.T...	..C.....C.	C.....	

Fig 2B

751		800
ZmEPSPS_Nat	GCTCCTTTGG CTC TTGGGA TGTGGAGATT GAAATCATTG ATAAATTAAT	
ZmEPSPS_ZM	..C...C... ..G..A.. C..C..A... ..G..... ..GC.G..	
801		850
ZmEPSPS_Nat	CTCCATTCCG TACGTCGAAA TGACATTGAG ATTGATGGAG CGTTTTGGTG	
ZmEPSPS_ZM	...T..C..T ..T..T..G. C.CC. TC.....A A.A..C..G.	
851		900
ZmEPSPS_Nat	TGAAAGCAGA GCATTCTGAT AGCTGGGACA GATTCTACAT TAAGGGAGGT	
ZmEPSPS_ZM	.C.....T.. ...C..C... TC..... ..G.....T.. C.....C..A	
901		950
ZmEPSPS_Nat	CAAAAATACA AGTCCCCTAA AAATGCCTAT GTTGAAGGTG ATGCCTCAAG	
ZmEPSPS_ZM	..G.....T. A..G.. G.....G..C ..C..G..A. .C..A..G..	
951		1000
ZmEPSPS_Nat	CGCAAGCTAT TTCTTGGCTG GTGCTGCAAT TACTGGAGGG ACTGTGACTG	
ZmEPSPS_ZM	...G..T..C ...C.T..G. .C.....C.. C..C..G..A ..C.....A.	

Fig 2D

1251 1300
ZmEPSPS_Nat GAGAGTAAAG GAGACCGAGA GGATGGTTGC GATCCGGACG GAGCTAACCA
ZmEPSPS_ZM ...G..C... ..A...C .T.....G... ..A... ..G..T.

1301 1350
ZmEPSPS_Nat AGCTGGGAGC ATCTGTTGAG GAAGGGCCGG ACTACTGCAT CATCAGGCCG
ZmEPSPS_ZM .A..C..G.. CAG...G... ..G..C.... .T..... A.....A..T

1351 1400
ZmEPSPS_Nat CCGGAGAAGC TGAACGTGAC GGCATCGAC ACGTACGACG ACCACAGGAT
ZmEPSPS_ZM ..A.....T C.. C..T..... ..A..... .T...C....

1401 1450
ZmEPSPS_Nat GGCCATGGCC TTCTCCCTTG CCGCCTGTGC CGAGGTCCCC GTCACCATCC
ZmEPSPS_ZM ...A..... ..TAG.T.G. .A..G..C... ..A..A..T ..G..T..AA

1451 1500
ZmEPSPS_Nat GGGACCCCTGG GTGCACCCGG AAGACCTTCC CCGACTACTT CGATGTGCTG
ZmEPSPS_ZM .A...T..A.. T.....C ..A..G..T. T... ..C..C..C

1501 1521
ZmEPSPS_Nat AGCACTTTTCG TCAAGAATTA A
ZmEPSPS_ZM TCA..C.... .G.....C.G .

Fig 2F

1	GmEPS_Nat	ATGGCCCAAG	TGAGCAGAGT	GCACAATCTT	GCTCAAAGCA	CTCAAATTTT	50
	GmEPS_GM	T..G.	.CTCTC.C..	T..T.....C	G..T.	.C..G..A...	
51	GmEPS_Nat	TGGCCATTCT	TCCAACCTCCA	ACAAACTCAA	ATCGGTGAAT	TCGGTTTCAT	100
	GmEPS_GM	C..A.....C	AGT.....A.	A..	G..T.....	AGT..A....C	
101	GmEPS_Nat	TGAGGCCACG	CCTTTGGGGG	GCCTCAAAAT	CTCGCATCCC	GATGCATAAA	150
	GmEPS_GM	.TC....T..	G..G.....A	..AAGT..GA	GC..T.....	T.....C..G	
151	GmEPS_Nat	AATGGAAGCT	TTATGGGAAA	TTTTAATGTG	GGGAAGGGAA	ATTCCGGCGT	200
	GmEPS_GM	..C..TTCG.	.C.....G..	C.....C..C	..C..A....	.C..A..T..	
201	GmEPS_Nat	GTTTAAGGTT	TCTGCATCGG	TCGCCGCCGC	AGAGAAGCCG	TCAACGTCGC	250
	GmEPS_GM	C..C..A..A	AGC...CAGC.	.A..T..G..	T.....C	AGT..T..T.	

Fig 3A

251	GmEPSPS_Nat	CGGAGATCGT	GTTGGAACCC	ATCAAGACT	TCTCGGGTAC	CATCACATTG	300
	GmEPSPS_GM	.T..A..T..	TC.T....G	..A..G..T.	A.....	G..T...C.A	
301	GmEPSPS_Nat	CCAGGGTCCA	AGTCTCTGTC	CAATCGAATT	TTGCTTCTTG	CTGCTCTCTC	350
	GmEPSPS_GM	..T..A..A.	..AG...C..	T...A.....	...T.G..C.	A.....G..	
351	GmEPSPS_Nat	TGAGGGAACA	ACTGTTGTAG	ACAACTTGTT	GTATAGTGAG	GATATTCAAT	400
	GmEPSPS_GM	C..A.....C	A..C.	.T...C.CC.	T.....C..A	A.....	
401	GmEPSPS_Nat	ACATGCTTGG	TGCATTAAAGG	ACCCTTGGAC	TGCGTGTGGA	AGATGACAAA	450
	GmEPSPS_GM	.T...T.G..	G..GC.C..A	..T.....G.	.AA.A..T..	G..C..T..G	
451	GmEPSPS_Nat	ACAACCAAAC	AAGCAATTGT	TGAAGGCTGT	GGGGGATTGT	TTCCCACTAG	500
	GmEPSPS_GM	..T..T....	T..C..	C.....T...	..C..G....	.C..T...TC	

Fig 3B

501		550
GmEPSPS_Nat	TAAGGAATCT AAAGATGAAA TCAATTTATT CCTTGGAAT GCTGGTATCG	
GmEPSPS_GM	. . . A . . . AG G . . . A . . C . . G . . T C . . C . . A . . A	
551		600
GmEPSPS_Nat	CAATGAAGTC CTTGACAGCA GCTGTGGTTG CTGCAGGTGG AAATGCAAGC	
GmEPSPS_GM	 AG . C . C . . C . . T C G . . G T . . C . . T . . T . .	
601		650
GmEPSPS_Nat	TACGTACTTG ATGGGGTGCC CCGAATGAGA GAGAGGCCAA TTGGGGATTT	
GmEPSPS_GM	 CT . A . . C . . C TA C A . . A . . T . . C . . T C . .	
651		700
GmEPSPS_Nat	GGTTGCTGGT CTTAAGCAAC TTGGTGCAGA TGTGATTGC TTTCTTGGCA	
GmEPSPS_GM	A . . G C . . A . . A . . G A C . . C T . . CT . G	
701		750
GmEPSPS_Nat	CAACTGTCC ACCTGTTTCGT GTAAATGGGA AGGAGGACT TCCTGGCGGA	
GmEPSPS_GM	 T . . C . . G . . C . . GA . A . . G . . C CT . G . . A T . .	

Fig 3C

18/52

	751		800
GmEPSPS_Nat	AAGGTGAAC	TGCTGGATC	AGTTAGCAGT CAATACTTGA CTGCTTTGCT
GmEPSPS_GM	T....	.A..C.....	G..CTCGTCA ..G...C.A. A.....
	801		850
GmEPSPS_Nat	TATGGCAGCT	CCTTTAGCTC	TTGGTGATGT GGAAATTGAG ATTGTTGATA
GmEPSPS_GM	C.....C..C	..GC.C...T	.G..G..C... ..G.....A ..C..C.....
	851		900
GmEPSPS_Nat	AACGTATTTC	TGTTCCATAT	GTTGAAATGA CTCTGAAGTT GATGGAGCGT
GmEPSPS_GM	.GT.....AG	C..G..T... ..G.....	.C..C..A.. ..A.G
	901		950
GmEPSPS_Nat	TTTGGAGTTT	CTGTGGAACA	CAGTGGTAAT TGGGATAGGT TCTTGGTCCA
GmEPSPS_GM	..C.....	.G..A.....	.TCC..G... ..C... .TC.T..A..
	951		1000
GmEPSPS_Nat	TGGAGGTCAA	AAGTACAAGT	CTCCTGGCAA TGCTTTTGT GAAGGTGATG
GmEPSPS_GM	C.....G...	AA	GC..A..... ..C..C..C

Fig 3D

19/52

1251		1300
GmEPSPS_Nat	TGCTGTTGTT	GCACTATTTG
GmEPSPS_GM	C.....C	..C..G..C.
		CTAATGGTCC
		CACTGCTATA
		AGAGATGTGG
		T..C..A...
		..G.....C.
1301		1350
GmEPSPS_Nat	CAAGTTGGAG	AGTTAAAGAG
GmEPSPS_GM	.TTCA...C.	T.....G..A
		..C..AC...
		C..T..
		T.....C..G
1351		1400
GmEPSPS_Nat	CTCAGAAAGC	TAGGAGCAAC
GmEPSPS_GM	T.GC.T....	.G..T.....
		G..G.....
		..A..A..C.
		.T..C.....
1401		1450
GmEPSPS_Nat	TACTCCACCT	GAGAAATTGA
GmEPSPS_GM	A..A..T...	..A..GC.C.
		G..C..
		..T.....T
		C..T.
1451		1500
GmEPSPS_Nat	ACAGAATGGC	CATGGCATTG
GmEPSPS_GM	T	A.....
		C..T..
		C..G..G..T

Fig 3F

1501
GmEPSPS_Nat ACCATCAAGG ATCCTGGTTG CACCAGGAAG ACATTTCCCTG ACTACTTTTGA 1550
GmEPSPS_GM ..G..... C..A..G.. T..T..... C..A.. .T.....

1551
GmEPSPS_Nat AGTCCTTGAG AGGTTAACAA AGCACTAA 1578
GmEPSPS_GM G..GT.G..A ..A..G....

Fig 3G

```

1          50
BAR1_Nat ATGAGCCCGAG AACGACGCCG GCGCGACATC CGCCGTGCCA CCGAGGCGGA
BAR1_ZM   ...TCG..T. .G..C..T.. T..T.....A A.A..C..T. ....A..
BAR1_AT   ....T.... .A.GA.A.. .T..T..T. .GA.A.... .A..T..

          100
BAR1_Nat CATGCCGGCG GTCTGCACCA TCGTCAACCA CTACATCGAG ACAAGCACGG
BAR1_ZM   ....T..T ..T..... .T..G..T.. .....TCT....
BAR1_AT   T.....T..T ..T..T..A. ....A..... T..T..... .CTCG..A.

          150
BAR1_Nat TCAACTTCCG TACCGAGCCG CAGGAACCGC AGGACTGGAC GGACGACCTC
BAR1_ZM   .A.....C..T.....T ..A..... .T.....C.....T...
BAR1_AT   .T..T..T.. C..T..... .G..A. ....T..... .T..G

          200
BAR1_Nat GTCCGTCTGC GGGAGCGCTA TCCCTGGCTC GTCGCCGAGG TGGACGGCGA
BAR1_ZM   ..G.....CA .A.....T.. ..G.....G ..T..A.... .....T..
BAR1_AT   .A...T.AA .A..A..T.. ..G.....A ..T..T.... .T.....A..

```

Fig 4A

201
 BAR1_Nat GGTGCGCGGC ATGCCTACG CGGGCCCTG GAAGGCACGC AACGCCTACG 250
 BAR1_ZM A..G..T..G

251
 BAR1_Nat ACTGGACGGC CGAGTCGACC GTGTACGTCT CCCCCCGCCA CCAGCGGACG 300
 BAR1_ZM .T.....T.. G....C..A ..C..... .A...A..A..C
 BAR1_AT A.. A..A..C..T ..C.....A G...T..T.. T..AA..A..C

301
 BAR1_Nat GGACTGGGCT CCACGCTCTA CACCACCTG CTGAAGTCCC TGGAGGCACA 350
 BAR1_ZM ..G..C.... .G..C.... T..G..T..C ..C.....T .A.....G..
 BAR1_AT ...T.A..GA G....T.G.. ...T..T..T T.A.....A.

351
 BAR1_Nat GGGCTTCAAG AGCGTGGTCG CTGTCATCGG GCTGCCCAAC GACCCGAGCG 400
 BAR1_ZM A TCT..A..G. .G..G..... CT....A... ..T..C..T..
 BAR1_AT A..... TCT..T..G. .A..T..T.. AT....A... ..T.....T..

