

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

(21) **PI 0713039-2 A2**

(22) Data de Depósito: 27/06/2007
(43) Data da Publicação: 17/07/2012
(RPI 2167)



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

(54) Título: MOLÉCULA DE ÁCIDO NUCLEÍCO E POLIPEPTÍDEO ISOLADOS, VETOR, CÉLULA HOSPEDEIRA, PLANTA E SEMENTE TRANSGÊNICAS E MÉTODOS PARA PRODUZIR UM POLIPEPTÍDEO COM ATIVIDADE DE RESISTÊNCIA A HERBICIDAS, PARA CONFERIR RESISTÊNCIA A UM HERBICIDA EM UMA PLANTA E PARA GERAR UM POLINUCLEOTÍDEO VARIANTE DE UM POLINUCLEOTÍDEO PARENTAL GRG33, GRG35, GRG37, GRG38, GRG39, OU GRGR50.

(30) Prioridade Unionista: 27/06/2006 US 60/816,676, 07/07/2006 US 60/819,119, 07/07/2006 US 60/819,122, 27/06/2006 US 60/816,676

(73) Titular(es): Athenix Corporation

(72) Inventor(es): Brian Carr, Brian Vande Berg, Cheryl Peters, Daniel John Tomso

(74) Procurador(es): Claudia Christina Schulz

(86) Pedido Internacional: PCT US2007072206 de 27/06/2007

(87) Publicação Internacional: WO 2008/002964 de 03/01/2008

(57) Resumo: MOLÉCULA DE ÁCIDO NUCLEÍCO E POLIPEPTÍDEO ISOLADOS, VETOR, CÉLULA HOSPEDEIRA, PLANTA E SEMENTE TRANSGÊNICAS E MÉTODOS PARA PRODUZIR UM POLIPEPTÍDEO COM ATIVIDADE DE RESISTÊNCIA A HERBICIDAS, PARA CONFERIR RESISTÊNCIA A UM HERBICIDA EM UMA PLANTA E PARA GERAR UM POLINUCLEOTÍDEO VARIANTE DE UM POLINUCLEOTÍDEO PARENTAL GRG33, GRG35, GRG36, GRG37, GRG38, GRG39, OU GRG50. São providos composições e métodos para conferir resistência a herbicidas para bactérias, plantas, células, tecidos e sementes vegetais. Composições incluem moléculas de ácidos nucleicos codificando polipeptídeos de resistência ou tolerância a herbicidas, vetores compreendendo tais moléculas de ácidos nucleicos e células hospedeiras compreendendo os vetores. As seqüências de nucleotídeos da invenção podem ser usadas em construções de DNA ou cassetes de expressão para transformação e expressão em organismos, incluindo microrganismos e plantas. Composições também compreendem bactérias, plantas, células tecidos e semente vegetais transformados. Em particular, a presente invenção provê moléculas de ácidos nucleicos isoladas compreendendo a seqüência de nucleotídeos apresentada em SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, uma seqüência de nucleotídeos codificando a seqüência de aminoácidos mostrada em SEQ ID NO: 2, 5, 8, 11, 14, 16, 19, ou 22, a seqüência de nucleotídeos de resistência a herbicidas depositada em um hospedeiro bacteriano como Nos. de Acesso NRRL B-30932, B-30933, B-30934, B-30945, B-30946, B-30947, ou B-30948, assim como variantes e fragmentos dessas.

MOLÉCULA DE ÁCIDO NUCLEÍCO E POLIPEPTÍDEO ISOLADOS, VETOR, CÉLULA HOSPEDEIRA, PLANTA E SEMENTE TRANSGÊNICAS E MÉTODOS PARA PRODUIR UM POLIPEPTÍDEO COM ATIVIDADE DE RESISTÊNCIA A HERBICIDAS, PARA CONFERIR RESISTÊNCIA A UM HERBICIDA EM
5 UMA PLANTA E PARA GERAR UM POLINUCLEOTÍDEO VARIANTE DE UM POLINUCLEOTÍDEO PARENTAL *GRG33*, *GRG35*, *GRG36*, *GRG37*, *GRG38*, *GRG39*, OU *GRG50*

CAMPO DA INVENÇÃO

10

Esta invenção provê novos genes codificando 5-enolpiruvilchiquimato-3-fosfato (EPSP) sintase que provêm resistência a herbicidas. Estes genes são úteis em biologia vegetal, melhoramento de cultivos, e cultura de células
15 vegetais.

FUNDAMENTOS DA INVENÇÃO

N-fosfonometilglicina, comumente referida como
20 glifosato, é um importante químico agrícola. Glifosato inibe a enzima que converte ácido fosfoenolpirúvico (PEP) e ácido 3-fosfochiquímico para ácido 5-enolpiruvil-3-fosfochiquímico. A inibição desta enzima (5-enolpiruvilchiquimato-3-fosfato sintase; aqui referida como
25 "EPSP sintase") mata células vegetais desligando a via de chiquimato, inibindo, dessa forma, a biossíntese de ácidos aromáticos.

Como os herbicidas de classe do glifosato inibem a biossíntese de aminoácidos aromáticos, eles podem não apenas matar células vegetais, mas são também tóxicos a células bacterianas. Glifosato inibe várias EPSP sintases bacterianas, e, assim, é tóxico para essas bactérias. Contudo, certas EPSP sintases bacterianas possuem uma alta tolerância a glifosato.

Células vegetais resistentes a toxicidade de glifosato podem ser produzidas transformando células vegetais para expressar sintases EPSP bacterianas resistentes a glifosato. Notavelmente, o gene bacteriano de *Agrobacterium tumefaciens* linhagem CP4 foi usado para conferir resistência a herbicidas em células vegetais em seguida da expressão em plantas. Uma EPSP sintase de *Salmonella typhimurium* linhagem CT7 confere resistência a glifosato em células bacterianas, e confere resistência a glifosato em células vegetais (Patentes Norte-Americanas Nos. 4.535.060; 4.769.061; e 5.094.945).

20

A Patente Norte-Americana 6.040.497 relata enzimas EPSP sintases de milho mutante tendo substituições de treonina para isoleucina na posição 102 e prolina para serina na posição 106 (a mutação "TIPS"). Tais alterações conferem resistência a glifosato na enzima de milho. Uma EPSP sintase mutada de *Salmonella typhimurium* linhagem CT7 confere resistência a glifosato em células bacterianas, e é relatado por conferir resistência a glifosato em células vegetais (Patentes Norte-Americanas Nos. 4.535.060;

4.769.061; e 5.094.945). He et al. ((2001) *Biochim et Biophysica Acta* 1568: 1-6) desenvolveram EPSP sintases com tolerância a glifosato aumentada por mutagênese e recombinação entre os genes de EPSP sintase de *E. coli* e
5 *Salmonella typhimurium*, e sugerem que mutações na posição 42 (T42M) e posição 230 (Q230K) são prováveis de serem responsáveis pela resistência observada.

Um trabalho subsequente (He et al. (2003) *Biosci.*
10 *Biotech. Biochem.* 67: 1405-1409) mostra que a mutação T42M (treonina para metionina) é suficiente para melhorar a tolerância de ambas enzimas de *E. coli* e *Salmonella typhimurium*. Essas enzimas contêm substituições de aminoácidos em seus sítios ativos que previnem a ligação de
15 glifosato sem afetar a ligação por PEP ou S3P. Mutações que ocorrem na região de junta entre os dois domínios globulares de EPSP sintase foram mostradas por alterar a afinidade de ligação de glifosato, mas não PEP (He et al. 2003, *supra*). Portanto, tais enzimas possuem atividade
20 catalítica elevada, mesmo na presença de glifosato.

Devido às várias vantagens que plantas com resistência a herbicidas fornecem, métodos para identificar genes de resistência a herbicidas com atividade de resistência a
25 glifosato são desejáveis.

SUMÁRIO DA INVENÇÃO

São providos composições e métodos para conferir resistência ou tolerância a herbicidas para bactérias, plantas, células, tecidos e sementes vegetais. Composições incluem moléculas de ácidos nucléicos codificando polipeptídeos de resistência ou tolerância a herbicidas, vetores compreendendo aquelas moléculas de ácidos nucléicos, e células hospedeiras compreendendo os vetores. Composições também incluem anticorpos para os polipeptídeos de resistência ou tolerância a herbicidas. Como notado, as seqüências de nucleotídeos da invenção podem ser usadas em construções de DNA ou cassetes de expressão para transformação e expressão em organismos, incluindo microorganismos e plantas. Composições também compreendem bactérias, plantas, células, tecidos, e sementes vegetais transformadas. Além disso, são providos métodos para produzir os polipeptídeos codificados pelos nucleotídeos sintéticos da invenção.

Em particular, moléculas de ácidos nucléicos isoladas e variantes dessas codificando polipeptídeos de resistência ou tolerância a herbicidas são providas. Adicionalmente, seqüências de aminoácidos e variantes dessas codificadas pelos polinucleotídeos que conferem resistência ou tolerância a herbicidas são englobadas. Em particular, a presente invenção provê moléculas de ácidos nucléicos isoladas compreendendo a seqüência de nucleotídeos apresentada em SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, uma seqüência de nucleotídeos codificando a seqüência de aminoácidos mostrada em SEQ ID

NO: 2, 5, 8, 11, 14, 16, 19, ou 22, a sequência de nucleotídeos de resistência a herbicidas depositada em um hospedeiro bacteriano como Nos. de Acesso NRRL B-30932, B-30933, B-30934, B-30945, B-30946, B-30947, ou B-30948, assim como variantes e fragmentos dessas. Sequências de nucleotídeos que são complementares a uma sequência de nucleotídeos da invenção, ou que hibridizam a uma sequência da invenção ou a um complemento de uma sequência da invenção são também englobadas.

10

DESCRIÇÃO DAS FIGURAS

A figura 1 mostra um alinhamento da sequência de aminoácidos de GRG33 (SEQ ID NO: 2) e GRG35 (SEQ ID NO:5) com sequências de EPSP sintase de *Streptomyces coelicolor* (SEQ ID NO: 24), *Streptomyces avermitilis* (SEQ ID NO: 25), *Zea mays* (SEQ ID NO: 38), e *E. coli* (SEQ ID NO: 37). O alinhamento mostra os resíduos de aminoácidos mais altamente conservados destacados em preto e os resíduos de aminoácidos altamente conservados destacados em cinza.

A figura 2 mostra um alinhamento da sequência de aminoácidos de GRG36 (SEQ ID NO:8) com sequências de EPSP sintase de *Bacillus halodurans* (SEQ ID NO: 26), *Bacillus clausii* (SEQ ID NO: 27), *Zea mays* (SEQ ID NO: 38), e *E. coli* (SEQ ID NO:37). O alinhamento mostra os resíduos de aminoácidos mais altamente conservados destacados em preto e os resíduos de aminoácidos altamente conservados destacados em cinza.

A figura 3 mostra um alinhamento de GRG38 (SEQ ID NO: 16) e GRG50 (SEQ ID NO: 22) com outras enzimas EPSP sintases, incluindo GRG8 (SEQ ID NO: 29), GRG12 (SEQ ID NO:30), GRG15 (SEQ ID NO:31), GRG5 (SEQ ID NO:32), GRG6 (SEQ ID NO: 33), GRG7 (SEQ ID NO: 34), GRG9 (SEQ ID NO: 35), GRG1 (SEQ ID NO:41), a EPSP sintase descrita em Pedido de Patente Internacional No. WO2005014820 (SEQ ID NO:36), e enzimas EPSP sintase de *E. coli* (SEQ ID NO: 37), *Zea mays* (SEQ ID NO: 38), *Agrobacterium tumefaciens* (SEQ ID NO: 39), e *Bacillus subtilis* (SEQ ID NO:40). O alinhamento mostra os resíduos de aminoácidos mais altamente conservados destacados em preto, e os resíduos de aminoácidos altamente conservados destacados em cinza.

15

DESCRIÇÃO DETALHADA

A presente invenção é projetada para composições e métodos para regular resistência a herbicidas em organismos, particularmente em plantas ou células vegetais. Os métodos envolvem transformar organismos com uma sequência de nucleotídeos codificando um gene de resistência a glifosato da invenção. Em particular, uma sequência de nucleotídeos da invenção é útil para preparar plantas que apresentam tolerância aumentada ao herbicida glifosato. Assim, bactérias, plantas, células vegetais, tecidos vegetais e sementes transformadas são providas. Composições incluem ácidos nucleicos e proteínas relacionando-se a tolerância a herbicidas em microorganismos e plantas, assim como bactérias, plantas,

tecidos vegetais e sementes transformadas. Mais particularmente, seqüências de nucleotídeos dos genes de resistência a glifosato (*grg33*, *syngrg33*, *grg35*, *syngrg35*, *grg36*, *syngrg36*, *grg37*, *syngrg37*, *grg38*, *syngrg38*, *grg39*, *syngrg39*, *grg50*, *syngrg50*) e as seqüências de aminoácidos das proteínas codificadas por estas são descritas. As seqüências são úteis na construção de vetores de expressão para transformação subsequente em plantas de interesse, como sondas para o isolamento de outros genes de resistência a glifosato, como marcadores selecionáveis, e similares. Assim, por "gene de resistência a glifosato da invenção" entende-se a seqüência de nucleotídeos apresentada em SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, e fragmentos e variantes dessas que codificam um polipeptídeo de resistência ou tolerância a glifosato. Da mesma forma, um "polipeptídeo de resistência a glifosato da invenção" é um polipeptídeo tendo a seqüência de aminoácidos apresentada em SEQ ID NO: 2, 5, 8, 11, 14, 16, 19, ou 22, e fragmentos e variantes dessas que conferem resistência ou tolerância a glifosato a uma célula hospedeira.

Plasmídeos contendo as seqüências de nucleotídeos de resistência a herbicidas da invenção foram depositados na coleção permanente da Coleção de Cultura do Serviço de Pesquisa Agrícola (*Agricultural Research Service Culture Collection*), Northern Regional Research Laboratory (NRRL), 1815 North University Street, Peoria, Illinois 61604, Estados Unidos da América, em 9 de junho de 2006, e

designadas como Nos. de Acesso NRRL B-30932 (para *grg33*), NRRL B-30933 (para *grg35*), e NRRL B-30934 (para *grg36*); e em 26 de junho de 2006, e designados como Nos. de Acesso NRRL B-30945 (para *grg37*), NRRL B-30946 (para *grg38*), NRRL B-30947 (para *grg39*), e NRRL B-30948 (para *grg50*). Este depósito será mantido sob os termos do Tratado de Budapeste sobre o Reconhecimento Internacional do Depósito de Microorganismos para Propósitos de Procedimento de Patentes (*International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure*). Este depósito foi feito meramente como uma conveniência para aqueles especialistas na arte e não é uma admissão de que um depósito é necessário sob 35 U.S.C. §112.

Por "glifosato" entende-se qualquer forma herbicida de N-fosfonometilglicina (incluindo qualquer sal desse) e outras formas que resultam na produção de ânion de glifosato em planta. Uma "proteína de resistência a herbicidas" ou uma proteína resultante da expressão de uma "molécula de ácido nucléico codificando resistência a herbicidas" inclui proteínas que conferem em uma célula a habilidade de tolerar uma concentração mais alta de um herbicida do que células que não expressam a proteína, ou tolerar uma certa concentração de um herbicida por um período de tempo maior do que células que não expressam a proteína. Uma "proteína de resistência a glifosato" inclui uma proteína que confere em uma célula a habilidade de tolerar uma concentração de glifosato mais alta do que células que não expressam a proteína, ou a tolerar uma

certa concentração de glifosato por um período de tempo maior do que células que não expressam a proteína. Por "tolerar" ou "tolerância" entende-se tanto sobreviver, ou conduzir funções celulares essenciais, tais como síntese de proteínas e respiração de uma maneira que não seja prontamente discernível de células não tratadas.

Moléculas de ácidos nucleicos isoladas, e variantes e fragmentos dessas

10

Um aspecto da invenção pertence a moléculas de ácidos nucleicos isoladas ou recombinantes compreendendo seqüências de nucleotídeos codificando proteínas e polipeptídeos de resistência a herbicidas ou porções biologicamente ativas dessas, assim como moléculas de ácidos nucleicos suficientes para uso como sondas de hibridização para identificar ácido nucleicos codificando resistência a herbicidas. Como aqui usado, o termo "molécula de ácido nucleico" deve incluir moléculas de DNA (e.g., DNAc, DNA recombinante, ou DNA genômico) e moléculas de RNA (e.g., RNAm) e análogos do DNA ou RNA geradas usando análogos de nucleotídeos. As moléculas de ácidos nucleicos podem ser de fita simples ou de dupla-fita, mas preferivelmente são DNA de dupla-fita.

25

Seqüências de nucleotídeos codificando as proteínas da presente invenção incluem as seqüências apresentadas em SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, as seqüências de nucleotídeos de resistência a

herbicidas depositadas em um hospedeiro bacteriano como Nos. de Acesso NRRL B-30932, B-30933, B-30934, B-30945, B-30946, B-30947, ou B-30948, e variantes, fragmentos, e complementos dessas. Por "complemento" entende-se uma

5 sequência de nucleotídeos que é suficientemente complementar a uma dada sequência de nucleotídeos de forma que esta hibridize à dada sequência de nucleotídeos para, dessa forma, formar um duplex estável. Em algumas concretizações, o complemento hibridiza pelo comprimento

10 total da sequência da invenção. Em outra concretização, o complemento hibridiza através de pelo menos cerca de 50% da sequência da invenção, pelo menos cerca de 55%, pelo menos cerca de 60%, pelo menos cerca de 65%, pelo menos cerca de 70%, pelo menos cerca de 75%, pelo menos cerca de 80%, pelo

15 menos cerca de 85%, pelo menos cerca de 90%, pelo menos cerca de 91%, pelo menos cerca de 92%, pelo menos cerca de 93%, pelo menos cerca de 94%, pelo menos cerca de 95%, pelo menos cerca de 96%, pelo menos cerca de 97%, pelo menos cerca de 98%, ou pelo menos cerca de 99% de uma sequência

20 da invenção. As sequências de aminoácidos correspondentes para as proteínas de resistência a herbicidas codificadas por essas sequências de nucleotídeos são apresentadas em SEQ ID NO: 2, 5, 8, 11, 14, 16, 19, ou 22. A invenção também engloba moléculas de ácidos nucléicos compreendendo

25 sequências de nucleotídeos codificando proteínas de resistência a herbicidas de comprimento parcial, e complementos dessa.

Em algumas concretizações, os polinucleotídeos da presente invenção codificam polipeptídeos que são enzimas EPSP sintase de Classe III. Para os propósitos da presente invenção, uma "enzima EPSP sintase de Classe III" é um

5 polipeptídeo tolerante a herbicida ou resistente a herbicida contendo uma ou mais dos domínios de seqüência de aminoácidos descritos no Pedido de Patente Norte-Americano No. 11/400.598, que é aqui incorporado por referência em sua íntegra.

10

Uma molécula de ácido nucléico "isolada" ou "purificada" ou proteína, ou porção biologicamente ativa dessa, é substancialmente livre de outro material celular, ou meio de cultura quando produzida por técnicas

15 recombinantes, ou substancialmente livre de precursores químicos ou outros químicos quando quimicamente sintetizadas. Preferencialmente, um ácido nucléico "isolado" é sem seqüências (preferencialmente seqüências codificando proteínas) que naturalmente flanqueiam o ácido

20 nucléico (i.e., seqüências localizadas nas extremidades 5' e 3' do ácido nucléico) no DNA genômico do organismo do qual o ácido nucléico é derivado. Para propósitos da invenção, "isolado" quando usado para referir-se a moléculas de ácidos nucléicos exclui cromossomos isolados.

25 Por exemplo, em várias concretizações, a molécula de ácido nucléico isolada codificando resistência a 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 naturalmente flanqueia a molécula de ácido nucléico no DNA

genômico da célula do qual o ácido nucléico é derivado. Uma proteína de resistência a herbicidas que é substancialmente livre de material celular inclui preparações de proteínas tendo menos do que cerca de 30%, 20%, 10%, ou 5% (por peso seco) de proteínas que não as de resistência a herbicidas (também referidas aqui como "proteína contaminante").

Moléculas de ácidos nucléicos que são fragmentos dessas seqüências de nucleotídeos codificando resistência a herbicidas são também englobadas pela presente invenção. Por "fragmento" entende-se uma porção de uma seqüência de nucleotídeos codificando uma proteína de resistência a herbicidas. Um fragmento de uma seqüência de nucleotídeos pode codificar uma porção biologicamente ativa de uma proteína de resistência a herbicidas, ou pode ser um fragmento que pode ser usado como uma sonda de hibridização ou *primer* de PCR usando métodos descritos abaixo. Moléculas de ácidos nucléicos que são fragmentos de uma seqüência de nucleotídeos de resistência a herbicidas 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 nucleotídeos contíguos, ou até o número de nucleotídeos presentes em uma seqüência de nucleotídeos de comprimento total codificando resistência a herbicidas aqui descrita (por exemplo, 1329 nucleotídeos para SEQ ID NO: 1; 1353 nucleotídeos para SEQ ID NO: 4; 1344 nucleotídeos para SEQ ID NO: 7, etc.) dependendo do uso pretendido. Por

nucleotídeos "contíguos" entende-se resíduos de nucleotídeos que são imediatamente adjacentes um ao outro.

