



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

(21) **PI0608940-2 A2**



(22) Data de Depósito: 07/04/2006
(43) Data da Publicação: 17/02/2010
(RPI 2041)

(51) *Int.Cl.:*
C12N 9/10 (2010.01)
C12N 15/82 (2010.01)

(54) Título: **IDENTIFICAÇÃO DE UMA NOVA CLASSE DE SINTASES EPSP**

(30) Prioridade Unionista: 08/04/2005 US 60/669,686, 06/05/2005 US 60/678,348, 29/06/2005 US 60/695,193, 08/04/2005 US 60/669,686, 08/04/2005 US 60/669,686, 06/05/2005 US 60/678,348

(73) Titular(es): Athenix Corporation

(72) Inventor(es): Brian Carr, Nadine Barbara Carozzi, Philp E. Hammer

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

(86) Pedido Internacional: PCT US2006013161 de 07/04/2006

(87) Publicação Internacional: WO 2006/110586 de 19/10/2006

(57) Resumo: IDENTIFICAÇÃO DE UMA NOVA CLASSE DE SINTASES EPSP. A presente invenção refere-se a composições e métodos para conferir tolerância ao glifosato em bactérias, plantas, células, tecidos e sementes de planta. Composições incluem uma nova classe de enzimas EPSPS, designadas Classe III, e polinucleotídeos que codificam tais enzimas, vetores que compreendem esses polinucleotídeos, e células hospedeiras que compreendem os vetores. As novas proteínas compreendem pelo menos um domínio de sequência selecionado a partir dos domínios Classe III aqui fornecidos. Esses domínios de sequência podem ser usados para identificar EPSP sintases com atividade de resistência ao glifosato.

Relatório Descritivo da Patente de Invenção para "**IDENTIFICAÇÃO DE UMA NOVA CLASSE DE SINTASES EPSP**".

REFERÊNCIA CRUZADA A PEDIDO DE PATENTE RELACIONADO

Esse pedido de patente reivindica o benefício dos Pedidos de patente Provisórios U.S. N^{os} 60/669.686, depositado em 8 de abril de 2005; 60/678.348 depositado em 6 de maio de 2005; 60/695.193 depositado em 29 de junho de 2005; e 60/725.182 depositado em 11 de outubro de 2005, cujos conteúdos estão incorporados aqui por referência em sua totalidade.

CAMPO DA INVENÇÃO

10 A presente invenção refere-se à biologia molecular de planta, particularmente a uma nova classe de EPSP sintases que conferem resistência ao herbicida glifosato.

ANTECEDENTES DA INVENÇÃO

15 N-fosfonometilglicina, comumente referida como glifosato, é um importante agrônomo químico. O glifosato inibe a enzima que converte o ácido fosfoenolpirúvico (PEP) e ácido 3-fosfoxiquímico (S3P) em ácido 5-enolpiruvil-3-fosfochiquímico. A inibição dessa enzima (5-enolpiruvilchiquimato-3-fosfato sintase; referida aqui como "EPSP sintase" ou "EPSPS") mata células de planta por bloquear a via do chiquimato, dessa maneira inibindo a biossíntese de aminoácidos aromáticos.

20 Já que herbicidas da classe do glifosato inibem a biossíntese de aminoácido aromático, eles não apenas matam células de planta, mas eles também são tóxicos para células bacterianas. O glifosato inibe várias EPSP sintases bacterianas, e assim é tóxico para essas bactérias. Entretanto, certas EPSP sintases bacterianas tem uma grande tolerância ao glifosato.

25 Células de planta resistentes a toxicidade do glifosato podem ser produzidas pela transformação de células de planta para expressar EPSP sintases bacterianas resistentes ao glifosato. Notavelmente, o gene bacteriano da cepa de *Agrobacterium tumefaciens* CP4 tem sido usado para conferir resistência a herbicida em células de planta após a expressão em plantas. Uma EPSP sintase mutada da cepa de *Salmonella typhimurium* CT7 confere resistência ao glifosato em células bacterianas e confere resistência ao glifo-

sato em células de planta (Patente U.S. Nºs 4.535.060; 4.769.061; e 5.094.945).

EPSP sintase (Mr 46.000) se dobra em dois domínios similares, cada um compreendendo três cópias de uma unidade de dobramento $\beta\alpha\beta\alpha\beta\beta$ (Stallings *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:5046–5050).

5 Lys-22, Arg-124, Asp-313, Arg-344, Arg-386, e Lys-411 são resíduos conservados da EPSP sintase de *E. coli* (Schönbrunn *et al.* (2001) *Proc. Natl. Acad. Sci. U.S.A.* 98:1376-1380). Resíduos conservados importantes para a atividade de EPSPS incluem também Arg-100, Asp-242, e Asp-384 (Selvapandiyan *et al.* (1995) *FEBS Letters* 374:253-256). Arg-27 se liga a S3P

10 (Shuttleworth *et al.* (1999) *Biochemistry* 38:296-302). Variantes da enzima EPSPS selvagem que foram isoladas são tolerantes ao glifosato como resultado de alterações na sequência codificante de aminoácidos de EPSPS (Kishore & Shah (1988) *Annu. Rev. Biochem.* 57:627-63; Wang *et al.* (2003) *J. Plant Res.* 116:455-60; Eschenburg *et al.* (2002) *Planta* 216:129-35). He *et al.* (2001, *Biochim et Biophysica Acta* 1568:1-6) desenvolveram EPSP sintases com tolerância aumentada ao glifosato por mutagênese e recombinação entre genes de EPSPS de *E. coli* e *Salmonella typhimurium* e sugeriram que

15 mutações na posição 42 (T42M) e na posição 230 (Q230K) são provavelmente responsáveis pela resistência observada. Um trabalho subsequente (He *et al.* (2003) *Biosci. Biotech. Biochem.* 67:1405-1409) mostra que a mutação T42M (treonina para metionina) é suficiente para melhorar a tolerância

20 de ambas as enzimas de *E. coli* e *Salmonella typhimurium*.

Devido às muitas vantagens que a resistência de plantas a herbicidas fornecem, métodos para identificação de genes resistentes a herbicida com atividade de resistência ao glifosato são desejáveis.

25

SUMÁRIO DA INVENÇÃO

Composições e métodos são fornecidos para conferir tolerância ao glifosato em bactérias, plantas, células de planta, tecidos e sementes. As composições incluem uma nova classe de enzimas EPSPS, designada

30 Classe III, e polinucleotídeos que codificam tais enzimas, vetores que compreendem estes polinucleotídeos e células hospedeiras que compreendem os vetores. As novas proteínas compreendem pelo menos um domínio de

seqüência selecionada a partir dos seguintes domínios (domínios Classe III):

Domínio I:

L-A-K-G-X₁-S-X₂-L-X₃-G-A-L-K-S-D-D-T (SEQ ID NO: 13), onde X₁ denota lisina ou treonina, onde X₂ denota arginina ou histidina, onde X₃ denota serina ou treonina;

Domínio Ia:

L-A-K-G-X₁, (SEQ ID NO: 14), onde X₁ denota lisina ou treonina;

Domínio Ib:

S-X₁-L-X₂ (SEQ ID NO: 15), onde X₁ denota arginina ou histidina, onde X₂ denota serina ou treonina;

Domínio Ic:

G-A-L-K-S-D-D-T (SEQ ID NO: 16);

Domínio II:

E-P-D-X₁-X₂-T-F-X₃-V-X₄-X₅-X₆-G (SEQ ID NO: 17), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina, onde X₄ denota treonina ou ácido glutâmico ou lisina, onde X₅ denota serina ou glicina, onde X₆ denota glutamina ou serina ou ácido glutâmico ou treonina;

Domínio IIa:

E-P-D-X₁-X₂-T-F-X₃-V (SEQ ID NO: 18), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina;

Domínio IIb:

X₁-X₂-X₃-G (SEQ ID NO: 19), onde X₁ denota treonina ou ácido glutâmico ou lisina, onde X₂ denota serina ou glicina, onde X₃ denota glutamina ou serina ou ácido glutâmico ou treonina;

Domínio III:

RFLTAA (SEQ ID NO: 20);

Domínio IV:

KRPI(G/M)P (SEQ ID NO: 21)

K-R-P-I-X₁-P, onde X₁ denota glicina ou metionina ou leucina;

Domínio V:

X₁-G-C-P-P-V (SEQ ID NO: 22), onde X₁ denota treonina ou serina;

Domínio VI:

- 5 I-G-A-X₁-G-Y-X₂-D-L-T (SEQ ID NO: 23), onde X₁ denota arginina ou lisina ou leucina, e onde X₂ denota isoleucina ou valina;

Domínio VII:

W-X₁-V-X₂-X₃-T-G (SEQ ID NO: 24), onde X₁ denota arginina ou lisina, onde X₂ denota alanina ou histidina ou ácido glutâmico ou serina, onde X₃ denota prolina ou alanina;

10 Domínio VIII:

E-P-D-A-S-A-A-T- Y-L-W-X₁-A-X₂-X₃-L (SEQ ID NO: 25), onde X₁ denota alanina ou glicina, onde X₂ denota ácido glutâmico ou glutamina, onde X₃ denota valina ou leucina ou alanina;

Domínio VIIIa:

- 15 E-P-D-A-S-A-A-T- Y-L-W (SEQ ID NO: 26);

Domínio IX:

I-D-X₁-G (SEQ ID NO: 27), onde X₁ denota isoleucina ou leucina;

Domínio X:

F-X₁-Q-P-D-A- K-A (SEQ ID NO: 28), onde X₁ denota treonina ou serina;

20 Domínio XI:

- X₁-F-P-X₂-X₃-X₄-A-X₅-X₆-X₇-G-S-Q-M-Q-D-A-I-P-T-X₈-A-V- X₉ A-A- F-N (SEQ ID NO: 29), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina, onde X₅ denota treonina ou ácido glutâmico ou valina, onde X₆ denota valina ou isoleucina, onde X₇ denota ácido aspártico ou valina, onde X₈ denota leucina ou isoleucina, onde X₉ denota leucina ou isoleucina;
- 25

Domínio XIa:

- X₁-F-P-X₂-X₃-X₄-A (SEQ ID NO: 30), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina;
- 30

Domínio XIb:

X_1 - X_2 - X_3 -G-**S-Q**-M-**Q**-D-A-I-P-T- X_4 -A- V- X_5 A-A- F-**N** (SEQ ID NO: 31), onde X_1 denota treonina ou ácido glutâmico ou valina, onde X_2 denota valina ou isoleucina, onde X_3 denota ácido aspártico ou valina, onde X_4 denota leucina ou isoleucina, onde X_5 denota leucina ou isoleucina;

Domínio XIc:

G-S-Q-M-Q-D-A-I-P-T (SEQ ID NO: 32);

Domínio XII:

P-V-R-F- X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C-D-R**- X_5 (SEQ ID NO: 33), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

Domínio Xlla:

P-V-R-F (SEQ ID NO: 34);

15 Domínio XIIb:

X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C-D-R**- X_5 (SEQ ID NO: 35), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

20 Domínio XIIc:

N-L-R-V-K-E-**C-D-R** (SEQ ID NO: 36);

Domínio XIII:

E-G-D-D-L- X_1 - X_2 (SEQ ID NO: 37), onde X_1 denota leucina ou isoleucina, onde X_2 denota valina ou isoleucina;

25 Domínio XIV:

X_1 -P- X_2 -L-A-G (SEQ ID NO: 38), onde X_1 denota ácido aspártico ou asparagina, onde X_2 denota alanina ou serina ou treonina;

Domínio XV:

A- X_1 -I-D- X_2 - X_3 - X_4 - **D-H-R** (SEQ ID NO: 39), onde X_1 denota leucina ou serina ou ácido glutâmico, onde X_2 denota treonina ou serina, onde X_3 denota histidina ou fenilalanina, onde X_4 denota alanina ou serina;

Domínio XVI:

F-A-L-A-X₁-L-K-X₂-X₃-G-I (SEQ ID NO: 40), onde X₁ denota glicina ou alanina, onde X₂ denota isoleucina ou valina, onde X₃ denota serina ou glicina ou alanina ou lisina;

5 Domínio XVIa:

F-A-L-A-X₁-L-K (SEQ ID NO: 41), onde X₁ denota glicina ou alanina;

Domínio XVIb:

L-K-X₁-X₂-G-I (SEQ ID NO: 42), onde X₁ denota isoleucina ou valina, onde X₂ denota serina ou glicina ou alanina ou lisina; e

10 Domínio XVII:

-X₁-P-X₂-C-V-X₃-K (SEQ ID NO: 43), onde X₁ denota asparagina ou ácido aspártico, onde X₂ denota alanina ou ácido aspártico, onde X₃ denota alanina ou glicina.

Domínio XVIII:

15 X₁-S-L-G-V (SEQ ID NO: 44), onde X₁ denota alanina ou serina ou prolina.

Os domínios descritos acima nas SEQ ID NOS: 13-44 foram identificados pelo alinhamento das seqüências Classe III que compartilham pelo menos 50% de identidade de seqüência. A presença de pelo menos um desses domínios de seqüências é preditiva de atividade de resistência ao

20 glifosato.

São fornecidas moléculas de ácido nucléico isoladas que correspondem a seqüências de ácido nucléico que conferem resistência a um herbicida. Além disso, são abrangidas seqüências de aminoácidos que correspondem aos polinucleotídeos. Em particular, a presente invenção fornece

25 moléculas de ácido nucléico isoladas que compreendem um domínio Classe III, incluindo a seqüência de nucleotídeos descrita em SEQ ID NOS: 9, 11, 55, 57 e 58, uma seqüência de nucleotídeos que codifica a seqüência de aminoácidos mostrada em SEQ ID NOS: 10, 12, 56 e 59, a seqüência de nucleotídeos da resistência a herbicida depositada em um hospedeiro bacteriano como N^os de Acesso B-30833 e B-30838, assim como variantes e

30 fragmentos dessas. Seqüências de nucleotídeos que são complementares a uma seqüência de nucleotídeos da invenção ou que hibridizam a uma se-

qüência da invenção também são abrangidas. As seqüências encontram uso na construção de vetores de expressão para transformação subsequente em plantas de interesse, como sondas para o isolamento de outros genes de resistência ao glifosato, como marcadores selecionáveis e similares.

- 5 As composições também incluem anticorpos para os polipeptídeos assim como polinucleotídeos sintéticos que codificam polipeptídeos de resistência a um herbicida. As seqüências codificantes podem ser usadas em construtos de DNA ou cassetes de expressão para transformação e expressão em organismos, incluindo microrganismos e plantas. As composi-
- 10 ções também compreendem bactérias, plantas, células de plantas, tecidos e sementes transformadas que são tolerantes ao glifosato pela introdução das composições da invenção no genoma do organismo. Onde o organismo é uma planta, a introdução da seqüência permite que herbicidas que contêm glifosato sejam aplicados à cultura para matar seletivamente ervas daninhas
- 15 sensíveis ao glifosato, mas não o organismo transformado.

- Métodos para identificar uma EPSP sintase com atividade de resistência ao glifosato são fornecidos adicionalmente. Os métodos compreendem obter uma seqüência de aminoácidos para uma EPSP sintase e identificar se a seqüência de aminoácidos compreende pelo menos um do-
- 20 mínio de seqüência da invenção.

As EPSP sintases aqui descritas representam uma nova classe de enzimas EPSPS, referidas daqui por diante como enzimas EPSPS Classe III.

DESCRIÇÃO DAS figuras

- 25 A figura 1 mostra um alinhamento de seqüências de aminoácidos de EPSP sintase. Os resíduos conservados que foram identificados como importantes para ligação de substrato e para atividade de EPSPS estão enquadrados. Os quadrados delineiam as localizações dos aminoácidos dos domínios de Classe III. Numerais romanos acima dos quadrados correspondem aos domínios de Classe III.
- 30

DESCRIÇÃO DETALHADA DA INVENÇÃO

A presente invenção é representada por composições e métodos

para conferir tolerância a herbicida, particularmente tolerância a glifosato, em organismos. Os métodos envolvem a transformação de organismos com seqüências de nucleotídeos que codificam um gene de tolerância ao glifosato da Classe III. Em particular, a presente invenção reconhece uma classe de

5 enzimas que confere tolerância ao glifosato e seqüências de nucleotídeos que codificam as tais enzimas. As seqüências encontram uso na preparação de plantas que mostram tolerância aumentada ao herbicida glifosato. Assim, são fornecidas bactérias, plantas, células de planta, tecidos de planta e sementes transformadas.

10 As enzimas Classe III são caracterizadas por terem pelo menos um domínio selecionado a partir dos domínios listados abaixo, referidos aqui como domínios Classe III:

Domínio I:

15 L-A-K-G-X₁-S-X₂-L-X₃-G-A-L-K-S-D-D-T (SEQ ID NO: 13), onde X₁ denota lisina ou treonina, onde X₂ denota arginina ou histidina, onde X₃ denota serina ou treonina;

Domínio Ia:

L-A-K-G-X₁, (SEQ ID NO: 14), onde X₁ denota lisina ou treonina;

Domínio Ib:

20 S-X₁-L-X₂ (SEQ ID NO: 15), onde X₁ denota arginina ou histidina, onde X₂ denota serina ou treonina;

Domínio Ic:

G-A-L-K-S-D-D-T (SEQ ID NO: 16);

Domínio II:

25 E-P-D-X₁-X₂-T-F-X₃-V-X₄-X₅-X₆-G (SEQ ID NO: 17), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina, onde X₄ denota treonina ou ácido glutâmico ou lisina, onde X₅ denota serina ou glicina, onde X₆ denota glutamina ou serina ou ácido glutâmico ou treonina;

30 Domínio IIa:

E-P-D-X₁-X₂-T-F-X₃-V (SEQ ID NO: 18), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleu-

cina;

Domínio IIb:

X_1 - X_2 - X_3 -G (SEQ ID NO: 19), onde X_1 denota treonina ou ácido glutâmico ou lisina, onde X_2 denota serina ou glicina, onde X_3 denota glutamina ou serina
5 ou ácido glutâmico ou treonina;

Domínio III:

RFLTAA (SEQ ID NO: 20);

Domínio IV:

KRPI(G/M)P (SEQ ID NO: 21)

10 K-R-P-I- X_1 -P, onde X_1 denota glicina ou metionina ou leucina;

Domínio V:

X_1 -G-C-P-P-V (SEQ ID NO: 22), onde X_1 denota treonina ou serina;

Domínio VI:

15 I-G-A- X_1 -G-Y- X_2 -D-L-T (SEQ ID NO: 23), onde X_1 denota arginina ou lisina ou leucina, e onde X_2 denota isoleucina ou valina;

Domínio VII:

W- X_1 -V- X_2 - X_3 -T-G (SEQ ID NO: 24), onde X_1 denota arginina ou lisina, onde X_2 denota alanina ou histidina ou ácido glutâmico ou serina, onde X_3 denota prolina ou alanina;

20 Domínio VIII:

E-P-D-A-S-A-A-T- Y-L-W- X_1 -A- X_2 - X_3 -L (SEQ ID NO: 25), onde X_1 denota alanina ou glicina, onde X_2 denota ácido glutâmico ou glutamina, onde X_3 denota valina ou leucina ou alanina;

Domínio VIIIa:

25 E-P-D-A-S-A-A-T- Y-L-W (SEQ ID NO: 26);

Domínio IX:

I-D- X_1 -G (SEQ ID NO: 27), onde X_1 denota isoleucina ou leucina;

Domínio X:

F- X_1 -Q-P-D-A- K-A (SEQ ID NO: 28), onde X_1 denota treonina ou serina;

30 Domínio XI:

X_1 -F-P- X_2 - X_3 - X_4 -A- X_5 - X_6 - X_7 -G-S-Q-M-Q-D-A-I-P-T- X_8 -A-V- X_9 A-A- F-N (SEQ ID NO: 29), onde X_1 denota glutamina ou lisina ou serina, onde X_2 denota

asparagina ou histidina, onde X_3 denota metionina ou leucina, onde X_4 denota prolina ou glutamina, onde X_5 denota treonina ou ácido glutâmico ou valina, onde X_6 denota valina ou isoleucina, onde X_7 denota ácido aspártico ou valina, onde X_8 denota leucina ou isoleucina, onde X_9 denota leucina ou isoleucina;

5 leucina;

Domínio XIa:

X_1 -**F-P**- X_2 - X_3 - X_4 -A (SEQ ID NO: 30), onde X_1 denota glutamina ou lisina ou serina, onde X_2 denota asparagina ou histidina, onde X_3 denota metionina ou leucina, onde X_4 denota prolina ou glutamina;

10 Domínio XIb:

X_1 - X_2 - X_3 -G-**S-Q**-M-**Q**-D-A-I-P-T- X_4 -A- V- X_5 A-A- F-**N** (SEQ ID NO: 31), onde X_1 denota treonina ou ácido glutâmico ou valina, onde X_2 denota valina ou isoleucina, onde X_3 denota ácido aspártico ou valina, onde X_4 denota leucina ou isoleucina, onde X_5 denota leucina ou isoleucina;

15 Domínio XIc:

G-S-Q-M-Q-D-A-I-P-T (SEQ ID NO: 32);

Domínio XII:

P-V-R-F- X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C-D-R**- X_5 (SEQ ID NO: 33), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

20

Domínio XIIa:

P-V-R-F (SEQ ID NO: 34);

Domínio XIIb:

25 X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C-D-R**- X_5 (SEQ ID NO: 35), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

Domínio XIIc:

30 N-L-R-V-K-E-**C-D-R** (SEQ ID NO: 36);

Domínio XIII:

E-G-D-D-L- X_1 - X_2 (SEQ ID NO: 37), onde X_1 denota leucina ou isoleucina,

onde X_2 denota valina ou isoleucina;

Domínio XIV:

X_1 -P- X_2 -L-A-G (SEQ ID NO: 38), onde X_1 denota ácido aspártico ou asparagina, onde X_2 denota alanina ou serina ou treonina;

5 Domínio XV:

A- X_1 -I-D- X_2 - X_3 - X_4 - D-H-R (SEQ ID NO: 39), onde X_1 denota leucina ou serina ou ácido glutâmico, onde X_2 denota treonina ou serina, onde X_3 denota histidina ou fenilalanina, onde X_4 denota alanina ou serina;

Domínio XVI:

- 10 F-A-L-A- X_1 -L-K- X_2 - X_3 -G-I (SEQ ID NO: 40), onde X_1 denota glicina ou alanina, onde X_2 denota isoleucina ou valina, onde X_3 denota serina ou glicina ou alanina ou lisina;

Domínio XVIa:

F-A-L-A- X_1 -L-K (SEQ ID NO: 41), onde X_1 denota glicina ou alanina;

15 Domínio XVIb:

L-K- X_1 - X_2 -G-I (SEQ ID NO: 42), onde X_1 denota isoleucina ou valina, onde X_2 denota serina ou glicina ou alanina ou lisina; e

Domínio XVII:

- 20 - X_1 -P- X_2 -C-V- X_3 -K (SEQ ID NO: 43), onde X_1 denota asparagina ou ácido aspártico, onde X_2 denota alanina ou ácido aspártico, onde X_3 denota alanina ou glicina.

Domínio XVIII:

X_1 -S-L-G-V (SEQ ID NO: 44), onde X_1 denota alanina ou serina ou prolina.

- 25 Os domínios descritos acima nas SEQ ID NOS: 13-44 foram identificados pelo alinhamento de seqüências Classe III que compartilham pelo menos 50% de identidade de seqüência. Em algumas modalidades pelo menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 ou 32 desses domínios de seqüência estão presentes.

- 30 Usando os métodos da invenção e os domínios identificados, proteínas adicionais (por exemplo, SEQ ID NOS: 2, 4, 46 e 48) que conferem tolerância ao glifosato podem ser identificadas. Essas proteínas incluem pro-

teínas conhecidas assim como as proteínas identificadas recentemente (por exemplo, SEQ ID NOS: 10, 12, 56 e 59).

Por “glifosato” entende-se qualquer forma herbicida de N-fosfometilglicina (incluindo qualquer sal desse) e outras formas que resultam na produção do ânion de glifosato em *planta*. Uma “proteína de resistência a herbicida”, “proteína tolerante a herbicida” ou uma proteína que resulta da expressão de um polinucleotídeo que codifica a “resistência a herbicida” ou a “tolerância a herbicida” inclui proteínas que conferem a uma célula a habilidade de tolerar uma concentração de um herbicida maior do que as células que não expressam a proteína ou tolerar uma certa concentração de um herbicida por um período de tempo maior do que as células que não expressam a proteína. Uma “proteína resistente ao glifosato” ou uma “proteína tolerante ao glifosato” inclui uma proteína que confere a uma célula a habilidade de tolerar uma concentração maior de glifosato do que as células que não expressam a proteína ou tolerar uma certa concentração de glifosato por um período de tempo maior do que as células que não expressam a proteína. Por “tolerar” ou “tolerância” entende-se tanto sobreviver, ou realizar funções celulares essenciais tais como síntese de proteína e respiração de uma maneira que não é prontamente discernível de células não tratadas.

20 Polinucleotídeos isolados e Variantes e Fragmentos Desses

Um aspecto da invenção diz respeito a polinucleotídeos isolados diferentes das seqüências de polinucleotídeos listadas em SEQ ID NOS: 1, 3, 7, 45, 47 e 53 que codificam enzimas EPSP sintase que tem pelo menos um domínio de seqüência Classe III da invenção. Por “diferente de” entende-se que a invenção não inclui as seqüências de polinucleotídeos descritas nas SEQ ID NOS citadas.

Os polinucleotídeos isolados da presente invenção compreendem seqüências de nucleotídeos que codificam proteínas e polipeptídeos resistentes a herbicida ou porções biologicamente ativas desses, assim como polinucleotídeos suficientes para uso como sondas de hibridização para identificar polinucleotídeos que codificam resistência a herbicida. Como usado aqui, o termo “polinucleotídeo” pretende incluir moléculas de DNA (por

exemplo, cDNA ou DNA genômico) e moléculas de RNA (por exemplo, mRNA) e análogos do DNA ou RNA gerados usando análogos de nucleotídeos. Os polinucleotídeos podem ser de DNA de filamento simples ou de filamento duplo.

5 Seqüências de nucleotídeos da invenção incluem aquelas caracterizadas pelos domínios incluídos acima. A informação usada na identificação desses domínios inclui alinhamento de seqüências de moléculas de EPSP sensíveis ao glifosato. Os alinhamentos de seqüências foram usados para identificar regiões de homologia entre as seqüências e para identificar
10 domínios Classe III que são característicos de enzimas EPSPS Classe III.

 Variantes dos domínios também são abrangidas dentro do escopo da invenção (por exemplo SEQ ID NO: 55). Tais variantes incluem seqüências que compartilham pelo menos 90%, pelo menos 95%, 96%, 97%, 98% ou 99% de identidade de seqüência e estão contidas em uma molécula
15 de DNA que confere tolerância ao glifosato.

 Um polinucleotídeo ou proteína "isolada" ou "purificada" ou uma porção biologicamente ativa desses, é substancialmente livre de outro material celular ou meio de cultura quando produzido por técnicas recombinantes, ou substancialmente livre de precursores químicos ou outros químicos
20 quando sintetizado quimicamente. Preferivelmente, um polinucleotídeo "isolado" é livre de seqüências (por exemplo, seqüências que codificam proteína) que flanqueiam naturalmente o polinucleotídeo (isto é, seqüências localizadas nas extremidades 5' e 3' do polinucleotídeo) no DNA genômico do organismo a partir do qual o polinucleotídeo é derivado. Para os propósitos
25 da invenção, "isolado" quando usado para se referir a polinucleotídeos exclui cromossomos isolados. Por exemplo, em várias modalidades, o polinucleotídeo isolado que codifica resistência ao glifosato pode conter menos do que cerca de 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb ou 0,1 kb de seqüência de nucleotídeos que flanqueiam naturalmente o polinucleotídeo no DNA genômico
30 da célula a partir da qual o polinucleotídeo é derivado. Uma proteína resistente ao glifosato que é substancialmente livre de material celular inclui preparações de proteína que tem menos do que 30%, 20%, 10% ou 5% (por

peso seco) de proteína de resistência a não-herbicida (também referidas aqui como “proteína contaminante”).

Polinucleotídeos que são fragmentos dessas seqüências de nucleotídeos que codificam resistência a herbicida também são abrangidos pela presente invenção (por exemplo, SEQ ID NOS: 57, 58 e 60). Por “fragmento” entende-se uma porção da seqüência de nucleotídeo que codifica uma proteína resistente a herbicida (por exemplo, SEQ ID NO: 59). Um fragmento de uma seqüência de nucleotídeos pode codificar uma porção biologicamente ativa de uma proteína de resistência a herbicida ou ela pode ser um fragmento que pode ser usado como uma sonda de hibridização ou iniciador de PCR usando os métodos descritos abaixo. Os polinucleotídeos que são fragmentos de uma seqüência de nucleotídeos de resistência a herbicida compreendem pelo menos cerca de 15, 20, 50, 75, 100, 200, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950 nucleotídeos contíguos ou até o número de nucleotídeos presentes em uma seqüência de nucleotídeos de extensão completa que codifica resistência a herbicida descrita aqui. Por nucleotídeos “contíguos” entende-se resíduos de nucleotídeos que são imediatamente adjacentes um ao outro.

Fragmentos das seqüências de nucleotídeos da presente invenção geralmente compreenderão pelo menos um dos domínios Classe III aqui descritos e codificarão fragmentos de proteína que retém a atividade biológica da proteína de resistência ao glifosato de extensão completa; isto é, atividade de resistência a herbicida. Por “retém a atividade de resistência a um herbicida” entende-se que o fragmento terá pelo menos cerca de 30%, pelo menos cerca de 50%, pelo menos cerca de 70% ou pelo menos cerca de 80% da atividade de resistência a um herbicida da proteína de resistência ao glifosato de extensão completa descrita aqui como SEQ ID NO: 6. Métodos para medir a atividade de resistência a um herbicida são bem-conhecidas na técnica. Veja, por exemplo, Patentes U.S. N^os 4.535.060 e 5.188.642, cada uma das quais está incorporada aqui por referência em sua totalidade.

Um fragmento de uma seqüência de nucleotídeos que codifica resistência a um herbicida que codifica uma porção biologicamente ativa de uma proteína da invenção codificará pelo menos 15, 25, 30, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400 aminoácidos contíguos ou até o número total de aminoácidos presentes em uma proteína de resistência ao herbicida de extensão completa da invenção. De forma importante, o fragmento compreenderá pelo menos um dos domínios Classe III descritos aqui.

Proteínas resistentes a herbicida da presente invenção são aquelas caracterizadas como Classe III ou fragmentos ou variantes dessas que retém atividade. O termo "suficientemente idênticas" é pretendido significar uma seqüência de aminoácidos ou nucleotídeos que tem pelo menos cerca de 60% ou 65% de identidade de seqüência, pelo menos cerca de 70% ou 75% de identidade de seqüência, pelo menos cerca de 80% ou 85% de identidade de seqüência ou pelo menos cerca de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% de identidade de seqüência comparada com uma seqüência de referência usando um dos programas de alinhamento aqui descritos usando parâmetros padronizados. Um versado na técnica reconhecerá que esses valores podem ser apropriadamente ajustados para determinar a identidade correspondente de proteínas codificadas por duas seqüências de nucleotídeos levando em consideração a degeneração de códon, similaridade de aminoácidos, posicionamento da fase de leitura e similares.

Para determinar o percentual de identidade de duas seqüências de aminoácidos ou de dois polinucleotídeos, as seqüências são alinhadas com o propósito de comparação ótima. O percentual de identidade entre as duas seqüências é uma função do número de posições idênticas compartilhadas pelas seqüências (isto é, percentual de identidade = número de posições idênticas/número total de posições (por exemplo, posições superpostas) x 100). Em uma modalidade, as duas seqüências têm a mesma extensão. O percentual de identidade entre as duas seqüências pode ser determinado usando técnicas similares àquelas descritas abaixo, com ou sem permitir intervalos. No cálculo de percentual de identidade, tipicamente são con-

tadas combinações exatas.

A determinação do percentual de identidade entre duas seqüências pode ser obtida usando um algoritmo matemático. Um exemplo não limitante de um algoritmo matemático utilizado para a comparação de duas seqüências é o algoritmo de Karlin & Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modificado como em Karlin & Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Tal algoritmo está incorporado nos programas BLASTN BLASTX de Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410. As pesquisas de nucleotídeo BLAST podem ser realizadas com o programa BLASTN, "score" = 100, "wordlength" = 12, para se obter seqüências de nucleotídeos homólogas aos polinucleotídeos semelhantes a GDC da invenção. As pesquisas de proteína BLAST podem ser realizadas com o programa BLASTX, "score" = 50, "wordlength" = 3, para se obter seqüências de aminoácidos homólogas as moléculas de proteína de resistência a herbicida da invenção. Para se obter alinhamentos com intervalos para fins de comparação, Gapped BLAST pode ser utilizado como descrito em Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-2402. Alternativamente, PSI-Blast pode ser usado para realizar uma pesquisa repetida que detecte relações distantes entre as moléculas. Veja Altschul *et al.* (1997) *supra*. Quando se utiliza os programas BLAST, Gapped BLAST e PSI-Blast, os parâmetros padrão dos respectivos programas (por exemplo, BLASTX e BLASTN) podem ser usados. Veja www.ncbi.nlm.nih.gov. Outro exemplo não limitante de um algoritmo matemático utilizado para a comparação de seqüências é o algoritmo ClustalW (Higgins *et al.* (1994) *Nucleic Acid Res.* 22:4673-4680). ClustalW compara seqüências e alinha a totalidade da seqüência dos aminoácidos ou DNA e assim pode fornecer dados sobre a conservação de seqüência da seqüência de aminoácidos inteira. O algoritmo ClustalW é usado em diversos pacotes de programas de análise de DNA/ aminoácidos disponíveis comercialmente, tais como o módulo ALIGNX do *Vector NTI Program Suite* (Invitrogen Corporation, Carlsbad, CA). Após o alinhamento das seqüências de aminoácidos com ClustalW, o percentual de identidade de aminoácidos pode ser obtido. Um exemplo não limitante de um programa útil para análise

dos alinhamentos de ClustalW é GENEDOC®. GENEDOC® (Karl Nicholas) permite a determinação de similaridade e identidade de aminoácido (ou DNA) entre várias proteínas. Outro exemplo não limitante de um algoritmo matemático utilizado para a comparação de seqüências é o algoritmo de Myers e Miller (1988) *CABIOS* 4;11-17. Tal algoritmo está incorporado no programa ALIGN (versão 2.0), que é parte do pacote de programas de alinhamento de seqüência GCG (disponibilizado por Accelrys, Inc., San Diego, CA). Quando se utiliza o programa ALIGN para a comparação de seqüências de aminoácidos, uma tabela de resíduo de peso PAM120, um "gap length" "penalty" de 12 e um gap "penalty" de 4.

A menos que estabelecido de outra maneira, GAP Versão 10, que usa o algoritmo de Needleman e Wunsch (1970) *J. Mol. Biol.* 48(3):443-453, será usado para determinar a identidade ou similaridade usando os seguintes parâmetros: % de identidade e % de similaridade de uma seqüência de nucleotídeos usando "GAP Weight de 50" e "Lenght Weight" de 3, e a matriz de pontuação nwsgapdna.cmp; % de identidade ou % de similaridade para uma seqüência de aminoácidos usando GAP weight de 8 e lenght weight de 2 e o programa de pontuação BLOSUM62. Programas equivalentes também podem ser usados. Por "programa equivalente" é entendido qualquer programa de comparação de seqüência que, para quaisquer duas seqüências em questão, gera um alinhamento que tem combinações de resíduo de nucleotídeo idênticas e um percentual de identidade de seqüência idêntico quando comparado ao alinhamento correspondente gerado pelo GAP Versão 10. A invenção também abrange polinucleotídeos variantes. "Variantes" das seqüências de nucleotídeos que codificam resistência a herbicida incluem aquelas seqüências que codificam a proteína resistente a herbicida descrita aqui mas que diferem conservativamente devido a degeneração do código genético, assim como aquelas que são suficientemente idênticas como discutido acima. Variantes alélicas que ocorrem naturalmente podem ser identificadas com o uso de técnicas de biologia molecular bem-conhecidas, tais como reação em cadeia da polimerase (PCR) e técnicas de

hibridização como delineado abaixo. Sequências de nucleotídeo variantes também incluem aquelas sequências de nucleotídeos derivadas sinteticamente que foram geradas, por exemplo, pelo uso de mutagênese direcionada ao sítio mas que ainda codificam as proteínas de resistência a herbicida descritas na presente invenção como discutido abaixo. As proteínas variantes abrangidas pela presente invenção são biologicamente ativas, isso é, elas retêm a atividade biológica desejada da proteína nativa, isso é, a atividade de resistência ao herbicida. Por "retém a atividade de resistência ao herbicida" entende-se que a variante terá pelo menos cerca de 30%, pelo menos cerca de 50%, pelo menos cerca de 70% ou pelo menos cerca de 80% da atividade de resistência ao herbicida da proteína nativa. Métodos para medir a atividade de resistência ao herbicida são bem-conhecidos na técnica. Veja, por exemplo, Patentes U.S. N^{os} 4.535.060 e 5.188.642, cada uma das quais está incorporada aqui por referência em sua totalidade.

O versado ainda apreciará que as alterações podem ser introduzidas por mutação nas sequências de nucleotídeos da invenção levando dessa maneira a alterações na sequência de aminoácidos das proteínas de resistência ao herbicida codificadas, sem alterar a atividade biológica das proteínas. Assim, polinucleotídeos variantes isolados podem ser criados pela introdução de uma ou mais substituições, adições ou deleções de nucleotídeos na sequência de nucleotídeos correspondente aqui descrita, tal que uma ou mais substituições, adições ou deleções de aminoácidos são introduzidas na proteína codificada. Mutações podem ser introduzidas por técnicas padronizadas, tais como mutagênese direcionada ao sítio e mutagênese mediada por PCR. Tais sequências variantes de nucleotídeos são também abrangidas pela presente invenção.

Por exemplo, substituições conservativas de aminoácidos podem ser feitas em um ou mais resíduos de aminoácidos preditos, não essenciais. Um resíduo de aminoácido "não essencial" é um resíduo que pode ser alterado a partir da sequência selvagem de uma proteína de resistência ao herbicida sem alterar a atividade biológica, enquanto que um resíduo de aminoácido "essencial" é necessário para a atividade biológica. Uma "substi-

tução conservativa de aminoácido” é uma na qual o resíduo de aminoácido é substituído por outro resíduo de aminoácido que tem uma cadeia lateral similar. Famílias de resíduos de aminoácidos que têm cadeias laterais similares foram bem-definidas na técnica. Essas famílias incluem aminoácidos com cadeias laterais básicas (por exemplo, lisina, arginina, histidina), cadeias laterais ácidas (por exemplo, ácido aspártico, ácido glutâmico), cadeias laterais polares não carregadas (p;ex., glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadeias laterais não polares (por exemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptofano), cadeias laterais beta-ramificadas (por exemplo, treonina, valina, isoleucina) e cadeias laterais aromáticas (por exemplo, tirosina, fenilalanina, triptofano, histidina). Substituições de aminoácidos podem ser feitas em regiões não conservadas que retem função. Em geral, tais substituições não deveriam ser feitas para resíduos de aminoácidos conservados ou para resíduos de aminoácidos que residem dentro de um motivo conservado, onde tais resíduos são essenciais para atividade da proteína. Entretanto, um versado na técnica poderá compreender que variantes funcionais podem ter alterações conservadas ou não conservadas menores nos resíduos conservados.