Fig 4B

401
 BAR1_Nat TGGCGCATGCA CGAGGCGCTC GGATATGCCC CCGCGGGCAT GCTGCGGGCG 450
 BAR1_ZM .A.A.....A... ..T..C..T. .TA.A..A.. ..CA.....
 BAR1_AT .T..A..... ..A.....T ..C..T. .A..A..T.. ..C..T..T

451
 BAR1_Nat GCCGGGCTTCA AGCACGGGAA CTGGCATGAC GTGGGTTTCT GGCAGCTGGA 500
 BAR1_ZM ..T..A.... ..T.....T.. T.....C... ..T..C.... ..A.....
 BAR1_AT A.... A..T..A.. T.....C... ..A.....T. A.....

501
 BAR1_Nat CTTCAGCCTG CCGGTACCGC CCCGTCCGGT CCTGCCCGTC ACCGAGATCT 550
 BAR1_ZM TCTT.. ..A..T..A. .T.....T.. G..A..... ..A.....
 BAR1_AT ...TTCA..T ..C..T..C. .TA.A..T.. A..T..A..T ..T..A.....

551
 BAR1_Nat GA
 BAR1_ZM AG
 BAR1_AT AG

Fig 4C

1
 CTP2CP4NAT ATGGCGCAAG TTAGCAGAAT CTGCAATGGT GTGCAGAACC CATCTCTTAT CTCCAATCTC 60
 CTP2CP4_ZM A..T..... C..C G..G..G.. C...
 CTP2CP4_AT C..G. .A..T..G.. ..T..C..A ..C..A.... .T..A..A.. A..G..C..G
 CTP2CP4_GM C..... C..... C.... ..T. A..A.. C..G

61
 CTP2CP4NAT TCGAAATCCA GTCAACGCAA ATCTCCCTTA TCGGTTTCTC TGAAGACGCA GCAGCATCCA 120
 CTP2CP4_ZM AGC..G.... .C..G..G.. G..G..GC.C CAGC. .C.....C.. A.....C..G
 CTP2CP4_AT ..A...AG.T C....A.A.. G..G..GC.T ..T..A..GT A..T.. A.....C..G
 CTP2CP4_GM ..C.....GT CA.....T.. G..G..A.... ..T...AGCT T.. ..A.....T

121
 CTP2CP4NAT CGAGCTTATC CGATTTCGTC GTCGTGGGGA TTGAAGAAGA GTGGGATGAC GTTAATTGGC 180
 CTP2CP4_ZM A.G..C..C. .T..CAGC.. A..C.....C C.C.....C C..... C.G..C...
 CTP2CP4_AT A.G..... .C..C..AAG C..C.....T C.A.....A. A..... AC.G..C..T
 CTP2CP4_GM ..C..A.... .T..A..AAG CAGT.....T AT CG..T..... C..G.....T

181
 CTP2CP4NAT TCTGAGCTTC GTCCTCTTAA GGTCATGTCT TCTGTTTCCA CGGCGTGCAT GCTTCACGGT 240
 CP2CP4_ZM AGC.....G. .G..A..C.. ..G.....C ..G..C..A. A.....A.....
 CTP2CP4_AT AGC..A..A. .A..G..G.. A.....C ..A..CAG.. .T.....T...A
 CTP2CP4_GM ..G..A...A .G..AT.G.. ..G..... ..A...AGT. .A..T.....

Fig 5A

241
 CTP2CP4NAT GCAAGCAGCC GGGCCGCAAC CGCCCGCAAA TCCTCTGGCC TTTCCGGAAC CGTCCGCATT 300
 CTP2CP4_ZM G.....T.....T.C....G....G...A.G....
 CTP2CP4_AT ...A..T.. T...A.G..G ..TAGC..G. .CAGT..C.. ..G.....C
 CTP2CP4_GM ..TTCCTC.A A.....G..TA.A..G AGT.....T .G..T.....T....

301
 CTP2CP4NAT CCCGGCGACA AGTCGATCTC CCACCGGTCC TTCATGTTCC GCGGTCTCGC GAGCGGTGAA 360
 CTP2CP4_ZM ..G....T. C..A.. G.....T.. C.....G
 CTP2CP4_AT ..T....T. .AGT..T.. A...A..AG. A..T.. T..T..A..G
 CTP2CP4_GM ..A.A.... .AAGC..TAG T.....C..T T. .T..G..G.. ATCT..A..G

361
 CTP2CP4NAT ACGCGCATCA CCGGCCTTCT GGAAGGCGAG GACGTCATCA ATACGGGCAA GGCCATGCAG 420
 CTP2CP4_ZM G....G...T....T... ..G....C..G..
 CTP2CP4_AT ..A.A.... .T..TT.G.. T..G.....A ..T..T.... .C..C..T.. ..G.....A
 CTP2CP4_GM T..T.....T.....A... ..A..G.. ..A.....

421
 CTP2CP4NAT GCGATGGGCG CCCGCATCCG TAAGGAAGGC GACACCTGGA TCATCGATGG CGTCGGCAAT 480
 CTP2CP4_ZM ..T....A. .G..T.....C.....T.....A....T..C.. ..T..G..
 CTP2CP4_AT ..A....T. ..A.A.....A..A..G... ..T..G....C.. T..T..T..
 CTP2CP4_GM ..T....T. T..T.. C.....T..T..T....A..C.. A..T..G.,

Fig 5B

	481		540			
CTP2CP4NAT	GGCGGCCTCC	TGGCGCCTGA	GGCGCCGCTC	GATTTCGGCA	ATGCCGCCAC	GGGCTGCCGC
CTP2CP4_ZM	 T . G .	. C . C	. . . C . T . .	. . . C	. . . G . G . .	 A . G
CTP2CP4_AT	. A . AT . G .	. C . T . C . .	A A . T	. C . T . G .	. C . A . T . .	. . . G T
CTP2CP4_GM	. T . . . T . A .	. T . A . . G . .	. . . T . T . . .	. C . T	. C . A	A . . G . . TA . A

	541		600
CTP2CP4NAT	CTGACGATGG GCCTCGTCGG GGCTACGAT TTCGACAGCA CCTTATCAGG CGACGCCTCG		
CTP2CP4_ZM	. . C . T A . G G C T G A		
CTP2CP4_AT	. . T . T A . G . A . . C . G . T . C . . T . . T C T T G A G C		
CTP2CP4_GM	. . T . T G T . . T T T C A T G . . T T		

CTP2CP4_NAT	CTCACAAAGC	GCCCGATGGG	CCGCGTGTG	AACCCGCTGC	GCGAAATGGG	CGTGCAGGTG	660
CTP2CP4_ZM		A.....	TC..	T...	G.....	A.....C	
CTP2CP4_AT	T...A	A..A.....	A..A...C..	..T..C...A	G..G....	T..C.....	
CTP2CP4_GM	T..A.	A.....	AA..A..CC.T	T..A	G..G....	A..A..T	
	601						

	661		720
CTP2CP4NAT	AAATCGGAAG	ACGGTGACCG	TCTTCCCGTT
CTP2CP4_ZM	C..G.	.T.....	..GC.....
CTP2CP4_AT	T..G.	.T.....	..TC....A.
CTP2CP4_GM	..G..C..G.	C...A.	AT.G.....

Fig 5C

CTP2CP4NAT	721	ATCACCTACC	GCGTGCCGAT	GGCCTCCGCA	CAGGTGAAGT	CCGCCGTGCT	GCTCGCCGGC	780
CTP2CP4_ZM		..T.....	A..	..A.....C	C...	..A.....	C.....A	
CTP2CP4_AT		G...A	..G..T....	G..A...	C....	..A..G..A..	C..G..G...	
CTP2CP4_GM		..T...A	..G..T..A..	..A..T..T	..A.....	..A..T..	T..A	
CTP2CP4NAT	781	CTCAACACGC	CCGGCATCAC	GACGGTCATC	GAGCCGATCA	TGACGCGCGA	TCATACGGAA	840
CTP2CP4_ZM		..G.....T	..G.....	..C.....G...	C...	CA.G..	C...	
CTP2CP4_AT		A	..T..A....	..A..C..G..T	..A..C....	TA.A..	C..C.....G	
CTP2CP4_GM		T.G.....A	..G..T..T..	..T..C..G..T	T...	T..T..	C..C..T..G	
CTP2CP4NAT	841	AAGATGCTGC	AGGGCTTTGG	CGCCAACCTT	ACCGTCGAGA	CGGATGCGGA	CGGCGTGCGC	900
CTP2CP4_ZM		T	G.....	G	..G.....	C..T..	CA.G	
CTP2CP4_AT		T	T..C..	T..T..A	..G.....A	..C..C..C..	A.G	
CTP2CP4_GM		T	T..C..	..T..T.....C	T..A	..A..C.....	T..A.G	
CTP2CP4NAT	901	ACCATCCGCC	TGGAAGGCCG	CGGCAAGCTC	ACCGGCCAAG	TCATCGACGT	GCCGGGCGAC	960
CTP2CP4_ZM		T	..T..G...A	..G..T..A..G	..T.....	T..	T.....A...	
CTP2CP4_AT		..A.....T	G...A	..A..T..A..G	..T.....	T..	T..A..T	
CTP2CP4_GM		T	G..AA	..G..A..A...	..T..T....	T.....	C..T..T	

Fig 5D

961 1020
 CTP2CP4NAT CCGTCCTCGA CGGCCCTTCCC GCTGGTTGCG GCCCTGCTTG TTCCGGGGCTC CGACGTCACC
 CTP2CP4_ZM G C GAG T G . . .
 CTP2CP4_AT C T C C T G . . .
 CTP2CP4_GM AGC G A C T C T A . . .

1021 1080
 CTP2CP4NAT ATCCTCAACG TGCTGATGAA CCCCACCCGC ACCGGCCTCA TCCTGACGCT GCAGGAAATG
 CTP2CP4_ZM G C G A . G G G . . .
 CTP2CP4_AT G T C TA . A C AT G . . .
 CTP2CP4_GM T A C T TA . A AT T

1081 1140
 CTP2CP4NAT GCGGCCGACA TCGAAGTCAT CAACCCGCGC CTTGCCGGCG GCGAAGACGT GGCGGACCTG
 CTP2CP4_ZM A G G G A T T . . .
 CTP2CP4_AT T T G TA . G T A G T . . .
 CTP2CP4_GM T T T TA . A A T G TT . .

1141 1200
 CTP2CP4NAT CGCGTTCGCT CCTCCACGCT GAAGGGCGTC ACGGTGCCGG AAGACCGCGC GCCTTCGATG
 CTP2CP4_ZM G T C G C T G C . . .
 CTP2CP4_AT G T TAGT A C G C T A . . .
 CTP2CP4_GM G GA . G T AT A T C A TA . G A

30/52

30/52

Fig 5E

CTP2CP4NAT	1201	ATCGACGAAT	ATCCGATTCT	CGCTGTGCGC	GCCGCCTTCG	CGGAAGGGGC	GACCGTGATG	1260
CTP2CP4_ZM		.A....G.	T....	G..C.....	..T.G....	.C.....	C..G..C...	
CTP2CP4_AT		G.	C..C.....	...C..T..T	..T.G..T.	.C..G..C..	A..T..A...	
CTP2CP4_GM		.T.....G.	A.....	T..C.....G	..T.....T.	.T..G..C..		
CTP2CP4NAT	1261	AACGGTCTGG	AGAACTCCG	CGTCAAGGAA	AGCGACCGCC	TCTCGGCCGT	CGCCAATGGC	1320
CTP2CP4_ZM		T.	.G.....	..G.....	TCG..T....	.G.....	G.....	
CTP2CP4_AT		C..T.	.G..GT.GA.	G..T.....G	..T...A.G.	.G..C..G..	G..G.....	
CTP2CP4_GM		..T..A..A.	.G..GT.GA.	A..G.....	TC....A.AT	.GAGC..A..	...T..C..A	
CTP2CP4NAT	1321	CTCAAGCTCA	ATGGCGTGGA	TTGCGATGAG	GGCGAGACGT	CGCTCGTCGT	GCGTGGCCGC	1380
CTP2CP4_ZM		..G.....	.C..T..T..	C.....C...	..T.....C.	.A.....	G.....	
CTP2CP4_AT		..G.....A.	.C.....	C.....C..A	..T..A....	.C..T..A..	C.....T...	
CTP2CP4_GM		..T..A....	T..	..T.....	..T.....TA	GCT.G..A..	C.....G..A	
CTP2CP4NAT	1381	CCTGACGGCA	AGGGGCTCGG	CAACGCCTCG	GGCGCCGCCG	TGCGCACCCA	TCTCGATCAC	1440
CTP2CP4_ZM		T....	C.....	AGT	..A..G....	G..	C.....T	
CTP2CP4_AT		..A.....G.	T.G..	G..T..T...	..A..T..T.	.G..G..G..	C..T.....T	
CTP2CP4_GM		..A.....A.	TT.G..	T...	..T..T....	.T..A..T..	CT.G.....T	

Fig 5F

1441 CTP2CP4NAT CGCATCGCCA TGAGCTTCCT CGTCATGGGC CTCGTGTCGG AAACCCCTGT CACGGTGGAC 1500
 CTP2CP4_ZM G. .TC....T. G..G.....T C..A. .G.....G.. G..C..C..T
 CTP2CP4_AT A.A..... ..TCA..T.. G..G.....A ..T..C..C. .G..T..G.. G..C..T...
 CTP2CP4_GM ..G..A..G. T..T.. G..G.....T AAGC. .G..T..... G..A..C...