Fragmentos das seqüências de nucleotídeos da presente
5 invenção geralmente codificarão fragmentos de proteínas que
retêm a atividade biológica da proteína de resistência a
glifosato de comprimento total; i.e., atividade de
resistência a herbicidas. Por "retém a atividade de
resistência a herbicidas" entende-se que o fragmento tenha
10 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 herbicidas da proteína de resistência a
glifosato de comprimento total aqui descrita como SEQ ID
NO: 2, 5, 8, 11, 14, 16, 19, ou 22. Métodos para medir a
15 atividade de resistência a herbicidas são bem conhecidos na
arte. Ver, por exemplo, Patentes Norte-Americanas Nos.
4.535.060, e 5.188.642, cada qual é aqui incorporada por
referência em sua íntegra.

20 Um fragmento de uma seqüência de nucleotídeos
codificando resistência a herbicidas que codifica uma
porção biologicamente ativa de uma proteína da invenção irá
codificar pelo menos cerca de 15, 25, 30, 50, 75, 100, 125,
150, 175, 200, 250, 300, 350, 400 aminoácidos contíguos, ou
25 até o número total de aminoácidos presente em uma proteína
de resistência a herbicidas da invenção de comprimento
total (por exemplo, 442 aminoácidos para SEQ ID NO: 2; 450
para SEQ ID NO: 5; 447 aminoácidos para SEQ ID NO: 8,
etc.).

Proteínas de resistência a herbicidas preferidas da presente invenção são codificadas por uma sequência de nucleotídeos suficientemente idêntica à sequência de nucleotídeos de SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23. O termo "suficientemente idêntico" deve significar um aminoácido ou sequência de nucleotídeos que possui pelo menos cerca de 60% ou 65% identidade de sequência, cerca de 70% ou 75% de identidade de sequência, cerca de 80% ou 85% de identidade de sequência, ou cerca de 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% de identidade de sequência comparada a uma sequência referência usando um dos programas de alinhamento aqui descritos usando parâmetros padrão. Um especialista na arte reconhecerá que esses valores podem ser apropriadamente ajustados para determinar a identidade correspondente de proteínas codificada por duas sequências de nucleotídeos levando-se em consideração a degeneração de códons, similaridade de aminoácidos, posicionamento do quadro de leitura, e similares.

20

Para determinar a identidade percentual de duas sequências de aminoácidos ou de dois ácidos nucléicos, as sequências são alinhadas para propósitos de comparação ótima. A identidade percentual entre as duas sequências é uma função do número de posições idênticas compartilhadas pelas sequências (i.e., identidade percentual = número de posições idênticas/número total de posições (e.g., posições se sobrepondo) x 100). Em uma concretização, as duas sequências são do mesmo comprimento. A identidade

percentual entre duas seqüências pode ser determinada usando técnicas similares àquelas descritas abaixo, permitindo ou não espaços vazios. Quando calculando a identidade percentual, são contadas tipicamente as correspondências exatas.

A determinação da identidade percentual entre duas seqüências pode ser conseguida usando um algoritmo matemático. Um exemplo não limitante de um algoritmo matemático utilizado para a comparação de duas seqüências é o algoritmo de Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264, modificado como em Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Tal algoritmo é incorporado nos programas BLASTN e BLASTX de Altschul et al. (1990) *J. Mol. Biol.* 215:403. Buscas de nucleotídeos BLAST podem ser realizadas com o programa BLASTN, pontuação = 100, tamanho de palavra = 12, para se obter seqüências de nucleotídeos homólogas a moléculas de ácidos nucléicos resistentes a glifosato da invenção. Buscas BLAST de proteínas podem ser efetuadas com o programa BLASTX, pontuação = 50, tamanho de palavra = 3, para se obter seqüências de aminoácidos homólogas a moléculas proteína de resistência a herbicidas da invenção. Para se obter alinhamentos com espaços para propósitos de comparação, *Gapped BLAST* pode ser utilizado como descrito em Altschul et al. (1997) *Nucleic Acids Res.* 25:3389. Alternativamente, PSI-Blast pode ser usado para efetuar uma busca iterada que detecte relações distantes entre moléculas. Ver Altschul et al. (1997) *supra*. Quando

utilizando os programas BLAST, Gapped BLAST, e PSI-Blast, os parâmetros padrão dos respectivos programas (e.g., BLASTX e BLASTN) podem ser usados. Ver www.ncbi.nlm.nih.gov. Outro exemplo não limitante de um

5 algoritmo matemático utilizado para a comparação de seqüências é o algoritmo de ClustalW (Higgins et al. (1994) *Nucleic Acids Res.* 22:4673-4680). ClustalW compara seqüências e alinha a totalidade da seqüência de aminoácidos ou de DNA e, assim, pode prover dados acerca da

10 conservação de seqüência de toda a seqüência de aminoácidos. O algoritmo de ClustalW é usado em diversos pacotes de software de análise de DNA/aminoácido comercialmente disponíveis, tais como o módulo ALIGNX de *Vector NTI Program Suite* (Invitrogen Corporation, Carlsbad,

15 CA). Após o alinhamento de seqüências de aminoácidos com ClustalW, a identidade percentual de aminoácidos pode ser avaliada. Um exemplo não limitante de um software útil para análise em alinhamentos de ClustalW é GENEDOC™. GENEDOC™ (Karl Nicholas) permite a obtenção da similaridade e

20 identidade de aminoácidos (ou DNA) entre múltiplas proteínas. Outro exemplo não limitante de um algoritmo matemático utilizado para a comparação de seqüências é o algoritmo de Myers and Miller (1988) *CABIOS* 4: 11-17. Tal algoritmo é incorporado no programa ALIGN (versão 2.0), que

25 é parte do pacote de software de alinhamento de seqüências GCG (disponibilizado pela Accelrys, Inc., 9865 Scranton Rd., San Diego, California, USA). Quando utilizando o programa ALIGN para comparar seqüências de aminoácidos, uma tabela de peso de resíduos PAM 120, uma penalidade por

comprimento de espaço (*gap length penalty*) de 12, e uma penalidade por espaço (*gap penalty*) de 4 podem ser usadas.

A menos que declarado ao contrário, GAP Versão 10, que
5 usa o algoritmo de Needleman and Wunsch (1970) *supra*, será
usado para determinar a identidade ou similaridade de
seqüências usando os seguintes parâmetros: % identidade e %
similaridade para uma seqüência de nucleotídeos usando peso
GAP (*GAP Weight*) de 50 e peso de comprimento (*Length*
10 *Weight*) de 3, e a matriz de pontuação de *nwsgapdna.cmp*; %
identidade ou % similaridade para uma seqüência de
aminoácidos usando peso GAP de 8 e peso de comprimento de
2, e o programa de pontuação *BLOSUM62*. Programas
equivalentes também podem ser usados. Por "programa
15 equivalente" entende-se qualquer programa de comparação de
seqüências que, para quaisquer duas seqüências em questão,
gere um alinhamento tendo correspondências de resíduos de
nucleotídeos idênticas e identidade de seqüência percentual
idêntica quando comparado ao alinhamento correspondente
20 gerado pelo programa GAP Versão 10.

A invenção também engloba moléculas de ácidos
nucléicos variantes. "Variantes" das seqüências de
nucleotídeos codificando resistência a herbicidas incluem
25 aquelas seqüências que codificam uma proteína de
resistência a herbicidas aqui descrita, mas que diferem
conservativamente por causa da degeneração do código
genético, assim como aquelas que são suficientemente
idênticas, como discutido acima. Variantes alélicos

naturalmente ocorrentes podem ser identificados com o uso de técnicas de biologia molecular bem conhecidas, tal como reação de cadeia de polimerase (PCR) e técnicas de hibridização, como destacado abaixo. Sequências de nucleotídeos variantes também incluem sequências de nucleotídeos derivadas sinteticamente que foram gerados, por exemplo, usando-se mutagênese sítio-direcionada, mas que ainda codificam as proteínas de resistência a herbicidas descrita na presente invenção, como discutido abaixo. Proteínas variantes englobadas pela presente invenção são biologicamente ativas, isto é, elas retêm a atividade biológica desejada da proteína nativa, ou seja, atividade de resistência a herbicidas. Por "retém a atividade de resistência a herbicidas" entende-se que o variante tenha 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 herbicidas da proteína nativa. Métodos para medir a atividade de resistência a herbicidas são bem conhecidos na arte. Ver, por exemplo, Patentes Norte-Americanas Nos. 4.535.060, e 5.188.642, cada qual é aqui incorporada por referência em sua íntegra.

O especialista na arte irá ainda ponderar que mudanças podem ser introduzidas por mutação nas sequências de nucleotídeos da invenção levando, dessa forma, a mudanças na sequência de aminoácidos da proteína de resistência a herbicidas codificada, sem alterar a atividade biológica da proteína. Assim, moléculas de ácidos nucléicos variantes isoladas podem ser criadas introduzindo-se um ou mais

substituições, adições ou deleções de nucleotídeos na sequência de nucleotídeos correspondente aqui descrita, de forma que um ou mais substituições, adições ou deleções de aminoácidos sejam introduzidas na proteína codificada.

5 Mutações podem ser introduzidas por técnicas padrão, tal como mutagênese sítio-direcionada e mutagênese mediada por PCR, ou técnicas de embaralhamento de genes. Tais sequências de nucleotídeos variantes são também englobadas pela presente invenção.

10

Por exemplo, substituições de aminoácidos conservativas podem ser feitas em um ou mais resíduos previstos, preferencialmente resíduos de aminoácido não essenciais. Um resíduo de aminoácido "não essencial" é um
15 resíduo que pode ser alterado a partir da sequência de tipo selvagem de uma proteína de resistência a herbicidas sem alterar a atividade biológica, enquanto que um resíduo de aminoácido "essencial" é necessário para a atividade biológica. Uma "substituição de aminoácido conservativa" é
20 uma na qual o resíduo de aminoácido é substituído por um resíduo de aminoácido tendo uma cadeia lateral similar. Famílias de resíduos de aminoácidos tendo cadeias laterais similares foram definidas na arte. Estas famílias incluem aminoácidos com cadeias laterais básicas (por exemplo,
25 lisina, arginina, histidina), cadeias laterais acídicas (por exemplo, ácido aspártico, ácido glutâmico), cadeias laterais polares sem carga (por exemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadeias laterais apolares (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 retenham a função. Em geral, tais substituições não seriam feitas para resíduos de aminoácidos conservados, ou para resíduos de aminoácidos residindo em um motivo conservado, onde tais resíduos são essenciais para a atividade da proteína. Entretanto, um especialista na arte entenderia que variantes funcionais podem ter pequenas alterações conservativas ou não conservativas ou modificações nos resíduos conservados. Exemplos de resíduos que são conservados e que podem ser essenciais para a atividade de proteína incluem, por exemplo, resíduos que são idênticos entre todas as proteínas contidas no alinhamento de Figuras 1, 2, ou 3. Exemplos de resíduos que são conservados, mas que podem permitir substituições de aminoácidos conservativas e ainda reter a atividade incluem, por exemplo, resíduos que possuem apenas substituições conservativas entre todas as proteínas contidas no alinhamento das Figuras 1, 2, ou 3.

Lis-22, Arg-124, Asp-313, Arg-344, Arg-386, e Lis-411, são resíduos conservados da EPSP sintase de *E. coli* (Schönbrunn et al. (2001) *Proc. Natl. Acad. Sci. USA* 98: 1376-1380). Resíduos conservados importantes para a atividade de EPSP sintase também incluem Arg-100, Asp-242,

e Asp-384 (Selvapandiyan et al. (1995) *FEBS Letters* 374: 253-256). Arg-27 se liga a S3P (Shuttleworth et al. (1999) *Biochemistry* 38:296-302).

5 Alternativamente, seqüências de nucleotídeos variantes podem ser feitas introduzindo-se mutações randomicamente ao longo de toda ou parte da seqüência codificadora, tal como por mutagênese de saturação, e os mutantes resultantes podem ser rastreados para a habilidade de conferir
10 atividade de resistência a herbicidas para identificar mutantes que retêm atividade. Após a mutagênese, a proteína codificada pode ser expressa de forma recombinante, e a atividade da proteína pode ser determinada usando técnicas de teste padrão.

15

 Usando métodos, tais como PCR, hibridização, e similares, seqüências de resistência a herbicidas correspondentes podem ser identificadas, tais seqüências possuem identidade substancial em relação às seqüências da
20 invenção. Ver, por exemplo, Sambrook J., and Russell, D.W. (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, NY).

25

 Em um método de hibridização, toda ou parte da seqüência de nucleotídeos de resistência a herbicidas pode ser usada para rastrear bibliotecas de DNAC ou genômico. Métodos para a construção de tais bibliotecas de DNAC ou

genômico são geralmente conhecidos na arte e são descritos em Sambrook and Russell, 2001, *supra*. As chamadas sondas de hibridização podem ser fragmentos de DNA genômico, fragmentos de DNAc, fragmentos de RNA, ou outros oligonucleotídeos, e podem ser rotuladas com um grupamento detectável, tal como ^{32}P , ou qualquer outro marcador detectável, tal como outros radioisótopos, um composto fluorescente, uma enzima, ou um co-fator enzimático. Sondas para hibridização podem ser feitas rotulando-se oligonucleotídeos sintéticos baseado na seqüência de nucleotídeos codificando resistência a herbicidas conhecida aqui descrita. *Primers* degenerados projetados com base nos nucleotídeos ou resíduos de aminoácidos conservados nas seqüências de nucleotídeos ou seqüências de aminoácidos codificadas podem ser adicionalmente usados. A sonda compreende tipicamente uma região de seqüência de nucleotídeos que hibridiza sob condições estringentes a pelo menos cerca de 12, pelo menos cerca de 25, pelo menos cerca de 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, 1200, 1300 nucleotídeos consecutivos de uma seqüência de nucleotídeos da invenção codificando resistência a herbicidas ou um fragmento ou variante dessa. Métodos para a preparação de sondas para hibridização são geralmente conhecidos na arte e são descritos em Sambrook and Russell (2001) *supra*, e Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), ambas são aqui incorporadas por referência.

Por exemplo, uma sequência inteira de resistência a herbicidas aqui descrita, ou uma ou mais porções dessa, pode ser usada como uma sonda capaz de hibridizar especificamente a sequências de resistência a herbicidas correspondentes e RNAs mensageiros. Para alcançar hibridização específica sob uma variedade de condições, tais sondas incluem sequências que são únicas e são de pelo menos cerca de 10 nucleotídeos em comprimento, ou de pelo menos cerca de 20 nucleotídeos em comprimento. Tais sondas podem ser usadas para amplificar sequências de resistência a herbicidas correspondentes por PCR de um organismo escolhido. Esta técnica pode ser usada para isolar sequências codificadoras adicionais de um organismo desejado ou como um teste diagnóstico para determinar a presença de sequências codificadoras em um organismo. Técnicas de hibridização incluem rastreamento de hibridização de bibliotecas de DNA plaqueado (tando placas como colônias; ver, por exemplo, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York)).

A hibridização de tais sequências pode ser conduzida sob condições estridentes. Por "condições estridentes" ou "condições de hibridização estridentes" entende-se condições sob as quais uma sonda hibridizará a sua sequência alvo em um grau detectavelmente maior do que a outras sequências (e.g., pelo menos 2 vezes sobre a base). Condições estridentes são dependentes de sequência e serão

diferentes em circunstâncias diferentes. Controlando a
estringência das condições de hibridização e/ou lavagem,
seqüências alvo que são 100% complementares à sonda podem
ser identificadas (sondagem homóloga). Alternativamente,
5 condições de estringência podem ser ajustadas para permitir
alguma falta de correspondência em seqüências, de forma que
níveis menores de similaridade sejam detectados (sondagem
heteróloga). Geralmente, uma sonda é de menos do que cerca
de 1000 nucleotídeos em comprimento, ou menos do que cerca
10 de 500 nucleotídeos em comprimento.

Tipicamente, condições estringentes serão aquelas nas
quais a concentração de sais é de menos do que cerca de 1,5
M de íon Na, tipicamente cerca de 0,01 a 1,0 M de
15 concentração de íons de 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 (e.g., 10 a 50 nucleotídeos) e pelo menos cerca de
60°C para sondas longas (e.g., maiores do que 50
nucleotídeos). Condições estringentes podem ser também
20 alcançadas com a adição de agentes desestabilizantes, tais
como formamida. Condições exemplares de baixa estringência
incluem hibridização com uma solução tampão de 30 a 35% de
formamida, NaCl 1 M, 1% SDS (sódio dodecil sulfato) a
37 °C, e uma lavagem em 1X a 2X SSC (20X SSC = NaCl 3,0 M /
25 triissódio citrato 0,3 M) a 50 a 55 °C. Condições
exemplares de estringência moderada incluem hibridização em
40 a 45% formamida, NaCl 1,0 M, 1% SDS a 37 °C, e uma
lavagem em 0,5X a 1X SSC a 55 a 60 °C. Condições
exemplares de estringência elevada incluem hibridização em

50% formamida, NaCl 1 M, 1% SDS a 37 °C, e uma lavagem em 0,1X SSC a 60 a 65 °C por pelo menos 30 minutos. Opcionalmente, tampões de lavagem podem compreender cerca de 0,1% a cerca de 1% SDS. A duração de hibridização é de
5 geralmente menos do que cerca de 24 horas, usualmente cerca de 4 a cerca de 12 horas. A menos que especificado ao ao contrário, as condições de hibridização estão sob alta estrringência.

10 A especificidade é tipicamente a função de lavagens pós hibridização, com os fatores críticos sendo a força iônica e temperatura da solução de lavagem final. Para híbridos DNA-DNA, a T_m pode ser aproximada a partir da equação de Meinkoth e Wahl (1984) *Anal. Biochem.* 138: 267-
15 284: $T_m = 81,5^{\circ}\text{C} + 16,6 (\log M) + 0,41 (\%GC) - 0,61 (\% \text{ form}) - 500/L$; onde M é a molaridade de cátions monovalentes, %GC é a porcentagem de nucleotídeos guanosina e citosina no DNA, % form é a porcentagem de formamida na solução de hibridização, e L é o comprimento do híbrido em
20 pares de bases. A T_m é a temperatura (sob definidas força iônica e pH) na qual 50% de uma seqüência alvo complementar hibridiza a uma sonda perfeitamente correspondente. A T_m é reduzida em cerca de 1 °C para cada 1% de falta de correspondência; assim, T_m e condições de hibridização, e/ou
25 lavagem podem ser ajustadas para hibridizar a seqüências da identidade desejada. Por exemplo, se seqüências com $\geq 90\%$ de identidade são procuradas, a T_m pode ser diminuída em 10 °C. Por exemplo, se seqüências com $\geq 90\%$ de identidade são procuradas, a T_m pode ser diminuída em 10°C. Geralmente,

condições estridentes são selecionadas para serem cerca de 5°C menores do que o ponto térmico de derretimento (T_m) para a sequência específica e seu complemento em uma definida força iônica e pH. Contudo, condições severamente

5 estridentes podem utilizar uma hibridização e/ou lavagem a 1, 2, 3, ou 4 °C a menos do que o ponto térmico de derretimento (T_m); condições moderadamente estridentes podem utilizar uma hibridização e/ou lavagem a 6, 7, 8, 9 ou 10°C a menos do que o ponto térmico de derretimento (T_m);

10 condições de baixa estridência podem utilizar uma hibridização e/ou lavagem a 11, 12, 13, 14, 15 ou 20°C a menos do que o ponto térmico de derretimento (T_m). Usando a equação, condições de hibridização e lavagem, e T_m desejada, aqueles de habilidade ordinária entenderão que variações na

15 estridência de soluções de hibridização e/ou lavagem são inerentemente descritas. Se o grau desejado de falta de correspondência resulta em uma T_m de menos do que 45 °C (solução aquosa) ou 32 °C (solução de formamida), é preferível aumentar a concentração de SSC de forma que uma

20 maior temperatura possa ser usada. Um guia extensivo para a hibridização de ácidos nucleicos é encontrado em Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes*, Part I, Chapter 2 (Elsevier, New York); e Ausubel, et al., eds. (1995) *Current Protocols in Molecular Biology*, Chapter 2 (Greene Publishing and Wiley-Interscience, New York). Ver, Sambrook, et al., (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York). Ver Sambrook et al. (1989) *Molecular*

Cloning: A Laboratory Manual (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

Proteínas Isoladas e Variantes e Fragmentos destas

5

Proteínas de resistência a herbicidas são também englobadas na presente invenção. Por "proteína de resistência a herbicidas" entende-se uma proteína tendo a seqüência de aminoácidos apresentada em SEQ ID NO: 2, 5, 8,
10 11, 14, 16, 19, ou 22. Fragmentos, porções biologicamente ativas, e variantes dessa são também providas, e podem ser usadas para praticar os métodos da presente invenção.

"Fragmentos" ou "porções biologicamente ativas"
15 incluem fragmentos de polipeptídeos compreendendo uma porção de uma seqüência de aminoácidos codificando uma proteína de resistência a herbicidas descrita em SEQ ID NO: 2, 5, 8, 11, 14, 16, 19, ou 22, e que retenha atividade de resistência a herbicidas. Uma porção biologicamente de uma
20 proteína de resistência a herbicidas pode ser um polipeptídeo que é, por exemplo, de 10, 25, 50, 100 ou mais aminoácidos em comprimento. Tais porções biologicamente ativas podem ser preparadas por técnicas de recombinação e avaliadas para atividade de resistência a herbicida.
25 Métodos para medir a atividade de resistência a herbicidas são bem conhecidos na arte. Ver, por exemplo, Patentes Norte-Americanas Nos. 4.535.060, e 5.188.642, cada qual é aqui incorporada por referência em sua íntegra. Como usado aqui, um fragmento compreende pelo menos 8 aminoácidos

contíguos de SEQ ID NO: 2, 5, 8, 11, 14, 16, 19, ou 22. A invenção engloba outros fragmentos, contudo, tal como qualquer fragmento na proteína maior do que cerca de 10, 20, 30, 50, 100, 150, 200, 250, 300, 350, ou 400
5 aminoácidos.