Alternativamente, seqüências variantes de nucleotídeos podem ser feitas pela introdução de mutações aleatoriamente ao longo ou em parte da seqüência codificante, tal como por mutagênese de saturação, e os mutantes resultantes podem ser rastreados quanto a habilidade de conferir atividade de resistência a um herbicida para identificar mutantes que retêm a atividade. Após a mutagênese, a proteína codificada pode ser expressa recombinantemente e a atividade da proteína pode ser determinada usando técnicas de ensaio padronizadas.

Usando métodos tais como PCR, hibridização e similares seqüências de resistência a herbicida correspondentes podem ser identificadas pela procura dos domínios conservados da invenção. Veja, por exemplo, Sambrook & Russell (2001) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) e Innis et al., (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic

Press, St. Louis, MO).

Em um método de hibridização, toda ou parte da seqüência do nucleotídeo de resistência a herbicida ou um domínio pode ser usado para rastrear bibliotecas de cDNA ou genômicas. Métodos para construção de tais bibliotecas de cDNA e genômicas são comumente conhecidas na técnica e estão descritos em Sambrook & Russell, 2001, *supra*. As assim chamadas sondas de hibridização podem ser fragmentos de DNA genômico, fragmentos de cDNA, fragmentos de RNA ou outros oligonucleotídeos e podem ser marcadas com um grupo detectável tal como ^{32}P ou qualquer outro marcador detectável, tais como radioisótopos, um composto fluorescente, uma enzima ou um co-fator de enzima. Sondas para hibridização podem ser feitas pela marcação de oligonucleotídeos sintéticos com base na seqüência de nucleotídeo que codifica resistência ao herbicida conhecida aqui descrita. Iniciadores degenerados desenhados com base nos nucleotídeos conservados ou resíduos de aminoácido na seqüência de nucleotídeos ou seqüência de aminoácidos codificada podem ser adicionalmente usados. A sonda compreende tipicamente uma região de seqüência de nucleotídeos que se hibridiza sob condições estridentes a pelo menos cerca de 12, a pelo menos cerca de 25 ou pelo menos cerca de 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800 nucleotídeos consecutivos de uma seqüência de nucleotídeos que codifica resistência a herbicida da invenção ou um fragmento ou variante dessa. Métodos para o preparo de sondas para hibridização são geralmente conhecidos na técnica e estão descritos em Sambrook & Russell, 2001, *supra* e Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), ambos os quais estão incorporados aqui por referência.

A hibridização de tais seqüências pode ser realizada sob condições estridentes. Por “condições estridentes” ou “condições estridentes de hibridização” entende-se condições sob as quais uma sonda hibridizará a sua seqüência alvo em um grau detectavelmente maior do que a outras seqüências (por exemplo, pelo menos 2 vezes sobre o basal). Condições es-

tringentes são dependentes da seqüência e serão diferentes em circunstâncias diferentes. Pelo controle da estringência da hibridização e/ou das condições de lavagem, as seqüências alvo que são 100% complementares a sonda podem ser identificadas (sondagem homóloga). Alternativamente, condições de estringência podem ser ajustadas para permitir alguns pareamentos errados nas seqüências tal que graus menores de similaridade sejam detectados (sondagem heteróloga). Geralmente, uma sonda é menor do que cerca de 1000 nucleotídeos de extensão ou menor do que 500 nucleotídeos em extensão.

10 Tipicamente, condições estringentes serão aquelas nas quais a concentração de sal é menor do que cerca de 1,5 M de íon Na, tipicamente cerca de 0,01 a 1,0 M de concentração de íon Na (ou outros sais) em pH 7.0 a 8.3 e a temperatura é de pelo menos cerca de 30°C para sondas curtas (por exemplo, 10 a 50 nucleotídeos) e pelo menos cerca de 60°C para sondas longas (por exemplo, maior do que 50 nucleotídeos). Condições estringentes também podem ser obtidas com a adição de agentes desestabilizantes tais como formamida. Condições de baixa estringência exemplares incluem hibridização com uma solução tampão de formamida 30 a 35% , NaCl 1 M, SDS (dodecil sulfato de sódio) 1% a 37°C e uma lavagem em SSC 1X a 2X (SSC 20X = NaCl 3,0 M / citrato de trisódio 0,3 M) a 50 a 55°C. Condições de estringência moderada exemplares incluem hibridização em formamida 40 a 45%, NaCl 1,0 M, SDS 1% a 37°C, e uma lavagem em SSC 0,5X a 1X em 55 a 60°C. Condições de alta estringência exemplares incluem hibridização em formamida 50%, NaCl 1 M, SDS 1% a 37°C, e uma lavagem em SSC 0,1X em 60 a 65°C. Opcionalmente, os tampões de lavagem podem compreender cerca de 0,1% a cerca de 1% de SDS. A duração da hibridização geralmente é menor do que 24 horas, comumente cerca de 4 a cerca de 12 horas.

30 Especificidade é tipicamente a função das lavagens pós-hibridização, os fatores críticos sendo a força iônica e a temperatura da solução final de lavagem. Para híbridos de DNA-DNA, a T_m pode ser aproximada a partir da equação de Meinkoth e Wahl (1984) *Anal. Biochem.*

138:267-284: $T_m = 81,5^{\circ}\text{C} + 16,6 (\log M) + 0,41 (\%GC) - 0,61 (\% \text{ form}) - 500/L$; onde M é a molaridade de cátions monovalentes, %GC é a porcentagem de nucleotídeos guanosina e citosina no DNA, % form é o percentual de formamida na solução de hibridização e L é a extensão do híbrido em pares de base. A T_m é a temperatura (sob força iônica e pH definidos) na qual 50% de uma sequência alvo complementar hibridiza com uma sonda perfeitamente pareada. A T_m é reduzida em cerca de 1°C para cada 1% de pareamento errado; assim, T_m , condições de hibridização e/ou lavagem podem ser ajustadas para hibridizar sequências da identidade desejada. Por exemplo, se

5 sequências com $\geq 90\%$ de identidade são buscadas, o T_m pode ser diminuído em 10°C . Geralmente, condições estridentes são selecionadas para serem cerca de 5°C menores do que o ponto de fusão térmica (T_m) para a sequência específica e seu complemento em uma força iônica e pH definidos. Entretanto, condições severamente estridentes podem utilizar uma hibridização e/ou lavagem 1, 2, 3 ou 4°C menores do que o ponto de fusão térmica (T_m); condições moderadamente estridentes podem utilizar uma hibridização e/ou lavagem a 6, 7, 8, 9 ou 10°C menores do que o ponto de fusão térmica (T_m); condições de baixa estringência podem utilizar uma hibridização e/ou lavagem 11, 12, 13, 14, 15 ou 20°C menores do que o ponto de

10 fusão térmica (T_m). Usando a equação, composições de hibridização e lavagem e o T_m desejado aqueles versados na técnica entenderão que variações na estringência de hibridização e/ou soluções de lavagem estão inerentemente descritas. Se o grau desejado de pareamentos errados resulta em um T_m de menos do que 45°C (solução aquosa) ou 32°C (solução de formamida), é preferido aumentar a concentração de SSC tal que uma temperatura

15 mais elevada possa ser usada. Um guia extenso para a hibridização de polinucleotídeos é encontrado em Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid Probes*, Parte I, Capítulo 2 (Elsevier, Nova Iorque); e Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Capítulo 2 (Greene Publishing & Wiley-Interscience, Nova Iorque). Veja Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold S-

20

25

30

pring Harbor, NY).

Proteínas Isoladas e Variantes e Fragmentos Dessas

Proteínas de resistência a herbicida também são abrangidas dentro da presente invenção. Por “proteína de resistência a herbicida” entende-se uma proteína Classe III que tem pelo menos um dos domínios descritos acima, incluindo, por exemplo, SEQ ID NOS: 10, 12, 56 e 59. Fragmentos, porções biologicamente ativas e variantes dessas também são fornecidos e podem ser usados para praticar os métodos da presente invenção.

“Fragmentos” ou “porções biologicamente ativas” incluem fragmentos de polipeptídeo que compreende uma porção de uma sequência de aminoácidos que codifica uma proteína de resistência a herbicida e que retém a atividade de resistência ao herbicida. Uma porção biologicamente ativa de uma proteína de resistência a herbicida pode ser um polipeptídeo que tem, por exemplo, 10, 25, 50, 100 ou mais aminoácidos de extensão. Tais porções biologicamente ativas podem ser preparadas por técnicas recombinantes e avaliadas quanto a resistência ao herbicida. Métodos para medir a atividade de resistência ao herbicida são bem-conhecidos na técnica. Veja, por exemplo, Patentes U.S. N^{os} 4.535.060 e 5.188.642, cada uma das quais está incorporada aqui por referência em sua totalidade.

Por “variantes” entende-se proteínas ou polipeptídeos que tem uma sequência de aminoácidos que é pelo menos cerca de 60%, 65%, cerca de 70%, 75%, cerca de 80%, 85%, ou pelo menos cerca de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% idêntica a uma enzima Classe III. Variantes também incluem polipeptídeos codificados por um polinucleotídeo que hibridiza ao polinucleotídeo que codifica uma enzima Classe III ou um complemento dessa, sob condições estringentes. Variantes incluem polipeptídeos que diferem na sequência de aminoácidos devido a mutagênese. Proteínas variantes abrangidas pela presente invenção são biologicamente ativas, ou seja continuam a possuir a atividade biológica desejada da proteína nativa que é reter a atividade de resistência ao herbicida. Métodos para medir a atividade de resistência ao herbicida são bem-conhecidos na técnica. Veja, por exemplo, Patentes U.S. N^{os} 4.535.060 e 5.188.642, cada uma das

quais está incorporada aqui por referência em sua totalidade.

Genes de bactérias quase sempre possuem múltiplos códons de iniciação de metionina na proximidade do início da fase aberta de leitura. Muitas vezes a iniciação da tradução em um ou mais desses códons de partida levarão a geração de uma proteína funcional. Esses códons de partida podem incluir códons ATG. Entretanto, bactérias tais como *Bacillus sp.* também reconhecem o códon GTG como um códon de partida e proteínas que iniciam a tradução em códons GTG contêm uma metionina no primeiro aminoácido. Além disso, muitas vezes não é determinado *a priori* quais desses

5 10 15

códons são usados naturalmente na bactéria. Dessa forma, é compreendido que o uso de um dos códons de metionina alternados pode levar a geração de variantes que conferem resistência ao herbicida (por exemplo, SEQ ID NO: 59 codificada pela SEQ ID NO: 58). Essas proteínas de resistência a herbicida são abrangidas na presente invenção e podem ser usadas nos métodos da presente invenção.

Anticorpos para os polipeptídeos da presente invenção ou para variantes ou fragmentos desses também são abrangidos. Métodos para produzir anticorpos são bem-conhecidos na técnica (veja, por exemplo, Harlow & Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; Patente U.S. Nº 4.196.265).

20

Transformação de Células Bacterianas e de Plantas

A transformação de células bacterianas pode ser obtida por uma das várias técnicas conhecidas na técnica, não limitadas a eletroporação ou transformação química (Veja, por exemplo, Ausubel (Ed.) (1994) *Current*

25 30

Protocols in Molecular Biology, John Wiley and Sons, Inc., Indianápolis, IN). Marcadores que conferem resistência a substâncias tóxicas são úteis na identificação de células transformadas (que incorporaram e expressaram o DNA de teste) das células não transformadas (aquelas que não contêm ou não expressam o DNA de teste). Em um aspecto da invenção, genes são úteis como um marcador para identificar a transformação de células bacterianas ou de planta.

A transformação de células de planta pode ser obtida de uma

maneira similar. Por "planta" entende-se plantas completas, órgãos de plantas (por exemplo, folhas, caules, raízes, etc.), sementes, células de planta, propágulos, embriões e progênie das mesmas. Células de planta podem ser diferenciadas ou indiferenciadas (por exemplo, calos, suspensões de cultura de células, protoplastos, células de folha, células da raiz, células de floema, pólen). "Plantas transgênicas" ou "plantas transformadas" ou plantas ou células ou tecidos "estavelmente transformados" se referem a plantas que incorporaram ou integraram seqüências de polinucleotídeos exógenos ou fragmentos de DNA na célula da planta. Por "transformação estável" é pretendido que o construto de nucleotídeo introduzido em uma planta se integre no genoma da planta e seja capaz de ser herdado pela progênie dessa.

Os genes da invenção podem ser modificados para se obter ou aumentar a expressão em células de plantas. As seqüências de resistência ao herbicida da invenção podem ser fornecidas em cassetes de expressão para expressão na planta de interesse. "Cassete de expressão em planta" inclui construtos de DNA que são capazes de resultar na expressão de uma proteína a partir de uma fase aberta de leitura em uma célula de planta. O cassete incluirá na direção de transcrição 5'-3', uma região de iniciação da transcrição (isto é, um promotor) operativamente ligado a uma seqüência de DNA da invenção e uma região de terminação da transcrição ou da tradução (isto é, região de terminação) funcional em plantas. Em algumas modalidades, a região de iniciação da transcrição causará a produção de uma seqüência de RNA que permite expressão suficiente da enzima EPSPS codificada para aumentar a tolerância ao glifosato de uma célula de planta transformada com o polinucleotídeo. Por "expressão suficiente" é pretendido que a região de iniciação da transcrição fornecerá a expressão de uma quantidade de polipeptídeo resistente ao glifosato da invenção (por exemplo, aqueles que contêm um domínio Classe III) que conferirá a planta ou célula a habilidade de tolerar uma concentração de glifosato maior do que plantas ou células que não contêm ou não expressam a proteína, ou tolerar uma certa concentração de glifosato por um tempo maior do que plantas ou células que não contêm ou não expressam a proteína.

O cassete pode, além disso, conter pelo menos um gene adicional a ser co-transformado no organismo, tal como um gene marcador selecionável. Alternativamente, o(s) gene(s) adicional(is) pode(m) ser fornecido(s) em cassetes de expressão múltiplos. Tal cassete de expressão é fornecido com uma pluralidade de sítios de restrição para inserção da sequência de resistência ao herbicida para estar sob a regulação transcricional das regiões regulatórias.

O promotor pode ser nativo ou análogo, ou estranho ou heterólogo, a planta hospedeira e/ou a sequência de DNA da invenção. Além disso, o promotor pode ser a sequência natural ou alternativamente uma sequência sintética. Onde o promotor é "nativo" ou "homólogo" à planta hospedeira, é pretendido que o promotor seja encontrado na planta nativa na qual o promotor é introduzido. Onde o promotor é "estranho" ou "heterólogo" à sequência de DNA da invenção, é pretendido que o promotor não seja o promotor nativo ou de ocorrência natural para a sequência de DNA operativamente ligada da invenção. "Heterólogo" geralmente se refere a sequências de polinucleotídeos que não são endógenas da célula ou parte do genoma nativo no qual eles estão presentes, e que foram adicionados à célula por infecção, transfecção, microinjeção, eletroporação, microprojeção ou semelhantes. Por "operativamente ligado" entende-se uma ligação funcional entre um promotor e uma segunda sequência, em que a sequência promotora inicia e media a transcrição da sequência de DNA que corresponde a segunda sequência. Geralmente, operativamente ligado significa que as sequências de polinucleotídeos a serem ligadas são contíguas e, onde necessário juntar duas regiões codificantes de proteína, contíguas e na mesma fase de leitura.

Freqüentemente, tais construtos também conterão regiões 5' e 3' não traduzidas. Tais construtos podem conter uma "sequência de sinal" ou uma "sequência líder" para facilitar o transporte co-traducional ou pós-traducional do peptídeo de interesse para certas estruturas intracelulares de interesse tais como cloroplastos (ou outro plastídeo), retículo endoplasmático ou aparelho de Golgi ou para ser secretado. Por exemplo, o gene pode ser manipulado para conter um peptídeo de sinal para facilitar a transferên-

cia do peptídeo para o retículo endoplasmático. Por “seqüência de sinal” entende-se uma seqüência que é conhecida ou suspeita de resultar em transporte de peptídeo co-traducional ou pós-traducional através da membrana celular. Em eucariotos, isso envolve tipicamente a secreção no aparelho de Golgi, com alguma glicosilação resultante. Por “seqüência líder” entende-se qualquer seqüência que quando traduzida, resulta em uma seqüência de aminoácidos suficiente para desencadear o transporte co-traducional da cadeia de peptídeo para uma organela subcelular. Assim, isso inclui seqüências líderes que desencadeiam transporte e/ou glicosilação pela passagem no retículo endoplasmático, passagem para vacúolos, plastídeos incluindo cloroplastos, mitocôndria e similares. O cassete de expressão em planta também pode ser manipulado para conter um íntron, tal que mRNA que processando do íntron seja necessário para expressão.

Por “região 3’ não traduzida” entende-se uma seqüência de nucleotídeo localizada a jusante de uma seqüência codificante. Seqüências de sinal de poliadenilação (por exemplo, nucleotídeos poliadenilados) e outras seqüências que codificam sinais regulatórios capazes de afetar a adição de tratos de ácido poliadenílico à extremidade 3’ do mRNA precursor são regiões 3’ não traduzidas. Por “região 5’ não traduzida” entende-se uma seqüência de nucleotídeo localizada a montante de uma seqüência codificante.

Outros elementos não traduzidos a montante e a jusante incluem intensificadores. Intensificadores são seqüências de nucleotídeos que agem para aumentar a expressão de uma região promotora. Intensificadores são bem-conhecidos na técnica e incluem, mas não são limitados a região intensificadora de SV40 e o elemento intensificador 35S.

A região de terminação pode ser nativa com a região de iniciação da tradução, pode ser nativa com a seqüência de resistência ao herbicida da presente invenção ou pode ser derivada de outra fonte. Regiões de terminação convenientes estão disponíveis no plasmídeo Ti de *A. tumefaciens*, tais como as regiões de terminação da octopina sintase e nopalina sintase. Veja também Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-

149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; e Joshi *et al.* (1987) *Nucleic Acid Res.* 15:9627-9639.

Onde apropriado, o(s) gene(s) pode(m) ser otimizado(s) para
 5 aumentar a expressão na célula hospedeira transformada. Ou seja, os genes
 podem ser sintetizados usando os códons preferidos da célula hospedeira
 para expressão melhorada, ou podem ser sintetizados usando códons em
 uma frequência de uso de códons preferido do hospedeiro. Geralmente o
 conteúdo de GC do gene estará aumentado. Veja, por exemplo, Campbell e
 10 Gowri (1990) *Plant Physiol.* 92:1-11 para uma discussão sobre o uso de có-
 dons preferido do hospedeiro. Métodos são conhecidos na técnica para sin-
 tetizar os genes preferidos do hospedeiro. Veja, por exemplo, Patentes U.S.
 N°s 6.320.100; 6.075.185; 5.380.831; e 5.436.391. Pedido Publicado U.S.
 N°s 20040005600 e 20010003849, e Murray *et al.* (1989) *Nucleic Acids Res.*
 15 17:477-498, aqui incorporados por referência.

Em uma modalidade, os polinucleotídeos de interesse são dirigi-
 dos para o cloroplasto para expressão. Dessa maneira, onde o polinucleotí-
 deo de interesse não está diretamente inserido no cloroplasto, o cassete de
 expressão conterá adicionalmente um polinucleotídeo que codifica um peptí-
 20 deo de trânsito para direcionar o produto de gene de interesse para os clo-
 roplastos. Tais peptídeos de transito são conhecidos na técnica. Veja, por
 exemplo, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep.* 9:104-126; Clark *et al.*
(1989) J. Biol. Chem. 264:17544-17550; Della-Cioppa *et al.* (1987) *Plant*
Physiol. 84:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res. Commun.*
 25 196:1414-1421; e Shah *et al.* (1986) *Science* 233:478-481. Em algumas mo-
 dalidades, os polinucleotídeos da invenção codificam um polipeptídeo de
 fusão que compreende um peptídeo de trânsito de cloroplasto na terminação
 amino e a enzima EPSPS. Um "polipeptídeo de fusão" pode ser gerado, por
 exemplo, pela remoção do códon de parada da sequência de polinucleotídeo
 30 que codifica um primeiro polipeptídeo, então acrescentar a sequência de
 polinucleotídeo que codifica um segundo polipeptídeo *in frame* tal que a se-
 quência de polinucleotídeo resultante será então expressa como um polipep-

tídeo único por uma célula.

Os polinucleotídeos de interesse a serem direcionados para o cloroplasto podem ser otimizados para expressão no cloroplasto para serem responsáveis pelas diferenças no uso de códon entre o núcleo da planta e
5 essa organela. Dessa maneira, os polinucleotídeos de interesse podem ser sintetizados usando os códons preferidos de cloroplasto. Veja, por exemplo, Patente U.S. Nº 5.380.831, incorporada aqui por referência.

Tipicamente esse “cassete de expressão em planta” será inserido em um “vetor de transformação de planta”. Por “vetor de transformação”
10 entende-se uma molécula de DNA que é necessária para a transformação eficiente de uma célula. Tal molécula pode consistir de um ou mais cassetes de expressão e pode estar organizada em mais de um “vetor” de molécula de DNA. Por exemplo, vetores binários são vetores de transformação de planta que utilizam dois vetores de DNA não contíguos para codificar todas
15 as funções atuantes cis- e trans- necessárias para transformação de células de planta (Hellens e Mullineaux (2000) *Trends in Plant Science* 5:446-451). “Vetor” se refere a um construto de polinucleotídeo designado para transferência entre células hospedeiras diferentes. “Vetor de expressão” se refere a um vetor que tem a habilidade de incorporar, integrar e expressar seqüên-
20 cias ou fragmentos de DNA heterólogos em uma célula estranha.

Esse vetor de transformação de planta pode ser compreendido de um ou mais vetores de DNA necessários para se obter a transformação da planta. Por exemplo, é uma prática comum na técnica utilizar vetores de transformação de planta que são constituídos por mais do que um segmento
25 de DNA contíguo. Esses vetores são geralmente referidos na técnica como “vetores binários”. Vetores binários assim como vetores com plasmídeos auxiliares são os mais freqüentemente usados para transformação mediada por *Agrobacterium*, onde o tamanho e a complexidade dos segmentos de DNA necessários para se obter a transformação eficiente são bastante grandes e é vantajoso separar funções em moléculas de DNA separadas. Veto-
30 res binários contêm tipicamente um plasmídeo vetor que contém as seqüências cis-atuantes necessárias para transferência de T-DNA (tal como a borda

esquerda e a borda direita), um marcador selecionável que é manipulado para ser capaz de expressão em uma célula de planta e um “gene de interesse” (um gene manipulado capaz de expressão em uma célula de planta para a qual a geração de plantas transgênicas é desejada). Também presentes nesse vetor de plasmídeo estão as seqüências necessárias para replicação bacteriana. As seqüências cis-atuantes são arranjadas de uma forma a permitir a transferência eficiente para células de planta e expressão nessas. Por exemplo, o gene marcador selecionável e o gene de interesse estão localizados entre as bordas esquerda e direita. Frequentemente um segundo vetor de plasmídeo contém os fatores trans-atuantes que mediam a transferência de T-DNA de *Agrobacterium* para células de planta. Esse plasmídeo frequentemente contém as funções de virulência (genes Vir) que permitem a infecção de células de planta pela *Agrobacterium* e transferência de DNA por clivagem nas seqüências de fronteira e transferência de DNA mediada por vir, como é compreendido na técnica (Hellens e Mullineaux (2000) *Trends in Plant Science* 5:446-451). Vários tipos de cepas de *Agrobacterium* (por exemplo, LBA 4404, GV3101, EHA101, EHA105, etc.) podem ser usadas para transformação de planta. O segundo vetor de plasmídeo não é necessário para a transformação de plantas por outros métodos tais como microprojeção, microinjeção, eletroporação, polietileno glicol, etc.

Transformação de planta

Métodos da invenção envolvem a introdução de um construto de nucleotídeo em uma planta. Por “introdução” é pretendido dar a planta o construto de nucleotídeo de tal maneira que o construto ganhe acesso ao interior de uma célula da planta. Os métodos da invenção não requerem que um método em particular para introdução de um construto de nucleotídeo em uma planta seja usado, apenas que o construto de nucleotídeo ganhe acesso ao interior de pelo menos uma célula da planta. Métodos para introdução de construtos de nucleotídeo em plantas são conhecidos na técnica incluindo, mas não limitado a métodos de transformação estável, métodos de transformação transitória e métodos mediados por vírus.

Em geral, métodos de transformação de planta envolvem a

transferência de DNA heterólogo para células alvo de plantas (por exemplo, embriões imaturos ou maduros, culturas em suspensão, calos indiferenciados, protoplastos, etc.), seguida pela aplicação de um nível limiar máximo de seleção apropriada (dependendo do gene marcador selecionável e nesse caso de "glifosato") para recuperar as células transformadas da planta a partir de um grupo de massa de células não transformadas. Em tais processos, polipeptídeos resistentes a glifosato que compreendem um ou mais domínios Classe III da presente invenção podem ser usados como marcador selecionável.

10 Explantos são transferidos tipicamente para um suprimento fresco do mesmo meio e cultivados rotineiramente. Subseqüentemente, as células transformadas são diferenciadas em brotos após colocadas em meio de regeneração suplementado com um nível limiar máximo do agente de seleção (por exemplo, "glifosato"). Os brotos são então transferidos para um
15 meio seletivo de enraizamento para recuperação de broto enraizado ou plântula. A plântula transgênica se desenvolve na planta madura e produz sementes férteis (por exemplo, Hiei *et al.* (1994) *The Plant Journal* 6:271-282; Ishida *et al.* (1996) *Nature Biotechnology* 14:745-750). Explantes são tipicamente transferidos para um suprimento fresco do mesmo meio e cultivados
20 rotineiramente. Uma descrição geral das técnicas e métodos para a geração de plantas transgênicas é encontrada em Ayres & Park (1994) *Critical Reviews in Plant Science* 13:219-239 e Bommineni e Jauhar (1997) *Maydica* 42:107-120. Já que o material transformado contém muitas células, tanto as células transformadas como as não transformadas estão presentes em qual-
25 quer pedaço do calo alvo ou tecido ou grupo de células em estudo. A habilidade de matar células não transformadas e permitir que células transformadas proliferem resulta em culturas de plantas transformadas. Frequentemente, a habilidade de remover células não transformadas é uma limitação à recuperação rápida de células transformadas de planta e a geração bem-sucedida de plantas transgênicas. Então, métodos moleculares e bioquímicos
30 serão usados para confirmar a presença do gene heterólogo integrado de interesse no genoma da planta transgênica.

A geração de plantas transgênicas pode ser realizada por um de vários métodos, incluindo mas não limitado a introdução de DNA heterólogo por *Agrobacterium* em células de planta (transformação mediada por *Agrobacterium*), bombardeio de células de planta com DNA estranho heterólogo aderido a partículas e vários outros métodos diretos não mediados por partículas (por exemplo, Hiei *et al.* (1994) *The Plant Journal* 6:271-282; Ishida *et al.* (1996) *Nature Biotechnology* 14:745-750; Ayres and Park (1994) *Critical Reviews in Plant Science* 13:219-239; Bommineni & Jauhar (1997) *Maydica* 42:107-120) para transferir DNA.

Métodos para transformação de cloroplastos são conhecidos na técnica. Veja, por exemplo, Svab *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530; Svab & Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917; Svab & Maliga (1993) *EMBO J.* 12:601-606. O método conta com a liberação de partículas de DNA que contém um marcador selecionável e que atinge o DNA do genoma do plastídeo através de recombinação homóloga. Além disso, a transformação do plastídeo pode ser obtida por transativação de um transgene silencioso gerado no plastídeo por expressão no tecido preferido de uma RNA polimerase codificada no núcleo e direcionada ao plastídeo. Tal sistema foi relatado em McBride *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:7301-7305.

As células que foram transformadas podem ser cultivadas em plantas de acordo com meios convencionais. Veja, por exemplo, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. Essas plantas podem então ser cultivadas e polinizadas tanto com a mesma cepa transformada quanto com cepas diferentes e o híbrido resultante que tem expressão constitutiva da característica fenotípica desejada, identificado. Duas ou mais gerações podem ser cultivadas para assegurar que a expressão da característica fenotípica desejada está estavelmente mantida e herdada e então as sementes colhidas para assegurar que a expressão da característica fenotípica desejada tenha sido obtida. Dessa maneira a presente invenção fornece semente transformada (também referida como "semente transgênica") que tem um construto de nucleotídeo da invenção, por exemplo, um cassete de expres-

são da invenção, estavelmente incorporado em seu genoma.

Plantas

A presente invenção pode ser usada para a transformação de qualquer espécie de planta, incluindo mas não limitado a monocotiledôneas e dicotiledôneas. Exemplos de plantas de interesse incluem, mas não são limitados a milho (milho graúdo), sorgo, trigo, girassol, tomate, crucíferas, pimentas, batata, algodão, arroz, soja, beterraba, cana de açúcar, tabaco, cevada, canola, *Brassica sp.*, alfafa, centeio, painço, açafrão, amendoim, batata doce, mandioca, café, coco, abacaxi, árvores cítricas, cacau, chá, banana, abacate, figo, goiaba, manga, azeitona, papaia, caju, macadâmia, amêndoa, aveia, vegetais, ornamentais e coníferas.

Vegetais incluem, mas não são limitados a tomates, alface, feijões verdes, feijão de lima, ervilhas e membros do gênero *Curcumis* tais como pepino, meloa e melão. As ornamentais incluem, mas não são limitadas a azaléia, hidrângea, hibisco, rosas, tulipas, narcisos, petúnias, cravo, poinsetia e crisântemo. Preferivelmente, as plantas da presente invenção são plantas de cultura (por exemplo, milho, sorgo, trigo, girassol, tomate, crucíferas, pimentas, batata, algodão, arroz, soja, beterraba, cana de açúcar, tabaco, cevada, canola, etc.).

Essa invenção é particularmente adequada para qualquer membro da família das plantas monocotiledôneas, incluindo mas não limitado a milho, arroz, cevada, aveia, trigo, sorgo, centeio, cana de açúcar, abacaxi, inhame, cebola, banana, coco e tâmaras.

Avaliação da Transformação de Planta

Após a introdução do DNA estranho heterólogo em células de planta, a transformação ou integração do(s) gene(s) heterólogo(s) no genoma da planta é confirmada por vários métodos tais como análise de polinucleotídeos, proteínas e metabólitos associados com o gene integrado.

A análise por PCR é um método rápido para rastrear células transformadas, tecidos ou brotos quanto a presença do(s) gene(s) incorporado(s) no estágio precoce antes do transplante para o solo (Sambrook & Russell (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, NY). PCR é realizado usando iniciadores de oligonucleotídeo específico para o gene de interesse ou vetor de *Agrobacterium*, etc.

5 A transformação da planta pode ser confirmada pela análise Southern blot de DNA genômico (Sambrook & Russell, 2001, *supra*). Em geral, o DNA total é extraído do transformante, digerido com as enzimas de restrição apropriadas, fracionado em um gel de agarose e transferido para membrana de nitrocelulose ou náilon. A membrana ou "blot" é então sondada com, por exemplo, fragmento de DNA alvo radiomarcado com ^{32}P para
10 confirmar a integração do gene introduzido no genoma da planta de acordo com técnicas padronizadas (Sambrook & Russell, 2001, *supra*).

Na análise Northern, o RNA é isolado de tecidos específicos do transformante, fracionado em gel de agarose formaldeído, e transferido para um filtro de náilon de acordo com procedimentos padronizados que são rotineiramente usados na técnica (Sambrook & Russell, 2001, *supra*). A expressão do RNA codificado pela seqüência da invenção é então testada pela hibridização do filtro com uma sonda radioativa derivada de um GDC, por métodos conhecidos na técnica (Sambrook e Russell, 2001, *supra*).
15

Western blot e ensaios bioquímicos e similares podem ser realizados nas plantas transgênicas para determinar a presença de proteína codificada pelo gene resistente ao herbicida por procedimentos padronizados (Sambrook e Russell, 2001, *supra*) usando anticorpos que se ligam a um ou mais epitopos presentes na proteína resistente ao herbicida.
20

A presente invenção também é direcionada a métodos para identificar uma EPSP sintase com atividade de resistência ao glifosato. Os métodos envolvem determinar se certos domínios de seqüência conservados estão presentes na seqüência de aminoácidos de uma EPSP sintase. A presença de um ou mais domínios listados acima.
25

Predição da Função da Proteína a partir da Seqüência

30 Usando os métodos da invenção e os domínios identificados, polipeptídeos adicionais (por exemplo, SEQ ID NOS: 2, 4, 46 e 48) que conferem tolerância ao glifosato podem ser identificados. Esses polipeptídeos

adicionais podem ser identificados pela pesquisa de bases de dados de sequência que contém seqüências de EPSP sintase e/ou por alinhamento de seqüências de polipeptídeos de EPSP sintase para pesquisar a presença de domínios Classe III. Esses polipeptídeos incluem polipeptídeos conhecidos assim como os novos polipeptídeos identificados. É entendido que alguma modificação desses domínios é tolerada na natureza sem perda da natureza que confere resistência ao glifosato desses domínios, e são portanto equivalentes aos domínios listados aqui.

Em geral, existem quatro níveis de estrutura de proteína: a estrutura primária, que consiste da cadeia linear de aminoácidos ou da seqüência de polipeptídeos; a estrutura secundária, que é dada pelas hélices α , filamentos β e voltas nas quais a proteína se dobra; a estrutura terciária, que é feita de motivos simples que se combinaram para formar domínios globulares compactos; e a estrutura quaternária, que pode compreender várias cadeias de aminoácidos ou subunidades. Quando se prediz a função a partir da seqüência, é importante identificar os motivos ou padrões funcionalmente importantes. Domínios de proteína com dobramentos similares frequentemente partilham a mesma função molecular (Hegyi & Gerstein (1999) *J. Mol. Biol.* 288:147-164; Moulton & Melamud (2000) *Curr. Opin. Struct. Biol.* 10:384-389; Shakhnovich *et al.* (2003) *J. Mol. Biol.* 326:1-9). A identificação de domínios importantes para a função da proteína pode ser feita por alinhamentos múltiplos de seqüência.

A estrutura tridimensional pode ser predita por modelagem por homologia, isto é, pelo uso de uma seqüência homóloga (> 25% de identidade de seqüência) com uma estrutura 3D experimentalmente determinada. A estrutura tridimensional de, por exemplo, EPSP sintase de *E. coli* (AroA) é bem-conhecida (Shönbrunn *et al.* (2001) *Proc. Natl. Acad. Sci. USA* 98:1375-1380). Essa estrutura é baseada na cristalização de AroA com glifosato e xiquimato 3-fosfato.

Os seguintes exemplos são oferecidos por meio de ilustração e não por meio de limitação.

EXPERIMENTAÇÃO

Exemplo 1. Isolamento de Genes de EPSPS

Genes que codificam enzimas EPSPS Classe III foram isolados a partir de sete bactérias diferentes (*Klebsiella pneumoniae*, *Agrobacterium radiobacter*, *Rhizobium* sp., *Brevundomonas vesicularis*, *Agrobacterium tumefaciens*, *Pseudomonas syringae*, e *Brucella/Ochrobactrum*).

A seqüência que codifica o DNA e a seqüência de aminoácidos da fase aberta de leitura de *grg8* são fornecidas no Pedido de Patente U.S Nº 60/640.195, depositado em 29 de dezembro de 2004 e fornecidas na
10 SEQ ID NO: 7 e SEQ ID NO:8 desse pedido de patente, respectivamente.

A seqüência que codifica o DNA e a seqüência de aminoácidos da fase aberta de leitura de *grg12* são fornecidas nas SEQ ID NOS: 57 e 10 e SEQ ID NOS: 58 e 59 desse pedido de patente, respectivamente.

A seqüência que codifica o DNA e a seqüência de aminoácidos da fase aberta de leitura de *grg15* são fornecidas nas SEQ ID NO: 11 e SEQ
15 ID NO:12 desse pedido de patente, respectivamente.

A seqüência que codifica o DNA e a seqüência de aminoácidos da fase aberta de leitura de *grg6* são fornecidas no Número de Acesso do GenBank AE016853, bases de 1.140.091 até 1.141.347 e fornecidas nas
20 SEQ ID NO: 1 e SEQ ID NO: 2 desse pedido de patente, respectivamente.

A seqüência que codifica o DNA e a seqüência de aminoácidos da fase aberta de leitura de *grg9* são fornecidas no Número de Acesso do GenBank NC_003304, bases de 628.398 até 629.675 e fornecidas nas SEQ
ID NO: 3 e SEQ ID NO: 4 desse pedido de patente, respectivamente.

25 A seqüência que codifica o DNA e a seqüência de aminoácidos da fase aberta de leitura de *grg7* são fornecidas nas SEQ ID NO: 45 e SEQ ID NO: 46 desse pedido de patente, respectivamente.

A seqüência que codifica o DNA e a seqüência de aminoácidos da fase aberta de leitura de *grg5* são fornecidas no Número de Acesso do
30 GenBank NC_005773, bases de 1 até 1257 e fornecidas nas SEQ ID NO: 47 e SEQ ID NO: 48 desse pedido de patente, respectivamente.

A seqüência que codifica o DNA e a seqüência de aminoácidos

da fase aberta de leitura de EPSPS do milho são fornecidas no Número de Acesso do GenBank X63374 (gi:1524383), bases de 1 até 1335 e a sequência da proteína é fornecida na SEQ ID NO: 50 desse pedido de patente.

A sequência que codifica o DNA e a sequência de aminoácidos de uma EPSPS bacteriana descrita no Pedido Internacional de Patente WO2005014820 são fornecidas na SEQ ID NO: 53 e SEQ ID NO: 54 desse pedido de patente, respectivamente.

A sequência que codifica o DNA e a sequência de aminoácidos da fase aberta de leitura de *grg7m1* são fornecidas nas SEQ ID NO: 55 e SEQ ID NO: 56 desse pedido de patente, respectivamente.