1501 CTP2CP4NAT GATGCCACGA TGATCGCCAC GAGCTTCCCG GAGTTCATGG ACCTGATGGC CGGGCTGGGC 1560
 CTP2CP4_ZM ..C..... ..A..G.. A T..... G..C..C..G
 CTP2CP4_AT T..C. CTC...T..T C..... A..CT....G
 CTP2CP4_GM A..T. T..TTC.....T TT.A..... A..A..A..T

1561 CTP2CP4NAT GCGAAGATCG AACTCTCCGA TACGAAGGCT GCCTGA 1596
 CTP4CP4_ZM ..C..... ..G..T.. C.....C ..T...
 CTP2CP4_AT ..C..... ..G..G..T.. ..T.....C ..T...
 CTP2CP4_GM ..A.....T. T.. ..C..A..G A.

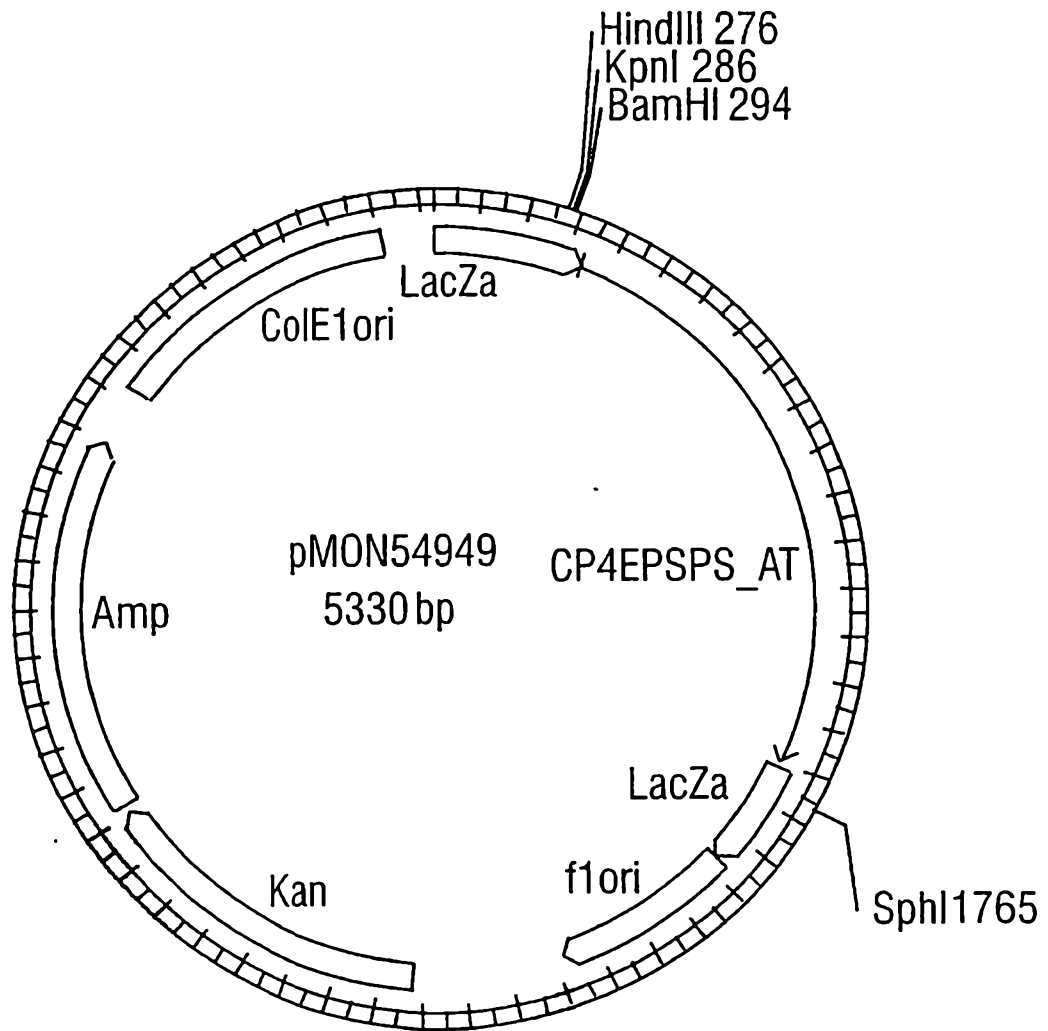


Fig 6

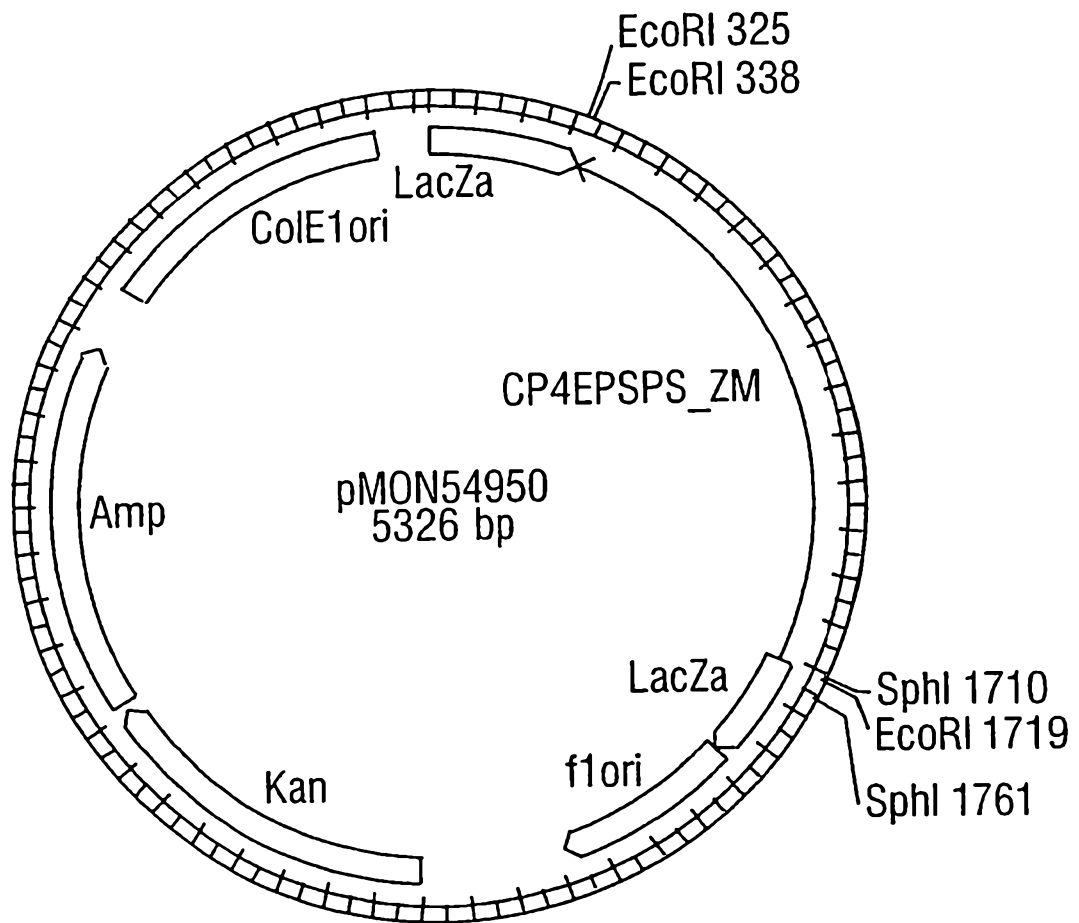


Fig 7

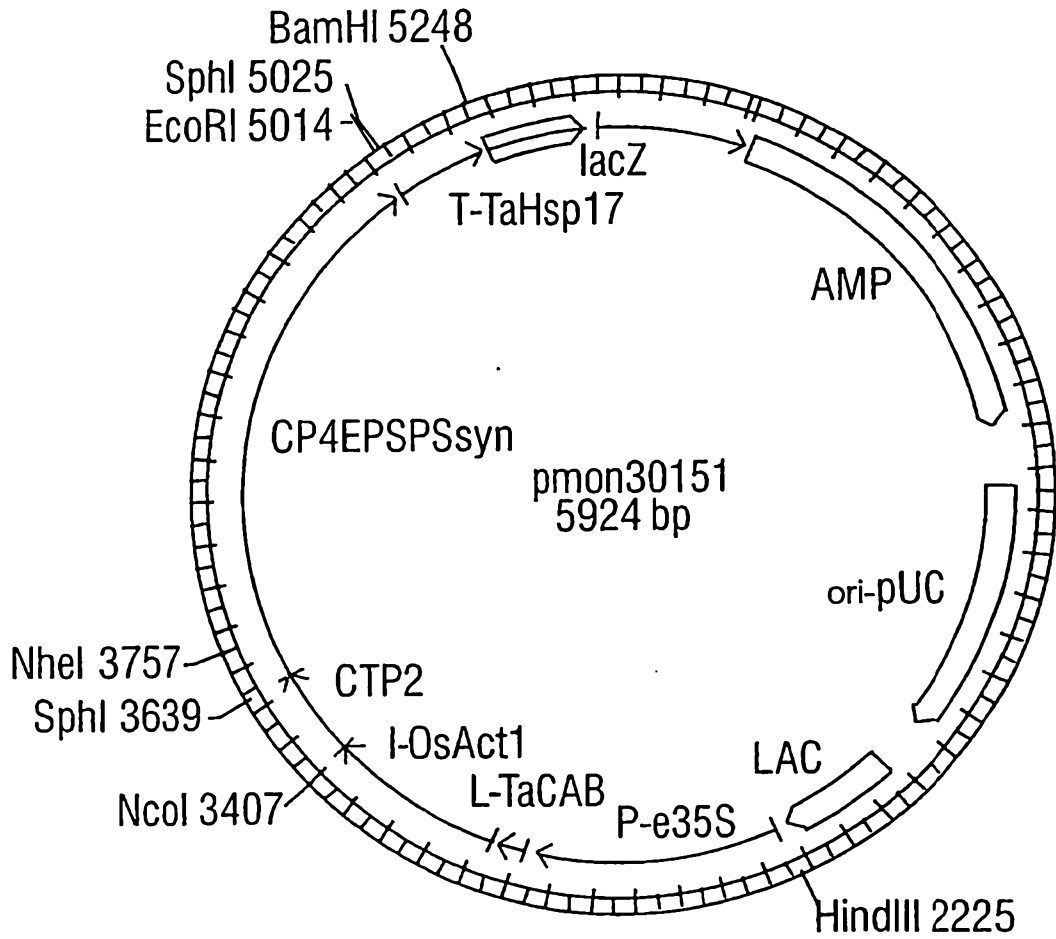


Fig 8

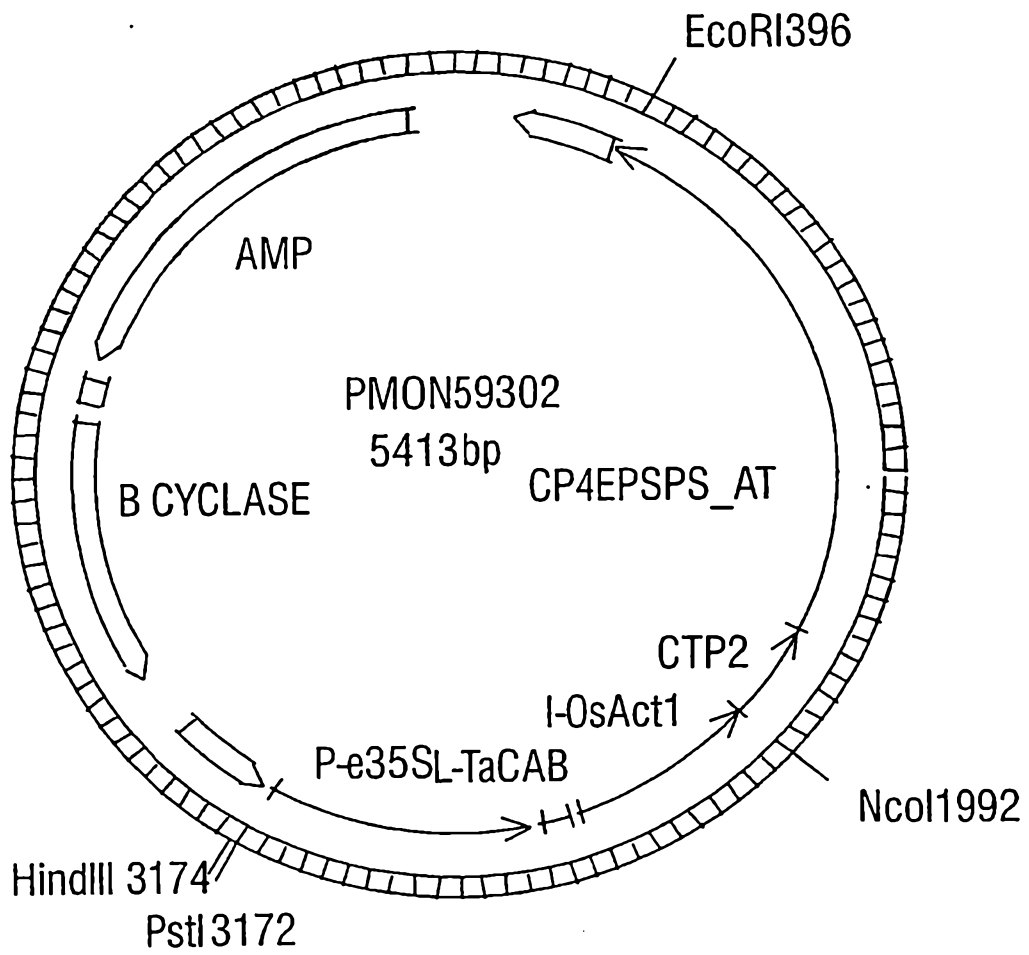


Fig 9

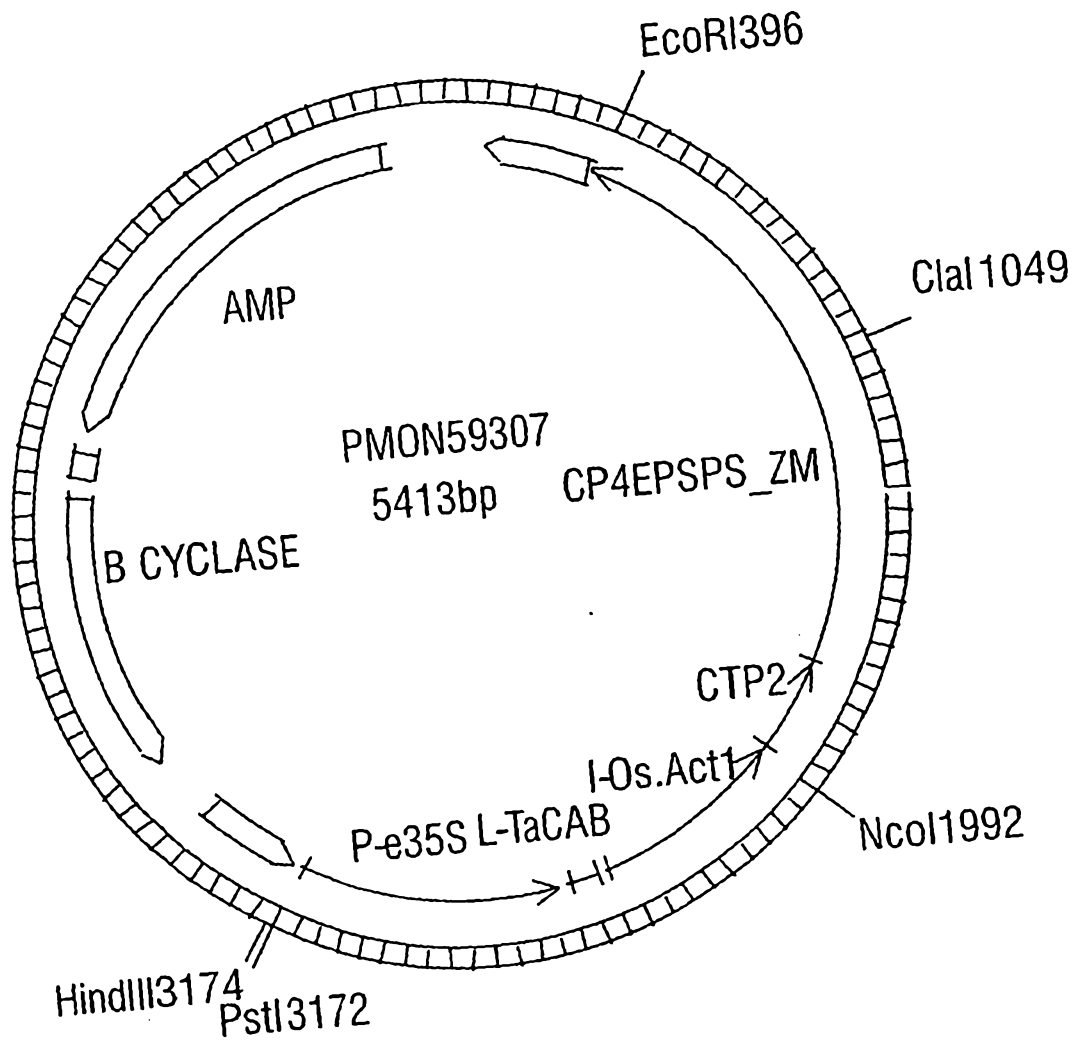


Fig 10

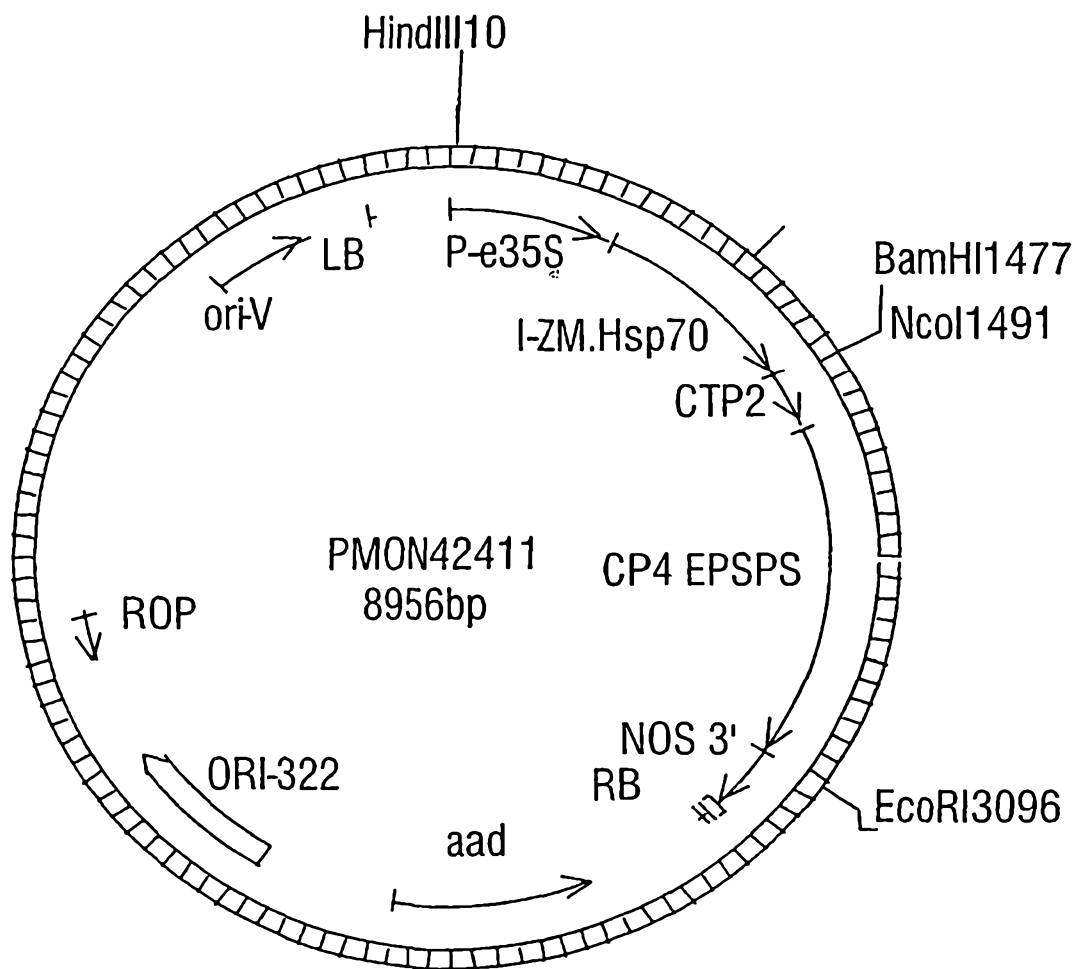


Fig 11

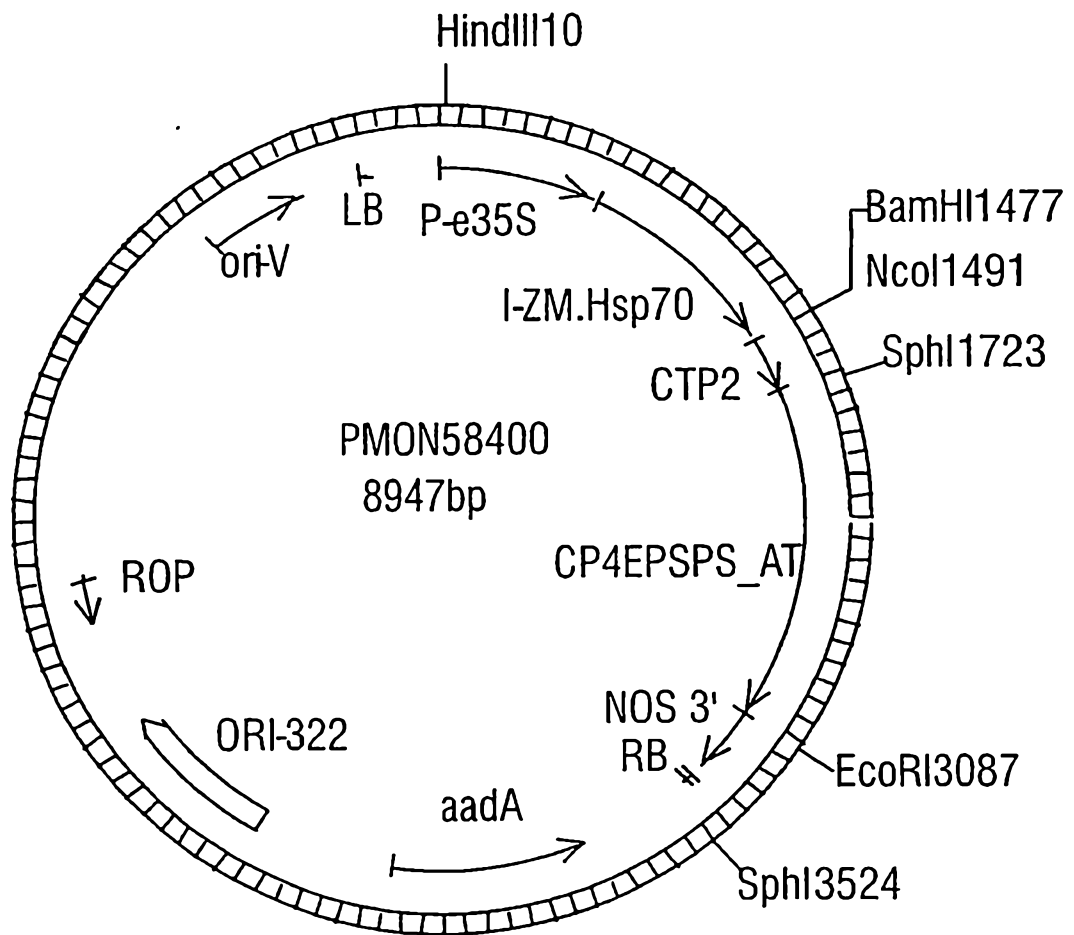


Fig 12