Por "variantes" entende-se proteínas ou polipeptídeos tendo uma seqüência de aminoácido que é pelo menos de cerca de 60%, 65%, cerca de 70%, 75%, cerca de 80%, 85%, 90%,
10 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% ou 99% idêntica à seqüência de aminoácidos de SEQ ID NO: 2, 5, 8, 11, 14, 16, 19, ou 22. Variantes também incluem polipeptídeos codificados por uma molécula de ácido nucléico que hibridiza à molécula de ácido nucléico de SEQ ID NO: 1, 3,
15 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, ou um complemento dessas, sob condições estridentes. Variantes incluem polipeptídeos que diferem em seqüência de aminoácidos devido a mutagênese. Proteínas variantes englobadas pela presente invenção são biologicamente
20 ativas, isto é, elas continuam a possuir a atividade biológica desejada da proteína nativa, ou seja, retém a atividade de resistência a herbicidas. Métodos para medir a atividade de resistência a herbicidas são bem conhecidos na arte. Ver, por exemplo, Patentes Norte-Americanas Nos.
25 4.535.060, e 5.188.642, cada qual é aqui incorporada por referência em sua íntegra.

Genes bacterianos, tais como os genes *grg* e *syngrg* desta invenção, muito freqüentemente possuem múltiplos

códons de iniciação de metionina em proximidade ao início do quadro de iniciação de leitura. Frequentemente, a iniciação de tradução em um ou mais desses códons de iniciação levarão à geração de uma proteína funcional.

5 Esses códons de iniciação podem incluir códons ATG. Contudo, bactérias tais como *Bacillus* sp. também reconhecem o códon GTG como códon de iniciação, e proteínas que iniciam a tradução em códons GTG contêm uma metionina no primeiro aminoácido. Adicionalmente, não é frequentemente

10 determinado *a priori* quais desses códons são usados naturalmente na bactéria. Assim, é entendido que usar um dos códons de metionina alternativos podem levar à geração de variantes de *grg* e *syngrg* que conferem resistência a herbicidas. Essas proteínas de resistência a herbicidas são

15 englobadas 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, são também

20 englobados. Métodos para produzir anticorpos são bem conhecidos na arte (ver, por exemplo, Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; Patente Norte-Americana No. 4.196.265).

25

Variantes Alterados ou Melhorados

É reconhecido que a sequência de DNAs dos genes *grg* ou *syngrg* da invenção pode ser alterada por vários métodos, e

que essas alterações podem resultar em seqüências de DNA codificando proteínas com seqüências de aminoácidos diferentes daquelas codificadas pelas seqüências *grg* ou *syngrg* aqui descritas. Essa proteína pode ser alterada de

5 vários jeitos, incluindo substituições, deleções, trunicações, e inserções de aminoácidos de um ou mais aminoácidos de SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, incluindo até cerca de 2, cerca de 3, cerca de 4, cerca de 5, cerca de 6, cerca de 7, cerca de

10 8, cerca de 9, cerca de 10, cerca de 15, cerca de 20, cerca de 25, cerca de 30, cerca de 35, cerca de 40, cerca de 45, cerca de 50, cerca de 55, cerca de 60, cerca de 65, cerca de 70, cerca de 75, cerca de 80, cerca de 85, cerca de 90, cerca de 100 ou mais substituições, deleções ou inserções

15 de aminoácidos.

Métodos para tais manipulações são geralmente conhecidos na arte. Por exemplo, variantes de seqüências de aminoácidos das proteínas GRG aqui descritas podem ser

20 preparados por mutações no DNA. Isso pode ser também conseguido por uma de diversas formas de mutagênese e/ou em evolução direcionada. Em alguns aspectos, as mudanças codificadas na seqüência de aminoácidos não afetarão substancialmente a função da proteína. Tais variantes

25 possuirão a atividade de resistência a herbicidas desejada. Contudo, é entendido que a habilidade das proteínas GRG aqui descritas de conferir resistência a herbicidas pode ser melhorada por uso de tais técnicas nas composições desta invenção. Por exemplo, pode-se expressar as

seqüências de *grg* ou *syngrg* aqui descritas em células hospedeiras que exibem altas taxas de falta de incorporação de bases durante a replicação de DNA, tais como XL-I Red (Stratagene, La Jolla, CA). Após a propagação em tais

5 linhagens, pode-se isolar o DNA da invenção (por exemplo preparando-se DNA de plasmídeo, ou amplificando-se por PCR e clonando o fragmento de PCR resultante em um vetor), cultivar as mutações *grg* em uma linhagem não mutagênica, e identificar genes mutados com resistência melhorada a um

10 herbicida, tal como glifosato, por exemplo cultivando-se células em concentrações crescentes de glifosato e testando clones que conferem habilidade de tolerar concentrações crescentes de glifosato.

15 Alternativamente, alterações podem ser feitas na seqüência da proteína de várias proteínas nos terminais amino ou carboxi sem afetar substancialmente a atividade. Isso pode incluir inserções, deleções, ou alterações introduzidas por métodos moleculares modernos, tais como

20 PCR, incluindo amplificações de PCR que alterar ou estendem a seqüência codificadora de proteína em virtude da inclusão de seqüências codificando aminoácidos nos oligonucleotídeos utilizados na amplificação por PCR. Alternativamente, as seqüências de proteínas adicionadas podem incluir as

25 seqüências codificadoras de proteínas inteiras, tais como aquelas usadas comumente na arte para gerar fusões de proteínas. Tais proteínas de fusão são freqüentemente usadas para (1) aumentar a expressão de uma proteína de interesse, (2) introduzir um domínio de ligação, atividade

enzimática, ou epitopo para facilitar tanto a purificação de proteína, detecção de proteína, ou outros usos experimentais conhecidos na arte, ou, (3) direcionar a secreção ou tradução de uma proteína para uma organela subcelular, tal como o espaço periplasmático de bactérias gram-negativas, ou o retículo endoplasmático de células eucariontes, este último freqüentemente resulta em glicosilação da proteína.

10 Sequências de nucleotídeos e de aminoácidos variantes da presente invenção também englobam seqüências derivadas de procedimentos mutagênicos e recombinogênicos, tais como embaralhamento de DNA. Com tal procedimento, um ou mais regiões codificadoras de proteína de resistência a herbicidas diferentes podem ser usadas para criar uma nova proteína de resistência a herbicidas possuindo as propriedades desejadas. Dessa maneira, bibliotecas de polinucleotídeos recombinantes são geradas a partir de uma população de polinucleotídeos de seqüência relacionados compreendendo regiões de seqüência que possuem identidade de seqüência substancial e podem ser homologamente recombinadas *in vitro* ou *in vivo*. Por exemplo, usando esta abordagem, motivos de seqüência codificando um domínio de interesse podem ser embaralhadas entre o gene de resistência a herbicidas da invenção e outros genes de resistência a herbicidas conhecidos para obter um novo gene codificando uma proteína com uma propriedade de interesse melhorada, tal como atividade de resistência a glifosato melhorada. Estratégias para tal embaralhamento de DNA são

conhecidas na arte. Ver, por exemplo, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91: 10747-10751; Stemmer (1994) *Nature* 370: 389-391; Crameri et al. (1997) *Nature Biotech.* 15: 436-438; Moore et al. (1997) *J. Mol. Biol.* 272: 336-347;
5 Zhang et al. (1997) *Proc. Natl. Acad. Sci. USA* 94: 4504-4509; Crameri et al. (1998) *Nature* 391: 288-291; e Patentes Norte-Americanas Nos. 5.605.793 e 5.837.458.

Transformação de Células Bacterianas ou Vegetais

10

Aqui providos estão novos genes isolados que conferem resistência a um herbicida. Também providas são seqüências de aminoácidos das proteínas GRG da invenção. A proteína resultante de tradução deste gene permite que células
15 funcionem na presença de concentrações de um herbicida que são, de outra forma, tóxicas a células incluindo células vegetais e células bacterianas. Em um aspecto da invenção, os genes *grg* ou *syngrg* são úteis como marcadores para avaliar a transformação de células bacterianas ou vegetais.

20 Métodos para detectar a presença de um transgene em uma planta, órgão vegetal (e.g., folhas, caules, raízes, etc.), sementes, célula vegetal, propágulo, embrião ou progênie da mesma são bem conhecidos na arte.

25 Manipulando-se os genes da invenção a serem expressos a partir de um promotor conhecido por estimular a transcrição no organismo a ser testado e corretamente traduzido para gerar um peptídeo GRG intacto, e posicionar as células em uma concentração de herbicida de outra forma

tóxica, pode-se identificar células que foram transformadas com o DNA em virtude de sua resistência a herbicidas. Por "promotor" entende-se uma sequência de ácidos nucleicos que funcione para direcionar a transcrição de uma sequência codificador abaixo. O promotor, junto com outras sequências de ácidos nucleicos reguladoras de transcrição e tradução, (também denominadas "sequências controle") é necessário para a expressão de uma sequência de DNA de interesse.

10 A transformação de células bacterianas é conseguida por uma de diversas técnicas conhecidas na arte, incluindo mas não se limitando a, eletroporação ou transformação química (ver, por exemplo, Ausubel, ed. (1994) *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc.,
15 Indianapolis, IN). Marcadores conferindo resistência a substâncias tóxicas são úteis para identificar células transformadas (tendo absorvido e expresso o DNA de teste) de células não transformadas (aquelas não contendo ou não expressando o DNA de teste). Em um aspecto da invenção, os
20 genes *grg* ou *syngrg* aqui descritos são úteis como marcadores para avaliar a transformação de células bacterianas ou vegetais.

A transformação de células vegetais pode ser
25 conseguida de modo similar. Por "planta" entende-se plantas inteiras, órgãos vegetais (por exemplo, folhas, caules, raízes, etc.), sementes, células vegetais, propágulos, embriões e descendentes das mesmas. Células vegetais podem ser diferenciadas ou indiferenciadas (por exemplo, calo,

células de cultura em suspensão, protoplastos, células de
folha, células de raiz, células de floema, pólen). "Plantas
transgênicas" ou "plantas transformadas" ou planta
"estavelmente transformadas" ou células ou tecidos,
5 referem-se a plantas que têm incorporado ou integrado
seqüências de ácidos nucleicos exógenas ou fragmentos de
DNA na célula vegetal. Por "transformação estável" entende-
se que a construção de nucleotídeo introduzida em uma
planta se integre no genoma da planta e seja capaz de ser
10 herdado pela progênie desta.

Os genes *grg* da invenção podem ser modificados para se
obter ou intensificar a expressão em células vegetais. As
seqüências de resistência a herbicidas da invenção podem
15 ser providas em cassetes de expressão para expressão na
planta de interesse. "Cassete de expressão vegetal" inclui
construções de DNA, incluindo construções de DNA
recombinantes, que são capazes de resultar na expressão de
uma proteína a partir de um quadro de iniciação de leitura
20 em uma célula vegetal. O cassete incluirá na direção de
transcrição 5'-3', uma região de iniciação transcricional
(i.e., promotor, particularmente um promotor heterólogo)
operacionalmente ligado a uma seqüência de DNA da invenção,
e/ou uma região de terminação transcricional e de tradução
25 (i.e., região de terminação) funcional em plantas. O
cassete pode conter adicionalmente pelo menos um gene
adicional a ser co-transformado no organismo, tal como um
gene marcador de seleção. Alternativamente, o(s) gene(s)
adicional(is) pode ser provido em múltiplos cassetes de

expressão. Tal cassete de expressão é provido com uma pluralidade de sítios de restrição para inserção da sequência de resistência a herbicidas a estar sob a regulação transcricional das regiões reguladoras.

5

O promotor pode ser nativo ou análogo, ou estrangeiro ou heterólogo, em relação ao hospedeiro vegetal e/ou em relação à sequência de DNA da invenção. Adicionalmente, o promotor pode ser a sequência natural ou alternativamente
10 uma sequência sintética. Onde o promotor é "nativo" ou "homólogo" em relação ao hospedeiro vegetal, entende-se que o promotor seja encontrado na planta nativa na qual o promotor é introduzido. Onde o promotor é "estrangeiro" ou "heterólogo" em relação à sequência de DNA da invenção,
15 entende-se que o promotor não seja o promotor nativo ou naturalmente ocorrente em relação à sequência de DNA da invenção operacionalmente ligada. "Heterólogo" geralmente refere-se às sequências de ácidos nucleicos que não são endógenas à célula ou parte do genoma nativo no qual estão
20 presentes, e foram adicionadas à célula por infecção transfecção, microinjeção, eletroporação, microprojeção, ou similares. Por "operacionalmente ligado" entende-se uma ligação funcional entre um promotor e uma segunda sequência, caracterizado pelo fato de que a sequência
25 promotora inicia e media a transcrição da sequência de DNA correspondente à segunda sequência. Geralmente, operacionalmente ligado significa que as sequências de ácidos nucleicos sendo ligadas são contíguas e, onde

necessário para juntar duas regiões codificadoras de proteínas, contíguas e no mesmo quadro de leitura.

Freqüentemente, tais construções também conterão
5 regiões 5' e 3' não traduzidas. Tais construções podem conter uma "seqüência sinal" ou "seqüência líder" para facilitar o transporte co-tradução ou pós-tradução do peptídeo de interesse para certas estruturas intracelulares, tais como o cloroplasto (ou outro
10 plastídeo), retículo endoplasmático, ou aparelho de Golgi, o para ser secretado. Por exemplo, o gene pode ser manipulado para conter um peptídeo sinal para facilitar a transferência do peptídeo para o retículo endoplasmático. Por "seqüência sinal" entende-se uma seqüência que é
15 conhecida ou suspeita de resultar em transporte co-tradução ou pós-tradução de peptídeo através da membrana celular. Em eucariontes, isso tipicamente envolve a secreção no aparelho e Golgi, com alguma glicosilação resultante. Por "seqüência líder" entende-se qualquer seqüência que, quando
20 traduzida, resulte em uma seqüência de aminoácidos suficiente para iniciar o transporte co-tradução da cadeia peptídica para uma organela sub-celular. Assim, isso inclui seqüências líder direcionando o transporte e/ou glicosilação por passagem no retículo endoplasmático,
25 passagem para vacúolos, plastídeo incluindo cloroplastos, mitocôndria, e similares. Pode ser também preferível construir o cassete de expressão de vegetal para conter um íntron, tal como o processamento de RNAm do íntron é necessário para a expressão.

Por "região 3' não traduzida" entende-se uma sequência de nucleotídeos localizada abaixo de uma sequência codificador. Sequências de sinal de poliadenilação e outras
5 sequências codificando sinais regulatórios capazes de afetar a adição de traços de ácido poliadenílico à extremidade 3' do precursor de RNAm são regiões 3' não traduzidas. Por "região 5' não traduzida" entende-se uma sequência de nucleotídeos localizada acima de uma sequência
10 codificadora.

Outros elementos não traduzidos acima ou abaixo incluem intensificadores. Intensificadores são sequências de nucleotídeos que atuam para aumentar a expressão de uma
15 região promotora. Intensificadores são bem conhecidos na arte e incluem, mas não estão limitados a, a região intensificadora SV40 e o elemento intensificador 35S.

A região de terminação pode ser nativa em relação à
20 região de iniciação transcricional, pode ser nativa em relação à sequência de resistência a herbicidas da presente invenção, ou pode ser derivada de outra fonte. Regiões de terminação convenientes são disponíveis do plasmídeo Ti de *A. tumefaciens*, tais como as regiões de terminação de
25 octopina sintase e nopalina sintase. Ver também Guerineau et al. (1991) *Mol. Gen. Genet.* 262: 141-144; Proudfoot (1991) *Cell* 64: 671-674; Sanfecon 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.

Em um aspecto da invenção, seqüências de DNA
5 sintéticas são projetadas para um dado polipeptídeo, tais
como os polipeptídeos da invenção. A expressão do quadro de
iniciação de leitura da seqüência de DNA sintética em uma
célula resulta na produção do polipeptídeo da invenção.
Seqüências de DNA sintéticas podem ser úteis para
10 simplesmente remover sítios de restrição de endonuclease
indesejados, para facilitar estratégias de clonagem de DNA,
para alterar ou remover qualquer tendência de códon em
potencial, para alterar ou melhorar o conteúdo de GC, para
remover ou alterar quadros de leitura alternados, e/ou para
15 alterar ou remover sítios de reconhecimento de *splice*
íntron/exon, sítios de poliadenilação, seqüências Shine-
Delgarno, elementos promotores indesejados e similares que
possam estar presentes na seqüência de DNA nativa. É também
possível que seqüências de DNA sintéticas possam ser
20 utilizadas para introduzir outras melhorias a uma seqüência
de DNA, tal como introdução de uma seqüência íntron,
criação de uma seqüência de DNA que é expressa como uma
proteína de fusão para seqüências direcionando organelas,
tais como peptídeos de trânsito de cloroplastos, peptídeos
25 direcionando apoplasto/vacúolo, ou seqüências peptídicas
que resultam em retenção do peptídeo resultante no retículo
endoplasmático. Genes sintéticos podem ser também
sintetizados usando códons preferidos de célula hospedeira
para expressão melhorada, ou podem ser sintetizados usando

códons em uma frequência de uso de códons preferida de hospedeiro. Ver, por exemplo, Campbell and Gowri (1990) *Plant Physiol.* 92: 1-11; Patentes Norte-Americanas Nos. 6.320.100; 6.075.185; 5.380.831; e 5.436.391, U.S. 5 Published Application Nos. 20040005600 e 20010003849, e Murray et al. (1989) *Nucleic Acids Res.* 17: 477-498, aqui incorporadas por referência.

Em uma concretização, os ácidos nucléicos de interesse 10 são direcionados para o cloroplasto para expressão. Dessa maneira, onde o ácido nucléico de interesse não é diretamente inserido no cloroplasto, o cassete de expressão conterà adicionalmente um ácido nucléico codificando um peptídeo de trânsito para direcionar o produto gênico de 15 interesse para os cloroplastos. Tais peptídeos de trânsito são conhecidos na arte. Ver, 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) 20 *Biochem. Biophys. Res. Commun.* 196: 1414-1421; e Shah et al. (1986) *Science* 233: 478-481.

Os ácidos nucléicos de interesse a serem direcionados para o cloroplasto podem ser otimizados para expressão no 25 cloroplasto para contabilizar diferenças no uso de códons entre o núcleo vegetal e esta organela. Dessa maneira, os ácidos nucléicos de interesse podem ser sintetizados usando códons preferidos de cloroplastos. Ver, por exemplo,

Patente Norte-Americana No. 5.380.831, aqui incorporada por referência.

Tipicamente este "cassete de expressão vegetal" será
5 inserido em um "vetor de transformação vegetal". Por "vetor de transformação" 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 uma molécula
10 de DNA "vetor". Por exemplo, vetores binários são vetores de transformação vegetais que utilizam dois vetores de DNA não contíguos para codificar todas as funções requisito cis- e trans- atuantes para a transformação de células vegetais (Hellens and Mullineaux (2000) *Trends in Plant*
15 *Science* 5: 446-451). "Vetor" refere-se a uma construção de ácidos nucleicos projetado para transferência entre células hospedeiras diferentes. "Vetor de expressão" refere-se a um vetor que possui a habilidade de incorporar, integrar e expressar seqüências de DNA heterólogas ou fragmentos em
20 uma célula estrangeira.

Este vetor de transformação vegetal pode ser compreendido de um ou mais vetores de DNA necessários para conseguir a transformação vegetal. Por exemplo, é uma
25 prática comum na arte utilizar vetores de transformação vegetais que são compreendidos de mais de um segmento de DNA contíguo. Esses vetores são freqüentemente na arte referidos com "vetores binários". Vetores binários, assim como vetores com plasmídeos ajudantes são mais

freqüentemente usados para transformação mediada por *Agrobacterium*, onde o tamanho e a complexidade de segmentos de DNA necessários para se conseguir transformação eficiente é bem grande, e é mais vantajoso separar funções em moléculas de DNA separadas. Vetores binários tipicamente contêm um vetor de plasmídeo que contém as seqüências cis-atuantes necessárias para a transferência de T-DNA (tal como limite esquerdo e limite direito), um marcador de seleção que é manipulado para ser capaz de expressão em uma célula vegetal, e um "gene de interesse" (um gene manipulado para ser capaz de expressão em uma célula vegetal para a qual a geração de plantas transgênicas é desejada). Também presentes no vetor de plasmídeo estão seqüências requeridas par a replicação bacteriana. As seqüências cis-atuantes são arranjadas do modo a permitir a transferência eficiente em células vegetais e expressão nas mesmas. Por exemplo, o gene marcador de seleção e o gene de interesse estão localizados entre os limites esquerdo e direito. Frequentemente um segundo vetor de plasmídeo contém os fatores trans-atuantes que mediam a transferência de T-DNA de *Agrobacterium* para células vegetais. Este plasmídeo freqüentemente contém funções de virulência (genes Vir) que permitem a infecção de células vegetais por *Agrobacterium*, e transferência de DNA por clivagem nas seqüências dos limites e transferência de mediada por vir, como é entendido na arte (Hellens and Mullineaux (2000) *Trends in Plant Science*, 5: 446-451). Diversos tipos de linhagens de *Agrobacterium* (e.g. LBA4404, GV3101, EHA101, EHA105, etc.) podem ser usados para a transformação

vegetal. O segundo vetor de plasmídeo não é necessário para transformar as plantas por outros métodos tais como microprojeção, microinjeção, eletroporação, polietileno glicol, etc.