Exemplo 2. Clonagem do gene de EPSP sintase da cepa DC3000 de *Pseudomonas syringae* pv *tomato*

A sequência codificante de EPSP sintase foi amplificada por PCR a partir do DNA genômico da cepa DC3000 de *Pseudomonas syringae* pv *tomato* (ATCC BAA-871) usando os seguintes iniciadores: CAGA-GATCTGGCATGCGACCTCAAGCCACTCTC (superior, SEQ ID NO: 61) e CAGGGCGCGCCTCAGCGCTGAACACTCACCC (inferior, SEQ ID NO: 62). O produto de PCR resultante de 1,3 kb foi digerido com *Bgl* II e *Asc* I, ligado em pUC18 modificado que foi digerido com *Bam*H I e *Asc* I, então eletroporado em células DH5 α . O plasmídeo de DNA foi preparado a partir de colônias resistentes a ampicilina e analisado por digestão de restrição. Um clone foi escolhido para análise posterior. A sequência de DNA do inserto foi determinada usando técnicas bem-conhecidas na técnica e foi descoberta ser 100% idêntica a sequência publicada para a cepa DC3000 (número de acesso do Genbank AE016853 bases de 1.140.091 até 1.141.347). Esse plasmídeo foi denominado de pAX703 e a ORF de EPSPS foi denominada de *grg6*.

O plasmídeo pAX703 foi transformado em células Δ *aroA* de *E. coli* e descoberto complementar a deleção. Isso demonstra que *grg6* codifica uma EPSP sintase funcional.

Exemplo 3. Clonagem do gene de EPSP sintase a partir da cepa C58 de *Agrobacterium radiobacter*

A sequência que codifica a EPSP sintase foi amplificada por PCR a partir do DNA genômico da cepa C58 de *Agrobacterium tumefaciens* (ATCC 33970) usando os seguintes iniciadores: CAGGGATCCGGCAT-GATCGAACTGACCATCACCC (superior, SEQ ID NO: 63) e CAGGGCGCGCCTCAGTGCTGCGGCTCGGCAGCG (inferior, SEQ ID NO: 64). O produto de PCR resultante de 1,3 kb foi digerido com *BamH I* e *Asc I*, ligado em pUC18 modificado que foi digerido com *BamH I* e *Asc I*, então eletroporado em células DH5 α . O plasmídeo de DNA foi preparado a partir de colônias resistentes a ampicilina e analisado por digestão de restrição. Um clone foi escolhido para análise posterior. A sequência de DNA do inserto foi determinada usando técnicas bem-conhecidas na técnica e foi descoberta ser 100% idêntica a sequência publicada para a cepa C58 (número de acesso do Genbank NC_003304 bases de 628398 até 629675). Esse plasmídeo foi denominado de pAX702 e a ORF de EPSPS de C58 foi denominada de *grg9*.

O plasmídeo pAX702 foi transformado em células Δ aroA de *E. coli* e encontrado complementar a deleção. Isso demonstra que *grg9* codifica uma EPSP sintase funcional.

Exemplo 4. Testando *grg6* e *grg9* quanto a resistência ao glifosato

Os plasmídeos pAX703 e pAX702, que contêm *grg6* e *grg9* respectivamente, foram transformados em células de *E. coli* e estriados sobre meio de ágar M63 contendo várias concentrações de glifosato. O plasmídeo vetor pUC18 foi usado como um controle sensível ao glifosato. Os resultados são apresentados na Tabela 1 abaixo e demonstram que a expressão de *grg6* ou *grg9* confere a resistência a níveis elevados de glifosato.

Tabela 1

Crescimento de <i>E. coli</i> que expressa <i>grg6</i> ou <i>grg9</i> na presença de glifosato.				
Plasmídeo	Gene	Concentração de Glifosato		
		0 mM	50 mM	100 mM
pUC18	(nenhum)	++	-	-
pAX703	<i>grg6</i>	++	++	++
pAX702	<i>grg9</i>	++	++	++

Exemplo 5. Clonagem do gene de EPSP sintase da cepa B728a de *Pseu-*

Pseudomonas syringae pv *syringae*.

A sequência que codifica a EPSP sintase foi amplificada por PCR a partir de uma colônia isolada de um único poço da cepa B728a de *Pseudomonas syringae* pv *syringae* usando os seguintes iniciadores:

5 CAGGGATCCGGCATGCGACCTCAAGCCACCCTC (superior, SEQ ID NO: 65) e CAGGGCGCGCCTCAGCGCTGAACACTCACAC (inferior, SEQ ID NO: 66). O produto de PCR resultante de 1,26 kb foi digerido com as enzimas de restrição apropriadas e ligado no vetor pRSF1b, então eletroporado em células de *E.coli*. O plasmídeo de DNA foi preparado a partir de colônias

10 resistentes a ampicilina e analisadas por digestão de restrição. Um clone foi escolhido para análise posterior e designado como pAX1923. A sequência de DNA da fase aberta de leitura em pAX1923 foi determinada e descoberta ser idêntica a sequência publicada da EPSPS da cepa B728a de *Pseudomonas syringae* pv *syringae*. Assim, designou-se essa fase aberta de leitura

15 como *grg7*.

Similarmente, um produto de PCR de 1,26 kb separado foi digerido com *BamH I* e *Asc I*, ligado em pUC18 modificado que foi digerido com *BamH I* e *Asc I*, então eletroporado em células DH5 α . O plasmídeo de DNA foi preparado a partir de colônias resistentes a ampicilina e analisado por

20 digestão de restrição. Um clone foi escolhido para análise posterior. A sequência de DNA do inserto de pAX712 foi determinada usando técnicas bem-conhecidas na técnica e foi encontrada conter 3 alterações de nucleotídeo quando comparada com a sequência de DNA publicada da cepa B728a de *Pseudomonas syringae* pv *syringae* (Número de Acesso do Genbank

25 NZ_AABP02000003, bases de 39.901 até 41.400). Essas três alterações de nucleotídeo do DNA resultam em uma proteína com uma alteração de aminoácido relativa a proteína hipotética codificada pela sequência publicada. A fase de leitura aberta da cepa B728a que codifica uma EPSPS como a identificada no plasmídeo pAX712 foi designada de *grg7m1* (SEQ ID NO: 55;

30 sequência da proteína *grg7m1* descrita na SEQ ID NO:56).

Exemplo 6. Clonagem do gene de EPSP sintase da cepa 1448a de *Pseudomonas syringae* pv *phaseolicola*

Os dados da sequência de EPSPS da cepa 1448a de *Pseudomonas syringae* pv *phaseolicola* foram obtidos do website www.tigr.org do *The Institute for Genomic Research*. A sequência que codifica a EPSP sinta-se da cepa 1448a de *Pseudomonas syringae* pv *phaseolicola* foi amplificada por PCR a partir do DNA genômico da cepa 1448a de *Pseudomonas syringae* pv *phaseolicola* (ATCC BAA-978) usando os seguintes iniciadores: CAGGGATCCGGCATGCGACCTCAAGCCACCCTC (superior, SEQ ID NO: 67) e AGAGGCGCGCCTCAGCGCTGAACACGCACC (inferior, SEQ ID NO: 68), designados baseado na sequência de DNA disponibilizada pelo *The Institute for Genomic Research* ("TIGR", comunicação pessoal). O produto de PCR resultante de 1,26 kb foi digerido com *BamH I* e *Asc I*, ligado em pUC18 modificado que foi digerido com *BamH I* e *Asc I*, então eletroporado em células DH5 α . O plasmídeo de DNA foi preparado a partir de colônias resistentes a ampicilina e analisado por digestão de restrição. Um clone foi escolhido para análise posterior e designado pAX713. A sequência de DNA do inserto de pAX713 foi determinada usando técnicas bem-conhecidas na técnica e foi descoberta ser 100% idêntica a sequência de DNA publicada da cepa 1448a de *Pseudomonas syringae* pv *phaseolicola* (realizada pelo *The Institute for Genomic Research* (TIGR) e disponibilizada online na forma eletrônica em www.tigr.org). Esse plasmídeo foi denominado de pAX713 e a fase de leitura aberta da cepa 1448a que codifica uma EPSPS como identificada no plasmídeo pAX713 foi designada como *grg5*.

Exemplo 7. Testando *grg5* e *grg7* quanto a resistência a glifosato

Os plasmídeos pAX713, pAX1923 e pAX712, que contêm *grg5* e *grg7* e *grg7m1*, respectivamente, foram transformados em células de *E.coli* e estriados sobre meio de ágar M63 contendo várias concentrações de glifosato. O plasmídeo vetor pUC18 foi usado como um controle sensível ao glifosato. Os resultados são apresentados na Tabela 2 abaixo e demonstram que a expressão de *grg5*, *grg7* ou *grg7m1* confere resistência a níveis elevados de glifosato.

Tabela 2.

Crescimento de <i>E. coli</i> que expressa <i>grg5</i> , <i>grg7</i> , ou <i>grg7m1</i> na presença de glifosato.				
Plasmídeo	Gene	Concentração de Glifosato		
		0 mM	50 mM	100 mM
pUC18	(nenhum)	++	-	-
pAX713	<i>grg5</i>	++	++	++
pAX1923	<i>grg7</i>	++	++	++
pAX712	<i>grg7m1</i>	++	++	++

Exemplo 8: Isolamento de ATX20019

ATX20019 foi isolado pelo plaqueamento de amostras de solo em Meio de Sais Minerais HEPES (HMSM) contendo glifosato como única fonte de fósforo. Já que HMSM não contém aminoácidos aromáticos, uma cepa deve ser resistente ao glifosato a fim de crescer nesse meio.

Dois gramas de solo foram suspensos em aproximadamente 10 ml de água, vortexados por 15 segundos e deixados assentar por 15 minutos. Um "loopful" de 10 µl dessa suspensão foi adicionado a 3 ml de HMSM suplementado com glifosato 10 mM (pH 7.0). HMSM contém (por litro): 10 g de glicose, 2 g de NH₄SO₄, 9,53 g de HEPES, 1,0 ml de MgSO₄ 0,8 M, 1,0 ml de CaCl₂ 0,1 M, 1,0 ml de Solução de Elementos Traço (Em 100 ml de solução 1000x: 0,1 g de FeSO₄·7H₂O, 0,5 mg de CuSO₄·5H₂O, 1,0 mg de H₃BO₃, 1,0 mg de MnSO₄·5H₂O, 7,0 mg de ZnSO₄·7H₂O, 1,0 mg de MoO₃, 4,0 g de KCl). A cultura foi cultivada em uma incubadora com agitação por quatro dias a 28°C e então 20 µl foram usados para inocular 2,5 ml de HMSM fresco contendo glifosato 10 mM como a única fonte de fósforo. Após dois dias, 20 µl foram usados para inocular outros 2,5 ml de cultura fresca. Após 5 dias, 20 µl foram usados para inocular 2,5 ml de cultura fresca. Após crescimento suficiente, a cultura foi plaqueada em meio sólido pelo estriamento de uma volta de 1 µl sobre a superfície da placa de ágar contendo ágar HMSM contendo glifosato 10 mM como única fonte de fósforo e armazenada a 28°C. A cultura foi então replaqueada para isolamento. Uma cepa em particular, designada ATX20019, foi selecionada devido a sua habilidade de crescer na presença de altas concentrações de glifosato. ATX20019 foi determinada ser um membro de *Ochrobactrum* sp./*Brucella* sp. pelo seqüenci-

amento do rDNA 16S e comparação contra uma base de dados.

Exemplo 9. Preparação e Rastreamento de Bibliotecas de Cosmídeos

DNA total foi extraído de uma cultura de ATX20019 usando métodos comumente conhecidos na técnica. O DNA foi parcialmente digerido com a enzima de restrição *Sau3A1* e ligado com o fragmento de vetor SuperCos (Stratagene) de acordo com as instruções do fabricante. Os produtos de ligação foram acondicionados em partículas de fagos usando o extrato de acondicionamento GigaPack III XL (Stratagene), transfectados em células de *E. coli aroA*. *E. coli aroA* é uma cepa na qual o gene *aroA* nativo, que codifica EPSP sintase, foi deletado. A cepa não pode crescer em meio M63 por que requer aminoácidos aromáticos suplementados exogenamente. A presença de um cosmídeo que contém um gene de EPSP sintase pode complementar geneticamente o fenótipo *aroA*, ou seja, permite que a cepa cresça no meio M63 sem aminoácidos aromáticos suplementados exogenamente.

As células transfectadas foram plaqueadas em meio M63 ágar que contém 50 µg/ml de canamicina meio M63 ágar contendo KH_2PO_4 100 mM, $(\text{NH}_4)_2\text{SO}_4$ 15 mM, CaCl_2 50 µM, FeSO_4 1 µM, MgCl_2 50 µM, glicose 55 mM, 25 mg/litro de L-prolina, 10 mg/litro de tiamina HCl, NaOH suficiente para ajustar o pH para 7.0, e 15 g/litro de ágar. Duas colônias que cresceram nesse meio foram identificadas. O cosmídeo de DNA foi preparado a partir de cada uma dessas colônias e retransformado em células de *E. coli aroA*. Em cada caso, as células retransformadas com o cosmídeo de DNA cresceram em meio M63 na presença de 0 ou 10 mM de glifosato enquanto que as células que continham o vetor SuperCos vazio não cresceram. Isso confirma que os cosmídeos são capazes de complementar o fenótipo *aroA* e capazes de conferir resistência ao glifosato. Esses cosmídeos foram denominados pAX1100 e pAX1101. Os cosmídeos pareceram ser idênticos pela análise da digestão por restrição usando duas enzimas diferentes. Um cosmídeo, pAX1101, foi selecionado para caracterização posterior.

Exemplo 10. Caracterização de *grg12* no Cosmídeo pAX1101

Para identificar o(s) gene(s) responsável(Responsáveis) pela re-

sistência ao glifosato mostrada pelo cosmídeo pAX1101, o DNA desse clone foi mutagenizado com elementos transponíveis. Nesse método, identificam-se clones que sofreram inserções com transposons e perderam a habilidade de conferir resistência ao glifosato. A localização das inserções de transposon identifica a fase aberta de leitura responsável pelo fenótipo de resistência ao glifosato.

O cosmídeo pAX1101 foi submetido a mutagênese com transposon *in vitro* usando um *EN::TN Insertion Kit* (Epicentre, Madison, WI) e o protocolo do fabricante. Esse processo insere aleatoriamente um fragmento de transposon no cosmídeo de DNA e assim interrompe a função de genes no cosmídeo. Os transposons contêm um gene que codifica resistência a um antibiótico, assim clones com inserção de transposon podem ser selecionados pela habilidade de crescer na presença daquele antibiótico. As localizações das inserções de transposon podem ser determinadas pelo mapeamento de fragmentos de restrição ou por seqüenciamento com iniciadores que se anelam no transposon.

Clones de pAX1101 com inserção de transposon foram transformados na cepa DH5 α de *E.coli* e plaqueados em meio M63 que contém glifosato. Foram encontrados múltiplos clones que tinham perdido a habilidade de crescer na presença de glifosato, indicando que o transposon interrompeu a gene responsável pela resistência.

A seqüência de DNA foi determinada para a região de pAX1101 que contém as inserções de transposon usando métodos de seqüenciamento bem-conhecidos na técnica e é apresentada como SEQ ID NO: 9. Uma fase de leitura aberta (ORF) foi identificada nas bases de 46 até 1380 da SEQ ID NO: 9. Essa seqüência de nucleotídeos é fornecida como SEQ ID NO: 57 e a seqüência de aminoácidos correspondente é fornecida como SEQ ID NO: 58. Análise de dados de seqüência de oito inserções de transposon seleciona aquela que perdeu a resistência ao glifosato revelando que todas estavam dentro da ORF. Isso indica que a ORF codifica a resistência ao glifosato conferida pelo cosmídeo. Esse gene foi denominado *grg12*. O cosmídeo pAX1101 que contém a ORF de *grg12* foi depositado no *Agricultu-*

ral Research Service Culture Collection (NRRL) em 4 de abril de 2005 e designado com Nº de Acesso B-30833.

Exemplo 11. Homologia de GRG12 com Outras Proteínas

GRG12 tem homologia com enzimas EPSP sintase. Um alinhamento da sequência de aminoácidos de GRG12 (SEQ ID NO: 10) com a sequência de aminoácidos de outras EPSP sintases é mostrado na figura 1. A Tabela 3 lista o percentual de identidade de GRG12 com várias EPSP sintases. O exame da sequência de aminoácidos (SEQ ID NO: 10) revelou que ela não contém os quatro domínios típicos de enzimas EPSP sintase Classe II. Assim ela é uma nova EPSP sintase não Classe II resistente ao glifosato.

GRG12 tem a maior homologia de aminoácidos com uma EPSPS descrita na WO2005014820 (SEQ ID NO: 54). GRG12 (SEQ ID NO: 10) tem 92% de identidade de aminoácidos com as EPSPS descritas na WO2005014820.

15 Tabela 3. Identidade de aminoácido de GRG12 a outras EPSP Sintases

Gene	Porcentagem de Identidade a GRG12
GRG8	61%
GRG7	65%
GRG6	65%
GRG9	60%
<i>D. psychrophila</i>	32%
<i>K. pneumoniae</i>	31%
<i>E. coli</i>	30%
<i>H. influenzae</i>	30%
<i>A. fulgidus</i>	22%
CP4	22%
<i>A. pernix</i>	21%
<i>C. perfringens</i>	21%
<i>B. subtilis</i>	17%
WO2005014820	92%

Análise adicional da sequência de DNA de *grg12* (SEQ ID NO: 9) revelou a presença de uma segunda fase aberta de leitura menor (SEQ ID NO: 58) começando com um códon de início GTG no nucleotídeo 142 de SEQ ID NO: 9. A tradução dessa fase aberta de leitura resulta em uma proteína (SEQ ID NO: 59) que é idêntica aos resíduos 33-444 de SEQ ID NO: 10, exceto que o códon de início de SEQ ID NO: 59 é uma metionina ao invés da valina presente em SEQ ID NO: 10. O alinhamento de SEQ ID NO:

59 com proteínas EPSPS conhecidas indicou que esta proteína contém todos os resíduos conhecidos por serem críticos para funcionar como uma EPSPS. Dessa forma, essa proteína provavelmente compreende uma enzima EPSPS resistente ao glifosato funcional, enquanto uma proteína que resulta do início da tradução de um códon de início interno aos domínios altamente conservados deve ser improvável de ser funcional. SEQ ID NO: 59 é 97% idêntica a EPSPS descrita em WO2005014820 (SEQ ID NO: 54).

A habilidade de ambas as fases abertas de leitura (SEQ ID NOS: 57 e 58) codificar atividade de EPSPS funcional pode ser testada pela amplificação de cada fase aberta de leitura por PCR, clonagem dos fragmentos de PCR resultantes em um vetor de plasmídeo sob o controle de um promotor adequado, indução da expressão de proteína a partir da fase aberta de leitura como conhecido na técnica, e comparação da habilidade da proteína expressa complementar o fenótipo *aroA* em *E. coli* e conferir resistência ao glifosato.

Exemplo 12. Manipulação de *grg12* para Expressão de Proteína GRG12 em *E. coli*

A fase aberta de leitura (ORF) de *grg12* foi amplificada por PCR, clonada em uma versão levemente modificada do vetor de plasmídeo pUC18 e transformada na cepa de *E. coli* DH5 α . As modificações a pUC18 foram como se segue: um códon de parada foi inserido na fase aberta de leitura de *lacZ* e um sítio de ligação de ribossomo (para otimizar a iniciação da tradução da ORF de *grg12* inserida) foi inserido a montante do sítio de restrição de *BamHI*. O DNA de plasmídeo foi preparado e a presença e orientação do inserto *grg12* foram determinadas por digestão de restrição. Um clone que continha a ORF na orientação de sentido direto com relação ao promotor *lac* no vetor foi nomeado pAX1106 e foi testado quanto a habilidade de conferir resistência ao glifosato. Células de *E. coli* que tinham pAX1106 ou plasmídeo de vetor vazio foram estriados sobre placas de ágar M63 contendo 0 a 200 mM de glifosato. Os resultados estão apresentados na Tabela 4 abaixo. Esses resultados demonstram que *grg12* confere resistência a altos níveis de glifosato.

Tabela 4. Crescimento de pAX1106 sobre Ágar de Glifosato M63

Concentração de Glifosato (mM)	pAX1106	Vetor Vazio
0	+	+
50	+	-
100	+	-
200	+	-

Exemplo 13. Purificação de GRG12 Expresso como uma Proteína 6xHis-tagged em *E. coli*

A região codificante de *grg12* é amplificada por PCR usando PFUULTRA® DNA polimerase (Stratagene). Os oligonucleotídeos usados para iniciar o PCR são desenhados para introduzir sítios de reconhecimento de enzima de restrição próximo às terminações 5' e 3' do produto de PCR resultante. O produto de PCR resultante é digerido com enzimas de restrição apropriadas e o produto digerido é clonado no vetor de expressão de 6xHis-tag, pRSF1b (Novagen). O clone resultante contém *grg12* na mesma fase de leitura translacional que, é imediatamente C-terminal ao 6xHis tag. Estratégias gerais para gerar tais clones, e para expressar proteínas que contêm 6xHis-tag são bem-conhecidas na técnica.

O nível de expressão da proteína GRG12 pode ser determinado em um gel de proteína SDS-PAGE. A proteína GRG12 pode ser isolada por purificação da proteína GRG12 induzida por cromatografia em, por exemplo, Resina Ni-NTA Superflow (Qiagen), seguindo as instruções do fabricante.

Exemplo 14: Isolamento de ATX4150

ATX4150 foi isolado pelo plaqueamento das amostras de solo em Meio Mínimo Enriquecido (EMM) contendo glifosato como a única fonte de fósforo. Já que EMM não contém aminoácidos aromáticos, uma cepa deve ser resistente ao glifosato para crescer nesse meio.

Dois gramas de solo foram suspensos em aproximadamente 30 ml de água, e sonificados por 30 segundos em um banho de água sonificador Aquasonic. A amostra foi vortexada por 5 segundos e deixada repousar por 60 segundos. Esse processo foi repetido 3 vezes. 100 µl dessa suspensão foram adicionados a 3 ml de EMM suplementado com 4 mM de glifosato (pH 6.0). EMM contém (por 900 mls): 10 g de sacarose, 2 g de NaNO₃, 1,0 ml de

MgSO₄ 0,8 M, 1,0 ml de CaCl₂ 0,1 M, 1,0 ml de Solução de Elementos Traço (Em 100 ml de Solução 1000x: 0,1 g de FeSO₄·7H₂O, 0,5 mg de CuSO₄·5H₂O, 1,0 mg de H₃BO₃, 1,0 mg de MnSO₄·5H₂O, 7,0 mg de ZnSO₄·7H₂O, 1,0 mg de MoO₃, 4,0 g de KCl). A cultura foi agitada em um tambor rotatório de cultura de tecido por dezesseis dias a 21°C e então 100 µl foram usados para inocular 3 ml de EMM fresco contendo 4 mM de glifosato como a única fonte de fósforo. Após cinco dias, 100 µl foram usados para inocular outra cultura de 3 ml fresca. Após alguns dias, a cultura foi plaqueada sobre meio sólido pelo estriamento de uma volta de 1 µl sobre a superfície da placa de ágar contendo ágar EMM contendo 5 mM de glifosato como a única fonte de fósforo. Após alguns dias, as colônias foram replaqueadas para isolamento em EMM contendo 5 mM de glifosato como a única fonte de fósforo. Uma cepa em particular, designada ATX4150, foi selecionada devido a sua habilidade de crescer na presença de altas concentrações de glifosato.

Exemplo 15. Preparo e Rastreamento de Bibliotecas de Cosmídeo

O DNA total foi extraído de uma cultura de ATX4150 usando métodos comumente usados na técnica. O DNA foi parcialmente digerido com enzima de restrição *Sau3A1* e ligado com o fragmento de vetor SuperCos (Stratagene) de acordo com as instruções do fabricante. Os produtos de ligação foram acondicionados dentro de partículas de fago usando *GigaPack III XL packaging extract* (Stratagene), transfectados em células *E. coli aroA-*. *E. coli aroA-*, é uma cepa na qual o gene *aroA* nativo, que codifica a EPSP sintase, foi deletado. Essa cepa não pode crescer em meio M63 porque ela requer aminoácidos aromáticos fornecidos exogenamente. A presença de um cosmídeo que contém um gene de EPSP sintase pode complementar geneticamente o fenótipo *aroA-*, isso é, ele permite que a cepa cresça em meio M63 sem aminoácidos aromáticos fornecidos exogenamente.

As células transfectadas foram plaqueadas em meio de ágar M63 contendo 50 µg/ml de canamicina. O meio de ágar M63 contém 100 mM de KH₂PO₄, 15 mM de (NH₄)₂SO₄, 50 µM de CaCl₂, 1 µM de FeSO₄, 50 µM de MgCl₂, 55 mM de glicose, 25 mg/litro de L-prolina, 10 mg/litro de tia-

mina HCl, NaOH suficiente para ajustar o pH para 7.0, e 15 g/litro de ágar. Cinco colônias que cresceram sobre esse meio foram identificadas. O DNA de cosmídeo foi preparado a partir de cada uma dessas colônias e retransformado em células de *E. coli aroA*-. Em cada caso, as células retransformadas com o DNA de cosmídeo cresceram sobre meio M63 na presença de 0 ou 10 mM de glifosato enquanto células contendo o vetor SuperCos vazio não. Isso confirma que os cosmídeos são capazes de complementar o fenótipo *aroA*- e são capazes de conferir resistência ao glifosato. Um cosmídeo foi selecionado para caracterização adicional e foi nomeado pAX305.

10 Exemplo 16. Identificação de *grg15* em Cosmídeo pAX305

Para identificar o(s) gene(s) responsável(is) pela resistência ao glifosato mostrada pelo cosmídeo pAX305, o DNA desse clone foi mutagenizado com elementos transponíveis. O cosmídeo pAX305 foi submetido a mutagênese de transposon *in vitro* usando um *EZ::TN Insertion Kit* (Epicentre, Madison, WI) e o protocolo do fabricante. Clones de inserção de transposon de pAX305 foram transformados em *E. coli* e plaqueados em meio M63 contendo glifosato. Foram encontrados múltiplos clones que perderam a habilidade de crescer na presença de glifosato, indicando que o transposon rompeu o gene responsável pela resistência.

20 A sequência de DNA foi determinada para a região de pAX305 que contém as inserções de transposon usando métodos de seqüenciamento bem-conhecidos na técnica e é apresentada em SEQ ID NO: 61. Uma fase aberta (ORF, bases de nucleotídeo 77 até 1354 de SEQ ID NO: 61) foi identificada. A análise dos dados de sequência de oito seleções de inserção de transposon que perderam a resistência ao glifosato revelou que todos estavam dentro da ORF. Isso indica que a ORF codifica a resistência ao glifosato conferida pelo cosmídeo. Esse gene foi nomeado *grg15*. O cosmídeo pAX305 contendo a ORF de *grg15* (SEQ ID NO: 11) foi depositado na *Agricultural Research Service Culture Collection* (NRRL) em 20 de abril de 2005 e designado o Nº de Acesso NRRL B-30838.

30 Exemplo 17. Homologia de GRG15 com Outras Proteínas

GRG15 tem homologia a enzimas EPSP sintase. Um alinhamen-

to da sequência de aminoácido de GRG15 (SEQ ID NO: 12) com as seqüências de aminoácido de outras EPSP sintases é mostrada na figura 1. A Tabela 5 lista o percentual de identidade de GRG15 a várias EPSP sintases. O exame da sequência de aminoácido (SEQ ID NO: 12) revelou que ela não contém os quatro domínios típicos de enzimas EPSP sintase Classe II. Dessa forma essa é uma nova EPSP sintase resistente ao glifosato, não-Classe II.

Tabela 5. Identidade de Aminoácido de GRG15 e outras EPSP Sintases

Gene	Percentual de Identidade a GRG15
GRG9	94%
GRG8	70%
GRG7	66%
GRG6	66%
GRG12	60%
<i>K. pneumoniae</i>	32%
<i>E. coli</i>	32%
<i>H. influenzae</i>	32%
<i>D. psychrophila</i>	31%
<i>A. fulgidus</i>	26%
<i>A. pernix</i>	26%
<i>C. perfringens</i>	24%
<i>B. subtilis</i>	24%
GRG1	23%
<i>Agrobacterium</i> Sp. CP4	22%

Exemplo 18. Blocos de Homologia entre enzimas não-Classe II resistentes ao Glifosato (Classe III)

10 Comparação das seqüências de aminoácido das proteínas GRG (proteínas resistentes ao glifosato não-classe II) mostra que as proteínas GRG têm significante homologia uma a outra, e são distintas de EPSP sintases resistentes ao glifosato identificadas anteriormente. (veja a figura 1).

Exemplo 19: Identificação de Domínios Conservados

15 As seqüências de aminoácido que sete EPSP sintases conhecidas foram analisadas para domínios conservados que não estavam presentes nem nas EPSP sintases Classe I nem nas classe II. Os seguintes domínios são encontrados nessas EPSP sintases: tipo em **negrito** denota resíduos conservados apenas entre essa classe, e em *itálico* denota resíduos conservados em todas as enzimas EPSPS.

20

Domínio I:

L-A-K-G-X₁-S-X₂-L-X₃-G-A-L-K-S-D-D-T (SEQ ID NO: 13), onde X₁ denota lisina ou treonina, onde X₂ denota arginina ou histidina, onde X₃ denota serina ou treonina.

Domínio Ia:

- 5 L-A-K-G-X₁, (SEQ ID NO: 14), onde X₁ denota lisina ou treonina.

Domínio Ib:

S-X₁-L-X₂ (SEQ ID NO: 15), onde X₁ denota arginina ou histidina, onde X₂ denota serina ou treonina.

Domínio Ic:

- 10 G-A-L-K-S-D-D-T (SEQ ID NO: 16)

Domínio II:

- 15 E-P-D-X₁-X₂-T-F-X₃-V-X₄-X₅-X₆-G (SEQ ID NO: 17), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina, onde X₄ denota treonina ou ácido glutâmico ou lisina, onde X₅ denota serina ou glicina, onde X₆ denota glutamina ou serina ou ácido glutâmico ou treonina.

Domínio IIa:

- 20 E-P-D-X₁-X₂-T-F-X₃-V (SEQ ID NO: 18), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina.

Domínio IIb:

X₁-X₂-X₃-G (SEQ ID NO: 19), onde X₁ denota treonina ou ácido glutâmico ou lisina, onde X₂ denota serina ou glicina, onde X₃ denota glutamina ou serina ou ácido glutâmico ou treonina.

- 25 Domínio III:

RFLTAA (SEQ ID NO: 20)

Domínio IV:

KRPI(G/M)P (SEQ ID NO: 21)

K-R-P-I-X₁-P, onde X₁ denota glicina ou metionina ou leucina.

- 30 Domínio V:

X₁-G-C-P-P-V (SEQ ID NO: 22), onde X₁ denota treonina ou serina.

Domínio VI:

I-**G-A**-X₁-**G**-Y-X₂-D-L-T (SEQ ID NO: 23), onde X₁ denota arginina ou lisina ou leucina, e onde X₂ denota isoleucina ou valina.

Domínio VII:

W-X₁-V-X₂-X₃-T-G (SEQ ID NO: 24), onde X₁ denota arginina ou lisina, onde
5 X₂ denota alanina ou histidina ou ácido glutâmico ou serina, onde X₃ denota prolina ou alanina.

Domínio VIII:

E-P-D-A-S-A-A-T- Y-L-W-X₁-A-X₂-X₃-L (SEQ ID NO: 25), onde X₁ denota alanina ou glicina, onde X₂ denota ácido glutâmico ou glutamina, onde X₃
10 denota valina ou leucina ou alanina.

Domínio VIIIa:

E-P-D-A-S-A-A-T- Y-L-W (SEQ ID NO: 26)

Domínio IX:

I-D-X₁-**G** (SEQ ID NO: 27), onde X₁ denota isoleucina ou leucina.

15 Domínio X:

F-X₁-Q-P-D-A- K-A (SEQ ID NO: 28), onde X₁ denota treonina ou serina,

Domínio XI:

X₁-**F-P**-X₂-X₃-X₄-A-X₅-X₆-X₇-G-**S-Q**-M-Q-D-A-I-P-T-X₈-A-V- X₉ A-A- **F-N** (SEQ ID NO: 29), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina, onde X₅ denota treonina ou ácido glutâmico ou valina, onde X₆ denota valina ou isoleucina, onde X₇ denota ácido aspártico ou valina, onde X₈ denota leucina ou isoleucina, onde X₉ denota leucina ou isoleucina.
20

25 Domínio XIa:

X₁-**F-P**-X₂-X₃-X₄-A (SEQ ID NO: 30), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina.

Domínio XIb:

30 X₁-X₂-X₃-G-**S-Q**-M-Q-D-A-I-P-T-X₄-A- V- X₅ A-A- **F-N** (SEQ ID NO: 31), onde X₁ denota treonina ou ácido glutâmico ou valina, onde X₂ denota valina ou isoleucina, onde X₃ denota ácido aspártico ou valina, onde X₄ denota leucina

ou isoleucina, onde X_5 denota leucina ou isoleucina.

Domínio XIc:

G-S-Q-M-Q-D-A-I-P-T (SEQ ID NO: 32)

Domínio XII:

- 5 P-V-R-F- X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-C-D-R- X_5 (SEQ ID NO: 33), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina.

Domínio XIIa:

- 10 P-V-R-F (SEQ ID NO: 34)

Domínio XIIb:

- X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-C-D-R- X_5 (SEQ ID NO: 35), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina.

Domínio XIIc:

N-L-R-V-K-E-C-D-R (SEQ ID NO: 36)

Domínio XIII:

- 20 E-G-D-D-L- X_1 - X_2 (SEQ ID NO: 37), onde X_1 denota leucina ou isoleucina, onde X_2 denota valina ou isoleucina.

Domínio XIV:

X_1 -P- X_2 -L-A-G (SEQ ID NO: 38), onde X_1 denota ácido aspártico ou asparagina, onde X_2 denota alanina ou serina ou treonina.

Domínio XV:

- 25 A- X_1 -I-D- X_2 - X_3 - X_4 -D-H-R- (SEQ ID NO: 39), onde X_1 denota leucina ou serina ou ácido glutâmico, onde X_2 denota treonina ou serina, onde X_3 denota histidina ou fenilalanina, onde X_4 denota alanina ou serina.

Domínio XVI:

- 30 F-A-L-A- X_1 -L-K- X_2 - X_3 -G-I (SEQ ID NO: 40), onde X_1 denota glicina ou alanina, onde X_2 denota isoleucina ou valina, onde X_3 denota serina ou glicina ou alanina ou lisina.

Domínio XVIa:

F-A-L-A-X₁-L-K (SEQ ID NO: 41), onde X₁ denota glicina ou alanina.

Domínio XVIb:

L-K-X₁-X₂-G-I (SEQ ID NO: 42), onde X₁ denota isoleucina ou valina, onde X₂ denota serina ou glicina ou alanina ou lisina.

5 Domínio XVII:

-X₁-P-X₂-C-V-X₃-K (SEQ ID NO: 43), onde X₁ denota asparagina ou ácido aspártico, onde X₂ denota alanina ou ácido aspártico, onde X₃ denota alanina ou glicina.

Domínio XVIII:

10 X₁-S-L-G-V (SEQ ID NO: 44), onde X₁ denota alanina ou serina ou prolina.

Exemplo 20. Manipulação de *grg12* para Transformação de Planta

A fase aberta de leitura (ORF) de *grg12* é amplificada por PCR a partir de um molde de cDNA de extensão completa. Sítios de restrição de *Hind* III são adicionados a cada extremidade da ORF durante o PCR. Adicionalmente, a sequência de nucleotídeo ACC é adicionada imediatamente 5' ao códon de início do gene para aumentar a eficiência traducional (Kozak (1987) *Nucleic Acids Research* 15:8125-8148; Joshi (1987) *Nucleic Acids Research* 15:6643-6653). O produto de PCR é clonado e seqüenciado, usando técnicas bem-conhecidas na técnica, para garantir que nenhuma mutação seja introduzida durante o PCR.

O plasmídeo contendo o produto de PCR de *grg12* é digerido com, por exemplo, *Hind* III e o fragmento contendo a ORF intacta é isolado. Nesse exemplo, o fragmento é clonado no sítio de *Hind* III de um plasmídeo, tal como pAX200, que é um vetor de expressão de planta que contém o promotor de actina do arroz (McElroy *et al.* (1991) *Molec. Gen. Genet.* 231:150-160), e o terminador PinII (An *et al.* (1989) *The Plant Cell* 1:115-122). O fragmento promotor - gene - terminador desse plasmídeo intermediário é então subclonado em um plasmídeo tal como pSB11 (Japan Tobacco, Inc.) para formar um plasmídeo final referido aqui como pSB11GRG12.

pSB11GRG12 é organizado tal que o fragmento de DNA que contém o construto do promotor - *grg12* - terminador pode ser cortado por enzimas de restrição apropriadas e também usando para transformação em plantas, por

exemplo, por injeção de um feixe de aerossol. A estrutura de pSB11GRG12 é verificada por digestão de restrição e eletroforese em gel e por seqüenciamento através das várias junções de clonagem.

- O plasmídeo é mobilizado em *Agrobacterium tumefaciens* cepa LBA4404 que também carrega o plasmídeo pSB1 (Japan Tobacco, Inc.), usando procedimentos de acasalamento triparentais bem-conhecidas na técnica, e plaqueamento sobre meio contendo antibiótico. O plasmídeo pSB11GRG12 carrega resistência a espectinomicina mas é um plasmídeo de estreita variação de hospedeiro e não pode replicar em *Agrobacterium*.
- Colônias resistentes a antibiótico surgem quando pSB11GRG12 se integra no plasmídeo de ampla variação de hospedeiro pSB1 através de recombinação homologa. O produto co-integrado resultante é verificado por hibridização Southern. A cepa de *Agrobacterium* que carrega o co-integrado pode ser usada para transformar milho, por exemplo, pelo método PureIntro (Japan Tobacco).

Exemplo 21. Transformação de *grg12* em Células de Planta

- Espigas de milho são melhor colhidas 8-12 dias após a polinização. Os embriões são isolados a partir das espigas, e esses embriões de 0,8-1,5 mm de tamanho são preferidos para uso na transformação. Os embriões são plaqueados com o escutelo para cima sobre um meio de incubação adequado, tal como meio DN62A5S (3,98 g/L de Sais N6; 1 mL/L (de estoque 1000x) de vitaminas N6; 800 mg/L de L-Asparagina; 100 mg/L de Mio-inositol; 1,4 g/L de L-Prolina; 100 mg/L de Casaminoácidos; 50 g/L de sacarose; 1 mL/L (de estoque a 1 mg/mL) de 2,4-D). Entretanto, o meio e os sais diferentes de DN62A5S são adequados e conhecidos na técnica. Os embriões são incubados durante a noite a 25°C no escuro. Entretanto, não é necessário *per se* incubar os embriões durante a noite.

- Os explantes resultantes são transferidos para "mesh squares" (30-40 por placa), transferidos para meio osmótico por cerca de 30-45 minutos, e então transferidos para uma placa de irradiação (veja, por exemplo, Publicação PCT Nº WO/0138514 e Patente U.S. Nº 5.240.842).