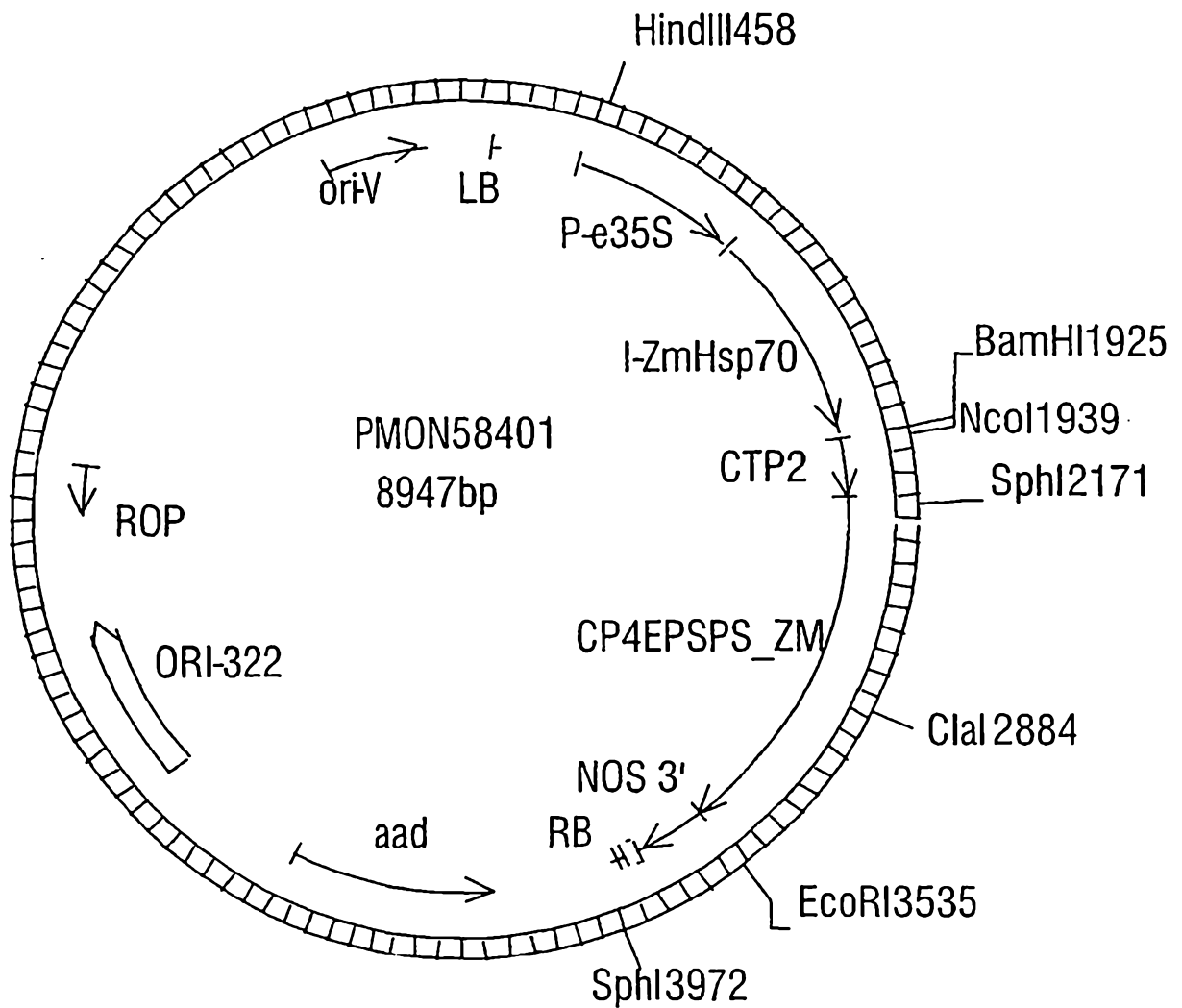


Fig 13

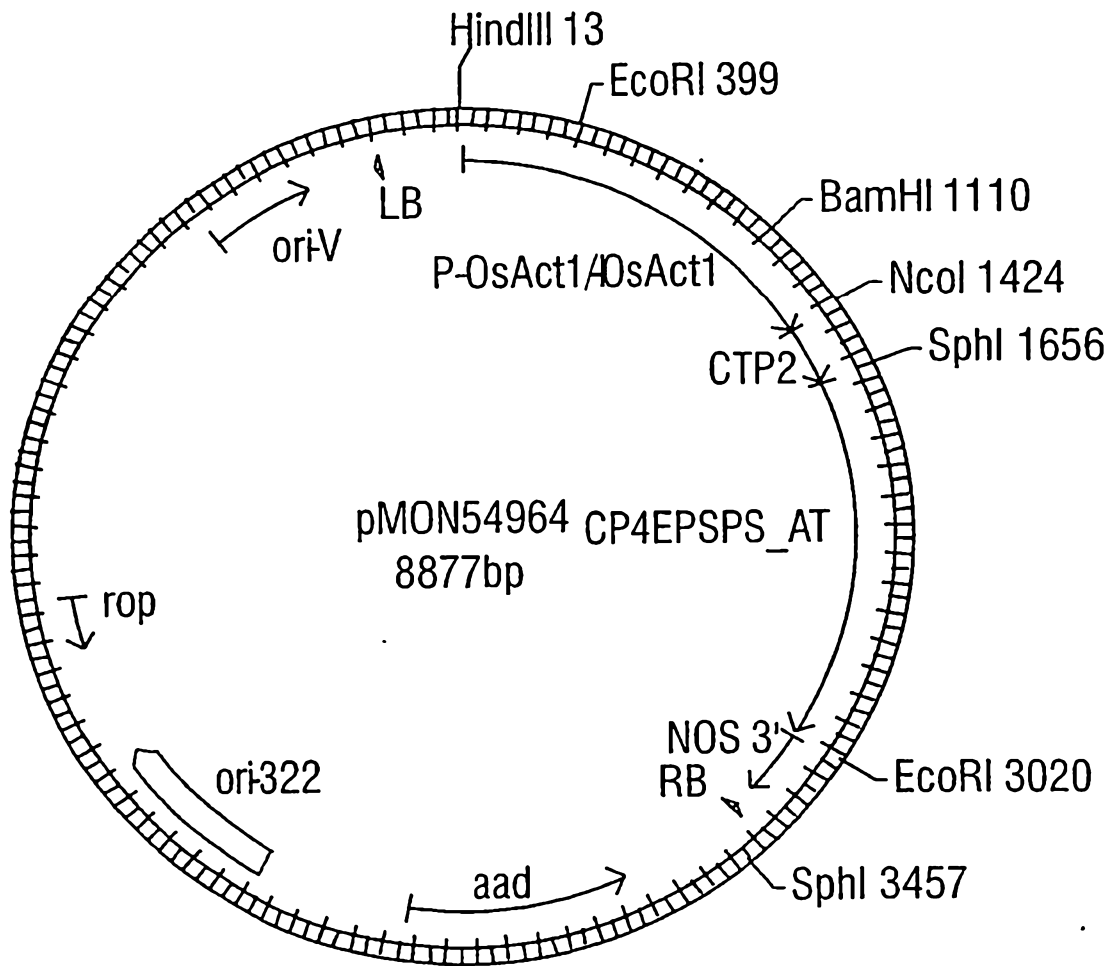


Fig 14

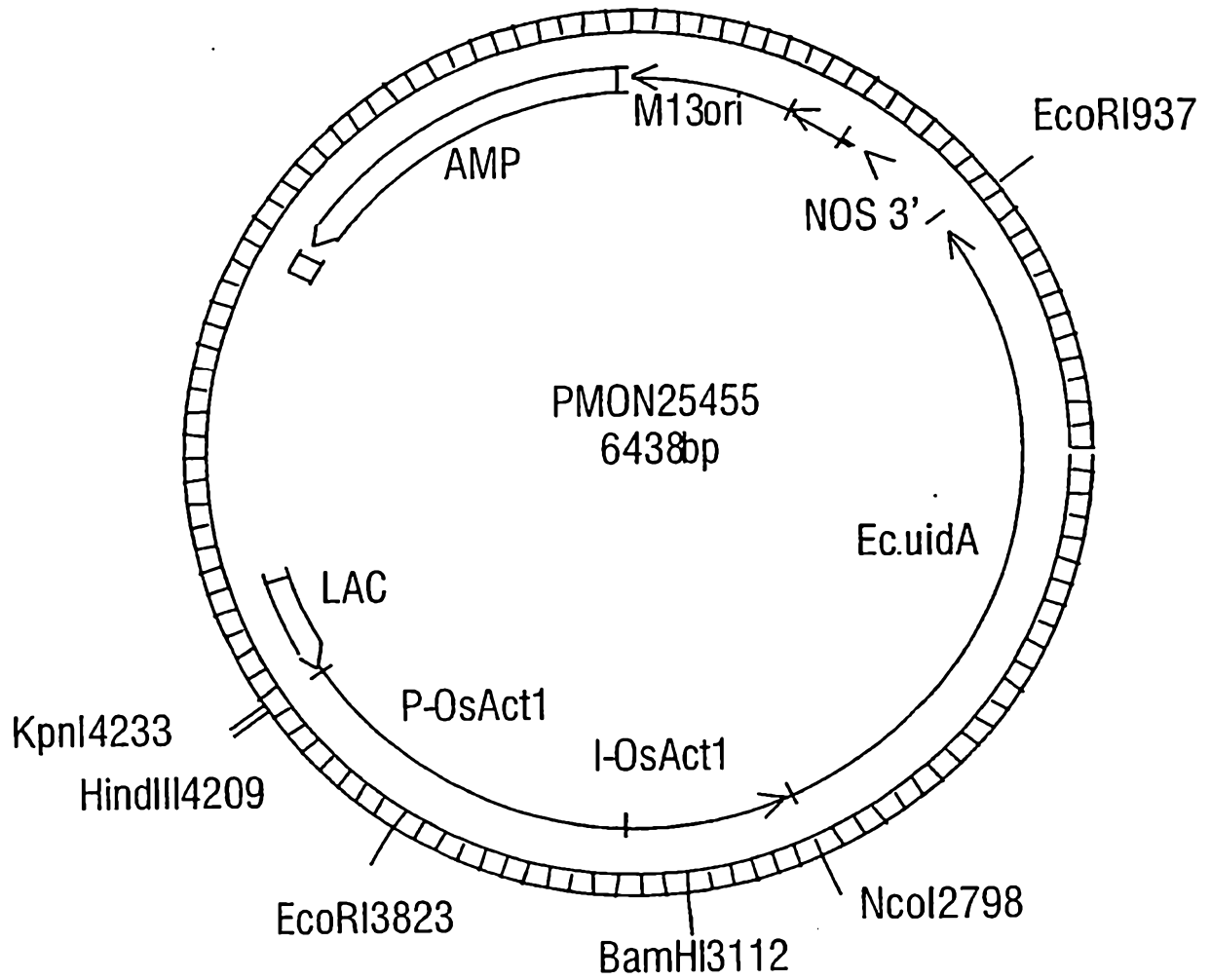


Fig 15

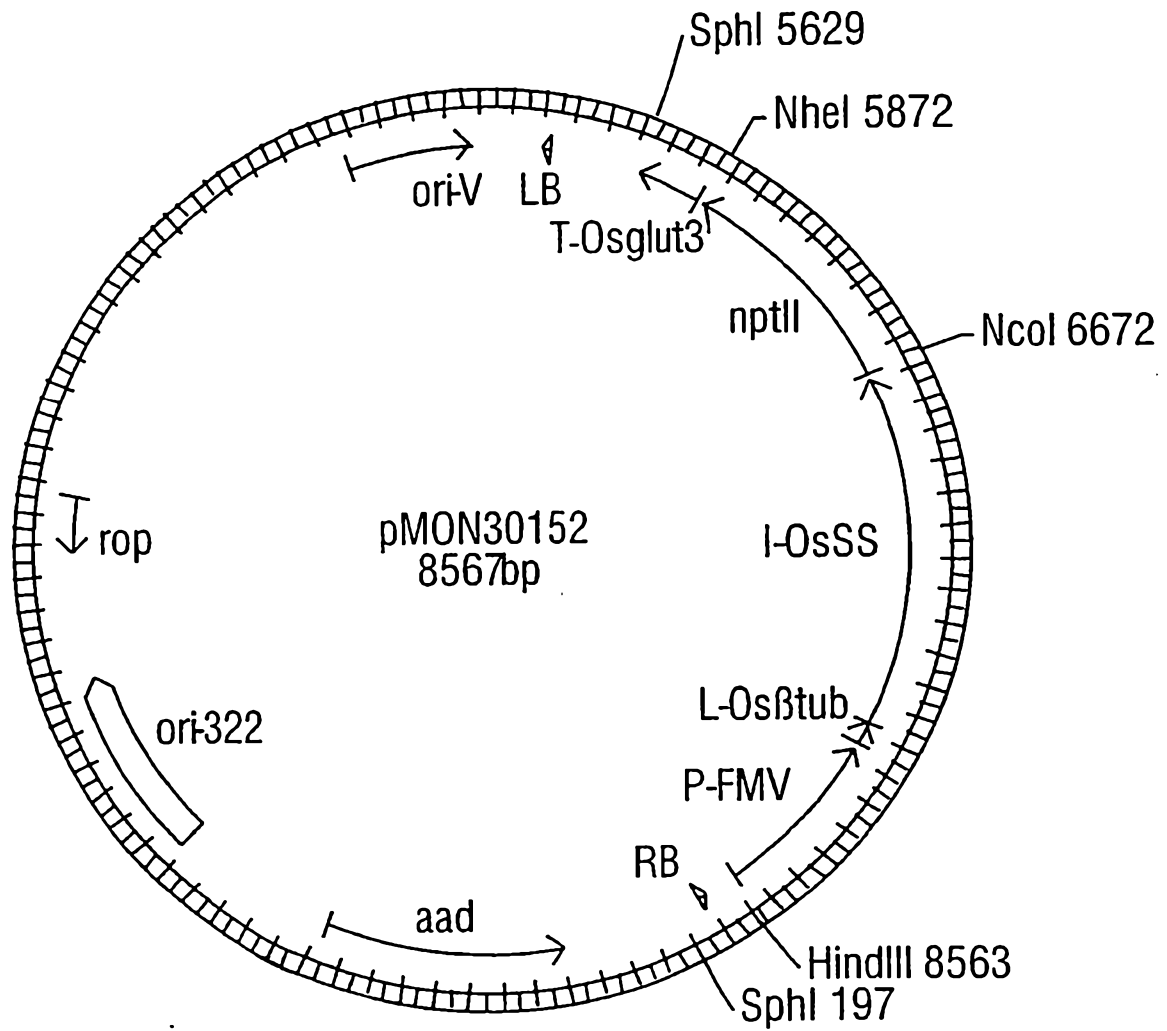


Fig 16

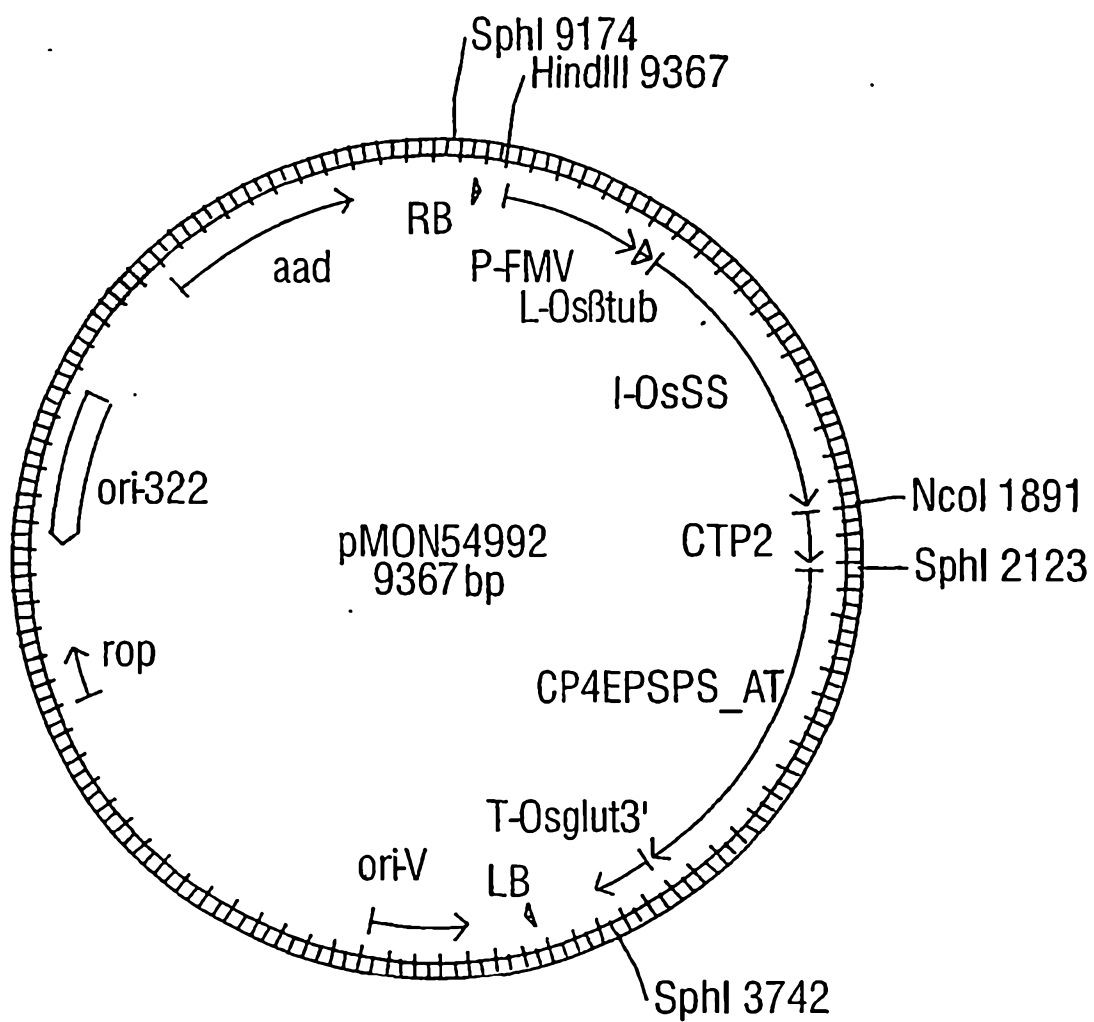


Fig 17

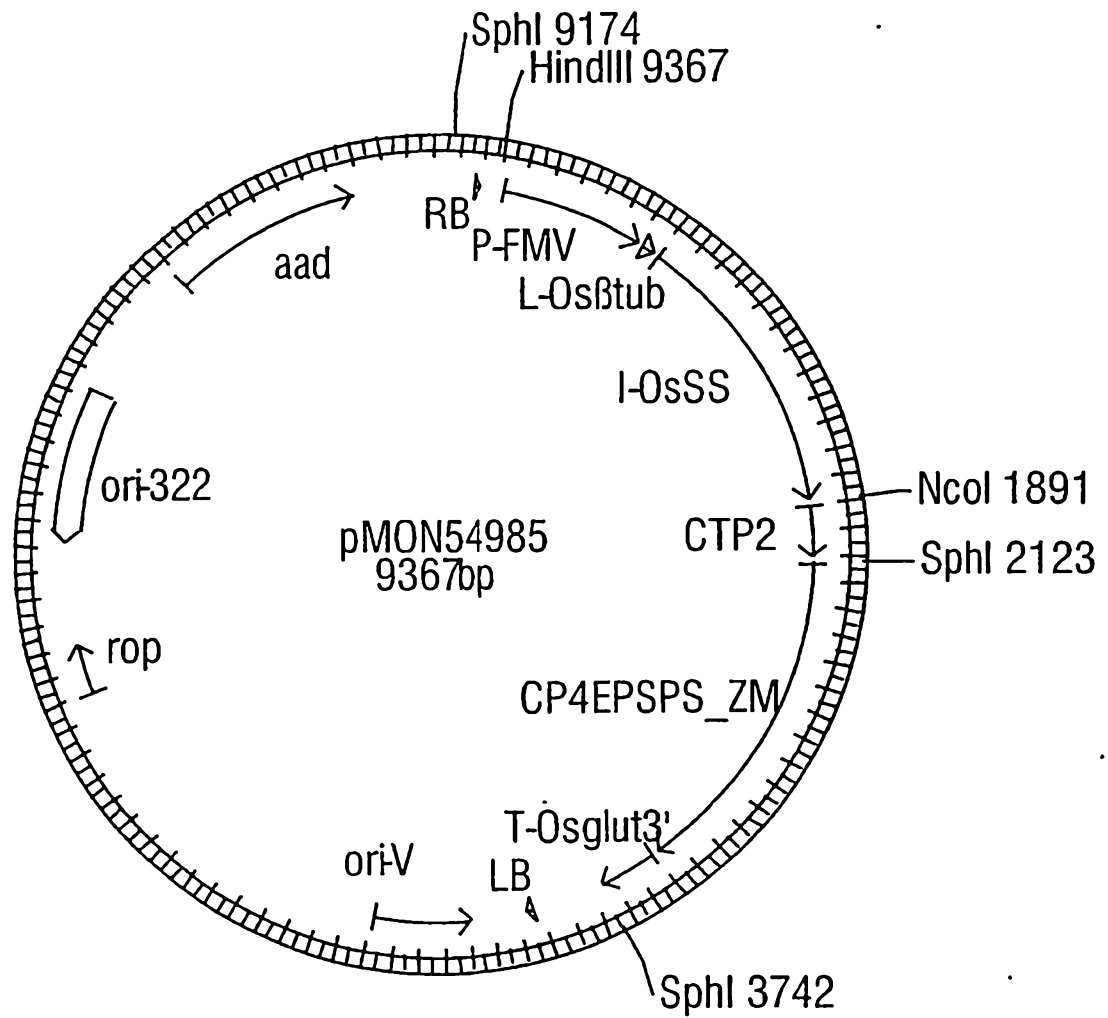


Fig 18

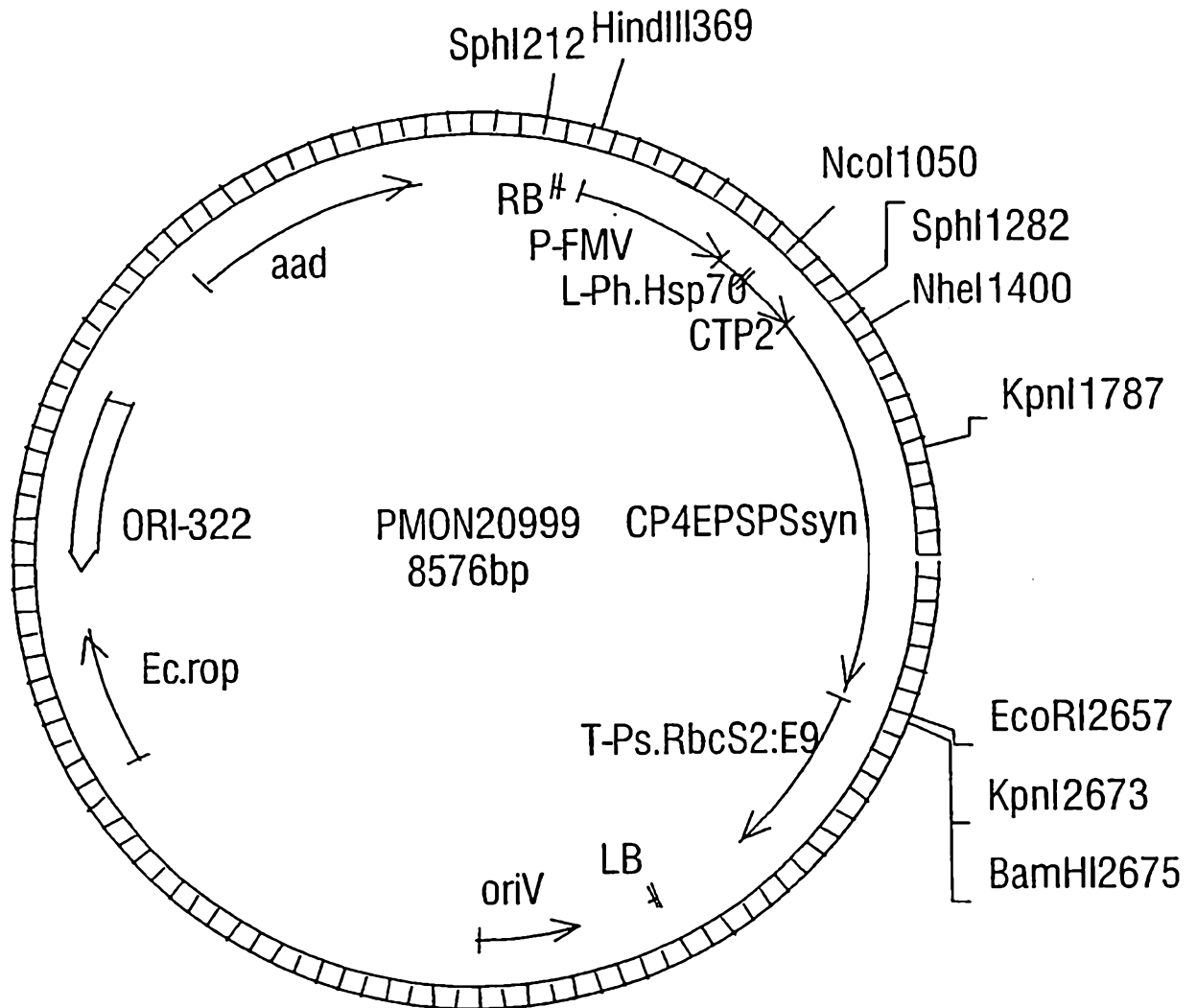


Fig 19

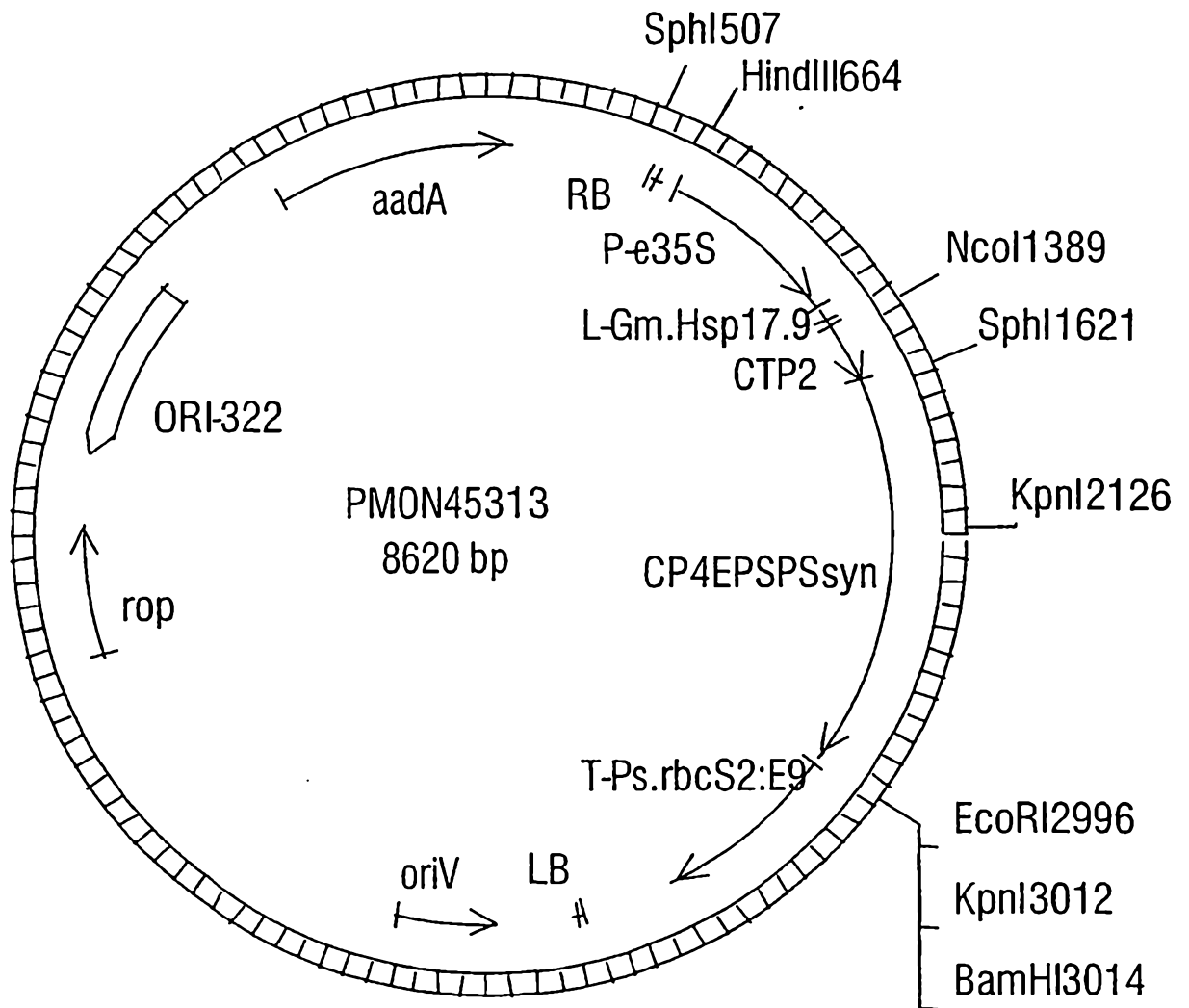