5

Transformação Vegetal

Métodos da invenção envolvem introduzir uma construção de nucleotídeo em uma planta. Por "introduzir" entende-se
10 apresentar à planta a construção de nucleotídeos de forma que a construção ganhe acesso ao interior de uma célula da planta. Os métodos da invenção não requerem que um método em particular para introdução de uma construção de nucleotídeos em uma planta seja usado, apenas que a
15 construção de nucleotídeos ganhe acesso ao interior de ao menos uma célula da planta. Métodos para introduzir construções de nucleotídeos em plantas são conhecidos na arte incluindo, mas não se limitando a, métodos de transformação estável, métodos de transformação transiente,
20 e métodos mediados por vírus.

Em geral, os métodos de transformação vegetal envolvem transferir DNA heterólogo para células vegetais alvo (e.g., embriões imaturos ou maduros, culturas em suspensão, calo
25 indiferenciado, protoplastos, etc.), seguido da aplicação de um nível limite máximo de seleção apropriado (dependendo do gene marcador selecionável e, neste caso, "glifosato") para se recuperar as células vegetais transformadas a partir de um grupo de massa de células não transformadas.

Explantos são tipicamente transferidos para um suprimento fresco do mesmo meio e cultivados rotineiramente. Subseqüentemente, as células transformadas são diferenciadas em brotos após a colocação em meio de regeneração suplementado com um nível limite máximo do agente de seleção (e.g. "glifosato"). Os brotos são então transferidos para um meio de enraizamento seletivo recuperar os brotos com raiz ou plântulas. As plântulas transgênicas então crescem em plantas maduras e produzem sementes férteis (por exemplo, Hiei et al. (1994) *Plant J.* 6: 271-282; Ishida et al. (1996) *Nat. Biotechnol.* 14: 745-750). Os explantes são tipicamente transferidos para um suprimento fresco do mesmo meio e cultivados rotineiramente. Uma descrição geral das técnicas e métodos para gerar plantas transgênicas é encontrada em Ayres and Park (1994) *Critical Reviews in Plant Science* 13: 219-239 e Bommineni and Jauhar (1997) *Maydica* 42: 107-120. Como o material transformado contém muitas células, tanto células transformadas e não transformadas estão presentes em qualquer pedaço do calo ou tecido ou grupo de células submetido a alvo. A habilidade de matar células não transformadas e permitir que as células transformadas se proliferem resulta em culturas vegetais transformadas. Frequentemente, a habilidade para remover células não transformadas é uma limitação à recuperação rápida de células vegetais transformadas e geração bem-sucedida de plantas transgênicas. Métodos moleculares e bioquímicos podem ser então usados para se confirmar a presença e do

polinucleotídeo(s) de interesse integrado no genoma da planta transgênica.

A geração de plantas transgênicas pode ser realizada por um dentre diversos métodos, incluindo, mas não se limitando a, introdução de DNA heterólogo por *Agrobacterium* em células vegetais (transformação mediada por *Agrobacterium*), bombardeamento de células vegetais com DNA estrangeiro heterólogo aderido a partículas, e vários outros métodos não mediados diretamente por partículas (por exemplo, Hiei et al. (1994) *Plant J.* 6: 271-282; Ishida et al. (1996) *Nat. Biotechnol* 14: 745-750; Ayres and Park (1994) *CRC Crit. Rev. Plant Sci.* 13: 219-239; Bommineni and Jauhar (1997) *Maydica* 42: 107-120) para transferência de DNA.

Métodos para a transformação de cloroplastos são conhecidos na arte. Ver, por exemplo, Svab et al. (1990) *Proc. Natl. Acad. Sci. USA* 87: 8526-8530; Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90: 913-917; Svab and Maliga (1993) *EMBO J.* 12: 601-606. O método confia na entrega de partículas por pistola de DNA contendo um marcador selecionável e o direcionamento do DNA para o genoma do plastídeo através de recombinação homóloga. Adicionalmente, transformação de plastídeo pode ser conseguida por transativação de um transgene silencioso transmitido por plastídeo por expressão preferencial de tecido de uma RNA polimerase codificada nuclearmente e direcionada por 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
5 cultivadas a plantas de acordo com meios convencionais.
Ver, por exemplo, McCormick *et al.* (1986) *Plant Cell Reports* 5: 81-84. Essas plantas podem ser, então,
cultivadas e tanto polinizadas com a mesma linhagem
transformada quanto com linhagens diferentes, e o híbrido
10 resultante tendo expressão constitutiva da característica
fenotípica desejada identificada. Duas ou mais gerações
podem ser cultivadas para assegurar que a expressão da
característica fenotípica desejada foi estavelmente mantida
e herdada e, então, as sementes colhidas para assegurar que
15 a expressão da característica fenotípica desejada foi
alcançada. Dessa maneira, a presente invenção provê
sementes transformadas (também referidas como "sementes
transgênicas") tendo uma construção de nucleotídeos da
invenção, por exemplo, um cassete de expressão da invenção,
20 estavelmente incorporado em seu genoma.

Plantas

A presente invenção pode ser usada para a
25 transformação de qualquer espécie de planta, incluindo, mas
não se limitando a, monocotiledôneas e dicotiledôneas.
Exemplos de plantas de interesse incluem, mas não são
limitados a, milho, sorgo, trigo, girassol, tomate,
crucíferas, pimentas, batata, algodão, arroz, soja,

beterraba, cana-de-açúcar, tabaco, cevada, e colzas, *Brassica* sp., alfafa, centeio, milheto, cártamo, amendoim, batata doce, mandioca, café, coco, abacaxi, frutas cítricas, cacau, chá, banana, abacate, figo, goiaba, manga, azeitona, mamão-papaia, caju, macadâmia, amêndoa, aveia, 5 verduras, plantas ornamentais, e coníferas.

Verduras incluem, mas não estão limitadas a, tomates, alface, feijão verde, feijões, ervilhas, e membros do 10 gênero *Curcumis* tais como pepino, cantalupo, e meloeiro. Plantas ornamentais incluem, mas não são limitadas a, azaléia, hortênsia, hibisco, rosas, tulipas, narcisos, petúncias, cravo, bico-de-papagaio, e crisântemo. Preferencialmente, plantas da presente invenção são plantas 15 de cultivo (por exemplo, milho, sorgo, trigo, girassol, tomate, crucíferas, pimentas, batatas, algodão, arroz, soja, beterraba, cana-de-açúcar, tabaco, cevada, colza, etc).

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

25

Avaliação da Transformação Vegetal

Após a introdução de DNA heterólogo estrangeiro em células vegetais, a transformação ou integração de gene

heterólogo no genoma da planta é confirmada por vários métodos, tais como análise de ácidos nucleicos, proteínas e metabólitos associados com o gene integrado.

5 Análise de PCR é um método rápido para rastrear células, tecidos ou partes aéreas transformados para a presença do gene incorporado no estágio inicial antes do transplante para o solo (Sambrook and Russell (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). PCR é conduzido usando *primers* de oligonucleotídeos específicos para o gene de interesse ou base de *Agrobacterium* vetor, etc.

15 A transformação vegetal pode ser confirmada por análise de *Southern blot* de DNA genômico (Sambrook and Russell, 2001, *supra*). Em geral, DNA total é extraído do transformante, digerido com enzimas de restrição apropriadas, fracionada em um gel de agarose e transferidas para uma membrana de nitrocelulose ou náilon. A membrana ou
20 "*blot*" é, então, sondada com, por exemplo, fragmentos de DNA alvo radio-rotulados com ^{32}P para confirmar a integração do gene introduzido no genoma vegetal de acordo com técnicas padrão (Sambrook and Russell, 2001, *supra*).

25 Na análise *Northern*, RNA é isolado de tecidos específicos de transformante, fracionado em um gel de agarose de formaldeído, *blotted* em um filtro de náilon de acordo com procedimentos padrão que são rotineiramente usados na arte (Sambrook and Russell, 2001, *supra*). A

expressão de RNA codificado pelos genes aqui descritos é, então, testada hibridizando o filtro a uma sonda radioativa derivada de um polinucleotídeo da invenção, por métodos conhecidos na arte (Sambrook and Russell, 2001, *supra*)

5

Western blot e testes bioquímicos e similares podem ser conduzidos nas plantas transgênicas para se determinar a presença de proteína codificada pelo gene de resistência a herbicidas por procedimentos padrão (Sambrook and
10 Russell, 2001, *supra*) usando anticorpos que se ligam a um ou mais epitopos presentes na proteína de resistência a herbicidas.

Métodos para aumentar a produção vegetal

15

Métodos para aumentar a produção vegetal são providos. Os métodos compreendem introduzir em uma planta ou célula vegetal um polinucleotídeo compreendendo uma sequência de *grg* aqui descrita. Como aqui definido, a "produção" da
20 planta refere-se à qualidade e/ou quantidade de biomassa produzida pela planta. Por "biomassa" entende-se qualquer produto vegetal mensurável. Um aumento na produção de biomassa é qualquer melhoria na produção do produto vegetal mensurável. Aumentar a produção vegetal possui diversas
25 aplicações comerciais. Por exemplo, aumentar a biomassa foliar pode aumentar a produção de verduras folhosas para consumo humano ou animal. Adicionalmente, aumentar a biomassa foliar pode ser usado para aumentar a produção de produtos farmacêuticos derivados de plantas ou produtos

industriais. Um aumento em produção pode compreender qualquer aumento estatisticamente significativo, incluindo, mas não se limitando a, pelo menos 1% de aumento, pelo menos 3% de aumento, pelo menos 5% de aumento, pelo menos 10% de aumento, pelo menos 20% de aumento, pelo menos 30%,
5 pelo menos 50%, pelo menos 70%, pelo menos 100% ou aumento maior.

Em métodos específicos, a planta é tratada com uma
10 concentração efetiva de um herbicida, onde a aplicação de herbicida resulta em produção vegetal. Por "concentração efetiva" entende-se a concentração que permite o aumento de produção na planta. Tais concentrações efetivas para herbicidas de interesse são geralmente conhecidas na arte.
15 O herbicida pode ser aplicado tanto pré ou após a emergência de acordo com técnicas usuais para aplicação de herbicidas a campos compreendendo cultivares que foram vistos por ser resistentes ao herbicida por expressão heteróloga de um gene *grg* da invenção.

20

Métodos para conferir resistência a herbicidas em uma planta ou parte de planta são também providos. Em tais métodos, um polinucleotídeo *grg* aqui descrito é introduzido na planta, caracterizado pelo fato de que a expressão do
25 polinucleotídeo resulta em tolerância ou resistência a glifosato. Plantas produzidas por este método podem ser tratadas com uma concentração efetiva de um herbicida e apresentam uma tolerância aumentada ao herbicida. Uma "concentração efetiva" de um herbicida nesta aplicação é

uma quantidade suficiente para diminuir ou parar o crescimento de plantas ou partes de plantas que não são naturalmente resistentes ou vistas como resistentes ao herbicida.

5

Em outra concretização, métodos para conferir resistência a herbicidas em uma planta ou parte de planta são providos, caracterizado pelo fato de que a planta ou parte de planta é cultivada sob temperaturas maiores ou
10 menores do que a temperatura ambiente como descrito *supra*. Enzimas EPSP sintases tolerantes a glifosato tendo estabilidade térmica em temperaturas maiores ou menores, ou possuem uma temperatura ótima em maiores ou menores temperaturas, são úteis para conferir tolerância a
15 glifosato em plantas que são cultivadas sob tais condições.

Métodos para controlar as ervas daninhas em um campo de cultivo

20

Métodos para se controlar seletivamente as ervas daninhas em um campo contendo uma planta também são providos. Em uma concretização, as sementes da planta ou as plantas são resistentes a glifosato, como resultado de um polinucleotídeo *grg* aqui descrito sendo inserido na semente
25 da planta ou na planta. Em métodos específicos, a planta é tratada com uma concentração efetiva de um herbicida, onde a aplicação de herbicidas resulta em um controle seletivo das ervas daninhas ou outras plantas não transformadas. Por "concentração efetiva" entende-se a concentração que

controla o crescimento ou dispersão das ervas daninhas ou outra planta não transformada sem afetar significativamente as plantas resistentes a herbicidas ou as sementes dessas plantas. Tais concentrações efetivas para os herbicidas de interesse são geralmente conhecidas na arte. O herbicida pode ser aplicado tanto pré- ou pós emergência de acordo com técnicas usuais para a aplicação de herbicida nos campos compreendendo plantas ou sementes de plantas que tenham se tornado resistentes ao herbicida.

Os exemplos a seguir são fornecidos a título de ilustração e não de limitação.

EXPERIMENTAL

Exemplo 1. Isolamento de EPSP sintases glifosato resistentes.

Linhagens capazes de crescer na presença de glifosato foram isoladas por plaqueamento de amostras de solo em meio de sais minerais HEPES "HEPES Mineral Salts Medium" (HMSM) contendo glifosato como a única fonte de fósforo. Desde que HMSM não contém nenhum aminoácido aromático, a linhagem deve ser resistente a glifosato para poder crescer neste meio.

Dois gramas de solo foram suspensos em aproximadamente 10 ml de água, agitados em vortex por 15 segundos e deixados para assentar por 15 minutos. Uma alçada de 10 μ l

desta suspensão foi adicionada para 3 ml de HMSM suplementado com 10 mM de glifosato (pH 7,0). HMSM contém (por litro): 10 g de glicose, 2 g de NH_4SO_4 , 9,53 g de HEPES, 1,0 ml de 0,8 M MgSO_4 , 1,0 ml de 0,1 M CaCl_2 , 1,0 ml de Solução de metais traços (em 100 ml de solução 1000x: 0,1 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0,5 mg de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 1,0 mg de H_3BO_3 , 1,0 mg de $\text{MnSO}_4 \cdot 5\text{H}_2\text{O}$, 7,0 mg de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 1,0 mg de MoO_3 , 4,0 g de KCl). A cultura foi cultivada em uma incubadora com agitação por quatro dias a 28°C e então 20 μl foi usado para inocular 2,5 ml de HMSM fresco contendo 10 mM de glifosato como a única fonte de fósforo. Após dois dias, 20 μl foi usado para inocular outros 2,5 ml de cultura fresca. Após 5 dias, 20 μl foi usado para inocular 2,5 ml de uma cultura fresca. Após crescimento suficiente, a cultura foi plaqueada em meio sólido estriando-se uma alçada de 1 μl sobre a superfície de placa agar contendo HMSM agar contendo 100 mM de glifosato como a única fonte de fósforo e estocado a 28°C. A cultura foi então replaqueada por isolamento. As linhagens listadas na Tabela 1 estavam entre as linhagens selecionadas devido à habilidade destas em crescer na presença de altas concentrações de glifosato.

Tabela 1.

Nome da linhagem	Identidade da linhagem	Nome do gene da EPS sintase
ATX21561	Desconhecido	<i>grg33</i>
ATX21563	Desconhecido	<i>grg35</i>
ATX21567	Desconhecido	<i>grg36</i>

Exemplo 2. Isolamento das EPSP sintases *grg37* e *grg39* resistentes a glifosato.

5 Linhagens capazes de crescer na presença de glifosato foram isoladas por plaqueamento das amostras de solo em vários meios de crescimento contendo glifosato. Algumas linhagens foram isoladas em meio de sais minerais suplementado com glifosato. Outras linhagens foram isoladas
10 sob meio rico na presença de glifosato e depois testadas em meio de sais minerais suplementado com glifosato. Desde que o meio de sais minerais não contenha nenhum aminoácido aromático, a linhagem deve ser resistente a glifosato para poder crescer neste meio.

15

As linhagens ATX21800 e ATX21804 foram isoladas por incubação sob condições ricas e suplementação com glifosato. Estas linhagens foram então testadas para suas habilidades em crescer na presença de glifosato sem
20 aminoácidos aromáticos. A linhagem ATX21804 foi isolada a partir de solo (0,01 gramas) que foi seco ao ar livre por dois dias e plaqueado em agar caldo nutriente suplementado com 100 mM de glifosato. Uma pequena quantidade (10 µl) foi então usada para inocular uma placa de agar eosina azul de
25 metileno contendo 300 mM de glifosato. ATX21800 foi isolada por incubação de 0,01 gramas de solo com 3 ml de caldo nutriente suplementado com 100 mM de glifosato. Após isolamento inicial, cada linhagem foi inoculada em placas Luria Bertani agar para confirmar tipo único de colônia.

Estas linhagens foram então testadas em meio mínimo Brunner contendo 100mM de glifosato e foram confirmadas para crescimento na presença de glifosato sem aminoácidos aromáticos.

5

As linhagens listadas na Tabela 2 estavam entre as linhagens selecionadas devido suas habilidades para crescer na presença de altas concentrações de glifosato.

10

Tabela 2.

Nome da linhagem	Identidade da linhagem	Nome do gene da EPS sintase
ATX21800	Desconhecido	<i>grg37</i>
ATX21804	Desconhecido	<i>grg39</i>

Exemplo 3. Isolamento das EPSP sintases glifosato resistentes *grg38* e *grg50*.

15

As Linhagens capazes de crescer na presença de glifosato foram isoladas por plaqueamento das amostras de solo em vários meios de crescimento contendo glifosato. Algumas linhagens foram isoladas em meio de sais minerais suplementado com glifosato. Outras linhagens foram isoladas sob meio rico na presença de glifosato e depois testadas em meio de sais minerais suplementado com glifosato. Desde que o meio de sais minerais não contenha nenhum aminoácido aromático, a linhagem deve ser resistente a glifosato para poder crescer neste meio.

20

A linhagem ATX20103 foi isolada suspendendo-se aproximadamente 2 gramas de solo em 10 ml de água, agitando-se em vortex por 15 segundos e deixando-se assentar por 15 minutos. Uma alçada de 10 μ l desta
5 suspensão foi adicionada em 3 ml de Tris MSM (TMSM) suplementado com 10 mM de glifosato (pH 7,0). TMSM contém (por litro): 10 g de glicose, 2 g de NH_4SO_4 , 12,12 g de Tris, 1,0 ml de 0,8 M MgSO_4 , 1,0 ml de 0,1 M CaCl_2 , 1,0 ml de solução de metais traços (Em 100 ml de solução 1000x:
10 0,1 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0,5 mg de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 1,0 mg de H_3BO_3 , 1,0 mg de $\text{MnSO}_4 \cdot 5\text{H}_2\text{O}$, 7,0 mg de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 1,0 mg de MoO_3 , 4,0 g de KCl). A cultura foi então incubada a 28°C para isolamento ao longo de repetidas ciclos de seleção e então inoculadas sobre agar Luria Bertani para confirmar tipo
15 único de colônia. ATX20103 foi então reconfirmada para crescimento em TMSM na presença de glifosato sem aminoácidos aromáticos. A linhagem ATX21806 foi isolada através de incubação sob condições ricas e suplementando-se com glifosato. Esta linhagem foi então testada para sua
20 habilidade em crescer na presença de glifosato sem aminoácidos aromáticos.

A linhagem ATX21806 foi isolada de solo (0,01 gramas) que tinha sido suspenso em 10 ml de água durante a noite.
25 Uma pequena quantidade (10 μ l) foi então usada para inocular uma placa de agar eosina azul de metileno contendo 300 mM de glifosato. Após cada isolamento inicial, a linhagem foi inoculada em placas agar Luria Bertani para confirmar tipo único de colônia. A linhagem foi então

testada no meio mínimo Brunner com 100 mM de glifosato e foi confirmada para crescimento na presença de glifosato sem aminoácidos aromáticos.

- 5 As linhagens listadas na Tabela 3 estavam entre as linhagens selecionadas devido suas habilidades para crescer na presença de altas concentrações de glifosato.

Tabela 3.

Nome da linhagem	Identidade da linhagem	Nome do gene da EPS sintase
ATX21806	Desconhecido	<i>grg38</i>
ATX20103	<i>Rhizobium luguminosarum</i>	<i>grg50</i>

10

Exemplo 4. Clonagem de EPSP Sintases resistentes a glifosato

- O DNA genômico foi extraído a partir das linhagens descritas nas Tabelas 1, 2, e 3, e o DNA resultante foi parcialmente digerido com a enzima de restrição *Sau3A* para produzir fragmentos de DNA de aproximadamente 5 quilobases em tamanho. Estas moléculas de DNA foram selecionadas por tamanho em géis de agarose, purificadas, e ligadas no braços do vetor LAMBDA ZAP® pré-digerido com *BamH* I. Os braços ligados foram então empacotados em partículas de fago, e a titulação de fago determinada como conhecido na arte. As bibliotecas resultantes foram amplificadas por métodos conhecidos na arte para gerar uma
- 15
- 20

titulação da biblioteca de entre 3×10^7 e 3×10^8 PFU/mL. Para cada biblioteca independente, *E. coli* (XL1 Blue MRF') foi então co-transfectada com fago de uma biblioteca amplificada assim como com fago M13 ajudante para permitir excisão de massa da biblioteca na forma de um DNA fita simples infectante e circular conforme conhecido na arte (Short et al. (1988) *Nucleic Acids Research* 16: 7583-7600). Após centrifugação das células co-infectadas, o sobrenadante contendo fago foi aquecido para 65-70°C por 15-20 minutos para incapacitar qualquer partícula residual de fago lambda. Diluições da biblioteca de plasmídeo DNA fita simples resultante foram transfectadas para culturas fresca de células de *E. coli* XL-Blue MRF' (aro^Δ) (XL1 Blue MRF') competentes. As células transfectadas resultantes foram plaqueadas em placas M63 contendo canamicina, 0,1 mM de IPTG e tanto 0 mM, 20 mM ou 50 mM de glifosato.