Os construtos de DNA projetados para expressar GRG12 em

células de planta são acelerados em tecido de planta usando um acelerador de feixe de aerossol, usando condições essencialmente como descritas na Publicação PCT Nº WO/0138514. Após a irradiação, os embriões são incubados por cerca de 30 minutos em meio osmótico, e colocados em meio de incubação durante a noite a 25°C no escuro. Para evitar danos excessivos ao explantes irradiados, eles são incubados por pelo menos 24 horas antes da transferência para meio de recuperação. Os embriões são espalhados sobre meio de período de recuperação, por cerca de 5 dias, 25°C no escuro, e então transferidos para um meio de seleção. Os explantes são incubados em meio de seleção por até oito semanas, dependendo da natureza e características da seleção em particular utilizada. Após o período de seleção, o calo resultante é transferido para meio de maturação de embrião, até a formação de embriões somáticos maduros ser observada. Os embriões somáticos maduros resultantes são então colocados sob luz leve, e o processo de regeneração é iniciado por métodos conhecidos na técnica. Os brotos resultantes são deixados enraizar em meio de enraizamento, e as plantas resultantes são transferidas para sementeiras e propagadas como plantas transgênicas.

Materiais

Tabela 6. Meio DN62A5S

Componentes	Por Litro	Fonte
Mistura Basal N6 de Chu (Prod. No. C 416)	3,98 g/L	Phytotechnology Labs
Solução de Vitamina N6 de Chu (Prod. No. C 149)	1 mL/L (de estoque 1000x)	Phytotechnology Labs
L-Asparagina	800 mg/L	Phytotechnology Labs
Mio-inositol	100 mg/L	Sigma
L-Prolina	1,4 g/L	Phytotechnology Labs
Casamino ácidos	100 mg/L	Fisher Scientific
Sacarose	50 g/L	Phytotechnology Labs
2,4-D (Prod. No. D-7299)	1 mL/L (de estoque 1 mg/mL)	Sigma

O pH da solução é ajustado para pH 5.8 com KOH 1N/KCl 1N, Gelrite (Sigma) é adicionado em uma concentração de até 3g/L, e o meio é autoclavado. Após o resfriamento para 50°C, 2 ml/L de uma solução estoque a 5 mg/ml de nitrato de prata (*Phytotechnology Labs*) é adicionado.

Exemplo 22. Transformação de *grg12* em Células de Milho por Transformação Mediada por *Agrobacterium*

Espigas são melhor colhidas 8-12 dias após a polinização. Os embriões são isolados a partir das espigas, e esses embriões de 0,8-1,5 mm de tamanho são preferidos para uso na transformação. Os embriões são plaqueados com o escutelo para cima em um meio de incubação adequado, e incubados durante a noite a 25°C no escuro. Entretanto, não é necessário *per se* incubar os embriões durante a noite. Os embriões são colocados em contato com uma cepa de *Agrobacterium* contendo os vetores apropriados para a transferência mediada pelo plasmídeo Ti por cerca de 5-10 min, e então plaqueados em meio de co-cultivo por cerca de 3 dias (25°C no escuro). Após o co-cultivo, os explantes são transferidos para meio de período de recuperação por cerca de cinco dias (a 25°C no escuro). Os explantes são incubados em meio de seleção por até oito semanas, dependendo da natureza e características da seleção em particular utilizada. Após o período de seleção, o calo resultante é transferido para meio de maturação de embrião, até a formação de embriões somáticos maduros ser observada. Os embriões somáticos maduros resultantes são então colocados sob luz baixa, e o processo de regeneração é iniciado como conhecido na técnica. Os brotos resultantes são deixados enraizar em meio de enraizamento, e as plantas resultantes são transferidas para sementeiras e propagadas como plantas transgênicas.

Todas as publicações e pedidos de patente mencionados no relatório descritivo são indicativos do nível de perícia dos versados na técnica a qual essa invenção pertence. Todas as publicações e pedidos de patente estão aqui incorporados por referência na da mesma forma que se cada publicação ou pedido de patente individual fosse específica e individualmente indicado para ser incorporado por referência.

Embora a invenção anterior tenha sido descrita em algum detalhe por meio de ilustração e exemplo para fins de clareza de entendimento, será óbvio que certas alterações e modificações podem ser praticadas dentro do escopo das reivindicações anexas.

Listagem de Seqüências

<110> Carozzi, Nadine
Carr, Brian
Hammer, Philip E.

<120> IDENTIFICAÇÃO DE UMA NOVA CLASSE DE EPSP SINTASES

<130> 045600/310289

<150> 60/669,686
<151> 2005-04-08

<150> 60/678,348
<151> 2005-05-06

<150> 60/695,193
<151> 2005-06-29

<150> 60/725,182
<151> 2005-10-11

<160> 68

<170> FastSEQ para Windows versão 4.0

<210> 1
<211> 1257
<212> DNA
<213> Pseudomonas syringae

<220>
<221> CDS
<222> (1) ... (1257)

<221> Região de interesse biológico
<222> (0) ... (0)
<223> Gene EPSPS da cepa DC3000

<400> 1

atg cga cct caa gcc act ctc act gtt atg cct gtc gag cgc ccg ctg	48
Met Arg Pro Gln Ala Thr Leu Thr Val Met Pro Val Glu Arg Pro Leu	
1 5 10 15	
gtc ggg cgc gtc agc ccg ccg gcc tcc aag tcg atc acc aac cgc gca	96
Val Gly Arg Val Ser Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala	
20 25 30	
ctg ctg ctg gcg ggg ctt gcc aaa ggc acc agc cgc ctg acc ggc gcg	144
Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala	
35 40 45	
ctg aag agt gac gat acc cgt gtg atg tcc gaa gca ctg cga ctg atg	192
Leu Lys Ser Asp Asp Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met	
50 55 60	
ggc gtg cag gtc gac gag ccg gat gac agc acc ttt gtg gtc acc agc	240
Gly Val Gln Val Asp Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser	
65 70 75 80	
agc ggc cac tgg caa gcg ccg cag caa gcg ctg ttt ctc ggc aac gcc	288
Ser Gly His Trp Gln Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala	
85 90 95	

999	acc	gcg	acg	cgc	ttt	ctg	acc	gct	gcg	ctg	gcc	aac	ttt	gaa	ggc	336
Gly	Thr	Ala	Thr	Arg	Phe	Leu	Thr	Ala	Ala	Leu	Ala	Asn	Phe	Glu	Gly	
			100					105					110			
gac	ttc	gtg	gtg	gat	ggc	gac	gag	tac	atg	cgt	aaa	cgc	ccg	atc	ggc	384
Asp	Phe	Val	Val	Asp	Gly	Asp	Glu	Tyr	Met	Arg	Lys	Arg	Pro	Ile	Gly	
		115					120					125				
ccg	ctg	gtc	gat	gcc	ttg	cag	cgc	atg	ggc	gtg	gaa	atc	agc	gcg	ccc	432
Pro	Leu	Val	Asp	Ala	Leu	Gln	Arg	Met	Gly	Val	Glu	Ile	Ser	Ala	Pro	
	130					135					140					
agc	ggt	tgc	ccg	ccc	gtg	gcg	atc	aag	ggc	aag	ggc	ggt	ctg	caa	gcc	480
Ser	Gly	Cys	Pro	Pro	Val	Ala	Ile	Lys	Gly	Lys	Gly	Gly	Leu	Gln	Ala	
145					150				155						160	
ggg	cgt	att	gaa	atc	gac	ggc	aac	ctg	tcc	agc	cag	tac	gta	tcg	gcg	528
Gly	Arg	Ile	Glu	Ile	Asp	Gly	Asn	Leu	Ser	Ser	Gln	Tyr	Val	Ser	Ala	
			165					170					175			
ctg	ctg	atg	gcc	gga	gcg	tgt	ggc	aaa	ggc	tcg	ctt	gaa	gtt	gct	ctg	576
Leu	Leu	Met	Ala	Gly	Ala	Cys	Gly	Lys	Gly	Ser	Leu	Glu	Val	Ala	Leu	
			180					185					190			
acc	ggt	agc	gag	atc	ggc	gca	cgc	ggc	tat	gtc	gac	ctg	acc	ctg	gcg	624
Thr	Gly	Ser	Glu	Ile	Gly	Ala	Arg	Gly	Tyr	Val	Asp	Leu	Thr	Leu	Ala	
		195					200					205				
gcg	atg	cag	gcc	ttc	ggc	gct	gaa	gtt	cag	gcc	atc	ggc	gac	gct	gcc	672
Ala	Met	Gln	Ala	Phe	Gly	Ala	Glu	Val	Gln	Ala	Ile	Gly	Asp	Ala	Ala	
	210					215					220					
tgg	aaa	gtc	tcg	gcc	acc	ggc	tac	cac	gct	acg	gat	ttc	cac	atc	gag	720
Trp	Lys	Val	Ser	Ala	Thr	Gly	Tyr	His	Ala	Thr	Asp	Phe	His	Ile	Glu	
225					230					235					240	
ccg	gat	gca	tcg	gca	gcc	acc	tac	ctc	tgg	gcc	gca	caa	gcc	ctg	acc	768
Pro	Asp	Ala	Ser	Ala	Ala	Thr	Tyr	Leu	Trp	Ala	Ala	Gln	Ala	Leu	Thr	
			245					250						255		
gaa	ggc	aac	atc	gac	ctg	ggc	gtg	gcc	agt	gat	gca	ttc	act	cag	cct	816
Glu	Gly	Asn	Ile	Asp	Leu	Gly	Val	Ala	Ser	Asp	Ala	Phe	Thr	Gln	Pro	
		260					265					270				
gac	gcg	ctg	gcc	agc	cag	atc	atc	gac	agc	ttc	ccg	aac	atg	ccg	gcg	864
Asp	Ala	Leu	Ala	Ser	Gln	Ile	Ile	Asp	Ser	Phe	Pro	Asn	Met	Pro	Ala	
		275				280						285				
gtg	atc	gac	ggt	tcg	caa	atg	cag	gac	gcc	att	ccg	acc	ctc	gca	gtg	912
Val	Ile	Asp	Gly	Ser	Gln	Met	Gln	Asp	Ala	Ile	Pro	Thr	Leu	Ala	Val	
	290					295					300					
ctc	gcg	gcc	ttc	aat	cgt	cag	ccg	gta	cgc	ttc	gtc	ggc	atc	gcc	aac	960
Leu	Ala	Ala	Phe	Asn	Arg	Gln	Pro	Val	Arg	Phe	Val	Gly	Ile	Ala	Asn	
305				310					315					320		
ctg	cgg	gtc	aag	gaa	tgt	gat	cgt	att	tca	gcg	ctg	tgt	gac	ggc	ctg	1008
Leu	Arg	Val	Lys	Glu	Cys	Asp	Arg	Ile	Ser	Ala	Leu	Cys	Asp	Gly	Leu	
			325					330					335			
tgc	gcc	atc	gca	ccg	ggc	ctt	gcg	gtt	gaa	gag	ggc	gac	gac	ctg	atc	1056
Cys	Ala	Ile	Ala	Pro	Gly	Leu	Ala	Val	Glu	Glu	Gly	Asp	Asp	Leu	Ile	
			340					345				350				
gtg	cac	gcc	aac	ccg	gcg	ctg	gca	ggg	acc	aca	gtc	aac	gca	ctg	atc	1104

Val His Ala Asn Pro Ala Leu Ala Gly Thr Thr Val Asn Ala Leu Ile
 355 360 365

gac acc cat tcc gac cat cgc atc gcc atg tgc ttt gcc ctg gct ggt 1152
 Asp Thr His Ser Asp His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly
 370 375 380

ttg aag atc aaa ggc atc cat att cag gat ccc gat tgc gtc gcc aag 1200
 Leu Lys Ile Lys Gly Ile His Ile Gln Asp Pro Asp Cys Val Ala Lys
 385 390 395 400

acc tat ccg ggt tac tgg gat gcg ctg gct tcg ctg ggg gtg agt gtt 1248
 Thr Tyr Pro Gly Tyr Trp Asp Ala Leu Ala Ser Leu Gly Val Ser Val
 405 410 415

cag cgc tga 1257
 Gln Arg *

<210> 2
 <211> 418
 <212> PRT
 <213> Pseudomonas syringae

<400> 2

Met Arg Pro Gln Ala Thr Leu Thr Val Met Pro Val Glu Arg Pro Leu
 1 5 10 15
 Val Gly Arg Val Ser Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala
 20 25 30
 Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala
 35 40 45
 Leu Lys Ser Asp Asp Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met
 50 55 60
 Gly Val Gln Val Asp Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser
 65 70 75 80
 Ser Gly His Trp Gln Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala
 85 90 95
 Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Leu Ala Asn Phe Glu Gly
 100 105 110
 Asp Phe Val Val Asp Gly Asp Glu Tyr Met Arg Lys Arg Pro Ile Gly
 115 120 125
 Pro Leu Val Asp Ala Leu Gln Arg Met Gly Val Glu Ile Ser Ala Pro
 130 135 140
 Ser Gly Cys Pro Pro Val Ala Ile Lys Gly Lys Gly Gly Leu Gln Ala
 145 150 155 160
 Gly Arg Ile Glu Ile Asp Gly Asn Leu Ser Ser Gln Tyr Val Ser Ala
 165 170 175
 Leu Leu Met Ala Gly Ala Cys Gly Lys Gly Ser Leu Glu Val Ala Leu
 180 185 190
 Thr Gly Ser Glu Ile Gly Ala Arg Gly Tyr Val Asp Leu Thr Leu Ala
 195 200 205
 Ala Met Gln Ala Phe Gly Ala Glu Val Gln Ala Ile Gly Asp Ala Ala
 210 215 220
 Trp Lys Val Ser Ala Thr Gly Tyr His Ala Thr Asp Phe His Ile Glu
 225 230 235 240
 Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Gln Ala Leu Thr
 245 250 255
 Glu Gly Asn Ile Asp Leu Gly Val Ala Ser Asp Ala Phe Thr Gln Pro
 260 265 270
 Asp Ala Leu Ala Ser Gln Ile Ile Asp Ser Phe Pro Asn Met Pro Ala
 275 280 285
 Val Ile Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val
 290 295 300
 Leu Ala Ala Phe Asn Arg Gln Pro Val Arg Phe Val Gly Ile Ala Asn
 305 310 315 320

<210>	3
<211>	1278
<212>	DNA
<213>	<i>Agrobacterium tumefaciens</i>

```
<220>
<221>      CDS
<222>      (1) ... (1278)

<221>      Região de interesse biológico
<222>      (0) ... (0)
<223>      EPSPS da cepa C58
```

<400>															3	
atg	atc	gaa	ctg	acc	atc	acc	ccg	ccc	ggc	cac	ccg	ctt	tcc	ggc	aag	48
Met	Ile	Glu	Leu	Thr	Ile	Thr	Pro	Pro	Gly	His	Pro	Leu	Ser	Gly	Lys	
1				5					10					15		
gtg	gag	ccg	ccc	ggc	tcc	aaa	tcc	atc	acc	aac	cgc	gcg	ctt	ctc	ttg	96
Val	Glu	Pro	Pro	Gly	Ser	Lys	Ser	Ile	Thr	Asn	Arg	Ala	Leu	Leu	Leu	
			20					25					30			
gcc	ggc	ctc	gcc	aag	ggc	aaa	agc	cac	ttg	tca	ggc	gcc	ctg	aaa	agc	144
Ala	Gly	Leu	Ala	Lys	Gly	Lys	Ser	His	Leu	Ser	Gly	Ala	Leu	Lys	Ser	
		35					40					45				
gac	gat	acg	ctt	tat	atg	gcc	gaa	gct	ctg	cgg	gag	atg	ggc	gtg	aag	192
Asp	Asp	Thr	Leu	Tyr	Met	Ala	Glu	Ala	Leu	Arg	Glu	Met	Gly	Val	Lys	
	50					55					60					
gtc	acc	gag	cct	gac	gcc	acc	acc	ttc	gtg	gtg	gag	gga	acg	ggc	gtg	240
Val	Thr	Glu	Pro	Asp	Ala	Thr	Thr	Phe	Val	Val	Glu	Gly	Thr	Gly	Val	
65					70				75						80	
ctg	cag	cag	ccg	gaa	aag	ccg	ctg	ttc	ctc	ggc	aat	gcc	ggc	acc	gcc	288
Leu	Gln	Gln	Pro	Glu	Lys	Pro	Leu	Phe	Leu	Gly	Asn	Ala	Gly	Thr	Ala	
				85					90					95		
acg	cgg	ttc	ctg	act	gcc	gcc	ggg	gca	ctt	gtc	gat	ggc	gcc	gtc	atc	336
Thr	Arg	Phe	Leu	Thr	Ala	Ala	Gly	Ala	Leu	Val	Asp	Gly	Ala	Val	Ile	
			100					105					110			
atc	gat	ggg	gac	gaa	cat	atg	cgc	aaa	cgc	ccg	ata	ctg	ccg	ctg	gtg	384
Ile	Asp	Gly	Asp	Glu	His	Met	Arg	Lys	Arg	Pro	Ile	Leu	Pro	Leu	Val	
		115					120					125				
cag	gcg	ctg	cgg	gct	ctc	ggc	gtg	gaa	gcg	gat	gcg	cca	acc	ggc	tgc	432
Gln	Ala	Leu	Arg	Ala	Leu	Gly	Val	Glu	Ala	Asp	Ala	Pro	Thr	Gly	Cys	
	130					135					140					

ccg	cct	gtc	acc	gtc	cgt	ggc	aag	ggg	atg	ggg	ttt	cca	aag	ggc	agc	480
Pro	Pro	Val	Thr	Val	Arg	Gly	Lys	Gly	Met	Gly	Phe	Pro	Lys	Gly	Ser	
145					150					155					160	
gtc	acc	atc	gac	gcc	aat	ctc	tcc	agc	cag	tat	gtg	tcg	gca	ctg	ttg	528
Val	Thr	Ile	Asp	Ala	Asn	Leu	Ser	Ser	Gln	Tyr	Val	Ser	Ala	Leu	Leu	
				165					170					175		
atg	gcc	gcc	gcc	tgc	ggg	gac	aag	ccc	gtg	gat	atc	atc	ctg	aag	ggg	576
Met	Ala	Ala	Ala	Cys	Gly	Asp	Lys	Pro	Val	Asp	Ile	Ile	Leu	Lys	Gly	
			180					185					190			
gag	gaa	atc	ggc	gcg	aag	ggc	tat	atc	gac	ctc	acc	aca	tcc	gcc	atg	624
Glu	Glu	Ile	Gly	Ala	Lys	Gly	Tyr	Ile	Asp	Leu	Thr	Thr	Ser	Ala	Met	
		195					200					205				
gaa	gcc	ttc	gga	gcg	aag	gtg	gag	cgg	gtc	agc	aac	gcc	atc	tgg	cgc	672
Glu	Ala	Phe	Gly	Ala	Lys	Val	Glu	Arg	Val	Ser	Asn	Ala	Ile	Trp	Arg	
	210					215					220					
gtg	cat	ccg	acc	ggg	tac	acg	gca	acc	gat	ttc	cac	atc	gag	ccg	gat	720
Val	His	Pro	Thr	Gly	Tyr	Thr	Ala	Thr	Asp	Phe	His	Ile	Glu	Pro	Asp	
225					230					235					240	
gcc	tcc	gcc	gcc	acc	tat	ctc	tgg	ggc	gcg	gag	ctt	ctg	acc	ggc	ggg	768
Ala	Ser	Ala	Ala	Thr	Tyr	Leu	Trp	Gly	Ala	Glu	Leu	Leu	Thr	Gly	Gly	
				245				250						255		
gct	atc	gat	atc	ggc	acg	ccc	gct	gac	aaa	ttc	act	cag	ccg	gat	gcc	816
Ala	Ile	Asp	Ile	Gly	Thr	Pro	Ala	Asp	Lys	Phe	Thr	Gln	Pro	Asp	Ala	
			260					265					270			
aag	gct	tat	gag	gtc	atg	gcg	cag	ttt	ccg	cat	ctg	ccc	gcc	gaa	atc	864
Lys	Ala	Tyr	Glu	Val	Met	Ala	Gln	Phe	Pro	His	Leu	Pro	Ala	Glu	Ile	
		275					280					285				
gac	ggg	tcg	cag	atg	cag	gac	gcc	att	cca	acc	atc	gcg	gtt	atc	gcc	912
Asp	Gly	Ser	Gln	Met	Gln	Asp	Ala	Ile	Pro	Thr	Ile	Ala	Val	Ile	Ala	
	290					295					300					
gcc	ttc	aac	gag	acg	ccg	gtg	cgt	ttc	gtc	ggc	atc	gcc	aat	ctg	cgc	960
Ala	Phe	Asn	Glu	Thr	Pro	Val	Arg	Phe	Val	Gly	Ile	Ala	Asn	Leu	Arg	
305					310					315					320	
gtc	aag	gaa	tgc	gac	cgc	atc	cgc	gcc	gtc	tcg	ctc	ggc	ctc	aac	gaa	1008
Val	Lys	Glu	Cys	Asp	Arg	Ile	Arg	Ala	Val	Ser	Leu	Gly	Leu	Asn	Glu	
				325					330					335		
atc	cgc	gag	ggg	ctg	gcg	cat	gaa	gag	ggg	gac	gac	ctg	atc	gta	cat	1056
Ile	Arg	Glu	Gly	Leu	Ala	His	Glu	Glu	Gly	Asp	Asp	Leu	Ile	Val	His	
			340					345					350			
gcc	gat	ccg	tcg	ctc	gcc	ggg	cag	acg	gtc	gat	gcc	tcc	atc	gat	acc	1104
Ala	Asp	Pro	Ser	Leu	Ala	Gly	Gln	Thr	Val	Asp	Ala	Ser	Ile	Asp	Thr	
		355					360					365				
ttt	gcc	gat	cac	cgc	atc	gcc	atg	agt	ttt	gca	ctg	gcg	gcg	ctc	aag	1152
Phe	Ala	Asp	His	Arg	Ile	Ala	Met	Ser	Phe	Ala	Leu	Ala	Ala	Leu	Lys	
	370					375					380					
atc	gga	ggc	atc	gcc	atc	cag	aat	ccc	gcc	tgc	gtg	gcc	aag	acc	tat	1200
Ile	Gly	Gly	Ile	Ala	Ile	Gln	Asn	Pro	Ala	Cys	Val	Ala	Lys	Thr	Tyr	
385					390					395					400	
ccg	ggc	tac	tgg	aaa	gcg	ctt	gcc	tcg	ctc	ggc	gtc	gac	tat	acc	gaa	1248
Pro	Gly	Tyr	Trp	Lys	Ala	Leu	Ala	Ser	Leu	Gly	Val	Asp	Tyr	Thr	Glu	

aag gaa agc gct gcc gag cgg cag cac tga
Lys Glu Ser Ala Ala Glu Pro Gln His *
420 425

1278

4
<211> 425
<212> PRT
<213> Agrobacterium tumefaciens

Met	Ile	Glu	Leu	Thr	Ile	Thr	Pro	Pro	Gly	His	Pro	Leu	Ser	Gly	Lys
1				5					10					15	
Val	Glu	Pro	Pro	Gly	Ser	Lys	Ser	Ile	Thr	Asn	Arg	Ala	Leu	Leu	Leu
			20					25					30		
Ala	Gly	Leu	Ala	Lys	Gly	Lys	Ser	His	Leu	Ser	Gly	Ala	Leu	Lys	Ser
			35					40				45			
Asp	Asp	Thr	Leu	Tyr	Met	Ala	Glu	Ala	Leu	Arg	Glu	Met	Gly	Val	Lys
	50					55					60				
Val	Thr	Glu	Pro	Asp	Ala	Thr	Thr	Phe	Val	Val	Glu	Gly	Thr	Gly	Val
	65				70					75				80	
Leu	Gln	Gln	Pro	Glu	Lys	Pro	Leu	Phe	Leu	Gly	Asn	Ala	Gly	Thr	Ala
				85					90					95	
Thr	Arg	Phe	Leu	Thr	Ala	Ala	Gly	Ala	Leu	Val	Asp	Gly	Ala	Val	Ile
			100					105					110		
Ile	Asp	Gly	Asp	Glu	His	Met	Arg	Lys	Arg	Pro	Ile	Leu	Pro	Leu	Val
	115							120					125		
Gln	Ala	Leu	Arg	Ala	Leu	Gly	Val	Glu	Ala	Asp	Ala	Pro	Thr	Gly	Cys
	130					135					140				
Pro	Pro	Val	Thr	Val	Arg	Gly	Lys	Gly	Met	Gly	Phe	Pro	Lys	Gly	Ser
	145				150					155					160
Val	Thr	Ile	Asp	Ala	Asn	Leu	Ser	Ser	Gln	Tyr	Val	Ser	Ala	Leu	Leu
			165							170				175	
Met	Ala	Ala	Ala	Cys	Gly	Asp	Lys	Pro	Val	Asp	Ile	Ile	Leu	Lys	Gly
			180					185					190		
Glu	Glu	Ile	Gly	Ala	Lys	Gly	Tyr	Ile	Asp	Leu	Thr	Thr	Ser	Ala	Met
	195						200					205			
Glu	Ala	Phe	Gly	Ala	Lys	Val	Glu	Arg	Val	Ser	Asn	Ala	Ile	Trp	Arg
	210					215					220				
Val	His	Pro	Thr	Gly	Tyr	Thr	Ala	Thr	Asp	Phe	His	Ile	Glu	Pro	Asp
	225				230					235					240
Ala	Ser	Ala	Ala	Thr	Tyr	Leu	Trp	Gly	Ala	Glu	Leu	Leu	Thr	Gly	Gly
			245						250					255	
Ala	Ile	Asp	Ile	Gly	Thr	Pro	Ala	Asp	Lys	Phe	Thr	Gln	Pro	Asp	Ala
			260					265					270		
Lys	Ala	Tyr	Glu	Val	Met	Ala	Gln	Phe	Pro	His	Leu	Pro	Ala	Glu	Ile
	275						280					285			
Asp	Gly	Ser	Gln	Met	Gln	Asp	Ala	Ile	Pro	Thr	Ile	Ala	Val	Ile	Ala
	290					295					300				
Ala	Phe	Asn	Glu	Thr	Pro	Val	Arg	Phe	Val	Gly	Ile	Ala	Asn	Leu	Arg
	305				310					315					320
Val	Lys	Glu	Cys	Asp	Arg	Ile	Arg	Ala	Val	Ser	Leu	Gly	Leu	Asn	Glu
			325						330					335	
Ile	Arg	Glu	Gly	Leu											

Lys Glu Ser Ala Ala Glu Pro Gln His

<210> 5
 <211> 1398
 <212> DNA
 <213> Enterobacteriaceas

<220>
 <221> CDS
 <222> (103) ... (1398)

<400> 5
 aaaaaaggaa atgaactatg tgttgctgga aaaagtaggg aaggaggagg tgaagagtat 60
 tccactgggt caattagaaa aaatcattca aggattacca aa gtg aaa gta aca 114
 Val Lys Val Thr
 1
 ata cag ccc gga gat ctg act gga att atc cag tca ccc gct tca aaa 162
 Ile Gln Pro Gly Asp Leu Thr Gly Ile Ile Gln Ser Pro Ala Ser Lys
 5 10 15 20
 agt tgc atg cag cga gct tgt gct gct gca ctg gtt gca aaa gga ata 210
 Ser Ser Met Gln Arg Ala Cys Ala Ala Leu Val Ala Lys Gly Ile
 25 30 35
 agt gag atc att aat ccc ggt cat agc aat gat gat aaa gct gcc agg 258
 Ser Glu Ile Ile Asn Pro Gly His Ser Asn Asp Asp Lys Ala Ala Arg
 40 45 50
 gat att gta agc cgg ctt ggt gcc agg ctt gaa gat cag cct gat ggt 306
 Asp Ile Val Ser Arg Leu Gly Ala Arg Leu Glu Asp Gln Pro Asp Gly
 55 60 65
 tct ttg cag ata aca agt gaa ggc gta aaa cct gtc gct cct ttt att 354
 Ser Leu Gln Ile Thr Ser Glu Gly Val Lys Pro Val Ala Pro Phe Ile
 70 75 80
 gac tgc ggt gaa tct ggt tta agt atc cgg atg ttt act ccg att gtt 402
 Asp Cys Gly Glu Ser Gly Leu Ser Ile Arg Met Phe Thr Pro Ile Val
 85 90 95 100
 gcg ttg agt aaa gaa gag gtg acg atc aaa gga tct gga agc ctt gtt 450
 Ala Leu Ser Lys Glu Val Thr Ile Lys Gly Ser Gly Ser Leu Val
 105 110 115
 aca aga cca atg gat ttc ttt gat gaa att ctt ccg cat ctc ggt gta 498
 Thr Arg Pro Met Asp Phe Phe Asp Glu Ile Leu Pro His Leu Gly Val
 120 125 130
 aaa gtt aaa tct aac cag ggt aaa ttg cct ctc gtt ata cag ggg cca 546
 Lys Val Lys Ser Asn Gln Gly Lys Leu Pro Leu Val Ile Gln Gly Pro
 135 140 145
 ttg aaa cca gca gac gtt acg gtt gat ggg tcc tta agc tct cag ttc 594
 Leu Lys Pro Ala Asp Val Thr Val Asp Gly Ser Leu Ser Ser Gln Phe
 150 155 160
 ctt aca ggt ttg ttg ctt gca tat gcg gcc gca gat gca agc gat gtt 642
 Leu Thr Gly Leu Leu Leu Ala Tyr Ala Ala Asp Ala Ser Asp Val
 165 170 175 180
 gcg ata aaa gta acg aat ctc aaa agc cgt ccg tat atc gat ctt aca 690
 Ala Ile Lys Val Thr Asn Leu Lys Ser Arg Pro Tyr Ile Asp Leu Thr
 185 190 195

ctg gat gtg atg aag cgg ttt ggt ttg aag act ccc gag aat cga aac 738
 Leu Asp Val Met Lys Arg Phe Gly Leu Lys Thr Pro Glu Asn Arg Asn
 200 205 210

tat gaa gag ttt tat ttc aaa gcc ggg aat gta tat gat gaa acg aaa 786
 Tyr Glu Glu Phe Tyr Phe Lys Ala Gly Asn Val Tyr Asp Glu Thr Lys
 215 220 225

atg caa cga tac acc gta gaa ggc gac tgg agc ggt ggt gct ttt tta 834
 Met Gln Arg Tyr Thr Val Glu Gly Asp Trp Ser Gly Gly Ala Phe Leu
 230 235 240

ctg gta gcg ggg gct att gcc ggg ccg atc acg gta aga ggt ttg gat 882
 Leu Val Ala Gly Ala Ile Ala Gly Pro Ile Thr Val Arg Gly Leu Asp
 245 250 255 260

ata gct tcg acg cag gct gat aaa gcg atc gtt cag gct ttg atg agt 930
 Ile Ala Ser Thr Gln Ala Asp Lys Ala Ile Val Gln Ala Leu Met Ser
 265 270 275

gcg aac gca ggt att gcg att gat gca aaa gag atc aaa ctt cat cct 978
 Ala Asn Ala Gly Ile Ala Ile Asp Ala Lys Glu Ile Lys Leu His Pro
 280 285 290

gct gat ctc aat gca ttt gaa ttt gat gct act gat tgc ccg gat ctt 1026
 Ala Asp Leu Asn Ala Phe Glu Phe Asp Ala Thr Asp Cys Pro Asp Leu
 295 300 305

ttt ccg cca ttg gtt gct ttg gcg tct tat tgc aaa gga gaa aca aag 1074
 Phe Pro Pro Leu Val Ala Leu Ala Ser Tyr Cys Lys Gly Glu Thr Lys
 310 315 320

atc aaa ggc gta agc agg ctg gcg cat aaa gaa agt gac aga gga ttg 1122
 Ile Lys Gly Val Ser Arg Leu Ala His Lys Glu Ser Asp Arg Gly Leu
 325 330 335 340

acg ctg cag gac gag ttc ggg aaa atg ggt gtt gaa atc cac ctt gag 1170
 Thr Leu Gln Asp Glu Phe Gly Lys Met Gly Val Glu Ile His Leu Glu
 345 350 355

gga gat ctg atg cgc gtg atc gga ggg aaa ggc gta aaa gga gct gaa 1218
 Gly Asp Leu Met Arg Val Ile Gly Gly Lys Gly Val Lys Gly Ala Glu
 360 365 370

gtt agt tca agg cac gat cat cgc att gcg atg gct tgc gcg gtg gct 1266
 Val Ser Ser Arg His Asp His Arg Ile Ala Met Ala Cys Ala Val Ala
 375 380 385

gct tta aaa gct gtg ggt gaa aca acc atc gaa cat gca gaa gcg gtg 1314
 Ala Leu Lys Ala Val Gly Glu Thr Thr Ile Glu His Ala Glu Ala Val
 390 395 400

aat aaa tcc tac ccg gat ttt tac agc gat ctt aaa caa ctt ggc ggt 1362
 Asn Lys Ser Tyr Pro Asp Phe Tyr Ser Asp Leu Lys Gln Leu Gly Gly
 405 410 415 420

gtt gta tct tta aac cat caa ttt aat ttc tca tga 1398
 Val Val Ser Leu Asn His Gln Phe Asn Phe Ser *
 425 430

<210> 6
 <211> 431
 <212> PRT
 <213> Enterobacteriaceas
 <400> 6

Val	Lys	Val	Thr	Ile	Gln	Pro	Gly	Asp	Leu	Thr	Gly	Ile	Ile	Gln	Ser
1				5					10					15	
Pro	Ala	Ser	Lys	Ser	Ser	Met	Gln	Arg	Ala	Cys	Ala	Ala	Ala	Leu	Val
			20					25						30	
Ala	Lys	Gly	Ile	Ser	Glu	Ile	Ile	Asn	Pro	Gly	His	Ser	Asn	Asp	Asp
		35					40					45			
Lys	Ala	Ala	Arg	Asp	Ile	Val	Ser	Arg	Leu	Gly	Ala	Arg	Leu	Glu	Asp
	50					55					60				
Gln	Pro	Asp	Gly	Ser	Leu	Gln	Ile	Thr	Ser	Glu	Gly	Val	Lys	Pro	Val
65					70					75				80	
Ala	Pro	Phe	Ile	Asp	Cys	Gly	Glu	Ser	Gly	Leu	Ser	Ile	Arg	Met	Phe
			85						90					95	
Thr	Pro	Ile	Val	Ala	Leu	Ser	Lys	Glu	Glu	Val	Thr	Ile	Lys	Gly	Ser
			100					105					110		
Gly	Ser	Leu	Val	Thr	Arg	Pro	Met	Asp	Phe	Phe	Asp	Glu	Ile	Leu	Pro
		115					120					125			
His	Leu	Gly	Val	Lys	Val	Lys	Ser	Asn	Gln	Gly	Lys	Leu	Pro	Leu	Val
	130					135					140				
Ile	Gln	Gly	Pro	Leu	Lys	Pro	Ala	Asp	Val	Thr	Val	Asp	Gly	Ser	Leu
145					150					155				160	
Ser	Ser	Gln	Phe	Leu	Thr	Gly	Leu	Leu	Leu	Ala	Tyr	Ala	Ala	Ala	Asp
			165						170					175	
Ala	Ser	Asp	Val	Ala	Ile	Lys	Val	Thr	Asn	Leu	Lys	Ser	Arg	Pro	Tyr
			180					185					190		
Ile	Asp	Leu	Thr	Leu	Asp	Val	Met	Lys	Arg	Phe	Gly	Leu	Lys	Thr	Pro
	195					200						205			
Glu	Asn	Arg	Asn	Tyr	Glu	Glu	Phe	Tyr	Phe	Lys	Ala	Gly	Asn	Val	Tyr
	210					215					220				
Asp	Glu	Thr	Lys	Met	Gln	Arg	Tyr	Thr	Val	Glu	Gly	Asp	Trp	Ser	Gly
225					230					235				240	
Gly	Ala	Phe	Leu	Leu	Val	Ala	Gly	Ala	Ile	Ala	Gly	Pro	Ile	Thr	Val
			245						250				255		
Arg	Gly	Leu	Asp	Ile	Ala	Ser	Thr	Gln	Ala	Asp	Lys	Ala	Ile	Val	Gln
		260						265					270		
Ala	Leu	Met	Ser	Ala	Asn	Ala	Gly	Ile	Ala	Ile	Asp	Ala	Lys	Glu	Ile
	275						280					285			
Lys	Leu	His	Pro	Ala	Asp	Leu	Asn	Ala	Phe	Glu	Phe	Asp	Ala	Thr	Asp
	290					295					300				
Cys	Pro	Asp	Leu	Phe	Pro	Pro	Leu	Val	Ala	Leu	Ala	Ser	Tyr	Cys	Lys
305					310					315				320	
Gly	Glu	Thr	Lys	Ile	Lys	Gly	Val	Ser	Arg	Leu	Ala	His	Lys	Glu	Ser
			325						330				335		
Asp	Arg	Gly	Leu	Thr	Leu	Gln	Asp	Glu	Phe	Gly	Lys	Met	Gly	Val	Glu
		340						345					350		
Ile	His	Leu	Glu	Gly	Asp	Leu	Met	Arg	Val	Ile	Gly	Gly	Lys	Gly	Val
	355					360						365			
Lys	Gly	Ala	Glu	Val	Ser	Ser	Arg	His	Asp	His	Arg	Ile	Ala	Met	Ala
	370					375					380				
Cys	Ala	Val	Ala	Ala	Leu	Lys	Ala	Val	Gly	Glu	Thr	Thr	Ile	Glu	His
385					390					395				400	
Ala	Glu	Ala	Val	Asn	Lys	Ser	Tyr	Pro	Asp	Phe	Tyr	Ser	Asp	Leu	Lys
			405						410					415	
Gln	Leu	Gly	Gly	Val	Val	Ser	Leu	Asn	His	Gln	Phe	Asn	Phe	Ser	
			420					425					430		

<210> 7
 <211> 1257
 <212> DNA
 <213> Brevundomonas vesicularis

<220>
 <221> CDS
 <222> (1) ... (1257)