Fig 20

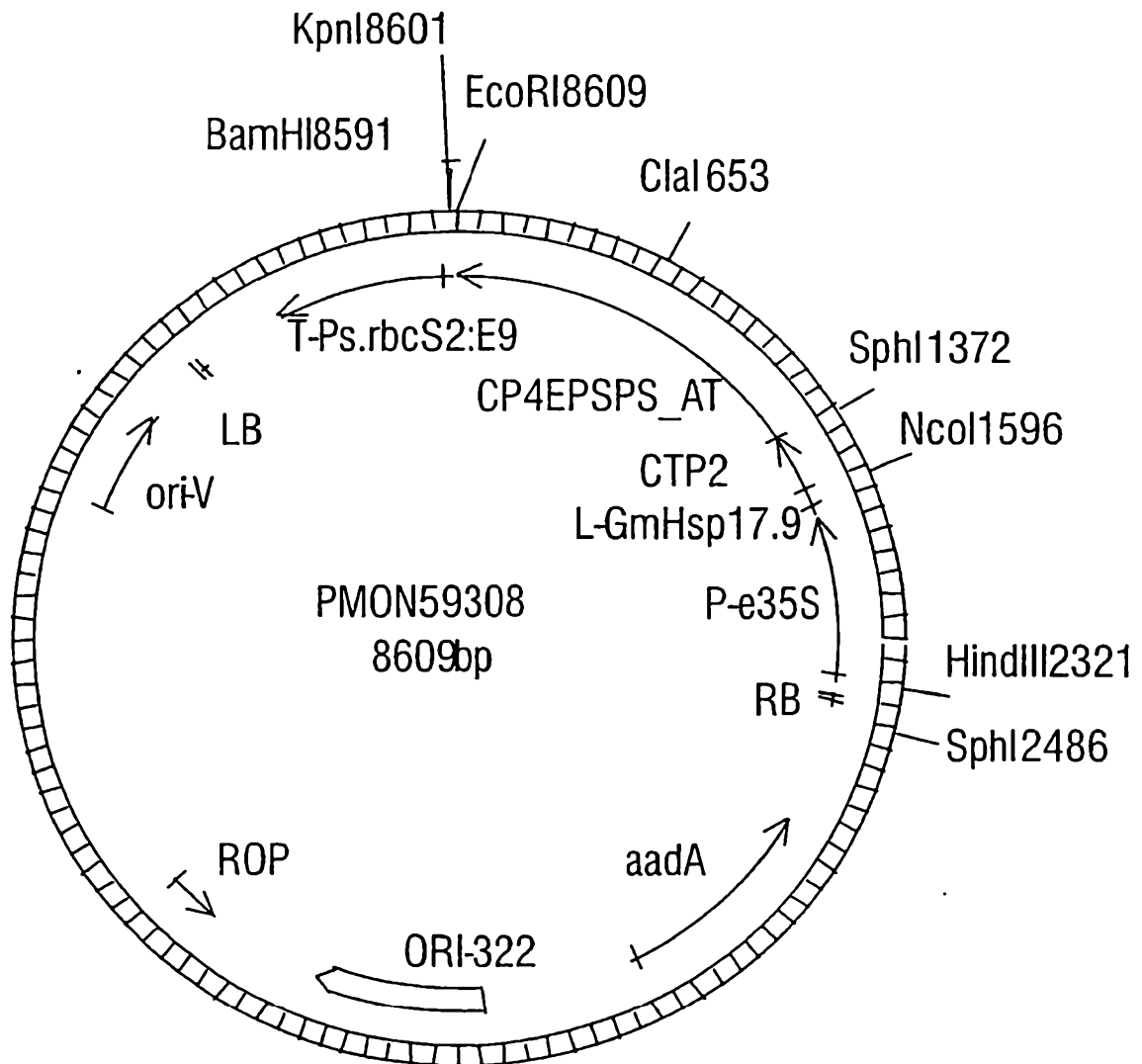


Fig 21

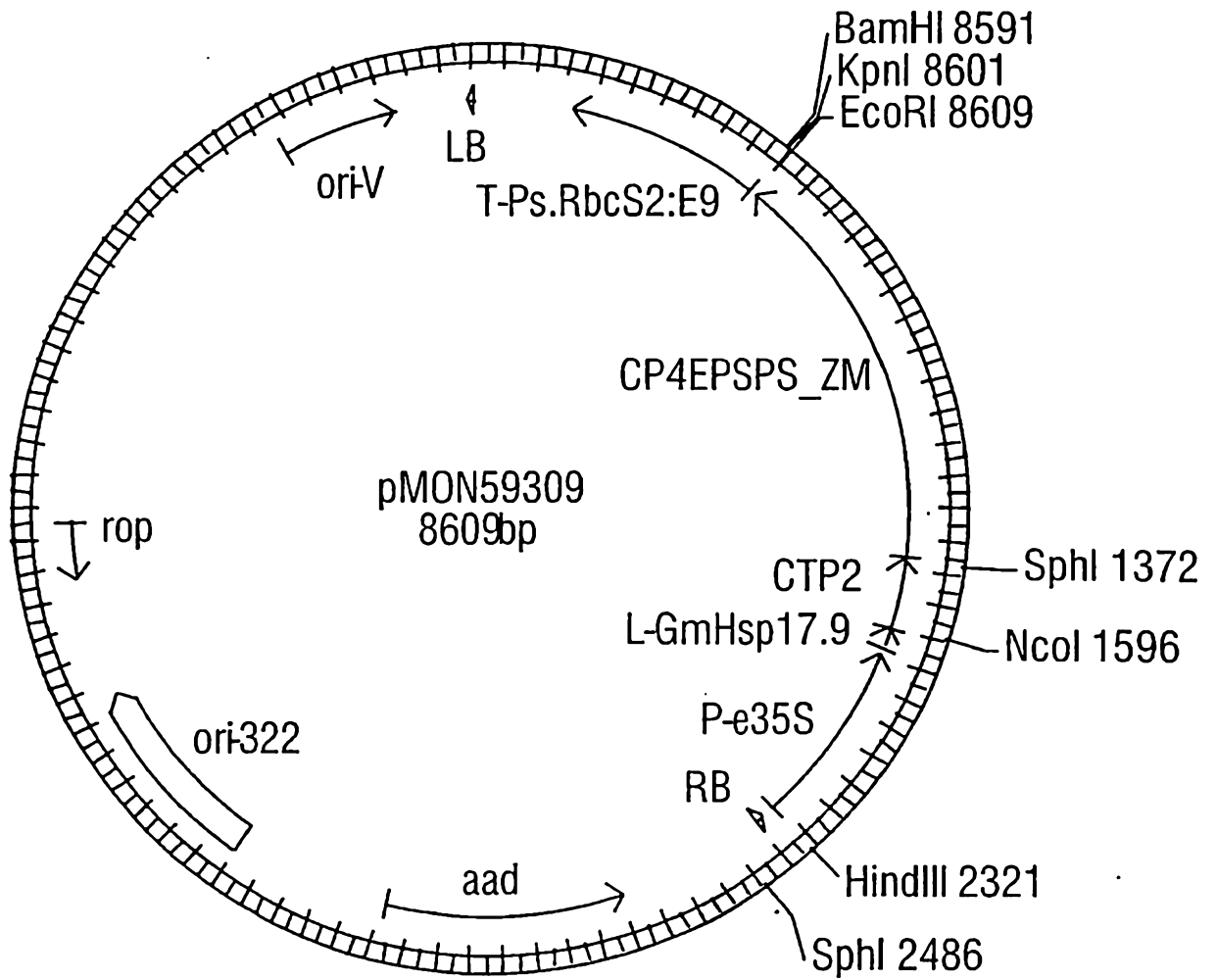


Fig 22

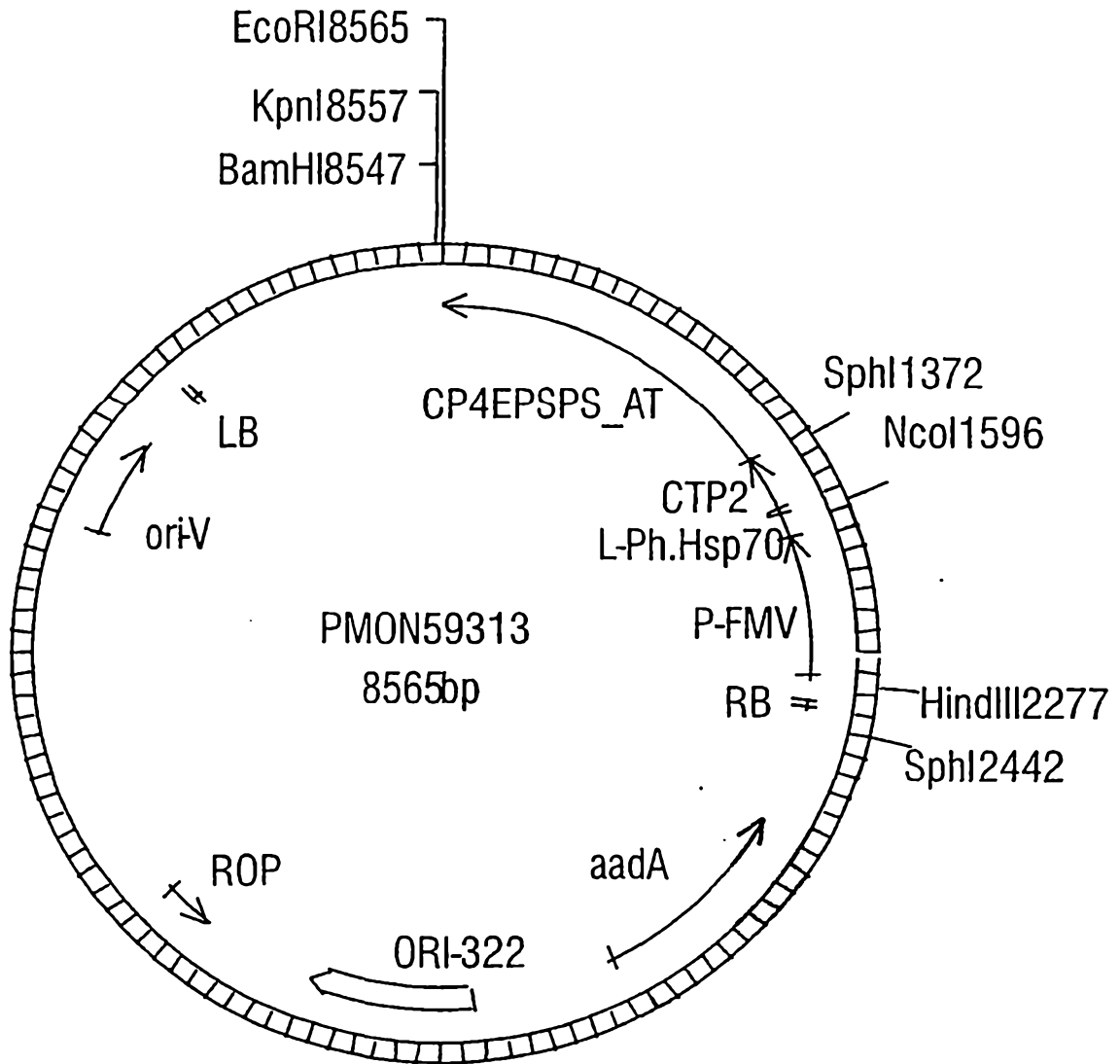


Fig 23

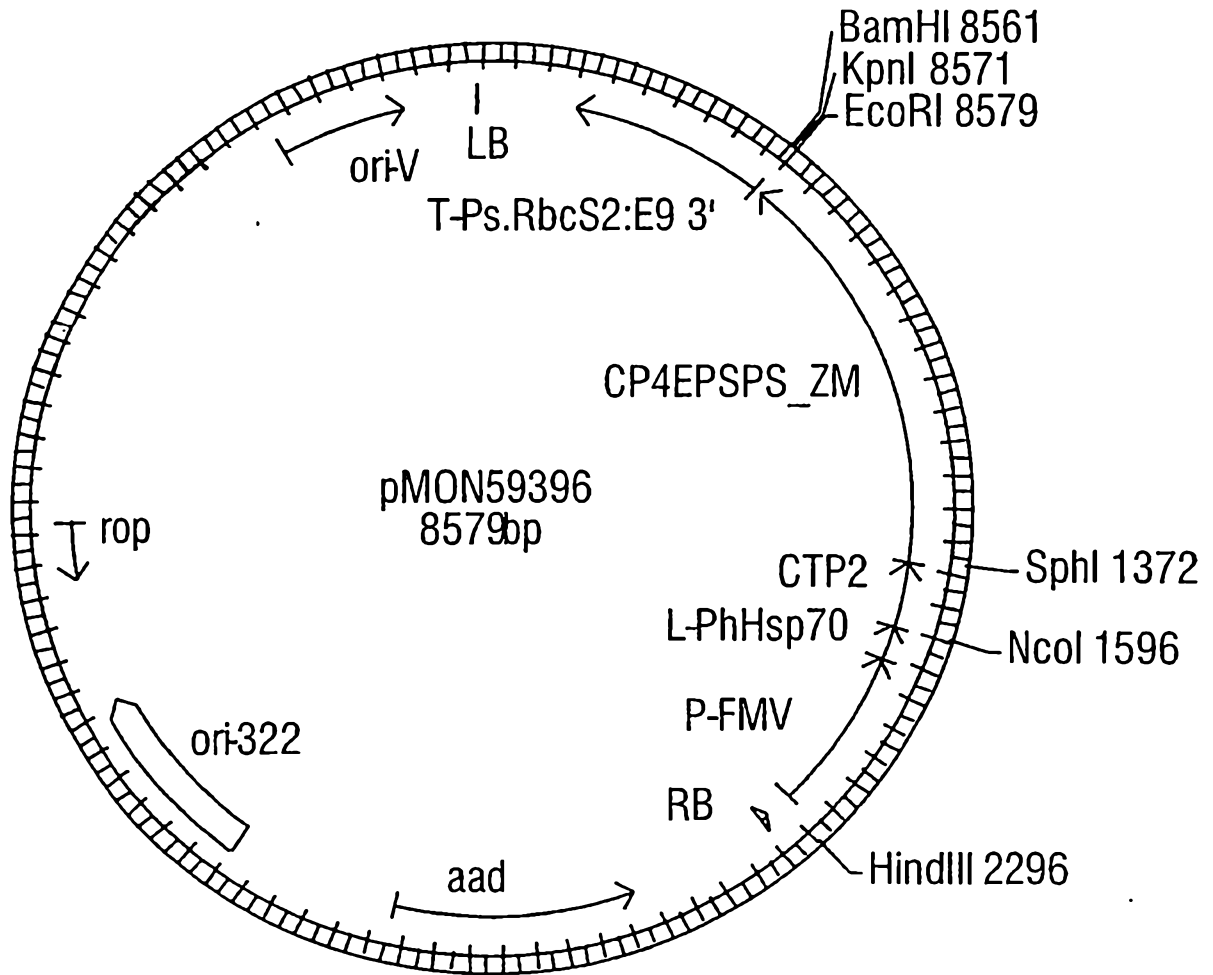


Fig 24

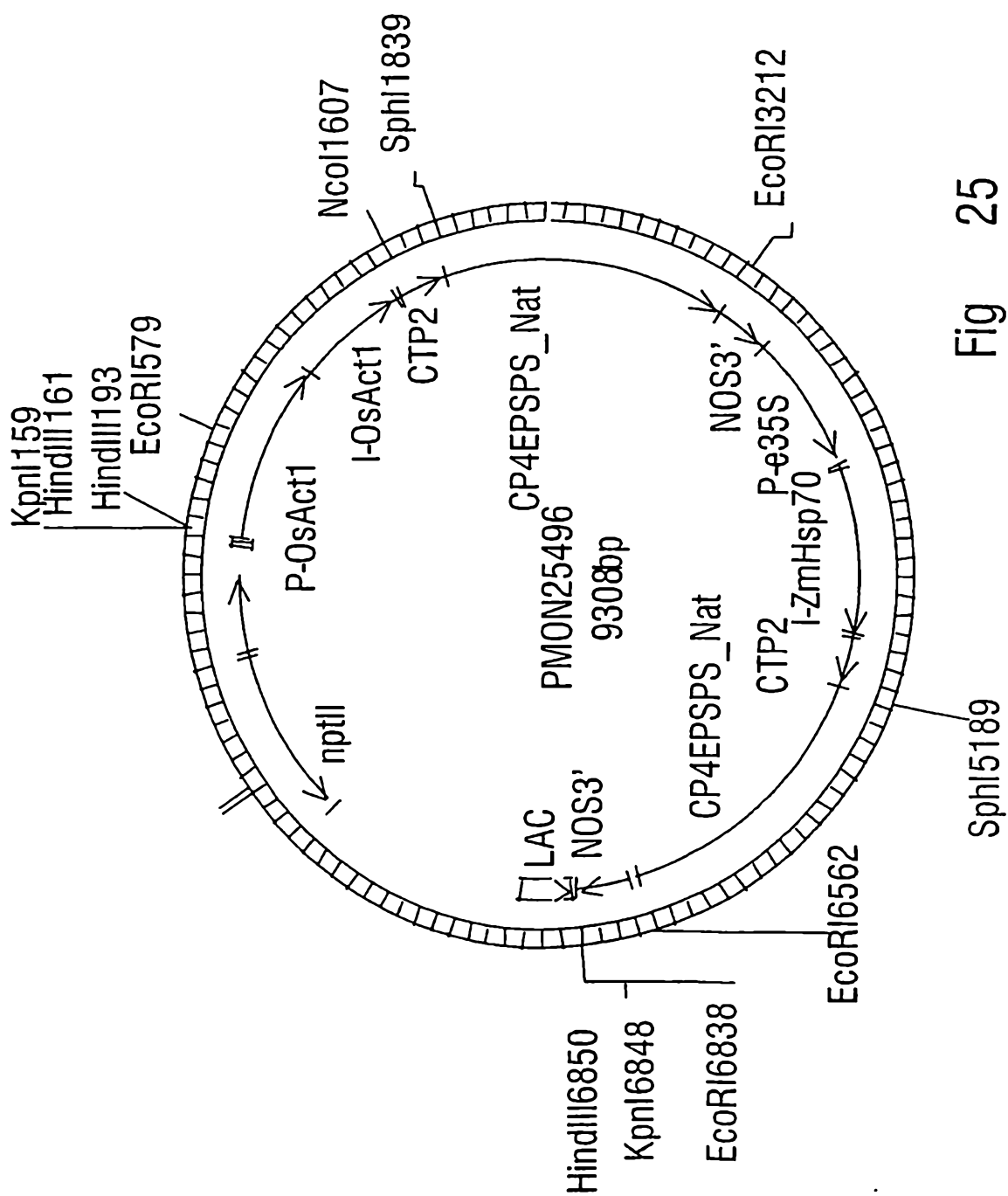


Fig 25

RESUMO

Patente de Invenção: **"MÉTODOS PARA REDUZIR SILENCIAMENTO DE TRANSGENE E PARA DETECTAR POLINUCLEOTÍDEO EM PLANTAS, MOLÉCULA E CONSTRUTO DE DNA, BEM COMO KIT DE DETECÇÃO DE DNA".**

5 A presente invenção refere-se a um método para reduzir o silenciamento de transgene em plantas transgênicas compreendendo as etapas de: a) obter um construto de DNA compreendendo um polinucleotídeo artificial que codifica um peptídeo de trânsito de cloroplasto operavelmente ligado à SEQ ID NO: 18; b) transformar o referido construto de DNA dentro de uma
10 célula vegetal; e c) regenerar a referida célula vegetal em uma planta transgênica fértil, bem como a uma molécula de polinucleotídeo artificial de SEQ ID NO: 18 e a um construto de DNA compreendendo a mesma. A presente invenção refere-se ainda a um método de detectar um polinucleotídeo artificial
15 em célula vegetal, planta ou sua progênie, e a kit de detecção de DNA.