A *E. coli* XL-Blue MRF' (*aroA*) usada para a transfecção expressa o pilus F, e também contem uma deleção do gene *aroA* codificando a enzima EPSP sintase endógena de *E. coli*. Esta linhagem também é referida aqui como ΔaroA . Esta linhagem ΔaroA é incapaz de crescer em meio mínimo sem aminoácidos aromáticos, ao menos que complementada por uma EPSP sintase funcional. Desde que o glifosato é um potente inibidor de EPSP sintases glifosato-sensitivas típicas, como a EPSP sintase tipo I, clones transfectados expressando uma EPSP sintase não resistente a glifosato seriam capazes de crescer em placas M63 sem glifosato, mas seriam incapazes de crescer em M63 contendo tanto 20 mM ou

50 mM de glifosato. Para poder crescer em placas de M63 contendo 20 mM ou 50 mM de glifosato, as células devem conter um plasmídeo que expresse uma EPSP sintase que seja tanto (1) capaz de complementar a mutação AaroA destas células, e (2) resistente a glifosato. Então, este método de rastreamento permite a identificação de clones contendo EPSP sintases resistentes a glifosato.

Colônias crescendo em 20 mM ou 50 mM de glifosato foram escolhidas e seus plasmídeos analisados por digestão com enzimas de restrição para identificação de plasmídeos com perfis de digestão compartilhados. Plasmídeos individuais foram seqüenciados por métodos conhecidos na arte.

15

Usando-se esta abordagem, como às vezes modificada para cada biblioteca como conhecido e apreciado na arte, os clones das bibliotecas contendo genes de EPSP sintase foram identificados para cada uma das linhagens listadas na Tabela 4.

20

Exemplo 5. Seqüências de DNA e proteína de EPSP sintases

As seqüências de DNA das EPSP sintases glifosato-resistente foram determinadas para cada um dos clones descritos acima por métodos bem conhecidos na arte.

25

grg33. A seqüência de DNA de *grg33* é fornecida aqui como SEQ ID NO: 1. O produto previsto da tradução de *grg33*

(GRG33) é fornecido aqui como SEQ ID NO: 2. Uma sequência sintética codificando GRG33 (*syngrg33*) também foi concebida e é fornecida aqui como SEQ ID NO:3.

5 *grg35*. A sequência de DNA de *grg35* é fornecida aqui como SEQ ID NO:4. O produto previsto da tradução de *grg35* (GRG35) é fornecido aqui como SEQ ID NO:5. Uma sequência sintética codificando GRG35 (*syngrg35*) também foi concebida e é fornecida aqui como SEQ ID NO:6.

10

grg36. A sequência de DNA de *grg36* é fornecida aqui como SEQ ID NO:7. O produto previsto da tradução de *grg36* (GRG36) é fornecido aqui como SEQ ID NO: 8. Uma sequência sintética codificando GRG36 (*syngrg36*) também foi concebida
15 e é fornecida aqui como SEQ ID NO:9.

grg37. A sequência de DNA de *grg37* é fornecida aqui como SEQ ID NO: 10. O produto previsto da tradução de *grg37* (GRG37) é fornecido aqui como SEQ ID NO: 11. Uma sequência
) sintética codificando GRG37 (*syngrg37*) também foi concebida
20 e é fornecida aqui como SEQ ID NO: 12.

grg38. A sequência de DNA de *grg38* é fornecida aqui como SEQ ID NO: 15. O produto previsto da tradução de *grg38*
25 (GRG38) é fornecido aqui como SEQ ID NO: 16. Uma sequência sintética codificando GRG38 (*syngrg38*) também foi concebida e é fornecida aqui como SEQ ID NO: 17.

grg39. A sequência de DNA de *grg39* é fornecida aqui como SEQ ID NO: 18. O produto previsto da tradução de *grg39* (GRG39) é fornecido aqui como SEQ ID NO: 19. Uma sequência sintética codificando GRG39 (*syngrg39*) também foi concebida e é fornecida aqui como SEQ ID NO: 20.

grg50. A sequência de DNA de *grg50* é fornecida aqui como SEQ ID NO: 21. O produto previsto da tradução de *grg50* (GRG50) é fornecido aqui como SEQ ID NO: 22. Uma sequência sintética codificando GRG50 (*syngrg50*) também foi concebida e é fornecida aqui como SEQ ID NO: 23.

Clones contendo cada um dos genes das EPSP sintase *grg33*, *grg35*, *grg36*, *grg37*, *grg38*, *grg39*, e *grg50* foram depositados no NRRL em 9 de junho de 2006 ou 26 de junho de 2007 e os números de depósito assinados conforme na Tabela 4.

Tabela 4. Clones contendo EPSP sintases resistentes a glifosato

EPSPS	Linhagens produzindo EPSPS	Originalmente isolada em pBKCMV	Número NRRL
GRG33	ATX21561	pAX1947	B-30932
GRG35	ATX21563	pAX1947	B-30933
GRG36	ATX21567	pAX1948	B-30934
GRG37	ATX21800	pAX1949	B-30945
GRG38	ATX21806	pAX1963	B-30946
GRG39	ATX21804	pAX1964	B-30947
GRG50	ATX20103	pAX1966	B-30948

Cada uma das proteínas GRG33, GRG35, e GRG36 demonstraram regiões de homologia para enzimas EPSP sintase no banco de dados NCBI por pesquisa de BLAST. A enzima
 5 EPSPS com a maior identidade de sequência de proteína para cada enzima GRG é listada na Tabela 5.

Tabela 5. Homologia de GRG33-GRG36 para EPSP sintases conhecidas

Proteína	Linhagens com EPS sintase homólogas	% de identidade
GRG33	GRG35 <i>S. coelicolor</i>	88%, 86%
GRG35	GRG35 de <i>S. coelicolor</i>	88%, 85%
GRG36	<i>Bacillus halodurans</i>	53%

10

As sequências de aminoácido de GRG33 e GRG35 são 88% idênticas. Uma pesquisa de banco de dados públicos de proteína com a sequência de aminoácido de GRG33 demonstrou
 15 que esta proteína é 86% idêntica sobre 430 aminoácidos à EPSP sintase de *Streptomyces coelicolor* (SEQ ID NO: 24 GENBANK® No. de Acesso NP 629359.1), e 82% idêntica sobre 430 aminoácidos à EPSP sintase de *Streptomyces avermitilis* (SEQ ID NO: 25; GENBANK® No. de Acesso NP824218.1).

20

Similarmente, a sequência de aminoácido de GRG35 é 85% idêntica sobre 434 aminoácidos a EPSP sintase de *Streptomyces coelicolor* (SEQ ID NO: 24; GENBANK® No. de Acesso NP 629359.1), e 81% idêntica sobre 441 aminoácidos a

EPSP sintase de *Streptomyces avermitilis* (SEQ ID NO: 25; GENBANK® No. de Acesso NP 824218.1).

Tabela 6. Identidade de aminoácido de GRG33 e GRG35 com

5 EPSP sintases de *Streptomyces*

EPS sintase	Identidade com GRG33	Identidade com GRG35
GRG33	-	88%
GRG35	88%	-
<i>Streptomyces</i> <i>coelicolor</i> A3(2)	84%	82%
<i>Streptomyces</i> <i>avermitilis</i> MA- 4680	80%	78%
<i>E. coli</i>	30%	29%
Milho	30%	29%

Uma pesquisa de banco de dados públicos de proteína com a sequência de aminoácido de GRG36 demonstra que esta proteína é relacionada a EPSP sintase de *Bacillus*
 10 *halodurans*, (64% idêntica sobre 441 aminoácidos, SEQ ID NO: 26; GENBANK® No. de Acesso BAB06432.1), e numa menor extensão para a EPSP sintase de *Bacillus clausii* (55% idêntica sobre 439 aminoácidos; SEQ ID NO: 27; GENBANK® No. de Acesso BAD63759.1)

Tabela 7. Identidade de aminoácido de GRG36 com EPSP sintases

EPSP sintases	Identidade com GRG36
<i>Bacillus halodurans</i>	62%
<i>Bacillus clausii</i>	54%
<i>E.coli</i>	31%
Milho	34%

Uma pesquisa de banco de dados de proteína públicos com a sequência de aminoácido de GRG37 demonstra que esta proteína é 81% idêntica a EPSP sintase de *Arthrobacter* sp. FB24 (SEQ ID NO: 28, GENBANK® No. de Acesso ZP 00413033.1)

O quadro de iniciação de leitura *grg37* tem dois potenciais códons de iniciação. O ATG acima (a sequência de aminoácido predita MTASPMGASADNS.....(correspondente as posições de aminoácido 1 até 13 da SEQ ID NO: 10)) contém o melhor sítio de ligação ao ribossomo na proximidade correta. Entretanto, um segundo ATG abaixo pode ser usado. Esta ORF produz a sequência aminoácido prevista MGASADNS (correspondente as posições de aminoácido 6 até 13 da SEQ ID NO: 10)). O ATG acima parece ter um sítio de ligação ao ribossomo ("RBS") que é mais próximo a sequência consenso do RBS. Entretanto, o quadro de leitura aberto originado a partir deste ATG acima parece ser acoplado por tradução a um quadro de leitura aberto anterior. O pareamento de tradução é uma estratégia conhecida na arte para ser empregada por bactérias para garantir a boa inicialização e poder substituir um sítio de ligação ao ribossomo. A sequência de nucleotídeo para o sítio de iniciação abaixo é

fornecida aqui como SEQ ID NO: 13, e a sequência de aminoácido codificada são fornecidas aqui como SEQ ID NO: 14.

5 GRG39 demonstra 96% de identidade de aminoácido a sequência da GRG30 EPSP sintase, e é altamente homóloga a sequência da GRG29 EPSP sintase descrita na Pedido de Patente Norte-Americano No. 11/760.570 preenchida em 8 de junho de 2007.

10

 GRG38 demonstra 94% de identidade de aminoácido a GRG 12 EPSP sintase descrita na Pedido de Patente Norte-Americano No. 11/400.598, preenchida em 7 de abril de 2006 (Tabela 8). GRG38 também contem os domínios das EPSP 15 sintases Classe III descritas na Pedido de Patente Norte-Americano No. 11/400.598.

 GRG50 demonstra 95% de identidade de aminoácido a GRG8 EPSP sintase descrita na Pedido de Patente Norte-Americano 20 No. 11/315.678 preenchida em 22 de dezembro de 2005 (Tabela 8). GRG50 também contem os domínios das EPSP sintases Classe III descritas na U.S. Application 11/400.598, preenchida em 7 de abril de 2006.

Tabela 8. Comparação com outras EPSP sintases Classe III

EPSPS	Identidade de aminoácido com GRG38	Identidade de aminoácido com GRG50
GRG38	-	65%
GRG50	65%	-
GRG8	65%	95%
GRG12	87%	62%
GRG6	67%	67%
GRG9	64%	70%
GRG15	64%	71%
GRG5	68%	68%
GRG37	67%	68%
E.coli	32%	34%
Milho	32%	31%

Exemplo 6. Clonagem de uma nova EPSP sintase glifosato-resistente em um vetor de expressão de *E. coli*

5

Os genes da EPSP sintase contidos nos clones da Tabela 4 foram sub-clonados no vetor de expressão de *E. coli* pRSF1b (Invitrogen). Os clones resultantes foram confirmados por sequenciamento de DNA, e usados para expressão induzida de cada EPSP sintase em *E. coli*. As proteínas com etiqueta His expressas foram então purificadas como conhecido na arte.

10

Exemplo 7. Resistência a Glifosato das EPSP sintases GRG33.

15

GRG35. e GRG36

Os clones em pRSF1b foram plaqueados em placas M63+ contendo antibióticos e tanto 0mM ou 50mM de glifosato. O crescimento recebeu uma pontuação após dois dias de crescimento a 37°C. Cada uma das três EPSP sintases foi observada para conferir resistência a 50 mM de glifosato em células de *E. coli* (Tabela 9).

Tabela 9. Rastreamento de Glifosato

EPSPS	Clone em pRSF1B	Crescimento em 50mM de glifosato
vetor	-	-
GRG33	pAX1951	+++
GRG35	pAX1952	+++
GRG36	pAX1953	+++

10 Exemplo 8. Resistência a Glifosato das EPSP sintases GRG37 e GRG39

Células contendo os clones plasmidiais demonstrados na Tabela 4 foram plaqueados em placas M63+ contendo antibiótico e tanto 0mM ou 20mM de glifosato. O crescimento foi pontuado após dois dias de crescimento a 37°C. Cada uma das EPSP sintases foi observada para conferir resistência a 20 mM de glifosato em células de *E. coli* (Tabela 10).

Tabela 10. Rastreamento de glifosato

EPSPS	Clone em pRSF1B	Crescimento em 50mM de glifosato
Vetor	-	-
GRG37	pAX1963	++
GRG39	pAX1965	++

Exemplo 9. Resistência a Glifosato das EPSP sintases GRG38 e GRG50

5

As células contendo os clones de plasmídeo mostrados na Tabela 4 foram plaqueadas em placas M63+ contendo antibiótico e tanto 0mM ou 20mM de glifosato. O crescimento foi pontuado após dois dias de crescimento a 37°C. Cada uma das EPSP sintases foi observada para conferir resistência a 20 mM de glifosato em células de *E. coli* (Tabela 11).

Tabela 11. Rastreamento de glifosato

EPSPS	Clone em pRSF1B	Crescimento em 50mM de glifosato
Vetor	-	-
GRG38	pAX1964	++
GRG50	pAX1966	++

15 Exemplo 10. Construindo *grg33*, *grg35*, *grg36*, *grg37*, *grg38*, *grg39*, *grg50*, *svngrg33*, *svngrg35*, *svngrg36*, *svngrg37*, *svngrg38*, *svngrg39*, e *svngrg50* para Transformação em Planta

O quadro de leitura aberto (ORF) para cada um dos genes *grg* é amplificado por PCR a partir do molde de cDNA

20

íntegro. Sítios de restrição *Hind* III são adicionados para cada final da orF durante o PCR. Adicionalmente, a sequência de nucleotídeo ACC é adicionada imediatamente a 5' do códon de início do gene para aumentar a eficiência de tradução (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 arte, para garantir que nenhuma mutação seja introduzida durante o PCR. O plasmídeo contendo o *grg* produto de PCR é digerido com, por exemplo, *Hind* III e o fragmento contendo o orF intacto é isolado.

Pode-se talvez se gerar construções similares que contenham uma sequência de cloroplasto alvo ligada a um polinucleotídeo da invenção por métodos conhecidos na arte.

Um fragmento de DNA contendo a EPSP sintase (e tanto contendo ou não contendo uma sequência de cloroplasto alvo) é clonado dentro de um plasmídeo, por exemplo no sítio *Hind* III de pAX200. pAX200 é um vetor de expressão em plantas contendo o promotor de actina de 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 (ou o fragmento promotor-líder-gene-terminador) deste plasmídeo intermediário é subclonado dentro de um plasmídeo como um pSB1 1 (Japan Tobacco, Inc.) para formar um plasmídeo final, referido aqui como, por exemplo, pSB1 1GRG33. pSB1 1GRG33 é organizado de forma que o fragmento de DNA contendo, por exemplo, a construção

promotor-*grg36*-terminador (ou a construção promotor-líder-*grg36*-terminador) pode ser retirado por enzimas de restrição apropriadas e também usados para transformação em plantas, por exemplo, por injeção de feixe de aerossol. A
5 estrutura do pSB1 1GRG33 é verificada por digestão com enzima de restrição e gel eletroforese, assim como por seqüenciamento ao longo das várias junções das clonagens. O mesmo método pode ser usado para gerar um plasmídeo final para cada um dos genes *grg* descritos aqui.

10

O plasmídeo é mobilizado dentro da linhagem de *Agrobacterium tumefaciens* LBA4404 que também carrega o plasmídeo pSB1 (Japan Tobacco, Inc.), usando procedimentos de conjugação triparental bem conhecidos na arte, e
15 plaqueamento em meio contendo antibiótico. O plasmídeo pSB1 1GRG36 carrega resistência a espectinomicina, mas é um plasmídeo de estreita faixa de hospedeiros e não pode se replicar em *Agrobacterium*. Colônias resistentes a antibiótico surgem quando pSB1 1GRG36 se integra dentro do
20 plasmídeo de vasta faixa de hospedeiros pSB1 através de recombinação homóloga. O produto co-integrado resultante é verificado por hibridação por *Southern*. A linhagem de *Agrobacterium* carregando o co-integrado pode ser usado para transformar milho, por exemplo, através do método de
25 *PureIntro* (Tabaco do Japão).

Exemplo 11. Transformação de *grg33*, *grg35*, *grg36*, *grg37*, *grg38*, *grg39*, *grg50*, *svngrg33*, *svngrg35*, *svngrg36*,

svnrg37, svnrg38, svnrg39, e svnrg50 em Células
vegetais

Espigas de milho são melhor colhidas entre 8-12 dias
5 após polinização. Os embriões são isolados a partir das
espigas, e os embriões de 0,8-1,5 mm em tamanho são
preferidos para uso na transformação. Os embriões são
plaqueados com lado do escutelo virado para cima em um meio
de incubação adequado, como o meio DN62A5S (3,98 g/L Sais
10 N6; 1 mL/L (de Estoque 1000x) 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 1 mg/mL) 2,4-D). Entretanto, outros meios e sais
além de DN62A5S são adequados e são conhecidos na arte. Os
15 embriões são incubados durante a noite a 25°C no escuro.
Entretanto, não é necessário por si se incubar os embriões
durante a noite.

Os explantes resultantes são transferidos para malha
20 quadrada (30-40 por placa), transferidos sobre meio
osmótico por cerca de 30-45 minutos, e então transferidos
para uma placa de bombardeamento (ver, por exemplo,
Publicação de Pedido PCT No. WO/0138514 e Patente Norte-
Americana No. 5,240,842).

25

Construções de DNA desenhadas para expressar as
proteínas GRG da presente invenção em células vegetais são
aceleradas para dentro do tecido vegetal usando um
acelerador de feixe de aerossol, usando as condições

essencialmente como descrito na Publicação de Pedido PCT No. WO/0138514. Após bombardeamento, os embriões são incubados por cerca de 30 min em meio osmótico, e plaqueados sobre meio de incubação durante a noite a 25°C no escuro. Para evitar explantes indevidamente danificados pelo bombardeamento, eles são incubados por ao menos 24 horas antes da transferência para o meio de recuperação. Os embriões são então espalhados sobre o meio de período de recuperação, por cerca de 5 dias, a 25°C no escuro, e então transferidos para o meio de seleção. Os explantes são incubados em meio de seleção por até oito semanas, dependendo da natureza e características da seleção particularmente utilizada. Após o período de seleção, o calo resultante é transferido para o meio de maturação do embrião, até a formação de embriões somáticos maduros ser observada. Os embriões somáticos maduros resultantes são então plaqueados sob luz baixa, e o processo de regeneração é iniciado por métodos conhecidos na arte. Os brotos resultantes são deixados para enraizar no meio de enraizamento, e as plantas resultantes são transferidas para os potes de viveiro e propagadas como plantas transgênicas.

Materiais

Tabela 12. Meio DN62A5S

Componentes	Por litro	Fonte
Mistura de Sais Basal N6 de Chu (Prod. No. C416)	3,98g/L	Phytotechnology Labs
Solução de vitaminas N6 de Chu (Prod. No. C149)	1mL/L (de estoque 1000X)	Phytotechnology Labs
L-Asparagina	800mg/L	Phytotechnology Labs
N-inositol	100mg/L	Sigma
L-Prolina	1,4g/L	Phytotechnology Labs
Casaminoácidos	100mg/L	Fisher Scientific
Sacarose	50g/L	Phytotechnology Labs
2,4D (Prod. No.D7299)	1mL/L(de estoque 1mg/L)	Sigma

O pH da solução é ajustado para pH 5,8 com 1N de KOH/1N KCl, Gelrite (Sigma) é adicionado na concentração de até 3 g/L, e o meio é autoclavado. Após resfriamento para 50°C, 2 ml/L de uma solução estoque de 5 mg/ml de nitrato de prata (Phytotechnology Labs) é adicionada.

Exemplo 12. A transformação de *grg33*, *grg35*, *grg36*, *grg37*, *grg38*, *grg39*, *grg50*, *svnrg33*, *svnrg35*, *svnrg36*, *svnrg37*, *svnrg38*, *svnrg39*, e *svnrg50* para Células vegetais de Milho por Transformação mediada por *Agrobacterium*

Espigas são melhor colhidas entre 8-12 dias após polinização. Os embriões são isolados a partir de espigas, e os embriões de 0.8-1.5 mm em tamanho são preferidos para uso na transformação. Os embriões são plaqueados com lado do escutelo virado para cima em um meio de incubação adequado, e incubados durante a noite a 25°C no escuro. Entretanto, não é necessário por si se incubar os embriões durante a noite. Os embriões são postos em contato com uma linhagem de *Agrobacterium* contendo os vetores apropriados para a transferência mediada por plasmídeo Ti por cerca de 5-10 min, e então plaqueados em meio de co-cultivação por cerca de 3 dias (25°C no escuro). Após 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 no meio de seleção por até oito semanas, dependendo da natureza e características da seleção particularmente utilizada. Após o período de seleção, o calo resultante é transferido para o meio de maturação de embrião, até a formação de embriões somáticos maduros ser observada. Os embriões somáticos maduros resultantes são então plaqueados sob luz baixa, e então o processo de regeneração é inicializado como conhecido na arte. Os brotos resultantes são deixados para enraizar em meio de enraizamento, e as plantas resultantes são transferidas para potes de viveiro e propagadas como plantas transgênicas.

Todas as publicações e aplicações de patentes mencionadas na especificação são indicativas do nível de

especialização daqueles especialistas na arte a qual esta invenção pertence. Todas as publicações e aplicações de patentes são aqui incorporadas por referência na mesma extensão como se cada publicação ou aplicação de patente
5 individual fosse especificamente e individualmente indicada para serem incorporadas por referência.