<400> 7

atg atg atg ggt aga gcc aaa ctc acg att atc ccg ccg ggc aag cct	48
Met Met Met Gly Arg Ala Lys Leu Thr Ile Ile Pro Pro Gly Lys Pro	
1 5 10 15	
ttg acc gga cgc gcc atg ccg ccg gga tcg aag tcg atc acc aac cgc	96
Leu Thr Gly Arg Ala Met Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg	
20 25 30	
gca ttg ctg ctc gcc ggc ctc gcc aag ggc acg agc cgg cta acc ggt	144
Ala Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly	
35 40 45	
gcg ctg aag agc gac gat acc cgc tat atg gcc gaa gcg ctg cgt gcg	192
Ala Leu Lys Ser Asp Asp Thr Arg Tyr Met Ala Glu Ala Leu Arg Ala	
50 55 60	
atg ggt gta acg atc gac gag ccc gac gac acc acg ttc atc gtc aaa	240
Met Gly Val Thr Ile Asp Glu Pro Asp Asp Thr Thr Phe Ile Val Lys	
65 70 75 80	
ggc agc ggc aag ctg cag ccg ccg gca gcc ccg ctt ttc ctc ggc aat	288
Gly Ser Gly Lys Leu Gln Pro Pro Ala Ala Pro Leu Phe Leu Gly Asn	
85 90 95	
gcc ggc acg gca acg cgc ttc ctg acg gcg gcc gcg gca ctg gtg gac	336
Ala Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Ala Ala Leu Val Asp	
100 105 110	
ggc aag gtc atc gtc gac ggc gat gcc cat atg cgc aag cgg ccg atc	384
Gly Lys Val Ile Val Asp Gly Asp Ala His Met Arg Lys Arg Pro Ile	
115 120 125	
gga ccg cta gtc gac gcg ttg cgc tcg ctc ggc atc gat gcc tcg gct	432
Gly Pro Leu Val Asp Ala Leu Arg Ser Leu Gly Ile Asp Ala Ser Ala	
130 135 140	
gaa acc ggc tgc ccg cca gtc acg atc aac ggc acc ggc cgc ttc gag	480
Glu Thr Gly Cys Pro Pro Val Thr Ile Asn Gly Thr Gly Arg Phe Glu	
145 150 155 160	
gca agc cgc gtg cag atc gat ggc ggc ctg tcc agc cag tat gtc tcg	528
Ala Ser Arg Val Gln Ile Asp Gly Gly Leu Ser Ser Gln Tyr Val Ser	
165 170 175	
gcg ctc ctg atg atg gcc gcc ggc ggc gat cgc gct gtc gat gtc gag	576
Ala Leu Leu Met Met Ala Ala Gly Gly Asp Arg Ala Val Asp Val Glu	
180 185 190	
ctt ctc ggc gaa cat atc ggc gct ctc ggc tat atc gac ctg acc gtt	624
Leu Leu Gly Glu His Ile Gly Ala Leu Gly Tyr Ile Asp Leu Thr Val	
195 200 205	
gcc gcc atg cgc gct ttc ggc gcg aag gtt gag cgt gtg agc ccg gtc	672
Ala Ala Met Arg Ala Phe Gly Ala Lys Val Glu Arg Val Ser Pro Val	
210 215 220	
gcc tgg cgc gtc gag ccc acc ggc tat cat gcg gcc gac ttc gtg atc	720
Ala Trp Arg Val Glu Pro Thr Gly Tyr His Ala Ala Asp Phe Val Ile	
225 230 235 240	
gag ccg gat gcc tct gct gcg acc tat ctc tgg gcc gcc gaa gtt ctg	768
Glu Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Glu Val Leu	
245 250 255	
agc ggc ggc aag atc gat ctc ggc acg ccg gcg gaa cag ttc tcg caa	816

Ser Gly Gly Lys Ile Asp Leu Gly Thr Pro Ala Glu Gln Phe Ser Gln
 260 265 270
 ccg gat gcg aaa gcc tat gat ctg att tcg aaa ttc ccg cat ctg cct 864
 Pro Asp Ala Lys Ala Tyr Asp Leu Ile Ser Lys Phe Pro His Leu Pro
 275 280 285
 gct gtc atc gac gcc tcg cag atg cag gac gcc atc ccg acg ctc gcc 912
 Ala Val Ile Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala
 290 295 300
 gtt ctc gcc gct ttc aac gaa atg cct gtg cgc ttc gtc ggt atc gaa 960
 Val Leu Ala Ala Phe Asn Glu Met Pro Val Arg Phe Val Gly Ile Glu
 305 310 315 320
 aac ctg cgc gtc aag gaa tgc gat cgt atc cgc gcg ctc tcg agc gcc 1008
 Asn Leu Arg Val Lys Glu Cys Asp Arg Ile Arg Ala Leu Ser Ser Gly
 325 330 335
 cta tcc cgc atc gtt ccg aac ctc gcc acg gaa gag gcc gac gat ctc 1056
 Leu Ser Arg Ile Val Pro Asn Leu Gly Thr Glu Glu Gly Asp Asp Leu
 340 345 350
 atc atc gcc tcc gat ccg agc ctt gcc gcc aaa atc ctg acc gca gag 1104
 Ile Ile Ala Ser Asp Pro Ser Leu Ala Gly Lys Ile Leu Thr Ala Glu
 355 360 365
 atc gat agc ttt gcc gat cac cgc atc gcc atg agc ttt gcg ctg gcc 1152
 Ile Asp Ser Phe Ala Asp His Arg Ile Ala Met Ser Phe Ala Leu Ala
 370 375 380
 gcc ctg aag atc gcc gcc att acc att ctc gac ccc gac tgc gtc gcc 1200
 Gly Leu Lys Ile Gly Gly Ile Thr Ile Leu Asp Pro Asp Cys Val Ala
 385 390 395 400
 aag aca ttc ccg tcc tac tgg aat gtg ctg tct tcg ctg ggg gtc gcc 1248
 Lys Thr Phe Pro Ser Tyr Trp Asn Val Leu Ser Ser Leu Gly Val Ala
 405 410 415
 tac gaa gac 1257
 Tyr Glu Asp

<210> 8
 <211> 419
 <212> PRT
 <213> Brevundomonas vesicularis

<400> 8
 Met Met Met Gly Arg Ala Lys Leu Thr Ile Ile Pro Pro Gly Lys Pro
 1 5 10 15
 Leu Thr Gly Arg Ala Met Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg
 20 25 30
 Ala Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly
 35 40 45
 Ala Leu Lys Ser Asp Asp Thr Arg Tyr Met Ala Glu Ala Leu Arg Ala
 50 55 60
 Met Gly Val Thr Ile Asp Glu Pro Asp Asp Thr Thr Phe Ile Val Lys
 65 70 75 80
 Gly Ser Gly Lys Leu Gln Pro Pro Ala Ala Pro Leu Phe Leu Gly Asn
 85 90 95
 Ala Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Ala Ala Leu Val Asp
 100 105 110
 Gly Lys Val Ile Val Asp Gly Asp Ala His Met Arg Lys Arg Pro Ile
 115 120 125

Gly Pro Leu Val Asp Ala Leu Arg Ser Leu Gly Ile Asp Ala Ser Ala
 130 135 140
 Glu Thr Gly Cys Pro Pro Val Thr Ile Asn Gly Thr Gly Arg Phe Glu
 145 150 155 160
 Ala Ser Arg Val Gln Ile Asp Gly Gly Leu Ser Ser Gln Tyr Val Ser
 165 170 175
 Ala Leu Leu Met Met Ala Ala Gly Gly Asp Arg Ala Val Asp Val Glu
 180 185 190
 Leu Leu Gly Glu His Ile Gly Ala Leu Gly Tyr Ile Asp Leu Thr Val
 195 200 205
 Ala Ala Met Arg Ala Phe Gly Ala Lys Val Glu Arg Val Ser Pro Val
 210 215 220
 Ala Trp Arg Val Glu Pro Thr Gly Tyr His Ala Ala Asp Phe Val Ile
 225 230 235 240
 Glu Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Glu Val Leu
 245 250 255
 Ser Gly Gly Lys Ile Asp Leu Gly Thr Pro Ala Glu Gln Phe Ser Gln
 260 265 270
 Pro Asp Ala Lys Ala Tyr Asp Leu Ile Ser Lys Phe Pro His Leu Pro
 275 280 285
 Ala Val Ile Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala
 290 295 300
 Val Leu Ala Ala Phe Asn Glu Met Pro Val Arg Phe Val Gly Ile Glu
 305 310 315 320
 Asn Leu Arg Val Lys Glu Cys Asp Arg Ile Arg Ala Leu Ser Ser Gly
 325 330 335
 Leu Ser Arg Ile Val Pro Asn Leu Gly Thr Glu Glu Gly Asp Asp Leu
 340 345 350
 Ile Ile Ala Ser Asp Pro Ser Leu Ala Gly Lys Ile Leu Thr Ala Glu
 355 360 365
 Ile Asp Ser Phe Ala Asp His Arg Ile Ala Met Ser Phe Ala Leu Ala
 370 375 380
 Gly Leu Lys Ile Gly Gly Ile Thr Ile Leu Asp Pro Asp Cys Val Ala
 385 390 395 400
 Lys Thr Phe Pro Ser Tyr Trp Asn Val Leu Ser Ser Leu Gly Val Ala
 405 410 415
 Tyr Glu Asp

<210> 9
 <211> 1385
 <212> DNA
 <213> Ochrobactrum/Brucella

<221> CDS
 <222> (142) ... (1380)

<400> 9
 ggcgcagccca ataaataacg cagatcaccc actcactcga caagcatggc gtgtttgcct 60
 gatgattcgg gtccgcacgt cggccactcc acgccacctt gccttgacca ggagccttgt 120
 accttgagtt cgcagaaaac c gtg acc gtt aca ccg ccc aac ttc ccc ctc 171
 Val Thr Val Thr Pro Pro Asn Phe Pro Leu
 1 5 10
 act ggc aag gtc gcg ccc ccc ggc tcc aaa tcc att acc aac cgt gcg 219
 Thr Gly Lys Val Ala Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala
 15 20 25
 ctg ttg ctc gcg gca ctg gcc aag ggc acc agc cgt ttg agc ggt gcg 267
 Leu Leu Leu Ala Ala Leu Ala Lys Gly Thr Ser Arg Leu Ser Gly Ala
 30 35 40
 ctc aaa agc gat gac acg cgc cac atg tcg gtc gcc ctg cgg cag atg 315
 Leu Lys Ser Asp Asp Thr Arg His Met Ser Val Ala Leu Arg Gln Met
 45 50 55

ggc gtc acc atc gac gag ccg gac gac acc acc ttt gtc gtc acc agc Gly Val Thr Ile Asp Glu Pro Asp Asp Thr Thr Phe Val Val Thr Ser 60 65 70	363
caa ggc tgc ctg caa ttg ccg gcc cag ccg ttg ttc ctc ggc aac gcc Gln Gly Ser Leu Gln Leu Pro Ala Gln Pro Leu Phe Leu Gly Asn Ala 75 80 85 90	411
ggc acc gcc atg cgc ttt ctc acg gct gcc gtc gcc acc gtg caa ggc Gly Thr Ala Met Arg Phe Leu Thr Ala Ala Val Ala Thr Val Gln Gly 95 100 105	459
acc gtg gta ctg gac ggt gac gag tac atg caa aaa cgc ccg atc ggc Thr Val Val Leu Asp Gly Asp Glu Tyr Met Gln Lys Arg Pro Ile Gly 110 115 120	507
ccg ctg ctg gcc acc ctg ggc cag aac ggc atc cag gtc gac agc ccc Pro Leu Leu Ala Thr Leu Gly Gln Asn Gly Ile Gln Val Asp Ser Pro 125 130 135	555
acc ggt tgc cca ccg gta acg gtg cat ggc gcg ggc aag gtc cag gcc Thr Gly Cys Pro Pro Val Thr Val His Gly Ala Gly Lys Val Gln Ala 140 145 150	603
agg cgt ttt gag att gac gga ggc ttg tcc agc cag tac gta tgc gcc Arg Arg Phe Glu Ile Asp Gly Gly Leu Ser Ser Gln Tyr Val Ser Ala 155 160 165 170	651
ctg ctg atg ctg gcg gcg tgc ggc gaa gca ccg att gaa gtg gcg ctg Leu Leu Met Leu Ala Ala Cys Gly Glu Ala Pro Ile Glu Val Ala Leu 175 180 185	699
acc ggc aag gac atc ggc gcc cgt ggc tat gtg gac ctg acc ctc gat Thr Gly Lys Asp Ile Gly Ala Arg Gly Tyr Val Asp Leu Thr Leu Asp 190 195 200	747
tgc atg cgt gcg ttc ggg gcc cag gta gac atc gtg gac gac acc acc Cys Met Arg Ala Phe Gly Ala Gln Val Asp Ile Val Asp Asp Thr Thr 205 210 215	795
tgg cgc gtg gcc ccc acg ggc tat acc gcc cat gat tac ctg atc gaa Trp Arg Val Ala Pro Thr Gly Tyr Thr Ala His Asp Tyr Leu Ile Glu 220 225 230	843
ccc gac gct tcc gcc gcc act tac ctg tgg gcc gca gaa gta ctg acc Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Glu Val Leu Thr 235 240 245 250	891
ggc ggc cgt atc gat att ggt gta gcc gcg cag gac ttc acc cag ccc Gly Gly Arg Ile Asp Ile Gly Val Ala Ala Gln Asp Phe Thr Gln Pro 255 260 265	939
gac gcc aag gca cag gcc gtg atc gcg caa ttc ccg aac atg cag gcc Asp Ala Lys Ala Gln Ala Val Ile Ala Gln Phe Pro Asn Met Gln Ala 270 275 280	987
acg gtg gtg ggt tca caa atg cag gat gcg atc ccg acc ctg gcg gtg Thr Val Val Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val 285 290 295	1035
ctc gcc gca ttc aac aat acc ccg gtg cgc ttc act gaa ctg gcg aac Leu Ala Ala Phe Asn Asn Thr Pro Val Arg Phe Thr Glu Leu Ala Asn 300 305 310	1083
ctg cgc gtc aag gaa tgt gac cgc gtg cag gcg ctg cac gat ggc ctc	1131

Leu Arg Val Lys Glu Cys Asp Arg Val Gln Ala Leu His Asp Gly Leu
 315 320 325 330
 aac gaa att cgc ccg ggc ctg gcg acc atc gaa ggt gat gac ctg ctg 1179
 Asn Glu Ile Arg Pro Gly Leu Ala Thr Ile Glu Gly Asp Asp Leu Leu
 335 340 345
 gtt gcc agc gac ccc gct ttg gct ggc acc gcc tgc acc gca ctg atc 1227
 Val Ala Ser Asp Pro Ala Leu Ala Gly Thr Ala Cys Thr Ala Leu Ile
 350 355 360
 gat acc cac gcc gac cat cgc atc gcc atg tgc ttt gcc ctg gcc ggg 1275
 Asp Thr His Ala Asp His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly
 365 370 375
 ctg aaa gtc tgc ggc att cgc atc caa gac cct gat tgc gta gcc aag 1323
 Leu Lys Val Ser Gly Ile Arg Ile Gln Asp Pro Asp Cys Val Ala Lys
 380 385 390
 acc tac cct gac tac tgg aaa gcg ctg gcc agc ctg ggc gtt cac tta 1371
 Thr Tyr Pro Asp Tyr Trp Lys Ala Leu Ala Ser Leu Gly Val His Leu
 395 400 405 410
 agc tac tga cacac 1385
 Ser Tyr *

<210> 10
 <211> 444
 <212> PRT
 <213> Ochrobactrum/Brucella

<400> 10
 Met Ala Cys Leu Pro Asp Asp Ser Gly Pro His Val Gly His Ser Thr
 1 5 10 15
 Pro Pro Cys Leu Asp Gln Glu Pro Cys Thr Leu Ser Ser Gln Lys Thr
 20 25 30
 Val Thr Val Thr Pro Pro Asn Phe Pro Leu Thr Gly Lys Val Ala Pro
 35 40 45
 Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Ala Leu
 50 55 60
 Ala Lys Gly Thr Ser Arg Leu Ser Gly Ala Leu Lys Ser Asp Asp Thr
 65 70 75 80
 Arg His Met Ser Val Ala Leu Arg Gln Met Gly Val Thr Ile Asp Glu
 85 90 95
 Pro Asp Asp Thr Thr Phe Val Val Thr Ser Gln Gly Ser Leu Gln Leu
 100 105 110
 Pro Ala Gln Pro Leu Phe Leu Gly Asn Ala Gly Thr Ala Met Arg Phe
 115 120 125
 Leu Thr Ala Ala Val Ala Thr Val Gln Gly Thr Val Val Leu Asp Gly
 130 135 140
 Asp Glu Tyr Met Gln Lys Arg Pro Ile Gly Pro Leu Leu Ala Thr Leu
 145 150 155 160
 Gly Gln Asn Gly Ile Gln Val Asp Ser Pro Thr Gly Cys Pro Pro Val
 165 170 175
 Thr Val His Gly Ala Gly Lys Val Gln Ala Arg Arg Phe Glu Ile Asp
 180 185 190
 Gly Gly Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Leu Ala Ala
 195 200 205
 Cys Gly Glu Ala Pro Ile Glu Val Ala Leu Thr Gly Lys Asp Ile Gly
 210 215 220
 Ala Arg Gly Tyr Val Asp Leu Thr Leu Asp Cys Met Arg Ala Phe Gly
 225 230 235 240
 Ala Gln Val Asp Ile Val Asp Asp Thr Thr Trp Arg Val Ala Pro Thr
 245 250 255

Gly Tyr Thr Ala His Asp Tyr Leu Ile Glu Pro Asp Ala Ser Ala Ala
 260 265 270
 Thr Tyr Leu Trp Ala Ala Glu Val Leu Thr Gly Gly Arg Ile Asp Ile
 275 280 285
 Gly Val Ala Ala Gln Asp Phe Thr Gln Pro Asp Ala Lys Ala Gln Ala
 290 295 300
 Val Ile Ala Gln Phe Pro Asn Met Gln Ala Thr Val Val Gly Ser Gln
 305 310 315 320
 Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn Asn
 325 330 335
 Thr Pro Val Arg Phe Thr Glu Leu Ala Asn Leu Arg Val Lys Glu Cys
 340 345 350
 Asp Arg Val Gln Ala Leu His Asp Gly Leu Asn Glu Ile Arg Pro Gly
 355 360 365
 Leu Ala Thr Ile Glu Gly Asp Asp Leu Leu Val Ala Ser Asp Pro Ala
 370 375 380
 Leu Ala Gly Thr Ala Cys Thr Ala Leu Ile Asp Thr His Ala Asp His
 385 390 395 400
 Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Val Ser Gly Ile
 405 410 415
 Arg Ile Gln Asp Pro Asp Cys Val Ala Lys Thr Tyr Pro Asp Tyr Trp
 420 425 430
 Lys Ala Leu Ala Ser Leu Gly Val His Leu Ser Tyr
 435 440

<210> 11
 <211> 1278
 <212> DNA
 <213> Desconhecido

<220>
 <223> Isolado da amostra de solo

<221> CDS
 <222> (1) ... (1278)

<400> 11
 atg atc gaa ctg acc atc acc ccg ccc ggc cac ccg ctt tcc ggc aag 48
 Met Ile Glu Leu Thr Ile Thr Pro Pro Gly His Pro Leu Ser Gly Lys
 1 5 10 15
 gtg gag ccg ccc ggt tcc aaa tcc att acc aac cgt gca ctt ctg ctg 96
 Val Glu Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu
 20 25 30
 gcc ggg ctc gcc aag ggc aaa agc cgt ctc acg ggc gcg ctg aaa agc 144
 Ala Gly Leu Ala Lys Gly Lys Ser Arg Leu Thr Gly Ala Leu Lys Ser
 35 40 45
 gac gat acg ctt tac atg gca gaa gcg ctg cgt gag atg ggt gtc aag 192
 Asp Asp Thr Leu Tyr Met Ala Glu Ala Leu Arg Glu Met Gly Val Lys
 50 55 60
 gta acc gag cct gac gcg acc acc ttc gtg gtg gag agt tca ggt ggg 240
 Val Thr Glu Pro Asp Ala Thr Thr Phe Val Val Glu Ser Ser Gly Gly
 65 70 75 80
 ttg cat cag ccg gaa aag ccg ctt ttc ctc ggc aat gcc ggc aca gct 288
 Leu His Gln Pro Glu Lys Pro Leu Phe Leu Gly Asn Ala Gly Thr Ala
 85 90 95
 acc cgc ttt ctc acc gcc gct gcc gcc ctt gtg gat ggc gcc gtc atc 336
 Thr Arg Phe Glu Thr Ala Ala Ala Leu Val Asp Gly Ala Val Ile
 100 105 110

atc gat ggc gac gag cat atg cgc aaa cgc ccg atc atg ccg ctg gtg Ile Asp Gly Asp Glu His Met Arg Lys Arg Pro Ile Met Pro Leu Val 115 120 125	384
gaa gcc ctg cgc tcc ctc ggc gtt gag gcg gag gcg ccg acc ggc tgc Glu Ala Leu Arg Ser Leu Gly Val Glu Ala Glu Ala Pro Thr Gly Cys 130 135 140	432
ccg ccc gtc acc gtc tgc ggt aag ggt act ggc ttc ccg aag ggc agc Pro Pro Val Thr Val Cys Gly Lys Gly Thr Gly Phe Pro Lys Gly Ser 145 150 155 160	480
gtc acg atc gac gcc aac ctt tcc agc cag tat gtg tcc gca ctt ctg Val Thr Ile Asp Ala Asn Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu 165 170 175	528
atg gcc gcc gcc tgc ggc gac aag cct gtc gat atc atc ctc aaa ggt Met Ala Ala Cys Gly Asp Lys Pro Val Asp Ile Ile Leu Lys Gly 180 185 190	576
gag gaa atc ggc gcg aag ggc tat atc gat ctc acc aca tcg gcc atg Glu Glu Ile Gly Ala Lys Gly Tyr Ile Asp Leu Thr Ser Ala Met 195 200 205	624
gaa gcc ttc ggc gca aag gtg gag cgg gtc agc aac gcc atc tgg cgc Glu Ala Phe Gly Ala Lys Val Glu Arg Val Ser Asn Ala Ile Trp Arg 210 215 220	672
gtg cat ccg acc ggc tac acg gcg acc gat ttc cat atc gag ccg gat Val His Pro Thr Gly Tyr Thr Ala Thr Asp Phe His Ile Glu Pro Asp 225 230 235 240	720
gcc tcg gcc gcc acc tat ctc tgg ggc gct gag ctt ttg acc ggc ggc Ala Ser Ala Ala Thr Tyr Leu Trp Gly Ala Glu Leu Leu Thr Gly Gly 245 250 255	768
gcc atc gat atc ggt acg ccg gcc gac aag ttc acc cag ccg gat gcc Ala Ile Asp Ile Gly Thr Pro Ala Asp Lys Phe Thr Gln Pro Asp Ala 260 265 270	816
aag gcc cat gag gtc atg gcg caa ttt ccg cat ctg ccc gcc gaa atc Lys Ala His Glu Val Met Ala Gln Phe Pro His Leu Pro Ala Glu Ile 275 280 285	864
gac ggt tcg cag atg cag gat gcc att ccc acc att gcc gtt ctc gcc Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Ile Ala Val Leu Ala 290 295 300	912
gcc ttt aac gaa acg ccg gtg cgt ttc gtc ggc atc gcc aat ctg cgc Ala Phe Asn Glu Thr Pro Val Arg Phe Val Gly Ile Ala Asn Leu Arg 305 310 315 320	960
gtc aag gag tgc gac cga atc cgc gcc gtc tca ctc ggc ctc aac gaa Val Lys Glu Cys Asp Arg Ile Arg Ala Val Ser Leu Gly Leu Asn Glu 325 330 335	1008
atc cgc gat ggt ctg gcg cat gag gaa ggc gac gac ttg atc gtg cat Ile Arg Asp Gly Leu Ala His Glu Glu Gly Asp Asp Leu Ile Val His 340 345 350	1056
tcc gat cct tcg ctt gcg ggc cag acg gtg aat gcc tcc atc gac act Ser Asp Pro Ser Leu Ala Gly Gln Thr Val Asn Ala Ser Ile Asp Thr 355 360 365	1104
ttc gcc gac cac cgt atc gcc atg agc ttt gcg ctg gcg gcg ctg aag Phe Ala Asp His Arg Ile Ala Met Ser Phe Ala Leu Ala Ala Leu Lys 370 375 380 385 390	1152

370 375 380
 atc ggc ggc att gcc atc cag aat ccg gcc tgc gtg ggc aag acc tat 1200
 ile Gly Gly ile Ala ile Gln Asn Pro Ala Cys Val Gly Lys Thr Tyr
 385 390 395 400

 ccc ggt tac tgg aag gcg ctc gcc tcg ctg gga gtc gaa tac tcg gaa 1248
 Pro Gly Tyr Trp Lys Ala Leu Ala Ser Leu Gly Val Glu Tyr Ser Glu
 405 410 415

 aag gaa acc gct gcc gag ccg cag cat tag 1278
 Lys Glu Thr Ala Ala Glu Pro Gln His *
 420 425

 <210> 12
 <211> 425
 <212> PRT
 <213> Desconhecido

 <220>
 <223> Isolado da amostra de solo

 <400> 12
 Met ile Glu leu Thr ile Thr Pro Pro Gly His Pro leu Ser Gly Lys
 1 5 10 15
 Val Glu Pro Pro Gly Ser Lys Ser ile Thr Asn Arg Ala leu leu leu
 20 25 30
 Ala Gly leu Ala Lys Gly Lys Ser Arg leu Thr Gly Ala leu Lys Ser
 35 40 45
 Asp Asp Thr leu Tyr Met Ala Glu Ala leu Arg Glu Met Gly Val Lys
 50 55 60
 Val Thr Glu Pro Asp Ala Thr Thr Phe Val Val Glu Ser Ser Gly Gly
 65 70 75 80
 leu His Gln Pro Glu Lys Pro leu Phe leu Gly Asn Ala Gly Thr Ala
 85 90 95
 Thr Arg Phe leu Thr Ala Ala Ala leu Val Asp Gly Ala Val ile
 100 105 110
 ile Asp Gly Asp Glu His Met Arg Lys Arg Pro ile Met Pro leu Val
 115 120 125
 Glu Ala leu Arg Ser leu Gly Val Glu Ala Glu Ala Pro Thr Gly Cys
 130 135 140
 Pro Pro Val Thr Val Cys Gly Lys Gly Thr Gly Phe Pro Lys Gly Ser
 145 150 155 160
 Val Thr ile Asp Ala Asn leu Ser Ser Gln Tyr Val Ser Ala leu leu
 165 170 175
 Met Ala Ala Ala Cys Gly Asp Lys Pro Val Asp ile ile leu Lys Gly
 180 185 190
 Glu Glu ile Gly Ala Lys Gly Tyr ile Asp leu Thr Thr Ser Ala Met
 195 200 205
 Glu Ala Phe Gly Ala Lys Val Glu Arg Val Ser Asn Ala ile Trp Arg
 210 215 220
 Val His Pro Thr Gly Tyr Thr Ala Thr Asp Phe His ile Glu Pro Asp
 225 230 235 240
 Ala Ser Ala Ala Thr Tyr leu Trp Gly Ala Glu leu leu Thr Gly Gly
 245 250 255
 Ala ile Asp ile Gly Thr Pro Ala Asp Lys Phe Thr Gln Pro Asp Ala
 260 265 270
 Lys Ala His Glu Val Met Ala Gln Phe Pro His leu Pro Ala Glu ile
 275 280 285
 Asp Gly Ser Gln Met Gln Asp Ala ile Pro Thr ile Ala Val leu Ala
 290 295 300
 Ala Phe Asn Glu Thr Pro Val Arg Phe Val Gly ile Ala Asn leu Arg
 305 310 315 320
 Val Lys Glu Cys Asp Arg ile Arg Ala Val Ser leu Gly leu Asn Glu
 325 330 335

```

Ile Arg Asp Gly Leu Ala His Glu Glu Gly Asp Asp Leu Ile Val His
      340      345      350
Ser Asp Pro Ser Leu Ala Gly Gln Thr Val Asn Ala Ser Ile Asp Thr
      355      360      365
Phe Ala Asp His Arg Ile Ala Met Ser Phe Ala Leu Ala Ala Leu Lys
      370      375      380
Ile Gly Gly Ile Ala Ile Gln Asn Pro Ala Cys Val Gly Lys Thr Tyr
      385      390      395      400
Pro Gly Tyr Trp Lys Ala Leu Ala Ser Leu Gly Val Glu Tyr Ser Glu
      405      410      415
Lys Glu Thr Ala Ala Glu Pro Gln His
      420      425

```

<210> 13
 <211> 17
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 5
 <223> Xaa= Lys ou Thr

<221> VARIANTE
 <222> 7
 <223> Xaa= Arg ou His

<221> VARIANTE
 <222> 9
 <223> Xaa= Ser ou Thr

<400> 13
 Leu Ala Lys Gly Xaa Ser Xaa Leu Xaa Gly Ala Leu Lys Ser Asp Asp
 1 5 10 15
 Thr

<210> 14
 <211> 5
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 5
 <223> Xaa= Lys ou Thr

<400> 14
 Leu Ala Lys Gly Xaa
 1 5

<210> 15
 <211> 4
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 2
 <223> Xaa= Arg ou His

```

<221>    VARIANTE
<222>    4
<223>    Xaa= Ser ou Thr
<400>    15
Ser Xaa Leu Xaa
1

<210>    16
<211>    8
<212>    PRT
<213>    Sequência artificial

<220>
<223>    Domínios conservados

<400>    16
Gly Ala Leu Lys Ser Asp Asp Thr
1                5

<210>    17
<211>    13
<212>    PRT
<213>    Sequência artificial

<220>
<223>    Domínios conservados

<221>    VARIANTE
<222>    4
<223>    Xaa= Asp ou Ala

<221>    VARIANTE
<222>    5
<223>    Xaa= Ser ou Thr

<221>    VARIANTE
<222>    8
<223>    Xaa= Val ou Ile

<221>    VARIANTE
<222>    10
<223>    Xaa= Thr, Glu ou Lys

<221>    VARIANTE
<222>    (12) ... (0)
<223>    Xaa= Gln, Ser, Glu ou Thr

<400>    17
Glu Pro Asp Xaa Xaa Thr Phe Xaa Val Xaa Xaa Xaa Gly
1                5                10

<210>    18
<211>    9
<212>    PRT
<213>    Sequência artificial

<220>
<223>    Domínios conservados

<221>    VARIANTE
<222>    4
<223>    Xaa= Asp ou Ala

```

<221> VARIANTE
 <222> 5
 <223> Xaa= Ser ou Thr

<221> VARIANTE
 <222> 8
 <223> Xaa= Val ou Ile

<400> 18
 Glu Pro Asp Xaa Xaa Thr Phe Xaa Val
 1 5

<210> 19
 <211> 4
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 1
 <223> Xaa= Thr, Glu ou Lys

<221> VARIANTE
 <222> 2
 <223> Xaa= Ser ou Gly

<221> VARIANTE
 <222> 3
 <223> Xaa= Gln, Ser, Glu ou Thr
 <400> 19

Xaa Xaa Xaa Gly
 1

<210> 20
 <211> 6
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<400> 20
 Arg Phe Leu Thr Ala Ala
 1 5

<210> 21
 <211> 6
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 5
 <223> Xaa= Gly, Met ou Leu

<400> 21
 Lys Arg Pro Ile Xaa Pro
 1 5

<210> 22
 <211> 6
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 1
 <223> Xaa= Thr ou Ser

<400> 22
 Xaa Gly Cys Pro Val
 1 5

<210> 23
 <211> 10
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 7
 <223> Xaa= Ile ou Val

<400> 23
 Ile Gly Ala Xaa Gly Tyr Xaa Asp Leu Thr
 1 5 10

<210> 24
 <211> 7
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 2
 <223> Xaa= Arg ou Lys

<221> VARIANTE
 <222> 4
 <223> Xaa= Ala, His, Glu ou Ser

<221> VARIANTE
 <222> 5
 <400> 24
 Trp Xaa Vall Xaa Xaa Thr Gly
 1 5

<210> 25
 <211> 16
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 12
 <223> Xaa= Ala ou Gly

<221> VARIANTE
 <222> 14
 <223> Xaa= Gly ou Gln

 <221> VARIANTE
 <222> 15
 <223> Xaa= Val, Leu ou Ala

 <400> 25
 Glu Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Xaa Ala Xaa Xaa Leu
 1 5 10 15

 <210> 26
 <211> 11
 <212> PRT
 <213> Sequência artificial

 <220>
 <223> Domínios conservados

 <400> 26
 Glu Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp
 1 5 10

 <210> 27
 <211> 4
 <212> PRT
 <213> Sequência artificial

 <220>
 <223> Domínios conservados

 <221> VARIANTE
 <222> 3
 <223> Xaa= Ile ou Leu

 <400> 27
 Ile Asp Xaa Gly
 1

 <210> 28
 <211> 8
 <212> PRT
 <213> Sequência artificial

 <220>
 <223> Domínios conservados

 <221> VARIANTE
 <222> 2
 <223> Xaa= Thr ou Ser

 <400> 28
 Phe Xaa Gln Pro Asp Ala Lys Ala
 1 5

 <210> 29
 <211> 28
 <212> PRT
 <213> Sequência artificial

 <220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 1
 <223> Xaa= Glu, Lys ou Ser

<221> VARIANTE
 <222> 4
 <223> Xaa= Asp ou His
 Trp Xaa Vall Xaa Xaa Thr Gly
 1 5

<221> VARIANTE
 <222> 5
 <223> Xaa= Met ou Leu

<221> VARIANTE
 <222> 6
 <223> Xaa= Pro ou Gln

<221> VARIANTE
 <222> 8
 <223> Xaa= Thr, Glu ou Val

<221> VARIANTE
 <222> (9) ... (0)
 <223> Xaa= Val ou Ile

<221> VARIANTE
 <222> (10) ... (0)
 <223> Xaa= Asp ou Val

<221> VARIANTE
 <222> (21) ... (0)
 <223> Xaa= Leu ou Ile

<221> VARIANTE
 <222> (24) ... (0)
 <223> Xaa= Leu ou Ile

<400> 29
 Xaa Phe Pro Xaa Xaa Xaa Ala Xaa Xaa Xaa Gly Ser Gln Met Gln Asp
 1 5 10 15
 Ala Ile Pro Thr Xaa Ala Val Xaa Ala Ala Phe Asn
 20 25

<210> 30
 <211> 7
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 1
 <223> Xaa= Gln, Lys ou Ser

<221> VARIANTE
 <222> 4
 <223> Xaa= Asp ou His

<221> VARIANTE
 <222> 5
 <223> Xaa= Met ou Leu

<221> VARIANTE
 <222> 6
 <223> Xaa= Pro ou Gln

<400> 30
 Xaa Phe Pro Xaa Xaa Xaa Ala
 1 5

<210> 31
 <211> 21
 <212> PRT
 <213> Seqüência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 1
 <223> Xaa= Thr Glu ou Val

<221> VARIANTE
 <222> 2
 <223> Xaa= Val ou Ile

<221> VARIANTE
 <222> 3
 <223> Xaa= Asp ou Val

<221> VARIANTE
 <222> 14
 <223> Xaa= Leu ou Ile

<221> VARIANTE
 <222> 17
 <223> Xaa= Leu ou Ile

<400> 31
 Xaa Xaa Xaa Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Xaa Ala Val
 1 5 10 15
 Xaa Ala Ala Phe Asn
 20

<210> 32
 <211> 10
 <212> PRT
 <213> Seqüência artificial

<220>
 <223> Domínios conservados

<400> 32
 Gly Ser Gln Met Gln Asp Ala Ile Pro Thr
 1 5 10

<210> 33
 <211> 18
 <212> PRT
 <213> Seqüência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 5

<223> Xaa= Val ou Thr
 <221> VARIANTE
 <222> 6
 <223> Xaa= Glu ou Gly
 <221> VARIANTE
 <222> 7
 <223> Xaa= Leu ou Ile
 <221> VARIANTE
 <222> 8
 <223> Xaa= Ala ou Glu
 <221> VARIANTE
 <222> 17
 <223> Xaa= Ile ou Val
 <400> 33
 Pro Val Arg Phe Xaa Xaa Xaa Asn Leu Arg Val Lys Glu Cys Asp
 1 5 10 15
 Arg Xaa
 <210> 34
 <211> 4
 <212> PRT
 <213> Sequência artificial
 <220>
 <223> Domínios conservados
 <400> 34
 Pro Val Arg Phe
 1
 <210> 35
 <211> 14
 <212> PRT
 <213> Sequência artificial
 <220>
 <223> Domínios conservados
 <221> VARIANTE
 <222> 1
 <223> Xaa= Val ou Thr
 <221> VARIANTE
 <222> 2
 <223> Xaa= Glu ou Gly
 <221> VARIANTE
 <222> 3
 <223> Xaa= Leu ou Ile
 <221> VARIANTE
 <222> 4
 <223> Xaa= Ala ou Glu
 <221> VARIANTE
 <222> 17
 <223> Xaa= Ile ou Val
 <400> 35

Xaa Xaa Xaa Xaa Asn Leu Arg Val Lys Glu Cys Asp Arg Xaa
 1 5 10

<210> 36
 <211> 9
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<400> 36
 Asn Leu Arg Val Lys Glu Cys asp Arg
 1 5

<210> 37
 <211> 7
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 6
 <223> Xaa= Leu ou Ile

<221> VARIANTE
 <222> 7
 <223> Xaa= Val ou Ile

<400> 37
 Glu Gly Asp Asp Leu Xaa Xaa
 1 5

<210> 38
 <211> 6
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE
 <222> 1
 <223> Xaa= Asp ou Asn

<221> VARIANTE
 <222> 3
 <223> Xaa= Ala, Ser ou Thr

<400> 38
 Xaa Pro Xaa Leu Ala Gly
 1 5

<210> 39
 <211> 10
 <212> PRT
 <213> Sequência artificial

<220>
 <223> Domínios conservados

<221> VARIANTE

<222> 2
 <223> Xaa= Leu Ser ou Glu

 <221> VARIANTE
 <222> 5
 <223> Xaa= Thr ou Ser

 <221> VARIANTE
 <222> 6
 <223> Xaa= His ou Phe

 <221> VARIANTE
 <222> 7
 <223> Xaa= ala ou Ser

 <400> 39
 Ala Xaa Ile Asp Xaa Xaa Xaa Asp His Arg
 1 5 10

 <210> 40
 <211> 11
 <212> PRT
 <213> Seqüência artificial

 <220>
 <223> Domínios conservados

 <221> VARIANTE
 <222> 5
 <223> Xaa= Gly ou Ala

 <221> VARIANTE
 <222> 8
 <223> Xaa= Ile ou Val

 <221> VARIANTE
 <222> 9
 <223> Xaa= Ser, Gly, Ala ou Lys

 <400> 40
 Phe Ala Leu Ala Xaa Leu Lys Xaa Xaa Gly Ile
 1 5 10

 <210> 41
 <211> 7
 <212> PRT
 <213> Seqüência artificial

 <220>
 <223> Domínios conservados

 <221> VARIANTE
 <222> 5

 <223> Xaa= Gly ou Ala

 <400> 41
 Phe Ala Leu Ala Xaa Leu Lys
 1 5

 <210> 42
 <211> 6
 <212> PRT
 <213> Seqüência artificial

<220>
 <223> Domínios conservados

 <221> VARIANTE
 <222> 3
 <223> Xaa= Ile ou Val

 <221> VARIANTE
 <222> 4
 <223> Xaa= Ser, Gly, Ala ou Lys

 <400> 42
 Leu Lys Xaa Xaa Gly Ile
 1 5

 <210> 43
 <211> 7
 <212> PRT
 <213> Sequência artificial

 <220>
 <223> Domínios conservados

 <221> VARIANTE
 <222> 1
 <223> Xaa= Asn ou Asp

 <221> VARIANTE
 <222> 3
 <223> Xaa= Ala ou Asp

 <221> VARIANTE
 <222> 6
 <223> Xaa= Ala ou Gly

 <400> 43
 Xaa Pro Xaa Cys Val Xaa Lys
 1 5

 <210> 44
 <211> 5
 <212> PRT
 <213> Sequência artificial

 <220>
 <223> Domínios conservados

 <221> VARIANTE
 <222> 1
 <223> Xaa= Ala, Ser ou Pro

 <400> 44
 Xaa Ser Leu Gly Val
 1 5

 <210> 45
 <211> 1500
 <212> DNA
 <213> Pseudomonas syringae

 <220>
 <221> CDS
 <222> (102) ... (1358)