Apesar de a invenção precedente ter sido descrita em algum detalhe por meio de ilustração e exemplo para
10 propósitos de clareza de entendimento, será óbvio que certas mudanças e modificações podem ser praticadas dentro do escopo das reivindicações em anexo.

LISTAGEM DE SEQUÊNCIAS

<110> Depositante: Athenix Corporation
 Cheryl Peters
 Brian Vande Berg
 Brian Carr
 Daniel John Tomso

<120> Título da Invenção: GRG33, GRG35 GRG36, GRG37, GRG38, GRG39,
 e GRG50: Novos genes de EPSP sintase
 conferindo resistência a herbicida

<130> Referência do documento: 45600/329203

<150> Número do Documento de Prioridade: 60/816.676
 <151> Data de Depósito do Documento de Prioridade: 27-06-2006

<150> Número do Documento de Prioridade: 60/819.122
 <151> Data de Depósito do Documento de Prioridade: 07-07-2006

<150> Número do Documento de Prioridade: 60/819.119
 <151> Data de Depósito do Documento de Prioridade: 07-07-2006

<160> Quantidade de SEQ ID NOs.: 41

<170> Software: FastSEQ for Windows Version 4.0

<210> SEQ ID NO: 1
 <211> Comprimento: 1329
 <212> Tipo: DNA
 <213> Organismo: Desconhecido

<220> Características:
 <223> Outras Informações: Isolado de amostra de solo

<221> Nome/Chave: CDS
 <222> Localização: (1)...(1329)

<400> Sequência: 1

atg acc gac agc aac ccg tcc acc gac ccg ctc acc ggc cgc tgg ccc	48
Met Thr Asp Ser Asn Pro Ser Thr Asp Pro Leu Thr Gly Arg Trp Pro	
1 5 10 15	
gcc ccg tac gcg gcc ggc gcc gtc gac gcc acc gtc acc gtg ccc gga	96
Ala Pro Tyr Ala Ala Gly Ala Val Asp Ala Thr Val Thr Val Pro Gly	
20 25 30	
tcg aag tcg gtc acc aac cgc gcc ctg gtg ctc gcc gcg ctg gcc gcc	144
Ser Lys Ser Val Thr Asn Arg Ala Leu Val Leu Ala Ala Leu Ala Ala	
35 40 45	
gag ccg ggc tgg gtg cgc cgc ccg ctg cgc tcg cgc gac acc ctg ctg	192
Glu Pro Gly Trp Val Arg Arg Pro Leu Arg Ser Arg Asp Thr Leu Leu	
50 55 60	
atg gcg gag gcg ctg cgc acc ctg ggc gtg aag atc gac gag ggc gtg	240
Met Ala Glu Ala Leu Arg Thr Leu Gly Val Lys Ile Asp Glu Gly Val	
65 70 75 80	

ggc ccg gac ggc acc ggc gag gcc tgg cgg atc atc ccg gcc ggg ctg Gly Pro Asp Gly Thr Gly Glu Ala Trp Arg Ile Ile Pro Ala Gly Leu	288
85 90 95	
cgc ggc ccg gcg acc gtc gac gtc ggc aac gcg ggc acg gtc atg cgc Arg Gly Pro Ala Thr Val Asp Val Gly Asn Ala Gly Thr Val Met Arg	336
100 105 110	
ttc ctg ccg ccg gtg gcg gcg ctc gcg aac ggc gcg gtg cgc ttc gac Phe Leu Pro Pro Val Ala Ala Leu Ala Asn Gly Ala Val Arg Phe Asp	384
115 120 125	
ggc gac ccg cgc tcc cac gag cgc ccg ctg cac ggg gtg atc gac gcg Gly Asp Pro Arg Ser His Glu Arg Pro Leu His Gly Val Ile Asp Ala	432
130 135 140	
ctg cgc gcg ctg ggc gcc cgg atc gac gac gac ggg cgc ggc gcg ctg Leu Arg Ala Leu Gly Ala Arg Ile Asp Asp Asp Gly Arg Gly Ala Leu	480
145 150 155 160	
ccg atg acc gtg cac ggc gcc ggc ggc ctg gag ggc ggg gtc gtg gag Pro Met Thr Val His Gly Ala Gly Gly Leu Glu Gly Gly Val Val Glu	528
165 170 175	
atc gac gcc tcc tcg tcc tcc cag ttc gtc agc gcg ctg ctg ctc tcc Ile Asp Ala Ser Ser Ser Ser Gln Phe Val Ser Ala Leu Leu Leu Ser	576
180 185 190	
ggc gcc cgc ttc aac cag ggc gtg gag gtg cgg cac gtc ggc acc cgg Gly Ala Arg Phe Asn Gln Gly Val Glu Val Arg His Val Gly Thr Arg	624
195 200 205	
ctg ccc tcg ctg ccg cac atc cgg atg acg gtc gac atg ctg cgc gcg Leu Pro Ser Leu Pro His Ile Arg Met Thr Val Asp Met Leu Arg Ala	672
210 215 220	
gtc ggc gcc cag gtc gac gag ccg gag cac ggc ggg cgt ccc gac gtg Val Gly Ala Gln Val Asp Glu Pro Glu His Gly Gly Arg Pro Asp Val	720
225 230 235 240	
tgg cgg gtc acc ccg tcc gcg ctg ctc ggc cgg gac ctg gtg gtg gag Trp Arg Val Thr Pro Ser Ala Leu Leu Gly Arg Asp Leu Val Val Glu	768
245 250 255	
ccg gac ctg tcg aac gcc cag ccg ttc ctg gcg gcg gcg ctg gtc acc Pro Asp Leu Ser Asn Ala Gln Pro Phe Leu Ala Ala Ala Leu Val Thr	816
260 265 270	
ggc ggc cgg gtc acc gtg ccg gac tgg ccg gcc agg acc acc cag ccc Gly Gly Arg Val Thr Val Pro Asp Trp Pro Ala Arg Thr Thr Gln Pro	864
275 280 285	
ggt gac gcg ctg cgg cag atc ttc acc gag atg ggt ggc tcc tgc gag Gly Asp Ala Leu Arg Gln Ile Phe Thr Glu Met Gly Gly Ser Cys Glu	912
290 295 300	
ctc acc gac cgg ggt ctc acc ttc acc gga acc ggc cgg atc cac ggc Leu Thr Asp Arg Gly Leu Thr Phe Thr Gly Thr Gly Arg Ile His Gly	960
305 310 315 320	

atc gac gtc gac ctc ggc gag gtc ggc gag ctg acc ccg ggc atc gcg Ile Asp Val Asp Leu Gly Glu Val Gly Glu Leu Thr Pro Gly Ile Ala 325 330 335	1008
gcg gtc gcc gcg ctc gcc gac tcc ccg tcc acc ctg cgc ggg gtg gcg Ala Val Ala Ala Leu Ala Asp Ser Pro Ser Thr Leu Arg Gly Val Ala 340 345 350	1056
cac ctg cgg ctg cac gag acc gac cgg ctc gcc gcg ctc acc cgg gag His Leu Arg Leu His Glu Thr Asp Arg Leu Ala Ala Leu Thr Arg Glu 355 360 365	1104
atc aac gcg ctg ggc ggc gac gtc acg gag acc gag gac ggc ctg cac Ile Asn Ala Leu Gly Gly Asp Val Thr Glu Thr Glu Asp Gly Leu His 370 375 380	1152
atc cgc ccg cgc ccg ctg cac ggc ggc ctc ttc cac acg tac cac gac Ile Arg Pro Arg Pro Leu His Gly Gly Leu Phe His Thr Tyr His Asp 385 390 395 400	1200
cac cgg atg gcg acc gcg ggc gcg ctc atc ggc ctg gcc gtg aag ggc His Arg Met Ala Thr Ala Gly Ala Leu Ile Gly Leu Ala Val Lys Gly 405 410 415	1248
gtg gag atc gag aac gtg aag acg acc gag aag acc ttg ccc gac ttc Val Glu Ile Glu Asn Val Lys Thr Thr Glu Lys Thr Leu Pro Asp Phe 420 425 430	1296
ccc agg atg tgg acc gaa atg ctc gga gtc tga Pro Arg Met Trp Thr Glu Met Leu Gly Val *	1329
435 440	

<210> SEQ ID NO: 2

<211> Comprimento: 442

<212> Tipo: PRT

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: Isolado de amostra de solo

<400> Sequência: 2

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

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

<210> SEQ ID NO: 3

<211> Comprimento: 1326

<212> Tipo: DNA

<213> Organismo: Sequência Artificial

<220> Características:

<223> Outras Informações: synrg33

<400> Sequência: 3

atg aca gac agc aac cca agc acc gac ccg ctc acc ggc cgc tgg ccg	48
gcg ccc tac gcc gcc ggc gcc gtc gac gcc acc gtc acc gtg cct gga	96
agc aag agc gtc acc aac agg gcg ctg gtg ctg gcg gcg ctg gct gct	144
gaa cct gga tgg gtg cgg cgg ccg ctg agg agc agg gac acc tta ttg	192
atg gcg gag gcg ctg agg acg ctc ggc gtc aag att gat gaa gga gtt	240
gga cct gat gga act gga gaa gca tgg agg atc atc ccc gcc ggc ctc	288
cgc ggc ccg gcg acg gtg gat gtt ggc aac gcc ggc acc gtg atg agg	336
ttc ctg ccg ccg gtg gcg gcg ctg gcc aac ggc gcc gtc cgc ttc gac	384
ggc gac cca aga agt cat gaa agg cct cta cat ggc gtc atc gac gcg	432
ctc cgc gcg ctg gga gca agg atc gac gac ggc cgc ggc gcg ctg	480
cca atg aca gtt cat ggc gcc ggc ggc ctg gag ggc ggc gtg gtg gag	528

```

att gat gca agc agc agc agc cag ttc gtc tcg gcg ctg ctg ctg agc 576
ggc gcg cgc ttc aac caa gga gtg gag gtg cgg cat gtt gga aca agg 624
ctg cca tca ttg ccg cac atc agg atg acg gtg gac atg ctg cgc gcc 672
gtc ggc gct caa gtt gat gag ccg gag cat gga gga agg cca gat gtt 720
tgg agg gtg acg ccg tcg gcg ctg ctg gga aga gat ctg gtg gtg gag 768
cca gat ctc tca aat gct caa ccc ttc ctg gcg gcg gcg ctg gtg acc 816
ggc ggc cgc gtc acc gtg cca gat tgg ccg gca agg acg acg cag ccc 864
ggc gac gcg ctg cgg cag atc ttc acc gag atg gga gga tca tgt gag 912
ctc acc gac cgc ggc ctc acc ttc acc ggc act gga agg att cat ggc 960
att gat gtg gac ctc ggc gag gtg gga gag ctg acg ccg ggc atc gcc 1008
gcc gtg gcg gcg ctc gcc gac tcg ccg tcg acg ctc cgc ggc gtg gct 1056
cac ctc cgc cta cat gaa aca gat cgg ctg gcg gcg ctg aca agg gag 1104
atc aac gcg ctc ggc ggc gac gtc acc gag aca gaa gat ggc ctc cac 1152
atc agg ccg cgg ccg cta cat gga ggc ctc ttc cac acc tac cat gat 1200
cac agg atg gcc acc gcc ggc gcg ctc atc ggc ctc gcc gtc aag ggc 1248
gtg gag att gaa aat gtg aag aca aca gag aag acg ctg ccg gac ttc 1296
cca agg atg tgg acg gag atg ctg ggc gtc 1326

```

<210> SEQ ID NO: 4

<211> Comprimento: 1353

<212> Tipo: DNA

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<221> Nome/Chave: CDS

<222> Localização: (1)...(1353)

<400> Sequência: 4

```

atg acc gca gct tcc ccc gcg cac ccc gat ccg tcg cag ccc gcc gtc 48
Met Thr Ala Ala Ser Pro Ala His Pro Asp Pro Ser Gln Pro Ala Val
1 5 10 15

ccc gcc ggc ccg gac ctc tgg ccc gcc ccg tac gcg agc ggc ccc gtc 96
Pro Ala Gly Pro Asp Leu Trp Pro Ala Pro Tyr Ala Ser Gly Pro Val
20 25 30

gac gcg acc gtg acc gtg ccc ggc tcg aag tcg gtc acc aac cgc gcc 144
Asp Ala Thr Val Thr Val Pro Gly Ser Lys Ser Val Thr Asn Arg Ala
35 40 45

ctg gtg ctc gcc gcg ctc gcc gcc gag ccc ggc tgg gtg cgg cgc ccg 192
Leu Val Leu Ala Ala Leu Ala Ala Glu Pro Gly Trp Val Arg Arg Pro
50 55 60

ctc cgc tcc cgc gac acc ctc ctg atg gcc gag ggc ctg cgc acc ctg 240
Leu Arg Ser Arg Asp Thr Leu Leu Met Ala Glu Gly Leu Arg Thr Leu
65 70 75 80

ggc gtg aag atc gag gag ggc gtg ggc ccg gac ggc acc ggc gag gcc 288
Gly Val Lys Ile Glu Glu Gly Val Gly Pro Asp Gly Thr Gly Glu Ala
85 90 95

tgg cgg atc atc ccg gcc ccg ctg cgc ggc ggc gcc acg gtc gac gtc 336
Trp Arg Ile Ile Pro Ala Pro Leu Arg Gly Gly Ala Thr Val Asp Val
100 105 110

ggc aac gcg ggc acg gtc atg cgc ttc ctg ccg ccg gtc gcc gcg ctc 384
Gly Asn Ala Gly Thr Val Met Arg Phe Leu Pro Pro Val Ala Ala Leu
115 120 125

```

gcc gac ggc ccg gtc cgc ttc gac ggc gac ccc cgc tcc cac gag cgc	432
Ala Asp Gly Pro Val Arg Phe Asp Gly Asp Pro Arg Ser His Glu Arg	
130 135 140	
ccg ctg cac ggg gtg atc gac gcg ctg cgc gcg ctc ggc gcc cgg atc	480
Pro Leu His Gly Val Ile Asp Ala Leu Arg Ala Leu Gly Ala Arg Ile	
145 150 155 160	
gac gac gag ggc cgc ggc gcc ctc ccg atg acc gtg cac ggc gcc ggc	528
Asp Asp Glu Gly Arg Gly Ala Leu Pro Met Thr Val His Gly Ala Gly	
165 170 175	
ggc ctg gag ggc ggc acg gtc gag atc gac gcc tcc tcg tcc tcc cag	576
Gly Leu Glu Gly Gly Thr Val Glu Ile Asp Ala Ser Ser Ser Ser Gln	
180 185 190	
ttc gtc agc gcc ctg ctg ctc tcc ggc gcc cgc ttc aac cag ggc gtc	624
Phe Val Ser Ala Leu Leu Leu Ser Gly Ala Arg Phe Asn Gln Gly Val	
195 200 205	
gag gtc cgg cac gcg ggc ggc cgg ctg ccg tcg atg ccg cac atc cgg	672
Glu Val Arg His Ala Gly Gly Arg Leu Pro Ser Met Pro His Ile Arg	
210 215 220	
atg acc gtg gac atg ctc cgc gcg gtg ggc gcg aag gtc gac gag ccg	720
Met Thr Val Asp Met Leu Arg Ala Val Gly Ala Lys Val Asp Glu Pro	
225 230 235 240	
gag cac ggc ggc gag ccg gac gtg tgg cgg gtg gcc ccg tcg gcg ctg	768
Glu His Gly Gly Glu Pro Asp Val Trp Arg Val Ala Pro Ser Ala Leu	
245 250 255	
cgc ggc cgc gac ctg gtc atc gag ccc gac ctg tcg aac gcc cag ccg	816
Arg Gly Arg Asp Leu Val Ile Glu Pro Asp Leu Ser Asn Ala Gln Pro	
260 265 270	
ttc ctg gcg gcc gcg ctg gtc acc ggc ggc cgg gtg acc gtc ccg gac	864
Phe Leu Ala Ala Ala Leu Val Thr Gly Gly Arg Val Thr Val Pro Asp	
275 280 285	
tgg ccg gcc cgc acc acc cag ccc ggc gac gag ctg cgc cgc atc ttc	912
Trp Pro Ala Arg Thr Thr Gln Pro Gly Asp Glu Leu Arg Arg Ile Phe	
290 295 300	
acc gag atg ggc ggc gcc tgc gag ctc acc gac gcc ggt ctc acc ttc	960
Thr Glu Met Gly Gly Ala Cys Glu Leu Thr Asp Ala Gly Leu Thr Phe	
305 310 315 320	
acc ggc acc ggc cgg atc cac ggc atc gac gtg gac ctc ggc gag gtc	1008
Thr Gly Thr Gly Arg Ile His Gly Ile Asp Val Asp Leu Gly Glu Val	
325 330 335	
ggc gag ctg acc ccc ggc atc gcg gcg gtc gcc gcc ctc gcc gac tcc	1056
Gly Glu Leu Thr Pro Gly Ile Ala Ala Val Ala Ala Leu Ala Asp Ser	
340 345 350	
ccc tcc acc ctg cgc ggc gtg gcc cac ctg cgg ctc cac gag acc gac	1104
Pro Ser Thr Leu Arg Gly Val Ala His Leu Arg Leu His Glu Thr Asp	
355 360 365	

```

cgg ctc gcg gcg ctc acc cgg gag atc aac ggt ctg ggc ggc gac gtc 1152
Arg Leu Ala Ala Leu Thr Arg Glu Ile Asn Gly Leu Gly Gly Asp Val
370 375 380

acc gag acc gag gac ggc ctg cac atc cgg ccg cgc ccg ctg acc ggc 1200
Thr Glu Thr Glu Asp Gly Leu His Ile Arg Pro Arg Pro Leu Thr Gly
385 390 395 400

ggc gtc ttc cac acg tac cac gac cac cgg atg gcg acc gcg ggc gcg 1248
Gly Val Phe His Thr Tyr His Asp His Arg Met Ala Thr Ala Gly Ala
405 410 415

atc atc ggc ctc gcc gtg aag ggc gtg gag atc gag gac gtg gcg acg 1296
Ile Ile Gly Leu Ala Val Lys Gly Val Glu Ile Glu Asp Val Ala Thr
420 425 430

acc gcc aag acc ttg ccg gac ttc ccg ggg atg tgg acc gaa atg ctc 1344
Thr Ala Lys Thr Leu Pro Asp Phe Pro Gly Met Trp Thr Glu Met Leu
435 440 445

gga gcc tga 1353
Gly Ala *
450

```

<210> SEQ ID NO: 5

<211> Comprimento: 450

<212> Tipo: PRT

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<400> Sequência: 5

```

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

```

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

<210> SEQ ID NO: 6

<211> Comprimento: 1326

<212> Tipo: DNA

<213> Organismo: Sequência Artificial

<220> Características:

<223> Outras Informações: synrg35

<400> Sequência: 6

gttgatccaa gccagccggc ggtgccggcg gggccggacc tctggccggc gccatatgct 60
 tctgggcccgg tggacgccac cgtcacctg cctggaagca agagcgtcac caacagggcg 120
 ctggtgctgg cggcgctggc tgctgaacct ggatgggtgc ggccggcgtc gaggagcagg 180
 gacaccttgc tgatggcaga agggctaagg acgctcggcg tcaagatcga ggagggcgctc 240
 ggccctgatg gaactggaga agcatggagg atcatcccgg cgcgctccg cggcggcgcc 300
 accgtcgacg tcggcaacgc cggcacctg atgaggttcc tgccgcccgtt ggcggcgctc 360
 gccgacgggc cggtagatt tgatggagat ccaagaagtc atgaaaggcc tctacatggc 420
 gtcacgcagc cgtcccgcc gctcggcgcc aggtatgatg atgaaggccg cggcgcgctg 480
 ccaatgacag ttcattggcg cggcggcctg gaaggaggaa cagtggagat tgatgcaagc 540
 agcagcagcc agttcgtctc ggcgctgctg ctgagcggcg cgcgcttcaa ccaaggagtg 600
 gaggtgcggc acgccggcgg ccggctgcc tcaatgcctc acatcaggat gacgggtggac 660
 atgctgcgcg ccgtcggcgc caaggtggat gagccggagc atggaggaga acctgatgtt 720
 tggaggggtg cgcgctcggc gctccgcgcc cgcgacctg tgattgagcc agatctctca 780
 aatgctcaac ccttcctggc ggccggcgctg gtgaccggcg gccgcgtcac cgtgccagat 840
 tggccggcaa ggacgacga gcccgcgac gagctaagga ggatcttcac cgagatgggc 900
 ggcgcctgcg agctcaccga cgcgggctc accttcacc gcaactggaag gattcatggc 960
 attgatgtgg acctcggcga ggtgggagag ctgacgcggc gcatcgccgc cgtggcgcg 1020
 ctcgccgact cgcgctcgac gctccgcgcc gtggctcacc tccgcctaca tgaaacagat 1080

```

cggctggcgg cgctgacaag ggagatcaat ggcctcggcg gcgacgtcac cgagacagaa 1140
gatggactac acatcaggcc gcggccgctc accggcggcg tcttccacac ctaccatgat 1200
cacaggatgg ccaccgccgg cgccatcatc ggcctcgccg tgaagggcgt ggagattgaa 1260
gatgtggcca ccaccgcca gacgctgccg gacttccctg ggatgtggac ggagatgctg 1320
ggcgcc

```

<210> SEQ ID NO: 7

<211> Comprimento: 1344

<212> Tipo: DNA

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<221> Nome/Chave: CDS

<222> Localização: (1)...(1344)

<400> Sequência: 7

```

atg aaa aat caa aat ttt gat gct aaa gcc cgt agc ccg tgg aca ccg      48
Met Lys Asn Gln Asn Phe Asp Ala Lys Ala Arg Ser Pro Trp Thr Pro
  1              5              10              15

tta aaa ggt gta aat aag ata agt gtt tca cct agt aaa gga aga ata      96
Leu Lys Gly Val Asn Lys Ile Ser Val Ser Pro Ser Lys Gly Arg Ile
      20              25              30

aat gga acc gtt act att cct ggt agc aag agt tta acc aac aga gct      144
Asn Gly Thr Val Thr Ile Pro Gly Ser Lys Ser Leu Thr Asn Arg Ala
      35              40              45

tta atc ata agt tct cta gct agt gga aag tca aaa gtg caa ggt att      192
Leu Ile Ile Ser Ser Leu Ala Ser Gly Lys Ser Lys Val Gln Gly Ile
      50              55              60

tta aaa agc gat gat tca ttt tgg tgt tta gac tcc ttg aaa aag cta      240
Leu Lys Ser Asp Asp Ser Phe Trp Cys Leu Asp Ser Leu Lys Lys Leu
      65              70              75              80

gga gta aat gtg aaa att caa gga gat aca gct ttt atc gaa gga aac      288
Gly Val Asn Val Lys Ile Gln Gly Asp Thr Ala Phe Ile Glu Gly Asn
      85              90              95

ggt ggt aaa tgg gaa tca ggt gat tta tat att ggt gca gca gga acg      336
Gly Gly Lys Trp Glu Ser Gly Asp Leu Tyr Ile Gly Ala Ala Gly Thr
      100             105             110

att gca cga ttt cta cct gga gca tta gca gtt tca ggt aca ggc ata      384
Ile Ala Arg Phe Leu Pro Gly Ala Leu Ala Val Ser Gly Thr Gly Ile
      115             120             125

tgg gag tta gaa gcc agt aaa agt atg agt aaa cga cct att tca ccc      432
Trp Glu Leu Glu Ala Ser Lys Ser Met Ser Lys Arg Pro Ile Ser Pro
      130             135             140

tta gta gat gct tta aaa gag ctt ggg gct gaa ata aca tat cta agc      480
Leu Val Asp Ala Leu Lys Glu Leu Gly Ala Glu Ile Thr Tyr Leu Ser
      145             150             155             160

gat caa ggc tac tat ccg ttg tta gtt aaa gga aaa caa cta aat ggg      528
Asp Gln Gly Tyr Tyr Pro Leu Leu Val Lys Gly Lys Gln Leu Asn Gly
      165             170             175

```

ggc gag gtt gaa ctc tca ggt aga att tct agt cag ttt ata agt ggt Gly Glu Val Glu Leu Ser Gly Arg Ile Ser Ser Gln Phe Ile Ser Gly 180 185 190	576
cta ttg att gcc tcg cct tat tta aat gat cca atc aag att aat att Leu Leu Ile Ala Ser Pro Tyr Leu Asn Asp Pro Ile Lys Ile Asn Ile 195 200 205	624
aaa gat cac atc gtt caa cac tca tat gta ctc tta act ttg gaa tta Lys Asp His Ile Val Gln His Ser Tyr Val Leu Leu Thr Leu Glu Leu 210 215 220	672
atg aaa aag ttt ggt gca aaa gtt aaa tac gat agt agc cta aaa gaa Met Lys Lys Phe Gly Ala Lys Val Lys Tyr Asp Ser Ser Leu Lys Glu 225 230 235 240	720
ata gtc gtc tat cca tct aag tac act cca caa gat ata aat tta gaa Ile Val Val Tyr Pro Ser Lys Tyr Thr Pro Gln Asp Ile Asn Leu Glu 245 250 255	768
gca gat gtt tcc act gca tgt tat ttt ctg gct ctt gct gca gtg acc Ala Asp Val Ser Thr Ala Cys Tyr Phe Leu Ala Leu Ala Val Thr 260 265 270	816
aac ggt aaa gta caa att gat aat cta act tat gaa aca aaa caa cca Asn Gly Lys Val Gln Ile Asp Asn Leu Thr Tyr Glu Thr Lys Gln Pro 275 280 285	864
gat ata aaa atg gtt gat atc ctt gaa cgg atg gga tgc aaa gta aca Asp Ile Lys Met Val Asp Ile Leu Glu Arg Met Gly Cys Lys Val Thr 290 295 300	912
aga ggt tct tca ttc att gaa ata gag gga gtt agt caa tta aaa ggt Arg Gly Ser Ser Phe Ile Glu Ile Glu Gly Val Ser Gln Leu Lys Gly 305 310 315 320	960
gga ttt gaa atc tct atg agg gaa atg tct gac caa gtg tta act cta Gly Phe Glu Ile Ser Met Arg Glu Met Ser Asp Gln Val Leu Thr Leu 325 330 335	1008
gca gca att gct cca ttc gca gat gaa cca ata acc ata aaa gac gtt Ala Ala Ile Ala Pro Phe Ala Asp Glu Pro Ile Thr Ile Lys Asp Val 340 345 350	1056
gaa cat ata cgc cat cac gaa tca aat cga atc agt gta cta gtt gat Glu His Ile Arg His His Glu Ser Asn Arg Ile Ser Val Leu Val Asp 355 360 365	1104
tca cta tct agg tta gga att ata gta gaa gaa ttt aaa gat gga cta Ser Leu Ser Arg Leu Gly Ile Ile Val Glu Glu Phe Lys Asp Gly Leu 370 375 380	1152
aaa gtg tat ccg ggt aat ccg aaa gcc act tta cta gat aca cac gat Lys Val Tyr Pro Gly Asn Pro Lys Ala Thr Leu Leu Asp Thr His Asp 385 390 395 400	1200
gat cat aga gtt gca atg gca tta tca ctt ata ggt tca aga gtt gaa Asp His Arg Val Ala Met Ala Leu Ser Leu Ile Gly Ser Arg Val Glu 405 410 415	1248

ggt ata caa ata aat gat cca gga tgt gta tct aaa act tgt cct cag 1296
 Gly Ile Gln Ile Asn Asp Pro Gly Cys Val Ser Lys Thr Cys Pro Gln
 420 425 430

tat ttt gaa tta ttg gaa aaa cta ggt ttg aat ata att aaa cat tga 1344
 Tyr Phe Glu Leu Leu Glu Lys Leu Gly Leu Asn Ile Ile Lys His *
 435 440 445

<210> SEQ ID NO: 8

<211> Comprimento: 447

<212> Tipo: PRT

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<400> Sequência: 8

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

Ala	Ala	Ile	Ala	Pro	Phe	Ala	Asp	Glu	Pro	Ile	Thr	Ile	Lys	Asp	Val
			340					345					350		
Glu	His	Ile	Arg	His	His	Glu	Ser	Asn	Arg	Ile	Ser	Val	Leu	Val	Asp
		355					360					365			
Ser	Leu	Ser	Arg	Leu	Gly	Ile	Ile	Val	Glu	Glu	Phe	Lys	Asp	Gly	Leu
		370				375					380				
Lys	Val	Tyr	Pro	Gly	Asn	Pro	Lys	Ala	Thr	Leu	Leu	Asp	Thr	His	Asp
385					390					395					400
Asp	His	Arg	Val	Ala	Met	Ala	Leu	Ser	Leu	Ile	Gly	Ser	Arg	Val	Glu
				405					410					415	
Gly	Ile	Gln	Ile	Asn	Asp	Pro	Gly	Cys	Val	Ser	Lys	Thr	Cys	Pro	Gln
			420					425					430		
Tyr	Phe	Glu	Leu	Leu	Glu	Lys	Leu	Gly	Leu	Asn	Ile	Ile	Lys	His	
		435					440					445			

```
<210> SEQ ID NO: 9
<211> Comprimento: 1344
<212> Tipo: DNA
<213> Organismo: Sequência Artificial
```

```
<220> Características:
<223> Outras Informações: syngrg36
```

<400> Sequência: 9						
atgaagaacc	agaacttcga	cgccaaggca	agaagcccct	ggacgccgct	caagggcgctc	60
aacaagatct	ccgtctcgcc	aagcaaagga	aggatcaatg	gcaccgtcac	catccctgga	120
agcaagagct	tgacaaatag	agctctcatc	atcagcagct	tagcaagcgg	caagagcaag	180
gttcaaggca	tcctcaagag	cgacgacagc	ttctggtgcc	tggattcatt	gaagaagctc	240
ggcgctcaatg	tgaagatcca	aggagacacc	gccttcattg	aaggaaatgg	agggaaatgg	300
gagagtggag	atctctacat	cggcgccgcc	ggcaccattg	caagatttct	tcctggcgcg	360
ctggcggtga	gcggcaccgg	catctgggag	ctggaggcaa	gcaagagcat	gagcaagagg	420
cccattctcac	cgctggtgga	tgtcttcaag	gagctcgggc	ccgagatcac	ctacctctct	480
gatcaaggct	actaccgcgt	gctggtgaag	ggcaagcagc	tcaatggagg	agaggtggag	540
ctctctggaa	ggatcagcag	ccagttcatc	agcggcctgc	tgattgcttc	accatatttg	600
aatgatccca	tcaagatcaa	catcaaggac	cacatcgctg	agcacagcta	tgtgctgctg	660
acgctggagc	tgatgaagaa	gttcggcgcc	aaggatgaat	acgacagcag	cttgaaggag	720
atcgctgctc	acccaagcaa	gtacacgcgc	caggacatca	acctggaagc	tgatgtttca	780
acagcatgct	acttctctgc	gctggcgggc	gtgacaaatg	gcaagggtgca	aattgacaac	840
ctcacctacg	agaccaagca	gccggacatc	aagattggtg	acatcctgga	gaggatgggc	900
tgcaagggtga	caagaggaag	cagcttcatc	gagattgaag	gagtgaagcca	gctgaagggc	960
ggcttcgaga	tctcaatgag	ggagatgagc	gaccaggtgc	tgacgctggc	ggccatcgcg	1020
ccatttgctg	atgagcccat	caccatcaag	gatgtggagc	acatccgcca	tcatgaaagc	1080
aacaggatct	ccgtgctggt	ggacagcctc	tcaaggctgg	gcatcatcgt	ggaggagttc	1140
aaggatggcc	tcaaggtgta	ccctggaaat	ccaaaggcga	cgctgctgga	cacctatgat	1200
gatcaccgcg	tggcaattgc	gctgagcttg	attggatcaa	gggtggaagg	catccagatc	1260
aatgatcctg	gctgcgtcag	caagacctgc	cctcaatatt	ttgagctgct	ggagaagctg	1320
ggcctcaaca	tcataagca	ctag				1344

```
<210> SEQ ID NO: 10
<211> Comprimento: 1407
<212> Tipo: DNA
<213> Organismo: Desconhecido
```

<220> Características:
<223> Outras Informações: isolado do solo

```
<221> Nome/Chave: CDS
<222> Localização: (1)...(1407)
```

<400> Sequência: 10

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

gac tcc acg ccc aac cac tgg att gtg ggc ccc ggg ccg atc cgc gcc Asp Ser Thr Pro Asn His Trp Ile Val Gly Pro Gly Pro Ile Arg Ala 245 250 255	768
ttc gac cag cgg atc gag cag gac ctg tcc aac gcc ggc ccg ttc ctg Phe Asp Gln Arg Ile Glu Gln Asp Leu Ser Asn Ala Gly Pro Phe Leu 260 265 270	816
gcg gcc gcg ctg gca acc ggc ggc acg gtg cgg att ccg gac tgg ccg Ala Ala Ala Leu Ala Thr Gly Gly Thr Val Arg Ile Pro Asp Trp Pro 275 280 285	864
gag cac acc acg cag gtg ggt gac atg tgg cgc agt atc ctc agt gac Glu His Thr Thr Gln Val Gly Asp Met Trp Arg Ser Ile Leu Ser Asp 290 295 300	912
atg ggg gca acc gtc acc ctg cag gac ggc acc ctg acc gtg acc ggc Met Gly Ala Thr Val Thr Leu Gln Asp Gly Thr Leu Thr Val Thr Gly 305 310 315 320	960
ggc agc aaa atc aac ggc gcg gac ttc gcc gag acc agc gaa ctg gcg Gly Ser Lys Ile Asn Gly Ala Asp Phe Ala Glu Thr Ser Glu Leu Ala 325 330 335	1008
ccc acc gtc gct gcg ctc tgc gcg ctg gca tcc ggt ccc tcg cgg ctg Pro Thr Val Ala Ala Leu Cys Ala Leu Ala Ser Gly Pro Ser Arg Leu 340 345 350	1056
acc ggc atc gcc cac ctc cgg gga cat gag acg gac cgg ctc gcc gcg Thr Gly Ile Ala His Leu Arg Gly His Glu Thr Asp Arg Leu Ala Ala 355 360 365	1104
ctc gtc acc gaa atc aac cgc ctc ggc ggc gat gcc gag gaa acc gcc Leu Val Thr Glu Ile Asn Arg Leu Gly Gly Asp Ala Glu Glu Thr Ala 370 375 380	1152
gac gga ctc gtc atc cgc ccc gcg gaa ctg cac gcg ggg gtg gtg cac Asp Gly Leu Val Ile Arg Pro Ala Glu Leu His Ala Gly Val Val His 385 390 395 400	1200
agc tac gcg gac cac cgg atg gcc aca gcc gga gcc atc ctc ggc ctg Ser Tyr Ala Asp His Arg Met Ala Thr Ala Gly Ala Ile Leu Gly Leu 405 410 415	1248
gcg gtc agg ggc gtg gaa gtg cag gac atc gcc acc aca tcc aag acc Ala Val Arg Gly Val Glu Val Gln Asp Ile Ala Thr Thr Ser Lys Thr 420 425 430	1296
atg ccg gac ttc ccc aag ctg tgg gcg gac atg ctc ggc cgg cag acc Met Pro Asp Phe Pro Lys Leu Trp Ala Asp Met Leu Gly Arg Gln Thr 435 440 445	1344
gat gca gct aca aac aca gcg ggt acc aac acc gcg ggg acc gcc ggt Asp Ala Ala Thr Asn Thr Ala Gly Thr Asn Thr Ala Gly Thr Ala Gly 450 455 460	1392
ggc acg cag cac tga Gly Thr Gln His *	1407
465	

<210> SEQ ID NO: 11
 <211> Comprimento: 468
 <212> Tipo: PRT
 <213> Organismo: Desconhecido

<220> Características:
 <223> Outras Informações: isolado do solo

<400> Sequência: 11

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

Ala Val Arg Gly Val Glu Val Gln Asp Ile Ala Thr Thr Ser Lys Thr
 420 425 430
 Met Pro Asp Phe Pro Lys Leu Trp Ala Asp Met Leu Gly Arg Gln Thr
 435 440 445
 Asp Ala Ala Thr Asn Thr Ala Gly Thr Asn Thr Ala Gly Thr Ala Gly
 450 455 460
 Gly Thr Gln His
 465

<210> SEQ ID NO: 12
 <211> Comprimento: 1404
 <212> Tipo: DNA
 <213> Organismo: Sequência Artificial

<220> Características:
 <223> Outras Informações: syngrg37

<400> Sequência: 12
 atgacagctt ctccaatggg cgcctccgcc gacaacagcg gcgcccgcgc gcactggccc 60
 gcgccccttcg ccgagcagcc ggtggacgcc accgtccgcg tgccctggaag caagagcctc 120
 accaacagat atttggtgct ggcggcgctg gctgatgggc catcaaggct ccgcgcgcgc 180
 ctgcattcaa gagattcggc gctgatgatc gccgcgctcc gtcagctcgg cgccgacatc 240
 agggaggtgc ccggcgacgg cgccttcggc cctgatcttg aggtgacgcc catcccggcc 300
 aacgccggcg gcgcccagct cgccatcgac tgcggcctcg ccggcaccgt gatgaggttc 360
 gtgcgcgcgc tggcggcgct gagatcagc accaccgtct ttgatggaga tcctcatgca 420
 agagaaaggc caatgggcac catcatcgag gcgctcgccg gcctcggcgt caccgtcagc 480
 ggcgaggacg gcggcacgcc ggcgtcgctg cccttcgcgc tggaaggaaac tgggtcaagt 540
 cgcgccggcc acctgggtgat tgatgcttca gcaagcagcc agttcgtctc ggcgctgctg 600
 ctggtgggag caagggttcac cgacggcctc caccgtggag acagcggcag caccgtgcca 660
 agcctggatc acatcaacat gaccatcgcc accctccgcg gcgcccggcg cgccgtggat 720
 gattcaacgc ccaaccatg gatcgtcggc ccgcgccccca tcagggcatt tgatcaaaga 780
 attgagcaag atctttcaaa tgctggaccc ttcttgccgc cggcgctggc caccggcgcc 840
 accgtgagga ttccagattg gccggagcac accacccaag ttggtgacat gtggaggagc 900
 atcctctccg acatgggcgc caccgtcacc cttcaagatg gcaccttgac cgtcaccggc 960
 ggcagcaaga tcaacggcgc cgacttcgcc gagacatcag agctggcgcc gacggtggcg 1020
 gcgctctgcg cgctggcttc tgggccaagc cgcctcaccg gcattgctca cctccgcggc 1080
 catgaaacag atcggctggc ggcgctggtg acagagatca acaggctcgg cggcgacgcc 1140
 gaggagaccg ccgacggcct ggtgatcagg ccggcggagc tacatgctgg agtgggtgcac 1200
 agctatgctg atcacaggat ggccaccgcc ggcccatcc tcggcctcgc cgtcagagga 1260
 gtggaggtgc aagatatagc aacaacaagc aagaccatgc cggacttccc caagctctgg 1320
 gccgacatgc tgggcccgcc aacagatgct gccaccaaca ccgccggcac caacaccgcc 1380
 ggcaccgcgc gcggcaccca gcac 1404

<210> SEQ ID NO: 13
 <211> Comprimento: 1392
 <212> Tipo: DNA
 <213> Organismo: Desconhecido

<220> Características:
 <223> Outras Informações: isolado do solo

<221> Nome/Chave: CDS
 <222> Localização: (1)...(1392)

<400> Sequência: 13
 atg ggc gcg tcc gct gac aac tcc ggc gct gct ccg cac tgg ccc gcg 48
 Met Gly Ala Ser Ala Asp Asn Ser Gly Ala Ala Pro His Trp Pro Ala
 1 5 10 15

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

gag cag gac ctg tcc aac gcc ggc ccg ttc ctg gcg gcc gcg ctg gca	816
Glu Gln Asp Leu Ser Asn Ala Gly Pro Phe Leu Ala Ala Ala Leu Ala	
260 265 270	
acc ggc ggc acg gtg cgg att ccg gac tgg ccg gag cac acc acg cag	864
Thr Gly Gly Thr Val Arg Ile Pro Asp Trp Pro Glu His Thr Thr Gln	
275 280 285	
gtg ggt gac atg tgg cgc agt atc ctc agt gac atg ggg gca acc gtc	912
Val Gly Asp Met Trp Arg Ser Ile Leu Ser Asp Met Gly Ala Thr Val	
290 295 300	
acc ctg cag gac ggc acc ctg acc gtg acc ggc ggc agc aaa atc aac	960
Thr Leu Gln Asp Gly Thr Leu Thr Val Thr Gly Gly Ser Lys Ile Asn	
305 310 315 320	
ggc gcg gac ttc gcc gag acc agc gaa ctg gcg ccc acc gtc gct gcg	1008
Gly Ala Asp Phe Ala Glu Thr Ser Glu Leu Ala Pro Thr Val Ala Ala	
325 330 335	
ctc tgc gcg ctg gca tcc ggt ccc tcg cgg ctg acc ggc atc gcc cac	1056
Leu Cys Ala Leu Ala Ser Gly Pro Ser Arg Leu Thr Gly Ile Ala His	
340 345 350	
ctc cgg gga cat gag acg gac cgg ctc gcc gcg ctc gtc acc gaa atc	1104
Leu Arg Gly His Glu Thr Asp Arg Leu Ala Ala Leu Val Thr Glu Ile	
355 360 365	
aac cgc ctc ggc ggc gat gcc gag gaa acc gcc gac gga ctc gtc atc	1152
Asn Arg Leu Gly Gly Asp Ala Glu Glu Thr Ala Asp Gly Leu Val Ile	
370 375 380	
cgc ccc gcg gaa ctg cac gcg ggg gtg gtg cac agc tac gcg gac cac	1200
Arg Pro Ala Glu Leu His Ala Gly Val Val His Ser Tyr Ala Asp His	
385 390 395 400	
cgg atg gcc aca gcc gga gcc atc ctc ggc ctg gcg gtc agg ggc gtg	1248
Arg Met Ala Thr Ala Gly Ala Ile Leu Gly Leu Ala Val Arg Gly Val	
405 410 415	
gaa gtg cag gac atc gcc acc aca tcc aag acc atg ccg gac ttc ccc	1296
Glu Val Gln Asp Ile Ala Thr Thr Ser Lys Thr Met Pro Asp Phe Pro	
420 425 430	
aag ctg tgg gcg gac atg ctc ggc cgg cag acc gat gca gct aca aac	1344
Lys Leu Trp Ala Asp Met Leu Gly Arg Gln Thr Asp Ala Ala Thr Asn	
435 440 445	
aca gcg ggt acc aac acc gcg ggg acc gcc ggt ggc acg cag cac tga	1392
Thr Ala Gly Thr Asn Thr Ala Gly Thr Ala Gly Thr Gln His *	
450 455 460	