 <221> Região de interesse biológico

<222> (0) ... (0)

<223> Cepa B728a de pv syringae

<400> 45

gctgaccgc gcccgggctt ggcgatgaat gaaaccggca ctgcgggctt gatcccgctc 60
 gtgctaggat cgcgatTTTT ccgctgacgc tcagggacat c atg cga cct caa gcc 116
 Met Arg Pro Gln Ala
 1 5

acc ctc act gtt ttg cct gtc gag cgc ccg ttg gtc ggg cgt gtc agc 164
 Thr Leu Thr Val Leu Pro Val Glu Arg Pro Leu Val Gly Arg Val Ser
 10 15 20

ccg ccg ggc tcc aag tcg atc acc aac cgt gca ttg ttg ctg gcc ggg 212
 Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Gly
 25 30 35

ctg gcc aaa ggc acc agc cgc ctg acc ggc gcg ctg aag agt gat gac 260
 Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala Leu Lys Ser Asp Asp
 40 45 50

acg cgc gtg atg tcc gaa gcc ttg cgt ctg atg ggc gtg cag gtc gac 308
 Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met Gly Val Gln Val Asp
 55 60 65

gag ccg gat gac agc acc ttc gtg gtc acc agc agc ggg cac tgg cag 356
 Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser Ser Gly His Trp Gln
 70 75 80 85

gcg ccg caa cag gcg ctg ttc ctc ggc aat gcc ggg act gcg aca cgc 404
 Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala Gly Thr Ala Thr Arg
 90 95 100

ttt ctg acc gcg gca ctg gcc aac ttc gaa ggc gac ttc gtg gtg gat 452
 Phe Leu Thr Ala Ala Leu Ala Asn Phe Glu Gly Asp Phe Val Val Asp
 105 110 115

ggc gac gaa tac atg cgc aag cgc ccg atc ggc ccg ttg gtc gat gcc 500
 Gly Asp Glu Tyr Met Arg Lys Arg Pro Ile Gly Pro Leu Val Asp Ala
 120 125 130

ttg cag cgc atg ggc gtg gag gtc agc gca ccc agt ggt tgc ccg ccg 548
 Leu Gln Arg Met Gly Val Glu Val Ser Ala Pro Ser Gly Cys Pro Pro
 135 140 145

gtg gcc atc aag ggc aag ggc ggt ctt gag gcc ggt cgt atc gaa atc 596
 Val Ala Ile Lys Gly Lys Gly Gly Leu Glu Ala Gly Arg Ile Glu Ile
 150 155 160 165

gac ggc aat ctg tcc agc cag tat gtg tcg gca ctg ctg atg gcc ggt 644
 Asp Gly Asn Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Ala Gly
 170 175 180

gcc tgt ggc aag ggg cct gtg gaa gtt gcc ctg aca ggc agc gag atc 692
 Ala Cys Gly Lys Gly Pro Val Glu Val Ala Leu Thr Gly Ser Glu Ile
 185 190 195

ggc gcg cgt ggt tac ctc gac ctc acg ctg gcg gcc atg ccg gcg ttc 740
 Gly Ala Arg Gly Tyr Leu Asp Leu Thr Leu Ala Ala Met Arg Ala Phe
 200 205 210

ggt gcc gag gtt cag gcc atc ggc gac gcc gcc tgg aaa gtc tcg gct 788
 Gly Ala Glu Val Gln Ala Ile Gly Asp Ala Ala Trp Lys Val Ser Ala
 215 220 225

acc ggt tat cgc gct acg gat ttc cac atc gaa ccg gat gcc tcg gcg 836
 Thr Gly Tyr Arg Ala Thr Asp Phe His Ile Glu Pro Asp Ala Ser Ala
 230 235 240 245

gcc acc tac ctt tgg gct gcg cag gcc ctg acc gag ggc gct atc gac 884
 Ala Thr Tyr Leu Trp Ala Ala Gln Ala Leu Thr Glu Gly Ala Ile Asp
 250 255 260
 ctg ggc gtg gcc agc aac gcg ttc act cag cct gat gca ctg gcc agt 932
 Leu Gly Val Ala Ser Asn Ala Phe Thr Gln Pro Asp Ala Leu Ala Ser
 265 270 275
 cag atc atc gcc agc ttc ccg aac atg ccg gcc gtg atc gac ggc tgc 980
 Gln Ile Ile Ala Ser Phe Pro Asn Met Pro Ala Val Ile Asp Gly Ser
 280 285 290
 cag atg cag gac gcg att ccc acg ctg gcc gta ctg gcc gcg ttc aat 1028
 Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn
 295 300 305
 cgt caa ccg gtg cgc ttt gtc ggc atc gcc aac ctg cgg gtc aag gag 1076
 Arg Gln Pro Val Arg Phe Val Gly Ile Ala Asn Leu Arg Val Lys Glu
 310 315 320 325
 tgc gac cgc atc tgc gca ctg tcc aac ggc ctg tgc gcc atc gca ccc 1124
 Cys Asp Arg Ile Ser Ala Leu Ser Asn Gly Leu Cys Ala Ile Ala Pro
 330 335 340
 ggc ctg gcg gtc gaa gag ggt gac gat ctg atc gtt acc gcc aac ccg 1172
 Gly Leu Ala Val Glu Glu Gly Asp Asp Leu Ile Val Thr Ala Asn Pro
 345 350 355
 acg ctg gca ggc act acg gtc gat gcc ttg atc gat acc cac tcc gac 1220
 Thr Leu Ala Gly Thr Thr Val Asp Ala Leu Ile Asp Thr His Ser Asp
 360 365 370
 cat cgg atc gcc atg tgc ttt gca ctg gcg ggc ctg aag att gcc ggc 1268
 His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Ile Ala Gly
 375 380 385
 atc cgc att ctc gac cct gat tgc gtc gcc aag acc tac ccg ggg tac 1316
 Ile Arg Ile Leu Asp Pro Asp Cys Val Ala Lys Thr Tyr Pro Gly Tyr
 390 395 400 405
 tgg gat gcg ctg gct tct ctg ggt gtg agt gtt cag cgc tga 1358
 Trp Asp Ala Leu Ala Ser Leu Gly Val Ser Val Gln Arg *
 410 415
 tagatcaggt ttatcggtc ggaagcctg taggatttcg agacatcttt agctggcagt 1418
 ctacgggtgt agatactatg ttatcacgtt ggggttgcat gcctcgtgtg aggtatctac 1478
 agtattgcta cgttatgagg tg 1500

<210> 46
 <211> 418
 <212> PRT
 <213> Pseudomonas syringae
 <220>
 <221> CDS
 <222> (0) ... (0)
 <223> Cepa B728a de pv syringae

<400> 46
 Met Arg Pro Gln Ala Thr Leu Thr Val Leu Pro Val Glu Arg Pro Leu
 1 5 10 15
 Val Gly Arg Val Ser Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala
 20 25 30
 Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala
 35 40 45
 Leu Lys Ser Asp Asp Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met
 50 55 60
 Gly Val Gln Val Asp Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser
 65 70 75 80

Ser Gly His Trp Gln Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala
 85 90 95
 Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Leu Ala Asn Phe Glu Gly
 100 105 110
 Asp Phe Val Val Asp Gly Asp Glu Tyr Met Arg Lys Arg Pro Ile Gly
 115 120 125
 Pro Leu Val Asp Ala Leu Gln Arg Met Gly Val Glu Val Ser Ala Pro
 130 135 140
 Ser Gly Cys Pro Pro Val Ala Ile Lys Gly Lys Gly Gly Leu Glu Ala
 145 150 155 160
 Gly Arg Ile Glu Ile Asp Gly Asn Leu Ser Ser Gln Tyr Val Ser Ala
 165 170 175
 Leu Leu Met Ala Gly Ala Cys Gly Lys Gly Pro Val Glu Val Ala Leu
 180 185 190
 Thr Gly Ser Glu Ile Gly Ala Arg Gly Tyr Leu Asp Leu Thr Leu Ala
 195 200 205
 Ala Met Arg Ala Phe Gly Ala Glu Val Gln Ala Ile Gly Asp Ala Ala
 210 215 220
 Trp Lys Val Ser Ala Thr Gly Tyr Arg Ala Thr Asp Phe His Ile Glu
 225 230 235 240
 Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Gln Ala Leu Thr
 245 250 255
 Glu Gly Ala Ile Asp Leu Gly Val Ala Ser Asn Ala Phe Thr Gln Pro
 260 265 270
 Asp Ala Leu Ala Ser Gln Ile Ile Ala Ser Phe Pro Asn Met Pro Ala
 275 280 285
 Val Ile Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val
 290 295 300
 Leu Ala Ala Phe Asn Arg Gln Pro Val Arg Phe Val Gly Ile Ala Asn
 305 310 315 320
 Leu Arg Val Lys Glu Cys Asp Arg Ile Ser Ala Leu Ser Asn Gly Leu
 325 330 335
 Cys Ala Ile Ala Pro Gly Leu Ala Val Glu Glu Gly Asp Asp Leu Ile
 340 345 350
 Val Thr Ala Asn Pro Thr Leu Ala Gly Thr Thr Val Asp Ala Leu Ile
 355 360 365
 Asp Thr His Ser Asp His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly
 370 375 380
 Leu Lys Ile Ala Gly Ile Arg Ile Leu Asp Pro Asp Cys Val Ala Lys
 385 390 395 400
 Thr Tyr Pro Gly Tyr Trp Asp Ala Leu Ala Ser Leu Gly Val Ser Val
 405 410 415
 Gln Arg

<210> 47
 <211> 1466
 <212> DNA
 <213> *Pseudomonas syringae*

<220>
 <221> CDS
 <222> (101) ... (1357)

<221> Região de interesse biológico
 <222> (0) ... (0)
 <223> Cepa 1448a de pv phaseolicola

<400> 47
 gtaaaccggct gtctggcccg ttggggcagc aaaaccgcac tgcgggcttg atcccgtctcg 60
 tgctaggatc gcgatttttc cgctgacgct cagggacatc atg cga cct caa gcc 115
 Met Arg Pro Gln Ala

acc ctc act gtt ttg cct gtc gag cgc ccg ctg gtc ggg cgc gtc agc	163
Thr Leu Thr Val Leu Pro Val Glu Arg Pro Leu Val Gly Arg Val Ser	
10 15 20	
ccg ccg ggc tcc aag tcg atc acc aac cgc gca ttg ttg ctg gcc ggg	211
Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Gly	
25 30 35	
ctg gcc aaa ggc acc agc cgc ctg acc ggc gcg ctg aag agt gac gac	259
Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala Leu Lys Ser Asp Asp	
40 45 50	
acc cgc gtg atg tcc gaa gca ctg cgt ctg atg ggc gtg cag gtc gac	307
Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met Gly Val Gln Val Asp	
55 60 65	
gag ccg gat gac agc acc ttc gtg gtc acc agc agc ggc cac tgg cag	355
Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser Ser Gly His Trp Gln	
70 75 80 85	
gca ccg cag cag gca ctg ttt ctc ggt aac gcc gga acc gca acg cgc	403
Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala Gly Thr Ala Thr Arg	
90 95 100	
ttt ctg acc gct gca ctg gcc aac ttc gaa ggc gac ttt gtg gtc gac	451
Phe Leu Thr Ala Ala Leu Ala Asn Phe Glu Gly Asp Phe Val Val Asp	
105 110 115	
ggc gac gag tac atg cgc aag cgc ccg atc ggc ccg ctg gtc gat gcc	499
Gly Asp Glu Tyr Met Arg Lys Arg Pro Ile Gly Pro Leu Val Asp Ala	
120 125 130	
ttg cag cgc atg ggt gtg gag gtc agt gcg ccc agt ggc tgc ccg ccg	547
Leu Gln Arg Met Gly Val Glu Val Ser Ala Pro Ser Gly Cys Pro Pro	
135 140 145	
gtg gcg atc aag ggc aaa ggc ggt ctg gaa gcc gga cgt att gaa atc	595
Val Ala Ile Lys Gly Lys Gly Gly Leu Glu Ala Gly Arg Ile Glu Ile	
150 155 160 165	
gat ggc aac ctg tcc agc cag tac gtg tca gcg ttg ctg atg gcc gga	643
Asp Gly Asn Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Ala Gly	
170 175 180	
gcc tgc ggc aag ggc ccg gtc gaa gtc gcc ttg acc ggc agc gag att	691
Ala Cys Gly Lys Gly Pro Val Glu Val Ala Leu Thr Gly Ser Glu Ile	
185 190 195	
ggc gca cgt ggt tac gtc gac ctt acc ctg gcg gcc atg cag gcc ttc	739
Gly Ala Arg Gly Tyr Val Asp Leu Thr Leu Ala Ala Met Gln Ala Phe	
200 205 210	
ggc gcc gag gtt cag gcc atc ggc gaa acc gcc tgg aaa gtc tcg gcc	787
Gly Ala Glu Val Gln Ala Ile Gly Glu Thr Ala Trp Lys Val Ser Ala	
215 220 225	
act ggt tat cgc gct acg gac ttc cat atc gaa ccg gat gca tcg gcg	835
Thr Gly Tyr Arg Ala Thr Asp Phe His Ile Glu Pro Asp Ala Ser Ala	
230 235 240 245	
gcc acc tac ctg tgg gcc gcc caa gca ctg acc gag ggc gat atc gac	883
Ala Thr Tyr Leu Trp Ala Ala Gln Ala Leu Thr Glu Gly Asp Ile Asp	
250 255 260	
ctc ggc gtg gcc agc gac gca ttc act cag cct gat gcc ctg gcc agt	931
Leu Gly Val Ala Ser Asp Ala Phe Thr Gln Pro Asp Ala Leu Ala Ser	
265 270 275	
cag atc atc gca agc ttc ccg aac atg cct gcc gtg atc gac ggt tcg	979
Gln Ile Ile Ala Ser Phe Pro Asn Met Pro Ala Val Ile Asp Gly Ser	
280 285 290	

caa atg cag gac gcg att ccg acg ctg gcg gta ctc gcc gcc ttc aac 1027
 Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn
 295 300 305

cgt cag cca gtg cgc ttt gtc ggc atc gcc aac ctg cgg gtc aag gaa 1075
 Arg Gln Pro Val Arg Phe Val Gly Ile Ala Asn Leu Arg Val Lys Glu
 310 315 320 325

tgt gac cgt att tca gca ctg tcc cac ggc ctg tgt gcc atc gcg ccg 1123
 Cys Asp Arg Ile Ser Ala Leu Ser His Gly Leu Cys Ala Ile Ala Pro
 330 335 340

ggc ctt gct gta gag gag ggc gac gac ctg ctg gtg cac gcc aac ccg 1171
 Gly Leu Ala Val Glu Glu Gly Asp Asp Leu Leu Val His Ala Asn Pro
 345 350 355

gcg ctg gca ggc acc acg gta gac gca ttg att gac acc cac tcc gac 1219
 Ala Leu Ala Gly Thr Thr Val Asp Ala Leu Ile Asp Thr His Ser Asp
 360 365 370

cat cgc atc gcc atg tgt ttt gcg ctg gca ggc ttg aag att gcc ggt 1267
 His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Ile Ala Gly
 375 380 385

att cgc att ctc gac cca gat tgc gtc ggc aag acc tac ccg ggt tac 1315
 Ile Arg Ile Leu Asp Pro Asp Cys Val Gly Lys Thr Tyr Pro Gly Tyr
 390 395 400 405

tgg gat gca ctg gct tcc ctg ggg gtg cgt gtt cag cgc tga 1357
 Trp Asp Ala Leu Ala Ser Leu Gly Val Arg Val Gln Arg *
 410 415

gagcagctat ggagccgttt ttcagatccc gaagaacggc tgcacatccc gcgccagcag 1417
 cccgtgaaag cgcagccttg attctgtcgc cagtacgcag atgcattccc 1466

<210> 48
 <211> 418
 <212> PRT
 <213> Pseudomonas syringae

<220>
 <221> VARIANTE
 <222> (0) ... (0)
 <223> Cepa 1448a de pv phaseolicola

<400> 48
 Met Arg Pro Gln Ala Thr Leu Thr Val Leu Pro Val Glu Arg Pro Leu
 1 5 10 15
 Val Gly Arg Val Ser Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala
 20 25 30
 Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala
 35 40 45
 Leu Lys Ser Asp Asp Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met
 50 55 60
 Gly Val Gln Val Asp Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser
 65 70 75 80
 Ser Gly His Trp Gln Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala
 85 90 95
 Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Leu Ala Asn Phe Glu Gly
 100 105 110
 Asp Phe Val Val Asp Gly Asp Glu Tyr Met Arg Lys Arg Pro Ile Gly
 115 120 125
 Pro Leu Val Asp Ala Leu Gln Arg Met Gly Val Glu Val Ser Ala Pro
 130 135 140
 Ser Gly Cys Pro Pro Val Ala Ile Lys Gly Lys Gly Gly Leu Glu Ala
 145 150 155 160

Gly Arg Ile Glu Ile Asp Gly Asn Leu Ser Ser Gln Tyr Val Ser Ala
 165 170 175
 Leu Leu Met Ala Gly Ala Cys Gly Lys Gly Pro Val Glu Val Ala Leu
 180 185 190
 Thr Gly Ser Glu Ile Gly Ala Arg Gly Tyr Val Asp Leu Thr Leu Ala
 195 200 205
 Ala Met Gln Ala Phe Gly Ala Glu Val Gln Ala Ile Gly Glu Thr Ala
 210 215 220
 Trp Lys Val Ser Ala Thr Gly Tyr Arg Ala Thr Asp Phe His Ile Glu
 225 230 235 240
 Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Gln Ala Leu Thr
 245 250 255
 Glu Gly Asp Ile Asp Leu Gly Val Ala Ser Asp Ala Phe Thr Gln Pro
 260 265 270
 Asp Ala Leu Ala Ser Gln Ile Ile Ala Ser Phe Pro Asn Met Pro Ala
 275 280 285
 Val Ile Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val
 290 295 300
 Leu Ala Ala Phe Asn Arg Gln Pro Val Arg Phe Val Gly Ile Ala Asn
 305 310 315 320
 Leu Arg Val Lys Glu Cys Asp Arg Ile Ser Ala Leu Ser His Gly Leu
 325 330 335
 Cys Ala Ile Ala Pro Gly Leu Ala Val Glu Glu Gly Asp Asp Leu Leu
 340 345 350
 Val His Ala Asn Pro Ala Leu Ala Gly Thr Thr Val Asp Ala Leu Ile
 355 360 365
 Asp Thr His Ser Asp His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly
 370 375 380
 Leu Lys Ile Ala Gly Ile Arg Ile Leu Asp Pro Asp Cys Val Gly Lys
 385 390 395 400
 Thr Tyr Pro Gly Tyr Trp Asp Ala Leu Ala Ser Leu Gly Val Arg Val
 405 410 415
 Gln Arg

<210> 49
 <211> 427
 <212> PRT
 <213> Escherichia coli

<400> 49
 Met Glu Ser Leu Thr Leu Gln Pro Ile Ala Arg Val Asp Gly Thr Ile
 1 5 10 15
 Asn Leu Pro Gly Ser Lys Thr Val Ser Asn Arg Ala Leu Leu Ala
 20 25 30
 Ala Leu Ala His Gly Lys Thr Val Leu Thr Asn Leu Leu Asp Ser Asp
 35 40 45
 Asp Val Arg His Met Leu Asn Ala Leu Thr Ala Leu Gly Val Ser Tyr
 50 55 60
 Thr Leu Ser Ala Asp Arg Thr Arg Cys Glu Ile Ile Gly Asn Gly Gly
 65 70 75 80
 Pro Leu His Ala Glu Gly Ala Leu Glu Leu Phe Leu Gly Asn Ala Gly
 85 90 95
 Thr Ala Met Arg Pro Leu Ala Ala Ala Leu Cys Leu Gly Ser Asn Asp
 100 105 110
 Ile Val Leu Thr Gly Glu Pro Arg Met Lys Glu Arg Pro Ile Gly His
 115 120 125
 Leu Val Asp Ala Leu Arg Leu Gly Gly Ala Lys Ile Thr Tyr Leu Glu
 130 135 140
 Gln Glu Asn Tyr Pro Pro Leu Arg Leu Gln Gly Gly Phe Thr Gly Gly
 145 150 155 160
 Asn Val Asp Val Asp Gly Ser Val Ser Ser Gln Phe Leu Thr Ala Leu
 165 170 175
 Leu Met Thr Ala Pro Leu Ala Pro Glu Asp Thr Val Ile Arg Ile Lys
 180 185 190
 Gly Asp Leu Val Ser Lys Pro Tyr Ile Asp Ile Thr Leu Asn Leu Met
 195 200 205
 Lys Thr Phe Gly Val Glu Ile Glu Asn Gln His Tyr Gln Gln Phe Val
 210 215 220

Val Lys Gly Gly Gln Ser Tyr Gln Ser Pro Gly Thr Tyr Leu Val Glu
 225 230 235 240
 Gly Asp Ala Ser Ser Ala Ser Tyr Phe Leu Ala Ala Ala Ala Ile Lys
 245 250 255
 Gly Gly Thr Val Lys Val Thr Gly Ile Gly Arg Asn Ser Met Gln Gly
 260 265 270
 Asp Ile Arg Phe Ala Asp Val Leu Glu Lys Met Gly Ala Thr Ile Cys
 275 280 285
 Trp Gly Asp Asp Tyr Ile Ser Cys Thr Arg Gly Glu Leu Asn Ala Ile
 290 295 300
 Asp Met Asp Met Asn His Ile Pro Asp Ala Ala Met Thr Ile Ala Thr
 305 310 315 320
 Ala Ala Leu Phe Ala Lys Gly Thr Thr Arg Leu Arg Asn Ile Tyr Asn
 325 330 335
 Trp Arg Val Lys Glu Thr Asp Arg Leu Phe Ala Met Ala Thr Glu Leu
 340 345 350
 Arg Lys Val Gly Ala Glu Val Glu Gly His Asp Tyr Ile Arg Ile
 355 360 365
 Thr Pro Pro Glu Lys Leu Asn Phe Ala Glu Ile Ala Thr Tyr Asn Asp
 370 375 380
 His Arg Met Ala Met Cys Phe Ser Leu Val Ala Leu Ser Asp Thr Pro
 385 390 395 400
 Val Thr Ile Leu Asp Pro Lys Cys Thr Ala Lys Thr Phe Pro Asp Tyr
 405 410 415
 Phe Glu Gln Leu Ala Arg Ile Ser Gln Ala Ala

420

425

<210> 50
 <211> 444
 <212> PRT
 <213> Zea mays

<400> 50

Ala Gly Ala Glu Glu Ile Val Leu Gln Pro Ile Lys Glu Ile Ser Gly
 1 5 10 15
 Thr Val Lys Leu Pro Gly Ser Lys Ser Leu Ser Asn Arg Ile Leu Leu
 20 25 30
 Leu Ala Ala Leu Ser Glu Gly Thr Thr Val Val Asp Asn Leu Leu Asn
 35 40 45
 Ser Glu Asp Val His Tyr Met Leu Gly Ala Leu Arg Thr Leu Gly Leu
 50 55 60
 Ser Val Glu Ala Asp Lys Ala Ala Lys Arg Ala Val Val Val Gly Cys
 65 70 75 80
 Gly Gly Lys Phe Pro Val Glu Asp Ala Lys Glu Glu Val Gln Leu Phe
 85 90 95
 Leu Gly Asn Ala Gly Thr Ala Met Arg Pro Leu Thr Ala Ala Val Thr
 100 105 110
 Ala Ala Gly Gly Asn Ala Thr Tyr Val Leu Asp Gly Val Pro Arg Met
 115 120 125
 Arg Glu Arg Pro Ile Gly Asp Leu Val Val Gly Leu Lys Gln Leu Gly
 130 135 140
 Ala Asp Val Asp Cys Phe Leu Gly Thr Asp Cys Pro Pro Val Arg Val
 145 150 155 160
 Asn Gly Ile Gly Gly Leu Pro Gly Gly Lys Val Lys Leu Ser Gly Ser
 165 170 175
 Ile Ser Ser Gln Tyr Leu Ser Ala Leu Leu Met Ala Ala Pro Leu Ala
 180 185 190

Leu Gly Asp Val Glu Ile Glu Ile Ile Asp Lys Leu Ile Ser Ile Pro
 195 200 205
 Tyr Val Glu Met Thr Leu Arg Leu Met Glu Arg Phe Gly Val Lys Ala
 210 215 220
 Glu His Ser Asp Ser Trp Asp Arg Phe Tyr Ile Lys Gly Gly Gln Lys
 225 230 235 240
 Tyr Lys Ser Pro Lys Asn Ala Tyr Val Glu Gly Asp Ala Ser Ser Ala
 245 250 255
 Ser Tyr Phe Leu Ala Gly Ala Ala Ile Thr Gly Gly Thr Val Thr Val
 260 265 270
 Glu Gly Cys Gly Thr Thr Ser Leu Gln Gly Asp Val Lys Phe Ala Glu
 275 280 285
 Val Leu Glu Met Met Gly Ala Lys Val Thr Trp Thr Glu Thr Ser Val
 290 295 300
 Thr Val Thr Gly Pro Pro Arg Glu Pro Phe Gly Arg Lys His Leu Lys
 305 310 315 320
 Ala Ile Asp Val Asn Met Asn Lys Met Pro Asp Val Ala Met Thr Leu
 325 330 335
 Ala Val Val Ala Leu Phe Ala Asp Gly Pro Thr Ala Ile Arg Asp Val
 340 345 350
 Ala Ser Trp Arg Val Lys Glu Thr Glu Arg Met Val Ala Ile Arg Thr
 355 360 365
 Glu Leu Thr Lys Leu Gly Ala Ser Val Glu Glu Gly Pro Asp Tyr Cys
 370 375 380
 Ile Ile Thr Pro Pro Glu Lys Leu Asn Val Thr Ala Ile Asp Thr Tyr
 385 390 395 400
 Asp Asp His Arg Met Ala Met Ala Phe Ser Leu Ala Ala Cys Ala Glu
 405 410 415
 Val Pro Val Thr Ile Arg Asp Pro Gly Cys Thr Arg Lys Thr Phe Pro
 420 425 430
 Asp Tyr Phe Asp Val Leu Ser Thr Phe Val Lys Asn
 435 440

<210> 51
 <211> 455
 <212> PRT
 <213> Agrobacterium sp.

<220>
 <221> VARIANTE
 <222> (0) ... (0)
 <223> Cepa CP4

<400> 51
 Met Ser His Gly Ala Ser Ser Arg Pro Ala Thr Ala Arg Lys Ser Ser
 1 5 10 15
 Gly Leu Ser Gly Thr Val Arg Ile Pro Gly Asp Lys Ser Ile Ser His
 20 25 30
 Arg Ser Phe Met Phe Gly Gly Leu Ala Ser Gly Glu Thr Arg Ile Thr
 35 40 45
 Gly Leu Leu Glu Gly Glu Asp Val Ile Asn Thr Gly Lys Ala Met Gln
 50 55 60
 Ala Met Gly Ala Arg Ile Arg Lys Glu Gly Asp Thr Trp Ile Ile Asp
 65 70 75 80
 Gly Val Gly Asn Gly Gly Leu Leu Ala Pro Glu Ala Pro Leu Asp Phe
 85 90 95
 Gly Asn Ala Ala Thr Gly Cys Arg Leu Thr Met Gly Leu Val Gly Val
 100 105 110
 Tyr Asp Phe Asp Ser Thr Phe Ile Gly Asp Ala Ser Leu Thr Lys Arg
 115 120 125
 Pro Met Gly Arg Val Leu Asn Pro Leu Arg Glu Met Gly Val Gln Val
 130 135 140
 Lys Ser Glu Asp Gly Asp Arg Leu Pro Val Thr Leu Arg Gly Pro Lys
 145 150 155 160
 Thr Pro Thr Pro Ile Thr Tyr Arg Val Pro Met Ala Ser Ala Gln Val
 165 170 175

Lys Ser Ala Val Leu Leu Ala Gly Leu Asn Thr Pro Gly Ile Thr Thr
 180 185 190
 Val Ile Glu Pro Ile Met Thr Arg Asp His Thr Glu Lys Met Leu Gln
 195 200 205
 Gly Phe Gly Ala Asn Leu Thr Val Glu Thr Asp Ala Asp Gly Val Arg
 210 215 220
 Thr Ile Arg Leu Glu Gly Arg Gly Lys Leu Thr Gly Gln Val Ile Asp
 225 230 235 240
 Val Pro Gly Asp Pro Ser Ser Thr Ala Phe Pro Leu Val Ala Ala Leu
 245 250 255
 Leu Val Pro Gly Ser Asp Val Thr Ile Leu Asn Val Leu Met Asn Pro
 260 265 270
 Thr Arg Thr Gly Leu Ile Leu Thr Leu Gln Glu Met Gly Ala Asp Ile
 275 280 285
 Glu Val Ile Asn Pro Arg Leu Ala Gly Gly Glu Asp Val Ala Asp Leu
 290 295 300
 Arg Val Arg Ser Ser Thr Leu Lys Gly Val Thr Val Pro Glu Asp Arg
 305 310 315 320
 Ala Pro Ser Met Ile Asp Glu Tyr Pro Ile Leu Ala Val Ala Ala Ala
 325 330 335
 Phe Ala Glu Gly Ala Thr Val Met Asn Gly Leu Glu Glu Leu Arg Val
 340 345 350
 Lys Glu Ser Asp Arg Leu Ser Ala Val Ala Asn Gly Leu Lys Leu Asn
 355 360 365
 Gly Val Asp Cys Asp Glu Gly Glu Thr Ser Leu Val Val Arg Gly Arg
 370 375 380
 Pro Asp Gly Lys Gly Leu Gly Asn Ala Ser Gly Ala Ala Val Ala Thr
 385 390 395 400
 His Leu Asp His Arg Ile Ala Met Ser Phe Leu Val Met Gly Leu Val
 405 410 415
 Ser Glu Asn Pro Val Thr Val Asp Asp Ala Thr Met Ile Ala Thr Ser
 420 425 430
 405 410 415
 Arg Ile Gln Asp Pro Asp Cys Val Ala Lys Thr Tyr Pro Asp Tyr Trp
 420 425 430
 Lys Ala Trp Pro Ser Leu Gly Val His Leu Asn Asp
 435 440
 Phe Pro Glu Phe Met Asp Leu Met Ala Gly Leu Gly Ala Lys Ile Glu
 435 440 445
 Leu Ser Asp Thr Lys Ala Ala
 450 455

<210> 52
 <211> 428
 <212> PRT
 <213> Bacillus subtilis

<220>
 <221> VARIANTE
 <222> (0) ... (0)
 <223> EPSPS-proteína

<400> 52
 Met Lys Arg Asp Lys Val Gln Thr Leu His Gly Glu Ile His Ile Pro
 1 5 10 15
 Gly Asp Lys Ser Ile Ser His Arg Ser Val Met Phe Gly Ala Leu Ala
 20 25 30
 Ala Gly Thr Thr Thr Val Lys Asn Phe Leu Pro Gly Ala Asp Cys Leu
 35 40 45
 Ser Thr Ile Asp Cys Phe Arg Lys Met Gly Val His Ile Glu Gln Ser
 50 55 60
 Ser Ser Asp Val Val Ile His Gly Lys Gly Ile Asp Ala Leu Lys Glu
 65 70 75 80
 Pro Glu Ser Leu Leu Asp Val Gly Asn Ser Gly Thr Thr Ile Arg Leu
 85 90 95

Met Leu Gly Ile Leu Ala Gly Arg Pro Phe Tyr Ser Ala Val Ala Gly
 100 105 110
 Asp Glu Ser Ile Ala Lys Arg Pro Met Lys Arg Val Thr Glu Pro Leu
 115 120 125
 Lys Lys Met Gly Ala Lys Ile Asp Gly Arg Ala Gly Gly Glu Phe Thr
 130 135 140
 Pro Leu Ser Val Ser Gly Ala Ser Leu Lys Gly Ile Asp Tyr Val Ser
 145 150 155 160
 Pro Val Ala Ser Ala Gln Ile Lys Ser Ala Val Leu Leu Ala Gly Leu
 165 170 175
 Gln Ala Glu Gly Thr Thr Thr Val Thr Glu Pro His Lys Ser Arg Asp
 180 185 190
 His Thr Glu Arg Met Leu Ser Ala Phe Gly Val Lys Leu Ser Glu Asp
 195 200 205
 Gln Thr Ser Val Ser Ile Ala Gly Gly Gln Lys Leu Thr Ala Ala Asp
 210 215 220
 Ile Phe Val Pro Gly Asp Ile Ser Ser Ala Ala Phe Phe Leu Ala Ala
 225 230 235 240
 Gly Ala Met Val Pro Asn Ser Arg Ile Val Leu Lys Asn Val Gly Leu
 245 250 255
 Asn Pro Thr Arg Thr Gly Ile Ile Asp Val Leu Gln Asn Met Gly Ala
 260 265 270
 Lys Leu Glu Ile Lys Pro Ser Ala Asp Ser Gly Ala Glu Pro Tyr Gly
 275 280 285
 Asp Leu Ile Ile Glu Thr Ser Ser Leu Lys Ala Val Glu Ile Gly Gly
 290 295 300
 Asp Ile Ile Pro Arg Leu Ile Asp Glu Ile Pro Ile Ile Ala Leu Leu
 305 310 315 320
 Ala Thr Gln Ala Glu Gly Thr Thr Val Ile Lys Asp Ala Ala Glu Leu
 325 330 335
 Lys Val Lys Glu Thr Asn Arg Ile Asp Thr Val Val Ser Glu Leu Arg
 340 345 350
 Lys Leu Gly Ala Glu Ile Glu Pro Thr Ala Asp Gly Met Lys Val Tyr
 355 360 365
 Gly Lys Gln Thr Leu Lys Gly Gly Ala Ala Val Ser Ser His Gly Asp
 370 375 380
 His Arg Ile Gly Met Met Leu Gly Ile Ala Ser Cys Ile Thr Glu Glu
 385 390 395 400
 Pro Ile Glu Ile Glu His Thr Asp Ala Ile His Val Ser Tyr Pro Thr
 405 410 415
 Phe Phe Glu His Leu Asn Lys Leu Ser Lys Lys Ser
 420 425

<210> 53
 <211> 1500
 <212> DNA
 <213> Desconhecida

<220>
 <223> Bactéria desconhecida

<221> CDS
 <222> (40) ... (1371)

<400> 53

ctctacagt tagggcaagt cccccaccac tcgacaagc atg gcg tgt ttg cct 54
 Met Ala Cys Leu Pro
 1 5

gat gat tcg ggt ccg cat gtc ggc cac tcc acg cca cct cgc ctt gac 102
 Asp Asp Ser Gly Pro His Val Gly His Ser Thr Pro Pro Arg Leu Asp
 10 15 20

cag gag cct tgt acc ttg agt tcg cag aaa acc gtg acc gtt aca ccg	150
Gln Glu Pro Cys Thr Leu Ser Ser Gln Lys Thr Val Thr Val Thr Pro	
25 30 35	
ccc aac ttc ccc ctc act ggc aag gtc gcg ccc ccc ggc tcc aaa tcc	198
Pro Asn Phe Pro Leu Thr Gly Lys Val Ala Pro Pro Gly Ser Lys Ser	
40 45 50	
att acc aac cgt gcg ctg ttg ctg gcg gca ttg gcc aag ggc acc agc	246
Ile Thr Asn Arg Ala Leu Leu Leu Ala Ala Leu Ala Lys Gly Thr Ser	
55 60 65	
cgt ttg agc ggt gcg ctc aaa agc gat gac acg cgc cac atg tcg gtc	294
Arg Leu Ser Gly Ala Leu Lys Ser Asp Asp Thr Arg His Met Ser Val	
70 75 80 85	
gcc ctg cgg cag atg ggc gtc acc atc gac gag ccg gac gac acc acc	342
Ala Leu Arg Gln Met Gly Val Thr Ile Asp Glu Pro Asp Asp Thr Thr	
90 95 100	
ttt gtg gtc acc agc caa ggc tcg ctg caa ttg ccg gcc cag ccg ttg	390
Phe Val Val Thr Ser Gln Gly Ser Leu Gln Leu Pro Ala Gln Pro Leu	
105 110 115	
ttc ctc ggc aac gct ggc acc gcc atg cgc ttt ctc acg gct gcc gtg	438
Phe Leu Gly Asn Ala Gly Thr Ala Met Arg Phe Leu Thr Ala Ala Val	
120 125 130	
gcc acc gtg caa ggc acc gtg gta ctg gac ggc gac gag tac atg caa	486
Ala Thr Val Gln Gly Thr Val Val Leu Asp Gly Asp Glu Tyr Met Gln	
135 140 145	
aaa cgc ccg att ggc ccg ctg ctg gct acc ctg ggc cag aac ggc atc	534
Lys Arg Pro Ile Gly Pro Leu Leu Ala Thr Leu Gly Gln Asn Gly Ile	
150 155 160 165	
cag gtc gac agc ccc acc ggt tgc cca ccg gtc acc gtg cac ggc atg	582
Gln Val Asp Ser Pro Thr Gly Cys Pro Pro Val Thr Val His Gly Met	
170 175 180	
ggc aag gtc cag gcc aag cgt ttc gag att gat ggt ggt ttg tcc agc	630
Gly Lys Val Gln Ala Lys Arg Phe Glu Ile Asp Gly Gly Leu Ser Ser	
185 190 195	
cag tac gta tcg gcc ctg ctg atg ctc gcg gcg tgc ggc gaa gcg ccg	678
Gln Tyr Val Ser Ala Leu Leu Met Leu Ala Ala Cys Gly Glu Ala Pro	
200 205 210	
att gaa gtg gcg ctg acc ggc aag gat atc ggt gcc cgt ggc tac gtg	726
Ile Glu Val Ala Leu Thr Gly Lys Asp Ile Gly Ala Arg Gly Tyr Val	
215 220 225	
gac ctg acc ctc gac tgc atg cgt gcc ttc ggg gcc cag gtg gac gcc	774
Asp Leu Thr Leu Asp Cys Met Arg Ala Phe Gly Ala Gln Val Asp Ala	
230 235 240 245	
gtg gac gac acc acc tgg cgc gtc gcc ccc acc ggc tat acc gcc cat	822
Val Asp Asp Thr Thr Trp Arg Val Ala Pro Thr Gly Tyr Thr Ala His	
250 255 260	
gat tac ctg atc gaa ccc gat gcg tcc gcc gcc acg tat ttg tgg gcc	870
Asp Tyr Leu Ile Glu Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala	
265 270 275	

gca gaa gtg ctg acc ggt ggg cgt atc gac atc ggc gta gcc gcg cag 918
Ala Glu Val Leu Thr Gly Gly Arg Ile Asp Ile Gly Val Ala Ala Gln
280 285 290

gac ttc acc cag ccc gac gcc aag gcc cag gcc gtg att gcg cag ttc 966
Asp Phe Thr Gln Pro Asp Ala Lys Ala Gln Ala Val Ile Ala Gln Phe
295 300 305

ccg aac atg caa gcc acg gtg gta ggc tcg caa atg cag gat gcg atc 1014
Pro Asn Met Gln Ala Thr Val Val Gly Ser Gln Met Gln Asp Ala Ile
310 315 320 325

ccg acc ctg gcg gtg ctc gcc gcg ttc aac aac acc ccg gtg cgt ttc 1062
Pro Thr Leu Ala Val Leu Ala Ala Phe Asn Asn Thr Pro Val Arg Phe
330 335 340

act gaa ctg gcg aac ctg cgc gtc aag gaa tgt gac cgc gtg cag gcg 1110
Thr Glu Leu Ala Asn Leu Arg Val Lys Glu Cys Asp Arg Val Gln Ala
345 350 355

ctg cac gat ggc ctc aac gaa att cgc ccg ggc ctg gcg acc atc gag 1158
Leu His Asp Gly Leu Asn Glu Ile Arg Pro Gly Leu Ala Thr Ile Glu
360 365 370

ggc gat gac ctg ctg gtc gcc agc gac ccg gcc ctg gca gcc acc gcc 1206
Gly Asp Asp Leu Leu Val Ala Ser Asp Pro Ala Leu Ala Gly Thr Ala
375 380 385

tgc acc gca ctg atc gac acc cac gcc gac cat cgc atc gcc atg tgc 1254
Cys Thr Ala Leu Ile Asp Thr His Ala Asp His Arg Ile Ala Met Cys
390 395 400 405

ttt gcc ctg gcc ggg ctt aaa gtc tcg ggc att cgc att caa gac ccg 1302
Phe Ala Leu Ala Gly Leu Lys Val Ser Gly Ile Arg Ile Gln Asp Pro
410 415 420

gac tgc gtg gcc aag acc tac cct gac tac tgg aaa gcc tgg ccc agc 1350
Asp Cys Val Ala Lys Thr Tyr Pro Asp Tyr Trp Lys Ala Trp Pro Ser
425 430 435

ctg ggc gtt cac cta aac gac tgacacacaa aacctgtagc agagcttgct 1401
Leu Gly Val His Leu Asn Asp
440

cgcgaaaaac gcacacgtgc cgcgtttgtt caggaaacac gcgttatcgt tgacgtttat 1461
cgagctaaagc tcgctcctac attttgacgc gagatcttg 1500

<210> 54
<211> 444
<212> PRT
<213> Desconhecida

<220>
<223> Bactéria desconhecida

<400> 54
Met Ala Cys Leu Pro Asp Asp Ser Gly Pro His Val Gly His Ser Thr
1 5 10 15
Pro Pro Arg Leu Asp Gln Glu Pro Cys Thr Leu Ser Ser Gln Lys Thr
20 25 30
Val Thr Val Thr Pro Pro Asn Phe Pro Leu Thr Gly Lys Val Ala Pro
35 40 45
Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Ala Leu
50 55 60
Ala Lys Gly Thr Ser Arg Leu Ser Gly Ala Leu Lys Ser Asp Asp Thr
65 70 75 80

Arg His Met Ser Val Ala Leu Arg Gln Met Gly Val Thr Ile Asp Glu
 85 90 95
 Pro Asp Asp Thr Phe Val Val Thr Ser Gln Gly Ser Leu Gln Leu
 100 105 110
 Pro Ala Gln Pro Leu Phe Leu Gly Asn Ala Gly Thr Ala Met Arg Phe
 115 120 125
 Leu Thr Ala Ala Val Ala Thr Val Gln Gly Thr Val Val Leu Asp Gly
 130 135 140
 Asp Glu Tyr Met Gln Lys Arg Pro Ile Gly Pro Leu Leu Ala Thr Leu
 145 150 155 160
 Gly Gln Asn Gly Ile Gln Val Asp Ser Pro Thr Gly Cys Pro Pro Val
 165 170 175
 Thr Val His Gly Met Gly Lys Val Gln Ala Lys Arg Phe Glu Ile Asp
 180 185 190
 Gly Gly Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Leu Ala Ala
 195 200 205
 Cys Gly Glu Ala Pro Ile Glu Val Ala Leu Thr Gly Lys Asp Ile Gly
 210 215 220
 Ala Arg Gly Tyr Val Asp Leu Thr Leu Asp Cys Met Arg Ala Phe Gly
 225 230 235 240
 Ala Gln Val Asp Ala Val Asp Asp Thr Thr Trp Arg Val Ala Pro Thr
 245 250 255
 Gly Tyr Thr Ala His Asp Tyr Leu Ile Glu Pro Asp Ala Ser Ala Ala
 260 265 270
 Thr Tyr Leu Trp Ala Ala Glu Val Leu Thr Gly Gly Arg Ile Asp Ile
 275 280 285
 Gly Val Ala Ala Gln Asp Phe Thr Gln Pro Asp Ala Lys Ala Gln Ala
 290 295 300
 Val Ile Ala Gln Phe Gln Asn Met Gln Ala Thr Val Val Gly Ser Gln
 305 310 315 320
 Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn Asn
 325 330 335
 Thr Pro Val Arg Phe Thr Glu Leu Ala Asn Leu Arg Val Lys Glu Cys
 340 345 350
 Asp Arg Val Gln Ala Leu His Asp Gly Leu Asn Glu Ile Arg Pro Gly
 355 360 365
 Leu Ala Thr Ile Glu Gly Asp Asp Leu Leu Val Ala Ser Asp Pro Ala
 370 375 380
 Leu Ala Gly Thr Ala Cys Thr Ala Leu Ile Asp Thr His Ala Asp His
 385 390 395 400
 Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Val Ser Gly Ile
 405 410 415
 Arg Ile Gln Asp Pro Asp Cys Val Ala Lys Thr Tyr Pro Asp Tyr Trp
 420 425 430
 Lys Ala Trp Pro Ser Leu Gly Val His Leu Asn Asp
 435 440

<210> 55
 <211> 1257
 <212> DNA
 <213> Pseudomonas syringae

<220>
 <221> Região de interesse biológico
 <222> (0) ... (0)
 <223> cepa B728a de pv syringae

<221> CDS
 <222> (1) ... (1257)

<400> 55
 atg cga cct caa gcc acc ctc act gtt ttg cct gtc gag cgc ccg ttg 48
 Met Arg Pro Gln Ala Thr Leu Thr Val Leu Pro Val Glu Arg Pro Leu
 1 5 10 15

gtc ggg cgt gtc agc ccg ccg ggc tcc aag tcg atc acc aac cgt gca	96
Val Gly Arg Val Ser Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala	
20 25 30	
ttg ttg ctg gcc ggg ctg gcc aaa ggc acc agc cgc ctg acc ggc gcg	144
Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala	
35 40 45	
ctg aag agt gat gac acg cgc gtg atg tcc gaa gcc ttg cgt ctg atg	192
Leu Lys Ser Asp Asp Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met	
50 55 60	
ggc gtg cag gtc gac gag ccg gat gac agc acc ttc gtg gtc acc agc	240
Gly Val Gln Val Asp Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser	
65 70 75 80	
agc ggg cac tgg cag gcg ccg caa cag gcg ctg ttc ctc ggc aat gcc	288
Ser Gly His Trp Gln Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala	
85 90 95	
ggg act gcg aca cgc ttt ctg acc gcg gca ctg gcc aac ttc gaa ggc	336
Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Leu Ala Asn Phe Glu Gly	
100 105 110	
gac ttc gtg gtg gat ggc gac gaa tac atg cgc aag cgc ccg atc ggc	384
Asp Phe Val Val Asp Gly Asp Glu Tyr Met Arg Lys Arg Pro Ile Gly	
115 120 125	
ccg ttg gtc gat gcc ttg cag cgc atg ggc gtg gag gtc agc gca ccc	432
Pro Leu Val Asp Ala Leu Gln Arg Met Gly Val Glu Val Ser Ala Pro	
130 135 140	
agt ggt tgc ccg ccg gtg gcc atc aag ggc aag ggc ggt ctt gag gcc	480
Ser Gly Cys Pro Pro Val Ala Ile Lys Gly Lys Gly Gly Leu Glu Ala	
145 150 155 160	
ggt cgt atc gaa atc gat ggc aat ctg tcc agc cag tat gtg tcg gca	528
Gly Arg Ile Glu Ile Asp Gly Asn Leu Ser Ser Gln Tyr Val Ser Ala	
165 170 175	
ctg ctg atg gcc ggt gcc tgt ggc aag ggg cct gtg gaa gtt gcc ctg	576
Leu Leu Met Ala Gly Ala Cys Gly Lys Gly Pro Val Glu Val Ala Leu	
180 185 190	
aca ggc agc gag atc ggc gcg cgt ggt tac ctc gac ctc acg ctg gcg	624
Thr Gly Ser Glu Ile Gly Ala Arg Gly Tyr Leu Asp Leu Thr Leu Ala	
195 200 205	
gcc atg ccg gcg ttc ggt gcc gag gtt cag gcc atc ggc gac gcc gcc	672
Ala Met Arg Ala Phe Gly Ala Glu Val Gln Ala Ile Gly Asp Ala Ala	
210 215 220	
tgg aaa gtc tcg gct acc ggt tat cgc gct acg gat ttc cac atc gaa	720
Trp Lys Val Ser Ala Thr Gly Tyr Arg Ala Thr Asp Phe His Ile Glu	
225 230 235 240	
ccg gat gcc tcg gcg gcc acc tac ctt tgg gct gcg cag gcc ctg acc	768
Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Gln Ala Leu Thr	
245 250 255	
gag ggc gct atc gac ctg ggc gtg gcc agc aac gcg ttc act cag cct	816
Glu Gly Ala Ile Asp Leu Gly Val Ala Ser Asn Ala Phe Thr Gln Pro	
260 265 270	
gat gca ctg gcc agt cag atc atc gcc agc ttc ccg aac atg ccg gcc	864
Asp Ala Leu Ala Ser Gln Ile Ile Ala Ser Phe Pro Asn Met Pro Ala	
275 280 285	

gtg atc gac ggc tgc cag atg cag gac gcg att ccc acg ctg gcc gta 912
 Val Ile Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val
 290 295 300
 ctg gcc gcg ttc aat cgt caa ccg gtg cgc ttt gtc ggc atc gcc aac 960
 Leu Ala Ala Phe Asn Arg Gln Pro Val Arg Phe Val Gly Ile Ala Asn
 305 310 315 320
 ctg cgg gtc aag gag tgc gac cgc atc tgc gca ctg tcc aac ggc ctg 1008
 Leu Arg Val Lys Glu Cys Asp Arg Ile Ser Ala Leu Ser Asn Gly Leu
 325 330 335
 tgc gcc atc gca ccc ggc ctg gcg gtc gaa gag ggt gac gat ctg atc 1056
 Cys Ala Ile Ala Pro Gly Leu Ala Val Glu Glu Gly Asp Asp Leu Ile
 340 345 350
 gtt acc gcc aac ccg acg ctg gca ggc act acg gtc gat gcc ttg atc 1104
 Val Thr Ala Asn Pro Thr Leu Ala Gly Thr Thr Val Asp Ala Leu Ile
 355 360 365
 gat acc cac tcc gac cat cgg atc gcc atg tgc ttt gca ctg gcg ggc 1152
 Asp Thr His Ser Asp His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly
 370 375 380
 ctg aag att gcc ggc atc cgc att ctc gac cct gac tgc gtc gcc aag 1200
 Leu Lys Ile Ala Gly Ile Arg Ile Leu Asp Pro Asp Cys Val Ala Lys
 385 390 395 400
 acc tac ccg ggg tac tgg gat gcg ctg gtt tct ctg ggt gtg agt gtt 1248
 Thr Tyr Pro Gly Tyr Trp Asp Ala Leu Val Ser Leu Gly Val Ser Val
 405 410 415
 cag cgc tga 1257
 Gln Arg *

<210> 56
 <211> 418
 <212> PRT
 <213> Pseudomonas syringae
 <220>
 <221> VARIANTE
 <222> (0) ... (0)
 <223> cepa B728a de pv syringae

<400> 56
 Met Arg Pro Gln Ala Thr Leu Thr Val Leu Pro Val Glu Arg Pro Leu
 1 5 10 15
 Val Gly Arg Val Ser Pro Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala
 20 25 30
 Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly Ala
 35 40 45
 Leu Lys Ser Asp Asp Thr Arg Val Met Ser Glu Ala Leu Arg Leu Met
 50 55 60
 Gly Val Gln Val Asp Glu Pro Asp Asp Ser Thr Phe Val Val Thr Ser
 65 70 75 80
 Ser Gly His Trp Gln Ala Pro Gln Gln Ala Leu Phe Leu Gly Asn Ala
 85 90 95
 Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Leu Ala Asn Phe Glu Gly
 100 105 110
 Asp Phe Val Val Asp Gly Asp Glu Tyr Met Arg Lys Arg Pro Ile Gly
 115 120 125
 Pro Leu Val Asp Ala Leu Gln Arg Met Gly Val Glu Val Ser Ala Pro
 130 135 140
 Ser Gly Cys Pro Pro Val Ala Ile Lys Gly Lys Gly Gly Leu Glu Ala
 145 150 155 160
 Gly Arg Ile Glu Ile Asp Gly Asn Leu Ser Ser Gln Tyr Val Ser Ala
 165 170 175

Leu Leu Met Ala Gly Ala Cys Gly Lys Gly Pro Val Glu Val Ala Leu
 180 185 190
 Thr Gly Ser Glu Ile Gly Ala Arg Gly Tyr Leu Asp Leu Thr Leu Ala
 195 200 205
 Ala Met Arg Ala Phe Gly Ala Glu Val Gln Ala Ile Gly Asp Ala Ala
 210 215 220
 Trp Lys Val Ser Ala Thr Gly Tyr Arg Ala Thr Asp Phe His Ile Glu
 225 230 235 240
 Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Gln Ala Leu Thr
 245 250 255
 Glu Gly Ala Ile Asp Leu Gly Val Ala Ser Asn Ala Phe Thr Gln Pro
 260 265 270
 Asp Ala Leu Ala Ser Gln Ile Ile Ala Ser Phe Pro Asn Met Pro Ala
 275 280 285
 Val Ile Asp Gly Ser Gln Met Gln Asp Ala Ile Pro Thr Leu Ala Val
 290 295 300
 Leu Ala Ala Phe Asn Arg Gln Pro Val Arg Phe Val Gly Ile Ala Asn
 305 310 315 320
 Leu Arg Val Lys Glu Cys Asp Arg Ile Ser Ala Leu Ser Asn Gly Leu
 325 330 335
 Cys Ala Ile Ala Pro Gly Leu Ala Val Glu Glu Gly Asp Asp Leu Ile
 340 345 350
 Val Thr Ala Asn Pro Thr Leu Ala Gly Thr Thr Val Asp Ala Leu Ile
 355 360 365
 Asp Thr His Ser Asp His Arg Ile Ala Met Cys Phe Ala Leu Ala Gly
 370 375 380
 Leu Lys Ile Ala Gly Ile Arg Ile Leu Asp Pro Asp Cys Val Ala Lys
 385 390 395 400
 Thr Tyr Pro Gly Tyr Trp Asp Ala Leu Val Ser Leu Gly Val Ser Val
 405 410 415
 Gln Arg

<210> 57
 <211> 1332
 <212> DNA
 <213> Ochrobactrum/Brucella

<221> CDS
 <222> (1) ... (1332)

<400> 57
 atg gcg tgt ttg cct gat gat tcg ggt ccg cac gtc ggc cac tcc acg 48
 Met Ala Cys Leu Pro Asp Asp Ser Gly Pro His Val Gly His Ser Thr
 1 5 10 15
 cca cct tgc ctt gac cag gag cct tgt acc ttg agt tcg cag aaa acc 96
 Pro Pro Cys Leu Asp Gln Glu Pro Cys Thr Leu Ser Ser Gln Lys Thr
 20 25 30
 gtg acc gtt aca ccg ccc aac ttc ccc ctc act ggc aag gtc gcg ccc 144
 Val Thr Val Thr Pro Pro Asn Phe Pro Leu Thr Gly Lys Val Ala Pro
 35 40 45
 ccc ggc tcc aaa tcc att acc aac cgt gcg ctg ttg ctc gcg gca ctg 192
 Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Ala Leu
 50 55 60
 gcc aag ggc acc agc cgt ttg agc ggt gcg ctc aaa agc gat gac acg 240
 Ala Lys Gly Thr Ser Arg Leu Ser Gly Ala Leu Lys Ser Asp Asp Thr
 65 70 75 80

cgc cac atg tgc gtc gcc ctg cgg cag atg ggc gtc acc atc gac gag	288
Arg His Met Ser Val Ala Leu Arg Gln Met Gly Val Thr Ile Asp Glu	
85 90 95	
ccg gac gac acc acc ttt gtc gtc acc agc caa ggc tgc ctg caa ttg	336
Pro Asp Asp Thr Thr Phe Val Val Thr Ser Gln Gly Ser Leu Gln Leu	
100 105 110	
ccg gcc cag ccg ttg ttc ctc ggc aac gcc ggc acc gcc atg cgc ttt	384
Pro Ala Gln Pro Leu Phe Leu Gly Asn Ala Gly Thr Ala Met Arg Phe	
115 120 125	
ctc acg gct gcc gtc gcc acc gtg caa ggc acc gtg gta ctg gac ggt	432
Leu Thr Ala Ala Val Ala Thr Val Gln Gly Thr Val Val Leu Asp Gly	
130 135 140	
gac gag tac atg caa aaa cgc ccg atc ggc ccg ctg ctg gcc acc ctg	480
Asp Glu Tyr Met Gln Lys Arg Pro Ile Gly Pro Leu Leu Ala Thr Leu	
145 150 155 160	
ggc cag aac ggc atc cag gtc gac agc ccc acc ggt tgc cca ccg gta	528
Gly Gln Asn Gly Ile Gln Val Asp Ser Pro Thr Gly Cys Pro Pro Val	
165 170 175	
acg gtg cat ggc gcg ggc aag gtc cag gcc agg cgt ttt gag att gac	576
Thr Val His Gly Ala Gly Lys Val Gln Ala Arg Arg Phe Glu Ile Asp	
180 185 190	
gga ggc ttg tcc agc cag tac gta tgc gcc ctg ctg atg ctg gcg gcg	624
Gly Gly Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Leu Ala Ala	
195 200 205	
tgc ggc gaa gca ccg att gaa gtg gcg ctg acc ggc aag gac atc ggc	672
Cys Gly Glu Ala Pro Ile Glu Val Ala Leu Thr Gly Lys Asp Ile Gly	
210 215 220	
gcc cgt ggc tat gtg gac ctg acc ctc gat tgc atg cgt gcg ttc ggg	720
Ala Arg Gly Tyr Val Asp Leu Thr Leu Asp Cys Met Arg Ala Phe Gly	
225 230 235 240	
gcc cag gta gac atc gtg gac gac acc acc tgg cgc gtg gcc ccc acg	768
Ala Gln Val Asp Ile Val Asp Asp Thr Thr Trp Arg Val Ala Pro Thr	
245 250 255	
ggc tat acc gcc cat gat tac ctg atc gaa ccc gac gct tcc gcc gcc	816
Gly Tyr Thr Ala His Asp Tyr Leu Ile Glu Pro Asp Ala Ser Ala Ala	
260 265 270	
act tac ctg tgg gcc gca gaa gta ctg acc ggt ggc cgt atc gat att	864
Thr Tyr Leu Trp Ala Ala Glu Val Leu Thr Gly Gly Arg Ile Asp Ile	
275 280 285	
ggt gta gcc gcg cag gac ttc acc cag ccc gac gcc aag gca cag gcc	912
Gly Val Ala Ala Gln Asp Phe Thr Gln Pro Asp Ala Lys Ala Gln Ala	
290 295 300	
gtg atc gcg caa ttc ccg aac atg cag gcc acg gtg gtg ggt tca caa	960
Val Ile Ala Gln Phe Pro Asn Met Gln Ala Thr Val Val Gly Ser Gln	
305 310 315 320	
atg cag gat gcg atc ccg acc ctg gcg gtg ctc gcc gca ttc aac aat	1008
Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn Asn	
325 330 335	
acc ccg gtg cgc ttc act gaa ctg gcg aac ctg cgc gtc aag gaa tgt	1056
Thr Pro Val Arg Phe Thr Glu Leu Ala Asn Leu Arg Val Lys Glu Cys	
340 345 350	

gac cgc gtg cag gcg ctg cac gat ggc ctc aac gaa att cgc ccg ggc 1104
 Asp Arg Val Gln Ala Leu His Asp Gly Leu Asn Glu Ile Arg Pro Gly
 355 360 365

ctg gcg acc atc gaa ggt gat gac ctg ctg gtt gcc agc gac ccc gct 1152
 Leu Ala Thr Ile Glu Gly Asp Asp Leu Leu Val Ala Ser Asp Pro Ala
 370 375 380

ttg gct ggc acc gcc tgc acc gca ctg atc gat acc cac gcc gac cat 1200
 Leu Ala Gly Thr Ala Cys Thr Ala Leu Ile Asp Thr His Ala Asp His
 385 390 395 400

cgc atc gcc atg tgc ttt gcc ctg gcc ggg ctg aaa gtc tcg ggc att 1248
 Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Val Ser Gly Ile
 405 410 415

cgc atc caa gac cct gat tgc gta gcc aag acc tac cct gac tac tgg 1296
 Arg Ile Gln Asp Pro Asp Cys Val Lys Thr Tyr Pro Asp Tyr Trp
 420 425 430

aaa gcg ctg gcc agc ctg ggc gtt cac tta agc tac 1332
 Lys Ala Leu Ala Ser Leu Gly Val His Leu Ser Tyr
 435 440

<210> 58
 <211> 1236
 <212> DNA
 <213> Ochrobactrum/Brucella

<221> CDS
 <222> (1) ... (1236)

<400> 58

gtg acc gtt aca ccg ccc aac ttc ccc ctc act ggc aag gtc gcg ccc 48
 Met Thr Val Thr Pro Pro Asn Phe Pro Leu Thr Gly Lys Val Ala Pro
 1 5 10 15

ccc ggc tcc aaa tcc att acc aac cgt gcg ctg ttg ctc gcg gca ctg 96
 Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Ala Leu
 20 25 30

gcc aag gcg acc agc cgt ttg agc ggt gcg ctc aaa agc gat gac acg 144
 Ala Lys Gly Thr Ser Arg Leu Ser Gly Ala Leu Lys Ser Asp Asp Thr
 35 40 45

cgc cac atg tcg gtc gcc ctg cgg cag atg ggc gtc acc atc gac gag 192
 Arg His Met Ser Val Ala Leu Arg Gln Met Gly Val Thr Ile Asp Glu
 50 55 60

ccg gac gac acc acc ttt gtc gtc acc agc caa ggc tcg ctg caa ttg 240
 Pro Asp Asp Thr Thr Phe Val Val Thr Ser Gln Gly Ser Leu Gln Leu
 65 70 75 80

ccg gcc cag ccg ttg ttc ctc ggc aac gcc ggc acc gcc atg cgc ttt 288
 Pro Ala Gln Pro Leu Phe Leu Gly Asn Ala Gly Thr Ala Met Arg Phe
 85 90 95

ctc acg gct gcc gtc gcc acc gtg caa ggc acc gtg gta ctg gac ggt 336
 Leu Thr Ala Ala Val Ala Thr Val Gln Gly Thr Val Val Leu Asp Gly
 100 105 110

gac gag tac atg caa aaa cgc ccg atc ggc ccg ctg ctg gcc acc ctg 384
 Asp Glu Tyr Met Gln Lys Arg Pro Ile Gly Pro Leu Leu Ala Thr Leu
 115 120 125

ggc cag aac ggc atc cag gtc gac agc ccc acc ggt tgc cca ccg gta	432
Gly Gln Asn Gly Ile Gln Val Asp Ser Pro Thr Gly Cys Pro Pro Val	
130 135 140	
acg gtg cat ggc gcg ggc aag gtc cag gcc agg cgt ttt gag att gac	480
Thr Val His Gly Ala Gly Lys Val Gln Ala Arg Arg Phe Glu Ile Asp	
145 150 155 160	
gga ggc ttg tcc agc cag tac gta tgc gcc ctg ctg atg ctg gcg gcg	528
Gly Gly Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Leu Ala Ala	
165 170 175	
tgc ggc gaa gca ccg att gaa gtg gcg ctg acc ggc aag gac atc ggc	576
Cys Gly Glu Ala Pro Ile Glu Val Ala Leu Thr Gly Lys Asp Ile Gly	
180 185 190	
gcc cgt ggc tat gtg gac ctg acc ctc gat tgc atg cgt gcg ttc ggg	624
Ala Arg Gly Tyr Val Asp Leu Thr Leu Asp Cys Met Arg Ala Phe Gly	
195 200 205	
gcc cag gta gac atc gtg gac gac acc acc tgg cgc gtg gcc ccc acg	672
Ala Gln Val Asp Ile Val Asp Asp Thr Thr Trp Arg Val Ala Pro Thr	
210 215 220	
ggc tat acc gcc cat gat tac ctg atc gaa ccc gac gct tcc gcc gcc	720
Gly Tyr Thr Ala His Asp Tyr Leu Ile Glu Pro Asp Ala Ser Ala Ala	
225 230 235 240	
act tac ctg tgg gcc gca gaa gta ctg acc ggt ggc cgt atc gat att	768
Thr Tyr Leu Trp Ala Ala Glu Val Leu Thr Gly Gly Arg Ile Asp Ile	
245 250 255	
ggg gta gcc gcg cag gac ttc acc cag ccc gac gcc aag gca cag gcc	816
Gly Val Ala Ala Gln Asp Phe Thr Gln Pro Asp Ala Lys Ala Gln Ala	
260 265 270	
gtg atc gcg caa ttc ccg aac atg cag gcc acg gtg gtg ggt tca caa	864
Val Ile Ala Gln Phe Pro Asn Met Gln Ala Thr Val Val Gly Ser Gln	
275 280 285	
atg cag gat gcg atc ccg acc ctg gcg gtg ctc gcc gca ttc aac aat	912
Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn Asn	
290 295 300	
acc ccg gtg cgc ttc act gaa ctg gcg aac ctg cgc gtc aag gaa tgt	960
Thr Pro Val Arg Phe Thr Glu Leu Ala Asn Leu Arg Val Lys Glu Cys	
305 310 315 320	
gac cgc gtg cag gcg ctg cac gat ggc ctc aac gaa att cgc ccg ggc	1008
Asp Arg Val Gln Ala Leu His Asp Gly Leu Asn Glu Ile Arg Pro Gly	
325 330 335	
ctg gcg acc atc gaa ggt gat gac ctg ctg gtt gcc agc gac ccc gct	1056
Leu Ala Thr Ile Glu Gly Asp Asp Leu Leu Val Ala Ser Asp Pro Ala	
340 345 350	
ttg gct ggc acc gcc tgc acc gca ctg atc gat acc cac gcc gac cat	1104
Leu Ala Gly Thr Ala Cys Thr Ala Leu Ile Asp Thr His Ala Asp His	
355 360 365	
cgc atc gcc atg tgc ttt gcc ctg gcc ggg ctg aaa gtc tgc ggc att	1152
Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Val Ser Gly Ile	
370 375 380	
cgc atc caa gac cct gat tgc gta gcc aag acc tac cct gac tac tgg	1200
Arg Ile Gln Asp Pro Asp Cys Val Ala Lys Thr Tyr Pro Asp Tyr Trp	
385 390 395 400	
aaa gcg ctg gcc agc ctg ggc gtt cac tta agc tac	1236
Lys Ala Leu Ala Ser Leu Gly Val His Leu Ser Tyr	
405 410	

<210> 59
 <211> 412
 <212> PRT
 <213> Ochrobactrum/Brucella

<400> 59
 Met Thr Val Thr Pro Pro Asn Phe Pro Leu Thr Gly Lys Val Ala Pro
 1 5 10 15
 Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Ala Leu
 20 25 30
 Ala Lys Gly Thr Ser Arg Leu Ser Gly Ala Leu Lys Ser Asp Asp Thr
 35 40 45
 Arg His Met Ser Val Ala Leu Arg Gln Met Gly Val Thr Ile Asp Glu
 50 55 60
 Pro Asp Asp Thr Thr Phe Val Val Thr Ser Gln Gly Ser Leu Gln Leu
 65 70 75 80
 Pro Ala Gln Pro Leu Phe Leu Gly Asn Ala Gly Thr Ala Met Arg Phe
 85 90 95
 Leu Thr Ala Ala Val Ala Thr Val Gln Gly Thr Val Val Leu Asp Gly
 100 105 110
 Asp Glu Tyr Met Gln Lys Arg Pro Ile Gly Pro Leu Leu Ala Thr Leu
 115 120 125
 Gly Gln Asn Gly Ile Gln Val Asp Ser Pro Thr Gly Cys Pro Pro Val
 130 135 140
 Thr Val His Gly Ala Gly Lys Val Gln Ala Arg Arg Phe Glu Ile Asp
 145 150 155 160
 Gly Gly Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Leu Ala Ala
 165 170 175
 Cys Gly Glu Ala Pro Ile Glu Val Ala Leu Thr Gly Lys Asp Ile Gly
 180 185 190
 Ala Arg Gly Tyr Val Asp Leu Thr Leu Asp Cys Met Arg Ala Phe Gly
 195 200 205
 Ala Gln Val Asp Ile Val Asp Asp Thr Thr Trp Arg Val Ala Pro Thr
 210 215 220
 Gly Tyr Thr Ala His Asp Tyr Leu Ile Glu Pro Asp Ala Ser Ala Ala
 225 230 235 240
 Thr Tyr Leu Trp Ala Ala Glu Val Leu Thr Gly Gly Arg Ile Asp Ile
 245 250 255
 Gly Val Ala Ala Gln Asp Phe Thr Gln Pro Asp Ala Lys Ala Gln Ala
 260 265 270
 Val Ile Ala Gln Phe Pro Asn Met Gln Ala Thr Val Val Gly Ser Gln
 275 280 285
 Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn Asn
 290 295 300
 Thr Pro Val Arg Phe Thr Glu Leu Ala Asn Leu Arg Val Lys Glu Cys
 305 310 315 320
 Asp Arg Val Gln Ala Leu His Asp Gly Leu Asn Glu Ile Arg Pro Gly
 325 330 335
 Leu Ala Thr Ile Glu Gly Asp Asp Leu Leu Val Ala Ser Asp Pro Ala
 340 345 350
 Leu Ala Gly Thr Ala Cys Thr Ala Leu Ile Asp Thr His Ala Asp His
 355 360 365
 Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Val Ser Gly Ile
 370 375 380
 Arg Ile Gln Asp Pro Asp Cys Val Ala Lys Thr Tyr Pro Asp Tyr Trp
 385 390 395 400
 Lys Ala Leu Ala Ser Leu Gly Val His Leu Ser Tyr
 405 410

<210> 60
 <211> 1496
 <212> DNA
 <213> Desconhecido

<220>
 <223> Isolado da amostra de solo

<400> 60

```

aagcacccggc tcacaaatc ccgcccgat tccgatagcc cggctttgcg ccgcattctt 60
tgacaagacg acgatcatga tcgaactgac catcaccgcc cccggccacc cgctttccgg 120
caaggtggag ccgcccgggt ccaaattccat taccaaccgt gcacttctgc tggccggggt 180
cgccaagggc aaaagccgtc tcacggggcg gctgaaaagc gacgatacgc ttacatggc 240
agaagcgctg cgtgagatgg gtgtcaaggt aaccgagcct gacgcgacca ccttcgtggt 300
ggagagtcca ggtgggttgc atcagccgga aaagccgctt ttctcggca atgcccggcac 360
agctaccgcg ttctcaccg ccgctgccgc ccttggtgat ggccgctca tcatcgatgg 420
cgacgagcat atgcgcaaac gcccgatcat gccgctggtg gaagccctgc gctccctcgg 480
cgttgaggcg gaggcgccga ccgctgccc gccgctcacc gtctgcggtg aggggtactgg 540
cttcccgaa ggcagcgta cgatcgacgc caacctttcc agccagtatg tgtccgcact 600
tctgatggcc gccgctgcg gcgacaagcc tgtcgatata atcctcaaag gtgaggaaat 660
cggcgcaag ggctatatcg atctcaccac atcggccatg gaagccttcg gcgcaaaagt 720
ggagcgggtc agcaacgcca tctggcgctg gcatccgacc ggctacacgg cgaccgattt 780
ccatatcgag ccggatgcct cggccgccac ctatctctgg ggcgctgagc ttttgaccgg 840
cggcgccatc gatatcggtg cggccggcga caagttcacc cagccggatg ccaaggccca 900
tgaggtcatg gcgcaatttc cgcattcgcc cgcggaaatc gacggttcgc agatgcagga 960
tgccattccc accattgcg ttctcgcgc ctttaacgaa acgcccgtgc gtttcgtcgg 1020
catcgccaat ctgcgcgtca aggagtgcga cgaatccgc gccgtctcac tcggcctcaa 1080
cgaaatccgc gatggtcttg cgcattgagga aggcgacgac ttgatcgtgc attccgatcc 1140
ttcgcttgcg ggccagacgg tgaatgcctc catcgacact ttccgcccacc accgtatcgc 1200
catgagcttt gcgctggcgg cgctgaagat cggcgccatt gccatccaga atccggcctg 1260
cgtgggcaag acctatccc gttactggaa ggcgctcgcc tcgctgggag tcgaatactc 1320
ggaaaaggaa accgctgccg agccgcagca ttagaagaac ggtattggtt ttccgtcatt 1380
ccggccctga gccggaatcc agtgcgatca agtccttgat cgcgaaagac tcttcaagcc 1440
gcgcagacgc gcggctgctg gataccgggt ctagcccgtg atgacggcaa tgactg 1496

```

<210> 61
 <211> 33
 <212> DNA
 <213> Sequência artificial

<220>
 <223> Oligonucleotídeo

<400> 61
 cagagatctg gcatgacgacc tcaagccact ctc 33

<210> 62
 <211> 31
 <212> DNA
 <213> Sequência artificial

<220>
 <223> Oligonucleotídeo

<400> 62
 cagggcgcg ctcagcgctg aacactcacc c 31

<210> 63
 <211> 34
 <212> DNA
 <213> Sequência artificial

<220>
 <223> Oligonucleotídeo

<400> 63
 cagggatccg gcatgatcga actgaccatc accc 34

<210> 64
 <211> 33
 <212> DNA
 <213> Sequência artificial

<220>
 <223> Oligonucleotídeo

 <400> 64
 cagggcgcg ctcagtgcgtg cggctcggca gcg 33

 <210> 65
 <211> 33
 <212> DNA
 <213> Sequência artificial

 <220>
 <223> Oligonucleotídeo

 <400> 65
 cagggatccg gcatgcgacc tcaagccacc etc 33

 <210> 66
 <211> 31
 <212> DNA
 <213> Sequência artificial

 <220>
 <223> Oligonucleotídeo

 <400> 66
 cagggcgcg ctcagcgctg aacactcaca c 31

 <210> 67
 <211> 33
 <212> DNA
 <213> Sequência artificial

 <220>
 <223> Oligonucleotídeo

 <400> 67
 cagggatccg gcatgcgacc tcaagccacc etc 33

 <210> 68
 <211> 30
 <212> DNA
 <213> Sequência artificial

 <220>
 <223> Oligonucleotídeo

 <400> 68
 agagggcgcg ctcagcgctg aacacgcacc 30

REIVINDICAÇÕES

1. Polinucleotídeo isolado diferente do polinucleotídeo listado nas SEQ ID NOS: 1, 3, 7, 45, 47, e 53 que codifica uma enzima EPSPS que tem pelo menos um domínio de sequência selecionado a partir do grupo que
5 consiste em:

Domínio I:

L-A-K-G-X₁-S-X₂-L-X₃-G-A-L-K-S-D-D-T (SEQ ID NO: 13), onde X₁ denota lisina ou treonina, onde X₂ denota arginina ou histidina, onde X₃ denota serina ou treonina;

10 Domínio Ia:

L-A-K-G-X₁, (SEQ ID NO: 14), onde X₁ denota lisina ou treonina;

Domínio Ib:

S-X₁-L-X₂ (SEQ ID NO: 15), onde X₁ denota arginina ou histidina, onde X₂ denota serina ou treonina;

15 Domínio Ic:

G-A-L-K-S-D-D-T (SEQ ID NO: 16);

Domínio II:

E-P-D-X₁-X₂-T-F-X₃-V-X₄-X₅-X₆-G (SEQ ID NO: 17), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃
20 denota valina ou isoleucina, onde X₄ denota treonina ou ácido glutâmico, ou lisina, onde X₅ denota serina ou glicina, onde X₆ denota glutamina ou serina ou ácido glutâmico ou treonina;

Domínio IIa:

E-P-D-X₁-X₂-T-F-X₃-V (SEQ ID NO: 18), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota va-
25 lina ou isoleucina;

Domínio IIb:

X₁-X₂-X₃-G (SEQ ID NO: 19), onde X₁ denota treonina ou ácido glutâmico ou lisina, onde X₂ denota serina ou glicina, onde X₃ denota gluta-
30 mina ou serina ou ácido glutâmico ou treonina;

Domínio III:

RFLTAA (SEQ ID NO: 20);

Domínio IV:

KRPI(G/M)**P** (SEQ ID NO: 21)

K-R-P-I-X₁-P, onde X₁ denota glicina ou metionina ou leucina;

Domínio V:

- 5 X₁-G-C-P-P-V (SEQ ID NO: 22), onde X₁ denota treonina ou serina;

Domínio VI:

I-G-A-X₁-G-Y-X₂-D-L-T (SEQ ID NO: 23), onde X₁ denota arginina ou lisina ou leucina, e onde X₂ denota isoleucina ou valina;

10 Domínio VII:

W-X₁-V-X₂-X₃-T-G (SEQ ID NO: 24), onde X₁ denota arginina ou lisina, onde X₂ denota alanina ou histidina ou ácido glutâmico ou serina, onde X₃ denota prolina ou alanina;

Domínio VIII:

- 15 E-P-D-A-S-A-A-T- Y-L-W-X₁-A-X₂-X₃-L (SEQ ID NO: 25), onde X₁ denota alanina ou glicina, onde X₂ denota ácido glutâmico ou glutamina, onde X₃ denota valina ou leucina ou alanina;

Domínio VIIIa:

E-P-D-A-S-A-A-T- Y-L-W (SEQ ID NO: 26);

20 Domínio IX:

I-D-X₁-G (SEQ ID NO: 27), onde X₁ denota isoleucina ou leucina;

Domínio X:

F-X₁-Q-P-D-A- K-A (SEQ ID NO: 28), onde X₁ denota treonina ou serina;

25 Domínio XI:

- X₁-F-P-X₂-X₃-X₄-A-X₅-X₆-X₇-G-S-Q-M-Q-D-A-I-P-T-X₈-A-V- X₉ A-A-F-N (SEQ ID NO: 29), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina, onde X₅ denota treonina ou ácido glutâmico ou valina, onde X₆ denota valina ou isoleucina, onde X₇ denota ácido aspártico ou valina, onde X₈ denota leucina ou isoleucina, onde X₉ denota leucina ou isoleucina;
- 30

Domínio XIa:

X_1 -**F-P**- X_2 - X_3 - X_4 -A (SEQ ID NO: 30), onde X_1 denota glutamina ou lisina ou serina, onde X_2 denota asparagina ou histidina, onde X_3 denota metionina ou leucina, onde X_4 denota prolina ou glutamina;

5 Domínio XIb:

X_1 - X_2 - X_3 -G-**S-Q**-M-**Q**-D-A-I-P-T- X_4 -A- V- X_5 A-A- **F-N** (SEQ ID NO: 31), onde X_1 denota treonina ou ácido glutâmico ou valina, onde X_2 denota valina ou isoleucina, onde X_3 denota ácido aspártico ou valina, onde X_4 denota leucina ou isoleucina, onde X_5 denota leucina ou isoleucina;

10 Domínio XIc:

G-S-Q-M-Q-D-A-I-P-T (SEQ ID NO: 32);

Domínio XII:

P-V-R-F- X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C-D-R**- X_5 (SEQ ID NO: 33), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

Domínio XIIIa:

P-V-R-F (SEQ ID NO: 34);

Domínio XIIIb:

20 X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C-D-R**- X_5 (SEQ ID NO: 35), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

Domínio XIIIc:

25 N-L-R-V-K-E-**C-D-R** (SEQ ID NO: 36);

Domínio XIII:

E-G-D-D-L- X_1 - X_2 (SEQ ID NO: 37), onde X_1 denota leucina ou isoleucina, onde X_2 denota valina ou isoleucina;

Domínio XIV:

30 X_1 -P- X_2 -L-A-G (SEQ ID NO: 38), onde X_1 denota ácido aspártico ou asparagina, onde X_2 denota alanina ou serina ou treonina;

Domínio XV:

A-X₁-I-D-X₂-X₃-X₄- D-H-R (SEQ ID NO: 39), onde X₁ denota leucina ou serina ou ácido glutâmico, onde X₂ denota treonina ou serina, onde X₃ denota histidina ou fenilalanina, onde X₄ denota alanina ou serina;

Domínio XVI:

- 5 F-A-L-A-X₁-L-K-X₂-X₃-G-I (SEQ ID NO: 40), onde X₁ denota glicina ou alanina, onde X₂ denota isoleucina ou valina, onde X₃ denota serina ou glicina ou alanina ou lisina;

Domínio XVIa:

- 10 F-A-L-A-X₁-L-K (SEQ ID NO: 41), onde X₁ denota glicina ou alanina;

Domínio XVIb:

L-K-X₁-X₂-G-I (SEQ ID NO: 42), onde X₁ denota isoleucina ou valina, onde X₂ denota serina ou glicina ou alanina ou lisina;

Domínio XVII:

- 15 -X₁-P-X₂-C-V-X₃-K (SEQ ID NO: 43), onde X₁ denota asparagina ou ácido aspártico, onde X₂ denota alanina ou ácido aspártico, onde X₃ denota alanina ou glicina;

Domínio XVIII:

- 20 X₁-S-L-G-V (SEQ ID NO: 44), onde X₁ denota alanina ou serina ou prolina.

2. Polinucleotídeo de acordo com a reivindicação 1, em que o dito polinucleotídeo é selecionado a partir do grupo que consiste em:

- a) a seqüência de nucleotídeo de SEQ ID NO: 9, 11, 55, 57, 58, ou 60, ou um complemento dessas;
- 25 b) uma seqüência de nucleotídeo que tem pelo menos 95% de identidade de seqüência com a seqüência de nucleotídeo de SEQ ID NO: 9, 11, 57, ou 59, ou 61, ou um complemento dessas;
- 30 c) a seqüência de nucleotídeo de resistência do inserto de DNA do plasmídeo depositado como N^os de Acesso B-30833 ou B-30838, ou um complemento desses;
- d) uma seqüência de nucleotídeo que codifica um polipeptí-

deo que compreende a seqüência de aminoácido de SEQ ID NO: 10, 12, 56, ou 59; e,

- e) uma seqüência de nucleotídeo que codifica um polipeptídeo que tem pelo menos 95% de identidade de seqüência de aminoácido com a seqüência de aminoácido de SEQ ID NO: 10, 12, ou 59.

3. Polinucleotídeo de acordo com a reivindicação 1, em que o dito polinucleotídeo está ligado a um promotor que funciona em células de planta para causar a produção de uma seqüência de RNA que permite expressão suficiente da enzima EPSPS codificada para aumentar a tolerância ao glifosato de uma célula de planta transformada com o polinucleotídeo.

4. Polinucleotídeo de acordo com a reivindicação 3 que ainda compreende uma região 3' não traduzida que funciona em células de planta para causar a adição de uma extensão de nucleotídeos poliadenilados à terminação 3' da seqüência de RNA.

5. Polinucleotídeo de acordo com a reivindicação 3, em que o dito promotor é heterólogo em relação ao polinucleotídeo.

6. Polinucleotídeo de acordo com a reivindicação 1, no qual o polinucleotídeo codifica um polipeptídeo de fusão que compreende um peptídeo de trânsito de cloroplasto amino-terminal e a enzima EPSPS.

7. Método de produzir plantas geneticamente transformadas que são tolerantes em relação ao herbicida glifosato, que compreende as etapas de:

a) inserir no genoma de uma célula de planta um polinucleotídeo que compreende:

- i) um promotor que funciona em células de planta para causar a produção de uma seqüência de RNA,
- ii) um polinucleotídeo diferente do polinucleotídeo listado na SEQ ID NOS: 7 ou 53 que causa a produção de uma seqüência de RNA que codifica uma enzima EPSPS que tem pelo menos um domínio de seqüência selecionado a partir do grupo que consiste em:

Domínio I:

L-A-K-G-X₁-S-X₂-L-X₃-G-A-L-K-S-D-D-T (SEQ ID NO: 13), onde X₁ denota lisina ou treonina, onde X₂ denota arginina ou histidina, onde X₃ denota serina ou treonina;

5 Domínio Ia:

L-A-K-G-X₁, (SEQ ID NO: 14), onde X₁ denota lisina ou treonina;

Domínio Ib:

S-X₁-L-X₂ (SEQ ID NO: 15), onde X₁ denota arginina ou histidina, onde X₂ denota serina ou treonina;

10 Domínio Ic:

G-A-L-K-S-D-D-T (SEQ ID NO: 16);

Domínio II:

E-P-D-X₁-X₂-T-F-X₃-V-X₄-X₅-X₆-G (SEQ ID NO: 17), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina, onde X₄ denota treonina ou ácido glutâmico, ou lisina, onde X₅ denota serina ou glicina, onde X₆ denota glutamina ou serina ou ácido glutâmico ou treonina;

Domínio IIa:

E-P-D-X₁-X₂-T-F-X₃-V (SEQ ID NO: 18), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina;

Domínio IIb:

X₁-X₂-X₃-G (SEQ ID NO: 19), onde X₁ denota treonina ou ácido glutâmico ou lisina, onde X₂ denota serina ou glicina, onde X₃ denota glutamina ou serina ou ácido glutâmico ou treonina;

Domínio III:

RFLTAA (SEQ ID NO: 20);

Domínio IV:

KRPI(G/M)P (SEQ ID NO: 21)
K-R-P-I-X₁-P, onde X₁ denota glicina ou metionina ou leucina;

Domínio V:

X₁-G-C-P-P-V (SEQ ID NO: 22), onde X₁ denota treonina ou se-

rina;

Domínio VI:

I-G-A-X₁-G-Y-X₂-D-L-T (SEQ ID NO: 23), onde X₁ denota arginina ou lisina ou leucina, e onde X₂ denota isoleucina ou valina;

5 Domínio VII:

W-X₁-V-X₂-X₃-T-G (SEQ ID NO: 24), onde X₁ denota arginina ou lisina, onde X₂ denota alanina ou histidina ou ácido glutâmico ou serina, onde X₃ denota prolina ou alanina;

Domínio VIII:

10 E-P-D-A-S-A-A-T- Y-L-W-X₁-A-X₂-X₃-L (SEQ ID NO: 25), onde X₁ denota alanina ou glicina, onde X₂ denota ácido glutâmico ou glutamina, onde X₃ denota valina ou leucina ou alanina;

Domínio VIIIa:

E-P-D-A-S-A-A-T- Y-L-W (SEQ ID NO: 26);

15 Domínio IX:

I-D-X₁-G (SEQ ID NO: 27), onde X₁ denota isoleucina ou leucina;

Domínio X:

F-X₁-Q-P-D-A- K-A (SEQ ID NO: 28), onde X₁ denota treonina ou serina;

20 Domínio XI:

X₁-F-P-X₂-X₃-X₄-A-X₅-X₆-X₇-G-S-Q-M-Q-D-A-I-P-T-X₈-A-V- X₉ A-A- F-N (SEQ ID NO: 29), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina, onde X₅ denota treonina ou ácido glutâmico ou valina, onde X₆ denota valina ou isoleucina, onde X₇ denota ácido aspártico ou valina, onde X₈ denota leucina ou isoleucina, onde X₉ denota leucina ou isoleucina;

Domínio XIa:

30 X₁-F-P-X₂-X₃-X₄-A (SEQ ID NO: 30), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina;

Domínio XIb:

X_1 - X_2 - X_3 -G-**S-Q**-M-**Q**-D-A-I-P-T- X_4 -A- V- X_5 A-A- F-N (SEQ ID NO: 31), onde X_1 denota treonina ou ácido glutâmico ou valina, onde X_2 denota valina ou isoleucina, onde X_3 denota ácido aspártico ou valina, onde X_4 denota leucina ou isoleucina, onde X_5 denota leucina ou isoleucina;

5 Domínio XIc:

G-S-Q-M-Q-D-A-I-P-T (SEQ ID NO: 32);

Domínio XII:

P -V-R-F- X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C**-D-R- X_5 (SEQ ID NO: 33), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

Domínio XIIa:

P-V-R-F (SEQ ID NO: 34);

Domínio XIIb:

15 X_1 - X_2 - X_3 - X_4 -N-L-R-V-K-E-**C**-D-R- X_5 (SEQ ID NO: 35), onde X_1 denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

Domínio XIIc:

20 N-L-R-V-K-E-**C**-D-R (SEQ ID NO: 36);

Domínio XIII:

E-G-D-D-L- X_1 - X_2 (SEQ ID NO: 37), onde X_1 denota leucina ou isoleucina, onde X_2 denota valina ou isoleucina;

Domínio XIV:

25 X_1 -P- X_2 -L-A-G (SEQ ID NO: 38), onde X_1 denota ácido aspártico ou asparagina, onde X_2 denota alanina ou serina ou treonina;

Domínio XV:

A- X_1 -I-D- X_2 - X_3 - X_4 - D-H-R (SEQ ID NO: 39), onde X_1 denota leucina ou serina ou ácido glutâmico, onde X_2 denota treonina ou serina, onde X_3 denota histidina ou fenilalanina, onde X_4 denota alanina ou serina;

Domínio XVI:

F-A-L-A- X_1 -L-K- X_2 - X_3 -G-I (SEQ ID NO: 40), onde X_1 denota gli-

cina ou alanina, onde X_2 denota isoleucina ou valina, onde X_3 denota serina ou glicina ou alanina ou lisina;

Domínio XVIa:

5 F-A-L-A- X_1 -L-K (SEQ ID NO: 41), onde X_1 denota glicina ou alanina;

Domínio XVIb:

L-K- X_1 - X_2 -G-I (SEQ ID NO: 42), onde X_1 denota isoleucina ou valina, onde X_2 denota serina ou glicina ou alanina ou lisina;

Domínio XVII:

10 - X_1 -P- X_2 -C-V- X_3 -K (SEQ ID NO: 43), onde X_1 denota asparagina ou ácido aspártico, onde X_2 denota alanina ou ácido aspártico, onde X_3 denota alanina ou glicina;

Domínio XVIII:

15 X_1 -S-L-G-V (SEQ ID NO: 44), onde X_1 denota alanina ou serina ou prolina.

iii) um polinucleotídeo 3' não traduzido que funciona em células de plantas para causar a adição de uma extensão de nucleotídeos poliadenilados no final de terminação 3' da sequência de RNA;

20 b) obter uma célula de planta transformada; e
c) regenerar a partir da célula de planta transformada uma planta geneticamente transformada que tem tolerância ao herbicida glifosato aumentada.

8. Método de acordo com a reivindicação 7, em que obter a planta transformada compreende rastrear por uma tolerância aumentada ao herbicida glifosato.

9. Método de acordo com a reivindicação 7, no qual o polinucleotídeo codifica um polipeptídeo de fusão que compreende um peptídeo de trânsito amino-terminal e a enzima EPSPS.

30 10. Célula de planta tolerante ao glifosato que compreende um polinucleotídeo heterólogo como definido na reivindicação 1.

11. Célula de planta tolerante ao glifosato de acordo com a reivin-

dicação 10, selecionada a partir do grupo que consiste em milho, trigo, arroz, cevada, soja, algodão, beterraba sacarina, colza, canola, linho, girassol, batata, tabaco, tomate, alfafa, choupo, pinho, eucalipto, maçã, alface, ervilhas, lentilhas, uva e gramíneas.

5 12. Planta tolerante ao glifosato que compreende células de planta como definidas na reivindicação 10.

13. Semente transformada da planta como definida na reivindicação 12.

10 14. Planta tolerante ao glifosato de acordo com a reivindicação 12, selecionada a partir do grupo que consiste em milho, trigo, arroz, cevada, soja, algodão, beterraba sacarina, colza, canola, linho, girassol, batata, tabaco, tomate, alfafa, choupo, pinho, eucalipto, maçã, alface, ervilhas, lentilhas, uva e gramíneas.

15 15. Polipeptídeo isolado selecionado a partir do grupo que consiste em:

- a) um polipeptídeo que compreende a seqüência de aminoácido de SEQ ID NO: 10, 12, 56, ou 59;
- b) um polipeptídeo codificado pela seqüência de nucleotídeo de SEQ ID NO: 9, 11, 55, 57, 58, ou 60;
- 20 c) um polipeptídeo que compreende uma seqüência de aminoácido que tem pelo menos 95% de identidade de seqüência com a seqüência de aminoácido de SEQ ID NO: 10, 12, ou 59, em que o dito polipeptídeo tem atividade de resistência a herbicida;
- 25 d) um polipeptídeo que é codificado por uma seqüência de nucleotídeo que é pelo menos 95% idêntica a seqüência de nucleotídeo de SEQ ID NO: 9, 11, 57, 58, ou 60, em que o dito polipeptídeo tem atividade de resistência a herbicida; e,
- 30 e) um polipeptídeo que é codificado pela seqüência de nucleotídeo de resistência a herbicida do inserto de DNA do plasmídeo depositado como N^{os} de Acesso B-30833 ou B-

30838.

16. Polipeptídeo de acordo com a reivindicação 15, compreendendo ainda uma seqüência de aminoácido heteróloga.

17. Método para controlar seletivamente ervas daninhas em um campo que tem sementes de plantas ou plantas que compreende as etapas de:

- a) plantar as sementes ou plantas que são tolerantes ao glifosato, as ditas plantas ou sementes tendo integrado nos seus genomas um polinucleotídeo que compreende um polinucleotídeo heterólogo que tem::
 - i) um promotor que funciona em células de planta para causar a produção e uma seqüência de RNA;
 - ii) um polinucleotídeo diferente do polinucleotídeo listado em SEQ ID NO: 7 ou 53 que causa a produção de uma seqüência de RNA que codifica uma enzima EPSPS que tem pelo menos um domínio de seqüência selecionado a partir do grupo que consiste em:

Domínio I:

L-A-K-G-X₁-S-X₂-L-X₃-G-A-L-K-S-D-D-T (SEQ ID NO: 13), onde X₁ denota lisina ou treonina, onde X₂ denota arginina ou histidina, onde X₃ denota serina ou treonina;

Domínio Ia:

L-A-K-G-X₁, (SEQ ID NO: 14), onde X₁ denota lisina ou treonina;

Domínio Ib:

S-X₁-L-X₂ (SEQ ID NO: 15), onde X₁ denota arginina ou histidina, onde X₂ denota serina ou treonina;

Domínio Ic:

G-A-L-K-S-D-D-T (SEQ ID NO: 16);

Domínio II:

E-P-D-X₁-X₂-T-F-X₃-V-X₄-X₅-X₆-G (SEQ ID NO: 17), onde X₁ denota ácido aspártico ou alanina, onde X₂ denota serina ou treonina, onde X₃ denota valina ou isoleucina, onde X₄ denota treonina ou ácido glutâmico, ou

lisina, onde X_5 denota serina ou glicina, onde X_6 denota glutamina ou serina ou ácido glutâmico ou treonina;

Domínio IIa:

5 E-P-D- X_1 - X_2 -T-F- X_3 -V (SEQ ID NO: 18), onde X_1 denota ácido aspártico ou alanina, onde X_2 denota serina ou treonina, onde X_3 denota valina ou isoleucina;

Domínio IIb:

10 X_1 - X_2 - X_3 -G (SEQ ID NO: 19), onde X_1 denota treonina ou ácido glutâmico ou lisina, onde X_2 denota serina ou glicina, onde X_3 denota glutamina ou serina ou ácido glutâmico ou treonina;

Domínio III:

RFLTAA (SEQ ID NO: 20);

Domínio IV:

15 KRPI(G/M)P (SEQ ID NO: 21)
K-R-P-I- X_1 -P, onde X_1 denota glicina ou metionina ou leucina;

Domínio V:

X_1 -G-C-P-P-V (SEQ ID NO: 22), onde X_1 denota treonina ou serina;

Domínio VI:

20 I-G-A- X_1 -G-Y- X_2 -D-L-T (SEQ ID NO: 23), onde X_1 denota arginina ou lisina ou leucina, e onde X_2 denota isoleucina ou valina;

Domínio VII:

25 W- X_1 -V- X_2 - X_3 -T-G (SEQ ID NO: 24), onde X_1 denota arginina ou lisina, onde X_2 denota alanina ou histidina ou ácido glutâmico ou serina, onde X_3 denota prolina ou alanina;

Domínio VIII:

E-P-D-A-S-A-A-T- Y-L-W- X_1 -A- X_2 - X_3 -L (SEQ ID NO: 25), onde X_1 denota alanina ou glicina, onde X_2 denota ácido glutâmico ou glutamina, onde X_3 denota valina ou leucina ou alanina;

30 Domínio VIIIa:

E-P-D-A-S-A-A-T- Y-L-W (SEQ ID NO: 26);

Domínio IX:

I-D-X₁-G (SEQ ID NO: 27), onde X₁ denota isoleucina ou leucina;

Domínio X:

F-X₁-Q-P-D-A- K-A (SEQ ID NO: 28), onde X₁ denota treonina ou serina;

5 Domínio XI:

X₁-F-P-X₂-X₃-X₄-A-X₅-X₆-X₇-G-S-Q-M-Q-D-A-I-P-T-X₈-A-V- X₉ A-A- F-N (SEQ ID NO: 29), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina, onde X₅ denota treonina ou ácido glutâmico ou valina, onde X₆ denota valina ou isoleucina, onde X₇ denota ácido aspártico ou valina, onde X₈ denota leucina ou isoleucina, onde X₉ denota leucina ou isoleucina;

Domínio XIa:

15 X₁-F-P-X₂-X₃-X₄-A (SEQ ID NO: 30), onde X₁ denota glutamina ou lisina ou serina, onde X₂ denota asparagina ou histidina, onde X₃ denota metionina ou leucina, onde X₄ denota prolina ou glutamina;

Domínio XIb:

20 X₁-X₂-X₃-G-S-Q-M-Q-D-A-I-P-T-X₄-A- V- X₅ A-A- F-N (SEQ ID NO: 31), onde X₁ denota treonina ou ácido glutâmico ou valina, onde X₂ denota valina ou isoleucina, onde X₃ denota ácido aspártico ou valina, onde X₄ denota leucina ou isoleucina, onde X₅ denota leucina ou isoleucina;

Domínio XIc:

G-S-Q-M-Q-D-A-I-P-T (SEQ ID NO: 32);

Domínio XII:

25 P-V-R-F-X₁-X₂-X₃-X₄-N-L-R-V-K-E-C-D-R-X₅ (SEQ ID NO: 33), onde X₁ denota valina ou treonina, onde X₂ denota ácido glutâmico ou glicina, onde X₃ denota leucina ou isoleucina, onde X₄ denota alanina ou ácido glutâmico, onde X₅ denota isoleucina ou valina;

Domínio XIIa:

30 P-V-R-F (SEQ ID NO: 34);

Domínio XIIb:

X₁-X₂-X₃-X₄-N-L-R-V-K-E-C-D-R-X₅ (SEQ ID NO: 35), onde X₁

denota valina ou treonina, onde X_2 denota ácido glutâmico ou glicina, onde X_3 denota leucina ou isoleucina, onde X_4 denota alanina ou ácido glutâmico, onde X_5 denota isoleucina ou valina;

Domínio XIIc:

5 N-L-R-V-K-E-C-D-R (SEQ ID NO: 36);

Domínio XIII:

E-G-D-D-L- X_1 - X_2 (SEQ ID NO: 37), onde X_1 denota leucina ou isoleucina, onde X_2 denota valina ou isoleucina;

Domínio XIV:

10 X_1 -P- X_2 -L-A-G (SEQ ID NO: 38), onde X_1 denota ácido aspártico ou asparagina, onde X_2 denota alanina ou serina ou treonina;

Domínio XV:

A- X_1 -I-D- X_2 - X_3 - X_4 -D-H-R (SEQ ID NO: 39), onde X_1 denota leucina ou serina ou ácido glutâmico, onde X_2 denota treonina ou serina, onde
15 X_3 denota histidina ou fenilalanina, onde X_4 denota alanina ou serina;

Domínio XVI:

F-A-L-A- X_1 -L-K- X_2 - X_3 -G-I (SEQ ID NO: 40), onde X_1 denota glicina ou alanina, onde X_2 denota isoleucina ou valina, onde X_3 denota serina ou glicina ou alanina ou lisina;

20 Domínio XVIa:

F-A-L-A- X_1 -L-K (SEQ ID NO: 41), onde X_1 denota glicina ou alanina;

Domínio XVIb:

L-K- X_1 - X_2 -G-I (SEQ ID NO: 42), onde X_1 denota isoleucina ou
25 valina, onde X_2 denota serina ou glicina ou alanina ou lisina;

Domínio XVII:

- X_1 -P- X_2 -C-V- X_3 -K (SEQ ID NO: 43), onde X_1 denota asparagina ou ácido aspártico, onde X_2 denota alanina ou ácido aspártico, onde X_3 denota alanina ou glicina;

30 Domínio XVIII:

X_1 -S-L-G-V (SEQ ID NO: 44), onde X_1 denota alanina ou serina ou prolina.

iii) um polinucleotídeo 3' não traduzido que funciona em células de planta para causar a adição de uma extensão de nucleotídeos poliadenilados à terminação 3' da sequência de RNA; e

5 b) aplicar às plantas e sementes no campo uma quantidade suficiente de herbicida glifosato para controlar as sementes sem afetar significativamente as plantas.

18. Método de acordo com a reivindicação 17, em que a dita planta ou sementes de planta compreendem uma planta de cultivo.

10 19. Método de acordo com a reivindicação 17, em que o dito promotor é heterólogo com relação ao polinucleotídeo e adaptado para causar expressão suficiente da enzima EPSPS para aumentar a tolerância ao glifosato da planta transformada com o polinucleotídeo.

15 20. Método de produzir uma planta que é tolerante em relação ao herbicida glifosato, que compreende as etapas de:

 a) inserir no genoma de uma célula de planta o polinucleotídeo como definido na reivindicação 1;
 b) obter células de planta transformadas; e
 c) regenerar a partir da célula de planta transformada uma
20 planta geneticamente transformada que tem tolerância em relação ao herbicida glifosato aumentada.

21. Método de produzir plantas geneticamente transformadas que são tolerantes ao glifosato, que compreende transformar células de planta com um construto de DNA que compreende um polinucleotídeo selecionado a partir do grupo que consiste em SEQ ID NOS: 1, 3, 9, 11, 45, 47,
25 55, 57, 58 ou 60, e regenerar uma planta transformada.

22. Método de acordo com a reivindicação 21, em que o dito construto de DNA ainda compreende um promotor que funciona em células de planta para causar a produção de uma sequência de RNA que permite a
30 expressão suficiente da enzima EPSPS codificada para aumentar a tolerância ao glifosato de uma célula de planta transformada com o polinucleotídeo.

23. Método de acordo com a reivindicação 21, no qual o polinu-

cleotídeo codifica um polipeptídeo de fusão que compreende um peptídeo de trânsito de cloroplasto amino-terminal e a enzima EPSPS.

24. Célula de planta tolerante ao glifosato que compreende um polinucleotídeo selecionado a partir do grupo que consiste em SEQ ID NOS:
5 1, 3, 9, 11, 45, 47, 55, 57, 58, e 60.

25. Célula de planta tolerante ao glifosato de acordo com a reivindicação 24, selecionada a partir do grupo que consiste em milho, trigo, arroz, cevada, soja, algodão, beterraba sacarina, colza, canola, linho, girassol, batata, tabaco, tomate, alfafa, choupó, pinho, eucalipto, maçã, alface,
10 ervilhas, lentilhas, uva e gramíneas.

26. Planta tolerante ao glifosato que compreende células de planta como definidas na reivindicação 24.

27. Planta tolerante ao glifosato de acordo com a reivindicação 26, selecionada a partir do grupo que consiste em milho, trigo, arroz, cevada,
15 soja, algodão, beterraba sacarina, colza, canola, linho, girassol, batata, tabaco, tomate, alfafa, choupó, pinho, eucalipto, maçã, alface, ervilhas, lentilhas, uva e gramíneas.

28. Semente transformada da planta como definida na reivindicação 26.

20 29. Método para controlar seletivamente ervas daninhas em um campo que tem sementes plantadas ou plantas que compreende as etapas de:

- a) plantar as plantas como definidas na reivindicação 26, ou a semente transformada derivada delas, e
- 25 b) aplicar às plantas e sementes no campo uma quantidade suficiente de herbicida glifosato para controlar as ervas daninhas sem afetar significativamente as plantas.

30. Método de acordo com a reivindicação 29, em que a dita planta ou sementes de planta compreende uma planta de cultivo.

30 31. Método de acordo com a reivindicação 29, no qual o polinucleotídeo codifica um polipeptídeo de fusão que compreende um peptídeo de trânsito de cloroplasto amino-terminal e a enzima EPSPS.

GRG8	-MMMGRAKLTIIPPGKPLTGRAMPPGSKSITNRALLLAGLAKGTSRLTGALKSDDTRYMA	59
GRG9	-----MIELTITPPGHPLSGKVEPPGSKSITNRALLLAGLAKGKSHLSGALKSDDTLYMA	55
GRG15	-----MIELTITPPGHPLSGKVEPPGSKSITNRALLLAGLAKGKSRLTGALKSDDTLYMA	55
GRG12	-----MTVTTPNFPLTGKVAAPPGSKSITNRALLLAALAKGTSRLSGALKSDDTRHMS	52
GRG6	---MRPQATLTVMVERPLVGRVSPPGSKSITNRALLLAGLAKGTSRLTGALKSDDTRVMS	58
GRG1	-----MKVTIQPGDLTGILQSPASKSSMQRACAAALVAKGISEIINPGHSNDKKAAR	52
E_coli	-----MESLTLOPIAR-VDGTINLPGSKTVSNRALLLAALAHGKTVLTNLLSDDVRHML	54
Agro_CP4	MSHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG	60
B_Subtilis	-----MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAAGTTTVKNFLPGADQLSTI	51

II

III

GRG8	EALRAMGVTIIDPDDTTTFIVKGS--GKLQPP-AAPLFLGNAGTATRFLTAALVDGKVI	116
GRG9	EALREMGVKVTEPDATTFVVEGT--GVLQPP-EKPLFLGNAGTATRFLTAALVDGAVI	112
GRG15	EALREMGVKVTEPDATTFVVESS--GGLHQP-EKPLFLGNAGTATRFLTAALVDGAVI	112
GRG12	VALRQMGVTIDEFDDTTTFVVTSS--GSLQLP-AQPLFLGNAGTAMRFLTAALVATVQGTIV	109
GRG6	EALRLMGVQVDEPDSTTFVVTSS--GHWQAP-QQALFLGNAGTATRFLTAALANFEGDFV	115
GRG1	DIVSRLGARLEDQPDGSLQITSEG---VKPVAFIDCGESGLSIRMFTPIVALSKEEVT	108
E_coli	NALTALGVSYTISADRTTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAALCLGSNDIV	114
Agro_CP4	KAMQAMGARIRKEGDTWIIDGVGN-GGLLAP-EAPLDFGNAATGCRMTMGLVGVYDFDST	118
B_Subtilis	DCFRKMGVHIECSSSDVVIHGKG-IDALKEP-ESLLDVGNSGTTIRMLGLIAGRPFFYSA	109

IV

V

GRG8	VDGDAHMKRPFGHLVDALRSLGIDASAEITGCPPVTINGTGR-FEASRVQIDGGLSSQYV	175
GRG9	IDGDEHMKRPFILELVQALRALGVEADAETGCPPVTVRGKGMGFPKGSVTIDANLSSQYV	172
GRG15	IDGDEHMKRPIMELVEALRSLGVEAEAEITGCPPVTVCCKGTGFPKGSVTIDANLSSQYV	172
GRG12	LDGDEYMKRPFGHLATLGONGIQVDSEITGCPPVTIVHGAGK-VQARRFEIDGGLSSQYV	168
GRG6	VDGDEYMKRPFGHLVDALQRMGVEISAPSGCPPVAIKGKGG-LQAGRIEIDGNLSSQYV	174
GRG1	IKSGSLVTRPMDEFDEILPHIGVKVKSNOG--KLPLVIQGP-LKPADVTVDGSLSSQFL	165
E_coli	LTGEPRMKERPIGHLVDALRLGGAKITYLEQENYFPLRLQGG-FTGGNVVDVGSVSSQFL	173
Agro_CP4	FIGDASLTKRPMGRVLNPLREMGVQVKSEOG-DRIPVTLRGP-KTPTPTITVRVPMASQV	176
B_Subtilis	VAGDESIKRPMKRVTETPLKKMGAKIDGRAGGEFTPLSVSGA--SLKGIDYVSPVASAQI	167

VI

VII

GRG8	SALLMMAAGGDRAVDVELLGEITGALGYIDLIVAAMRAFGAKVERVS---PVAWRVEPTG	232
GRG9	SALLMAAACGDKPVDIILKGEEIGAKGYIDLITSAMEAFGAKVERVS---NAIWRVHPTG	229
GRG15	SALLMAAACGDKPVDIILKGEEIGAKGYIDLITSAMEAFGAKVERVS---NAIWRVHPTG	229
GRG12	SALLMLAACGEAPIEVALTGKDIGARGYVDLTLD CMRAFGAQVDIVD---DTTWRVAPTG	225
GRG6	SALLMAGACGKGSLEVALTGSEIGARGYVDLTLAAMQAFGAEVQAIG---DAAWKVSATG	231
GRG1	TGLLLAYAAAD-ASDVAIKVTNLKSRPYIDLTLDMKRFGLKTPENRNYEEFYFKAGNVY	224
E_coli	TALLMTAPLAPEDTVIRIKG-DLVSKPYIDITLNLMTFGVEIENQH---YQCFVVKGGQ	229
Agro_CP4	KSAVLLAGLNTPGITTVEIP-IMTRDHTKMLQGFGANLTVETDADG---VRTIRLEGRG	232
B_Subtilis	KSAVLLAGLQAEGTTTVTEP---HKSRDHTERMLSAFGVKLSEDO---TSVSIAGGQK	219

Fig. 1A

	VIII	IX	X	XI	
GRG8	--YHAADFVIEPDASAATYLWAAEVL	SGGKIDIGTPAEQFSQPLAKAYDLISKEPH---			286
GRG9	--YTATDFHIEPDASAATYLWGAELITGG	AIDIGTPADKFTQPD	AKAYEVMAQEPH---		283
GRG15	--YTATDFHIEPDASAATYLWGAELITGG	AIDIGTPADKFTQPD	AKAHEVMAQEPH---		283
GRG12	--YTAHDYLIIEPDASAATYLWAAEVL	TGGRIDIGVAAQDFTQPD	AKAQAVIAQEPN---		279
GRG6	--YHATDFHIEPDASAATYLWAAQALTEG	NIDIGVASDAFTQPD	ALASQIIDSEPN---		275
GRG1	DETKMQRYTVEGDWSSGAFLLVAGAIAGP-	ITVRGLDIAS	TQADKAI	VOALMSANAGIAI	283
E_coli	SYQSPGTYLVEGDASSASYFLAAAALKGG	TVKVTGIGRNSMQGD	IRFADVLEKMG	ATICW	289
Agro_CP4	K-LTGQVIDVPGDPSSTAFPLVAALIVPG-	SDVTILNVIMNP	TRTGLILTLQEM	GADIEV	290
B_Subtilis	--LTAADIFVPGDISSAAFFLAAGAMVPN-	SRIVLKNVGLNP	TRTGII	DVLQNMGA	LEI 276

	XI (continuação)	XII	
GRG8	-----LPAVIDGSQMQDAIPTLAVLA	AFNEMPVRFVGIENI	322
GRG9	-----LPAEIDGSQMQDAIPTIAVIA	AFNETPVRFVGIANI	319
GRG15	-----LPAEIDGSQMQDAIPTIAVLA	AFNETPVRFVGIANI	319
GRG12	-----MQATVVGSQMQDAIPTLAVLA	AFNTPVRFTELANI	315
GRG6	-----MPAVIDGSQMQDAIPTLAVLA	AFNRQPVRFVGIANI	311
GRG1	DAKEIKLHP-----ADLNAFEF	DATDCPDLFPPLVALASYCKGETKIKGVSRI	331
E_coli	GDDYIS-----CTRGELNAID	MDMNHIPDAAMTIATAALFAKGTTRLRNIYNW	337
Agro_CP4	INPRLAGGEDVADLRVRSSTLKGVTVP	EDRAPSMIDEYPILAVAAAFACATVMNGLEEL	350
B_Subtilis	KPSADSGAEPYGDLIETSSLKAVEIGGDI	IPRLIDEIPIALLATQAEETT	VIKDAEI 336

	XII (cont.)	XIII	XIV	XV	XVI		
GRG8	RVKECDRI	RALSSGLSRIVPNLGT	EEGDDLI	IASDPSL-AGKIL	TAEDSFADHRIAMSE	381	
GRG9	RVKECDRI	RAVSLGLNEIREGLAHEEGDDLI	VHADPSL-AGQTV	DA	SIDTFADHRIAMSE	378	
GRG15	RVKECDRI	RAVSLGLNEIRDGLAHEEGDDLI	VHSDPSL-AGQTV	NA	SIDTFADHRIAMSE	378	
GRG12	RVKECDRI	QALHDGLNEIRPGLATIEGDDLI	VASOPAL-AGTACT	ALIDTHADHRIAMCE		374	
GRG6	RVKECDRI	SALCDGLCAIAPGLAVEEGDDLI	VHANPAL-AGTTV	NALIDTHSDHRIAMCE		370	
GRG1	AHKESDRGL	TLQDEFGKMGVEIHLEGDLMRV	IGGKG----	VKGA	EVSSRHDHRIAMAC	385	
E_coli	RVKETDRI	FAMATEL	RKVG--AEVEEGHDY	IRITPP----	EKLNF	AEIATYNDHRMAMCE	391
Agro_CP4	RVKESDRI	SAVANGKLNGVDCDEGETSLV	VRGRPDGKGLGNASGA	AAVATHL	DHRIAMSE	410	
B_Subtilis	KVKETNR	IDTVVSEL	RKLGAETIEPTADGMKVY	GKQTLK----	GGA	VSSHGDHRIAMML	391

	XVI (cont.)	XVII	XVIII	
GRG8	ALAGLKIGG-	ITILDPDCVAKTFPSYWNVLSS	LGVAYED-----	419
GRG9	ALAALKIGG-	IAIONPACVAKTYPGYWKALASL	GVDYTEKESAAEPQH	425
GRG15	ALAALKIGG-	IAIONPACVGKTYPGYWKALASL	GVEYSEKETAAEPQH	425
GRG12	ALAGLKVSG-	IRIQDPDCVAKTYPDYWKALASL	GVHLSY-----	412
GRG6	ALAGLKIKG-	IHIQDPDCVAKTYPGYWDALASL	GVSVQR-----	418
GRG1	AVAALKAVGETT	IEHAEAVNKSYPDFYS	DLKQLGGVVS	LNHQFNFS-- 431
E_coli	SLVALSDTP-	VTILDPKCTAKTFPDYFEQ	LARISQAA-----	427
Agro_CP4	LVMGLVSEN	PVTVDATMIATSFPEFMDL	MAGLGA	KIELSDTKAA--- 455
B_Subtilis	GIASCITEEPI	IEIHTDAIHVSYP	TTFFHLNKL	SKKS----- 428

Fig. 1B

RESUMO

Patente de Invenção: "IDENTIFICAÇÃO DE UMA NOVA CLASSE DE SINTASES EPSP".

A presente invenção refere-se a composições e métodos para
 5 conferir tolerância ao glifosato em bactérias, plantas, células, tecidos e se-
 mentes de planta. Composições incluem uma nova classe de enzimas
 EPSPS, designadas Classe III, e polinucleotídeos que codificam tais enzi-
 mas, vetores que compreendem esses polinucleotídeos, e células hospedei-
 ras que compreendem os vetores. As novas proteínas compreendem pelo
 10 menos um domínio de seqüência selecionado a partir dos domínios Classe
 III aqui fornecidos. Esses domínios de seqüência podem ser usados para
 identificar EPSP sintases com atividade de resistência ao glifosato.