<210> SEQ ID NO: 14

<211> Comprimento: 463

<212> Tipo: PRT

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<400> Sequência: 14

Met	Gly	Ala	Ser	Ala	Asp	Asn	Ser	Gly	Ala	Ala	Pro	His	Trp	Pro	Ala
1				5					10					15	
Pro	Phe	Ala	Glu	Gln	Pro	Val	Asp	Ala	Thr	Val	Arg	Val	Pro	Gly	Ser
			20					25					30		
Lys	Ser	Leu	Thr	Asn	Arg	Tyr	Leu	Val	Leu	Ala	Ala	Leu	Ala	Asp	Gly
		35					40					45			
Pro	Ser	Arg	Leu	Arg	Ala	Pro	Leu	His	Ser	Arg	Asp	Ser	Ala	Leu	Met
	50					55				60					
Ile	Ala	Ala	Leu	Arg	Gln	Leu	Gly	Ala	Asp	Ile	Arg	Glu	Val	Pro	Gly
65					70				75					80	
Asp	Gly	Ala	Phe	Gly	Pro	Asp	Leu	Glu	Val	Thr	Pro	Ile	Pro	Ala	Asn
			85					90						95	
Ala	Gly	Gly	Ala	Asp	Val	Ala	Ile	Asp	Cys	Gly	Leu	Ala	Gly	Thr	Val
			100					105					110		
Met	Arg	Phe	Val	Pro	Pro	Leu	Ala	Ala	Leu	Arg	Ser	Gly	Thr	Thr	Val
		115					120					125			
Phe	Asp	Gly	Asp	Pro	His	Ala	Arg	Glu	Arg	Pro	Met	Gly	Thr	Ile	Ile
	130					135					140				
Glu	Ala	Leu	Ala	Gly	Leu	Gly	Val	Thr	Val	Ser	Gly	Glu	Asp	Gly	Gly
145					150					155				160	
Thr	Pro	Ala	Ser	Leu	Pro	Phe	Arg	Val	Glu	Gly	Thr	Gly	Gln	Val	Arg
				165					170					175	
Gly	Gly	His	Leu	Val	Ile	Asp	Ala	Ser	Ala	Ser	Ser	Gln	Phe	Val	Ser
			180					185					190		
Ala	Leu	Leu	Leu	Val	Gly	Ala	Arg	Phe	Thr	Asp	Gly	Leu	His	Leu	Glu
		195					200				205				
His	Ser	Gly	Ser	Thr	Val	Pro	Ser	Leu	Asp	His	Ile	Asn	Met	Thr	Ile
	210					215					220				
Ala	Thr	Leu	Arg	Gly	Ala	Gly	Val	Ala	Val	Asp	Asp	Ser	Thr	Pro	Asn
225					230					235					240
His	Trp	Ile	Val	Gly	Pro	Gly	Pro	Ile	Arg	Ala	Phe	Asp	Gln	Arg	Ile
				245					250					255	
Glu	Gln	Asp	Leu	Ser	Asn	Ala	Gly	Pro	Phe	Leu	Ala	Ala	Ala	Leu	Ala
			260					265					270		
Thr	Gly	Gly	Thr	Val	Arg	Ile	Pro	Asp	Trp	Pro	Glu	His	Thr	Thr	Gln
		275					280					285			
Val	Gly	Asp	Met	Trp	Arg	Ser	Ile	Leu	Ser	Asp	Met	Gly	Ala	Thr	Val
	290					295					300				
Thr	Leu	Gln	Asp	Gly	Thr	Leu	Thr	Val	Thr	Gly	Gly	Ser	Lys	Ile	Asn
305					310					315				320	
Gly	Ala	Asp	Phe	Ala	Glu	Thr	Ser	Glu	Leu	Ala	Pro	Thr	Val	Ala	Ala
			325						330					335	
Leu	Cys	Ala	Leu	Ala	Ser	Gly	Pro	Ser	Arg	Leu	Thr	Gly	Ile	Ala	His
			340					345					350		
Leu	Arg	Gly	His	Glu	Thr	Asp	Arg	Leu	Ala	Ala	Leu	Val	Thr	Glu	Ile
		355					360					365			
Asn	Arg	Leu	Gly	Gly	Asp	Ala	Glu	Glu	Thr	Ala	Asp	Gly	Leu	Val	Ile
	370					375					380				
Arg	Pro	Ala	Glu	Leu	His	Ala	Gly	Val	Val	His	Ser	Tyr	Ala	Asp	His
385					390					395				400	
Arg	Met	Ala	Thr	Ala	Gly	Ala	Ile	Leu	Gly	Leu	Ala	Val	Arg	Gly	Val
				405					410					415	
Glu	Val	Gln	Asp	Ile	Ala	Thr	Thr	Ser	Lys	Thr	Met	Pro	Asp	Phe	Pro
			420					425					430		
Lys	Leu	Trp	Ala	Asp	Met	Leu	Gly	Arg	Gln	Thr	Asp	Ala	Ala	Thr	Asn
		435					440					445			
Thr	Ala	Gly	Thr	Asn	Thr	Ala	Gly	Thr	Ala	Gly	Gly	Thr	Gln	His	
	450					455					460				

<210> SEQ ID NO: 15
 <211> Comprimento: 1239
 <212> Tipo: DNA
 <213> Organismo: Desconhecido

 <220> Características:
 <223> Outras Informações: isolado do solo

 <221> Nome/Chave: CDS
 <222> Localização: (1)...(1239)

<400> Sequência: 15

gtg acc gtt aca ccg ccc aat ttc ccc ctc aat ggc aag gtc gcg ccc	48
Val Thr Val Thr Pro Pro Asn Phe Pro Leu Asn Gly Lys Val Ala Pro	
1 5 10 15	
ccc ggc tcc aaa tcc att acc aac cgc gcc ctg ttg ctg gcg gcc ctg	96
Pro Gly Ser Lys Ser Ile Thr Asn Arg Ala Leu Leu Leu Ala Ala Leu	
20 25 30	
gcc aag ggc acc agc cgt ctc agc ggc gcg ctc aag agc gac gac acc	144
Ala Lys Gly Thr Ser Arg Leu Ser Gly Ala Leu Lys Ser Asp Asp Thr	
35 40 45	
cgc cat atg tcg gtg gcc ctg cgc caa atg ggc gtg acc atc gat 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 ttc gtt gtc acc ggc aac ggc aaa ctg cac ctg	240
Pro Asp Asp Thr Thr Phe Val Val Thr Gly Asn Gly Lys Leu His Leu	
65 70 75 80	
ccg tcg caa ccg ctg ttc ctc ggc aat gcc ggt act gcc atg cgc ttt	288
Pro Ser Gln Pro Leu Phe Leu Gly Asn Ala Gly Thr Ala Met Arg Phe	
85 90 95	
ctc acg gct gcg gtg gcc acg gtc gaa ggc acc gtg gtg ctg acc ggc	336
Leu Thr Ala Ala Val Ala Thr Val Glu Gly Thr Val Val Leu Thr Gly	
100 105 110	
gac gac tac atg caa aaa cgc ccg atc ggc ccg ttg ctg gcg acc ctc	384
Asp Asp 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 ccg acc ggt tgc ccg ccg gtc	432
Gly Gln Asn Gly Ile Gln Val Asp Ser Pro Thr Gly Cys Pro Pro Val	
130 135 140	
acc gtg cat ggc gtg ggc aag atc aag gcc aag cgc ttc gag atc gac	480
Thr Val His Gly Val Gly Lys Ile Lys Ala Lys Arg Phe Glu Ile Asp	
145 150 155 160	
ggc ggc ctg tcc agc cag tac gtg tcg gcg ctg ctg atg ctc gcg gcc	528
Gly Gly Leu Ser Ser Gln Tyr Val Ser Ala Leu Leu Met Leu Ala Ala	
165 170 175	
tgc ggc gaa gcg ccg att gaa gtg gcg ctg acc ggc aag gac atc ggt	576
Cys Gly Glu Ala Pro Ile Glu Val Ala Leu Thr Gly Lys Asp Ile Gly	
180 185 190	

gcc cgt ggc tat gtc gac ctg acc ctg gac tgc atg cgc gcc ttc ggt Ala Arg Gly Tyr Val Asp Leu Thr Leu Asp Cys Met Arg Ala Phe Gly 195 200 205	624
gcc cag gtt gaa gcc gtt gac gac acc acc tgg tgc gtg gcc ccg acc Ala Gln Val Glu Ala Val Asp Asp Thr Thr Trp Cys Val Ala Pro Thr 210 215 220	672
ggc tac atc gcc cac gac tac ctg atc gaa ccc gac gcc tcc gcc gcc Gly Tyr Ile Ala His Asp Tyr Leu Ile Glu Pro Asp Ala Ser Ala Ala 225 230 235 240	720
act tac ctg tgg gcc gct gaa gtg ttg acc ggg ggg cgc atc gac atc Thr Tyr Leu Trp Ala Ala Glu Val Leu Thr Gly Gly Arg Ile Asp Ile 245 250 255	768
ggc gtg gcc gcg cag gat ttc acc cag ccg gac gcc aag gcc cag gcc Gly Val Ala Ala Gln Asp Phe Thr Gln Pro Asp Ala Lys Ala Gln Ala 260 265 270	816
gtg atc gcg cag ttc ccg cat atg caa gcc acg gtg gtc ggc tcg cag Val Ile Ala Gln Phe Pro His Met Gln Ala Thr Val Val Gly Ser Gln 275 280 285	864
atg cag gac gcc atc ccg acc ctg gcg gtg ctc gcg gct ttc aac aac Met Gln Asp Ala Ile Pro Thr Leu Ala Val Leu Ala Ala Phe Asn Asn 290 295 300	912
acg ccg gtg cgt ttc acc gag ttg gcc aac ctg ccg gtc aag gag tgc Thr Pro Val Arg Phe Thr Glu Leu Ala Asn Leu Arg Val Lys Glu Cys 305 310 315 320	960
gac cgt gtg cag gca ctg cat gac ggc ctc aat gca atc cgc ccg ggc Asp Arg Val Gln Ala Leu His Asp Gly Leu Asn Ala Ile Arg Pro Gly 325 330 335	1008
ctg gcg acc atc gaa ggc gac gac ctg ctg gta gcc agc gac ccc gca Leu Ala Thr Ile Glu Gly Asp Asp Leu Leu Val Ala Ser Asp Pro Ala 340 345 350	1056
ttg gcc ggc acc gcg tgc acc gcg ctg atc gac acc cac gcc gac cac Leu Ala Gly Thr Ala Cys Thr Ala Leu Ile Asp Thr His Ala Asp His 355 360 365	1104
cgc atc gcc atg tgc ttt gcc ctg gcc ggg ctg aag gtt tcg ggt att Arg Ile Ala Met Cys Phe Ala Leu Ala Gly Leu Lys Val Ser Gly Ile 370 375 380	1152
cgc atc cag gac ccg gat tgc gtg gca aag acc tat ccc gag tac tgg Arg Ile Gln Asp Pro Asp Cys Val Ala Lys Thr Tyr Pro Glu Tyr Trp 385 390 395 400	1200
aag gcc ctg ggc agt ctc ggg gtg cag ctg agc tat taa Lys Ala Leu Gly Ser Leu Gly Val Gln Leu Ser Tyr *	1239
405 410	

<210> SEQ ID NO: 16

<211> Comprimento: 412

<212> Tipo: PRT

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<400> Sequência: 16

Val	Thr	Val	Thr	Pro	Pro	Asn	Phe	Pro	Leu	Asn	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	Gly	Asn	Gly	Lys	Leu	His	Leu
65					70					75				80	
Pro	Ser	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	Glu	Gly	Thr	Val	Val	Leu	Thr	Gly
			100					105					110		
Asp	Asp	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	Val	Gly	Lys	Ile	Lys	Ala	Lys	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	Glu	Ala	Val	Asp	Asp	Thr	Thr	Trp	Cys	Val	Ala	Pro	Thr
	210					215					220				
Gly	Tyr	Ile	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	His	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	Ala	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	Glu	Tyr	Trp
385					390					395					400
Lys	Ala	Leu	Gly	Ser	Leu	Gly	Val	Gln	Leu	Ser	Tyr				
				405					410						

<210> SEQ ID NO: 17

<211> Comprimento: 1236

<212> Tipo: DNA

<213> Organismo: Sequência Artificial

<220> Características:

<223> Outras Informações: syngrg38

<400> Sequência: 17

```

atgacggtga cgccgcaaaa cttcccgctc aatggcaagg tggcgccgcc gggcagcaag 60
agcatcacca acagggcgct gctgctggcg gcgctggcca agggcacctc aaggctgagc 120
ggcgcgctga agagcgacga caccgcgcac atgagcgctg cgctgcggca gatgggcgtc 180
accattgatg aacctgatga caccaccttc gtcgtcaccg gcaatggcaa gctgcatctt 240
ccatcacagc cgctcttcct cggcaacgcc ggcaccgcca tgaggttctt gacggcgggc 300
gtggccaccg tggaaggaa acgtggtgctc actggagatg actacatgca aaagaggcca 360
attggacctc tactggcgac gctggggccaa aatggcatcc aggtggactc gccgacgggc 420
tgcccgcggg tgacagttca tggcgtcggc aagatcaagg ccaagagatt tgagatcgac 480
ggcggcctca gcagccaata tgtttcagcg ctgctgatgc tggcggcctg cggcgaggcg 540
cccatcgagg tggcgctcac cggcaaggac atcggcgcgc gcggctatgt ggacctcacc 600
ttggactgca tgagggcctt cggcgctcaa gtggaggcgg tggatgacac cacttggtgc 660
gtggcgccga cgggctacat tgctcatgac tacctcatcg agccagatgc ctccgcgcgc 720
acctacctat gggcggcgga ggtgctcacc ggcggcagga tcgacatcgg cgtcgccgct 780
caagatttca cccaacctga tgccaaggct caagctgtga ttgctcaatt tctcaccatg 840
caagcaacag tggtgggcag ccagatgcaa gatgccatcc cgacgctggc ggtgctggcg 900
gccttcaaca acacgccggg gaggttcacc gagctggcca acctacgagt gaaggaatgt 960
gacaggggtc aagctcttca tgatggcctc aacgccatca ggccagggtt gccaccatt 1020
gaaggagatg atctgctggt ggcttcagat ccggcgctcg ccggcaccgc ctgcacggcg 1080
ctcatcgaca ccacgccga ccaccgcac gccatgtgct tcgcgctcgc cggcctcaag 1140
gtgagcggca tcaggattca agatccagat tgtgtggcca agacctacc gccagtactg 1200
aaggcgctgg gcagcctcgg cgtgcagctg agctac 1236

```

<210> SEQ ID NO: 18

<211> Comprimento: 1296

<212> Tipo: DNA

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<221> Nome/Chave: CDS

<222> Localização: (1)...(1296)

<400> Sequência: 18

```

atg gac gtt atc gtt aaa cca acc cca tcc ctg aac ggg gaa att gga 48
Met Asp Val Ile Val Lys Pro Thr Pro Ser Leu Asn Gly Glu Ile Gly
1 5 10 15

gct ttg tct tcc aaa aac tac acc aca cgc tac ttg cta gct gct gcg 96
Ala Leu Ser Ser Lys Asn Tyr Thr Thr Arg Tyr Leu Leu Ala Ala Ala
20 25 30

ctg gca gaa ggc aca agt aca atc cat tac ccg gct cat agt gaa gat 144
Leu Ala Glu Gly Thr Ser Thr Ile His Tyr Pro Ala His Ser Glu Asp
35 40 45

agt gat gct atg cgc aga tgt att cgt gac ctt ggg gcg gtg ctt gaa 192
Ser Asp Ala Met Arg Arg Cys Ile Arg Asp Leu Gly Ala Val Leu Glu
50 55 60

gaa gat gat agc aaa atc gtt atc caa gga ttc ggc agc cat ccg cgt 240
Glu Asp Asp Ser Lys Ile Val Ile Gln Gly Phe Gly Ser His Pro Arg
65 70 75 80

```

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

gcg gaa ggc acc tca cgg ttc tat aat gta gag aac tta cgt tac aag	1008
Ala Glu Gly Thr Ser Arg Phe Tyr Asn Val Glu Asn Leu Arg Tyr Lys	
325 330 335	
gaa tgt gac cga att acg gat tat ttg aac gaa ctg cgg aag gca gga	1056
Glu Cys Asp Arg Ile Thr Asp Tyr Leu Asn Glu Leu Arg Lys Ala Gly	
340 345 350	
gcc aac gta gaa gaa cgt cag gcc gag att atc gta cat ggt cgt ccg	1104
Ala Asn Val Glu Glu Arg Gln Ala Glu Ile Ile Val His Gly Arg Pro	
355 360 365	
gaa ggc gtc gaa ggc ggc gtt gag att aac gct cac tac gat cat cgc	1152
Glu Gly Val Glu Gly Gly Val Glu Ile Asn Ala His Tyr Asp His Arg	
370 375 380	
gta att atg gca ctg acc gtt gtt ggt ttg cgt tcc aaa gaa ccg ctt	1200
Val Ile Met Ala Leu Thr Val Val Gly Leu Arg Ser Lys Glu Pro Leu	
385 390 395 400	
cgt att cgg gat gca cac cat gta gcg aag tct tat cca caa tat ttc	1248
Arg Ile Arg Asp Ala His His Val Ala Lys Ser Tyr Pro Gln Tyr Phe	
405 410 415	
gat cat ttg cag gcg ctt ggc gcc tcg gtt caa tgg gta aaa gag taa	1296
Asp His Leu Gln Ala Leu Gly Ala Ser Val Gln Trp Val Lys Glu *	
420 425 430	

<210> SEQ ID NO: 19

<211> Comprimento: 431

<212> Tipo: PRT

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: isolado do solo

<400> Sequência: 19

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

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

<210> SEQ ID NO: 20

<211> Comprimento: 1293

<212> Tipo: DNA

<213> Organismo: Sequência Artificial

<220> Características:

<223> Outras Informações: syngrg39

<400> Sequência: 20

atggatgtca tcgtcaagcc gacgccaagc ctcaatggag agatcggcgc gctgagcagc 60
 aagaactaca caacaagata tttgctggcg gcggcgctgg cagaaggaac aagcaccatc 120
 cactaccctg ctcatcaga agattctgat gccatgagaa gatgcatcag ggacctcggc 180
 gccgtgctgg aagaagatga cagcaagatc gtcattccaag gcttcggctc acatccaaga 240
 gatgtgaggg agctcaatgt tggaaatgct ggccgctgc tgcgcttctt gatgggcgtg 300
 acggcgctct gcccgaggt gaccttcgtc aacacctacc ccgacagcct cggcaagagg 360
 cctcatgatg acctcatcga cgcgctggcg cagctaggag tggaggtgca gcatgagcaa 420
 ggaaggctgc ccatcaccat caagggcggc caagcaaaag gaggccacat cagagtttct 480
 ggaagcgctca gctctcagta cctctcggcg ctgctcttcg tgacgccgct gctggcagaa 540
 gattccacca tcgaggtgct aaatgacctc aagagcaagg tggatgatcg ccagacgctg 600
 gaggtgctgg agcaagctgg catcgatcag catgcttcag atgactacat gagcttcaga 660
 gttcctggag gacaagccta caagccgcag acctacaccg tccaaggaga ttatcctgga 720
 agcgcgcgcg tgctggccgc cgccgcgctc acccaaagtg atgtgaagat cctccggctg 780
 atggagcaat caaagcaagg agaaagggcc atcgtggatg tgctgaggat gatggaggtg 840
 ccgctcacc atgaaaatga tgtggtgcat gttcaaggaa atggcacctt gaaggcggtg 900
 gagtttgatg gagatgctgc aacagatgct gtgctggcaa tgggtggcggc ggcggtgtt 960
 gctgaaggaa catcaagatt ctacaatgtg gagaacttga gatacaagga atgtgacagg 1020
 atcaccgact acctcaacga gctgaggaag gccggcgcca atgtggagga gaggcaagct 1080

gagatcattg ttcattggaag gccagaagga gtggagggcg gcgtggagat caatgctcac 1140
 tacgaccacc gcgtcatcat ggcgctaaca gtgggtggggc tgaggagcaa ggagccgctg 1200
 aggatcagag atgctcatca tgtggccaag agctaccctc aatattttga tcattcttcaa 1260
 gctctcggcg cctccgtgca gtgggtgaag gag 1293

<210> SEQ ID NO: 21

<211> Comprimento: 1260

<212> Tipo: DNA

<213> Organismo: *Rhizobium leguminosarum*

<220> Características:

<221> Nome/Chave: CDS

<222> Localização: (1)...(1260)

<400> Sequência: 21

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

gcg ctg ctg ctg gca ggc ctt gcc aag ggc acg agc cgg ctg acc ggc 144
 Ala Leu Leu Leu Ala Gly Leu Ala Lys Gly Thr Ser Arg Leu Thr Gly
 35 40 45

gcg ctg aaa agc gac gat acc cgc tat atg gcc gaa gcg ttg cgc gcc 192
 Ala Leu Lys Ser Asp Asp Thr Arg Tyr Met Ala Glu Ala Leu Arg Ala
 50 55 60

atg ggc gtt gcg atc gac gag ccc gac gac acc acc ttc atc gtg aca 240
 Met Gly Val Ala Ile Asp Glu Pro Asp Asp Thr Thr Phe Ile Val Thr
 65 70 75 80

ggc agt ggc aag ctg cag gcg ccg gca gcc cct ctt ttc ctc ggc aat 288
 Gly Ser Gly Lys Leu Gln Ala Pro Ala Ala Pro Leu Phe Leu Gly Asn
 85 90 95

gcc ggc acg gca acg cgc ttc ctg aca gca gca gca gcg cta gtc gaa 336
 Ala Gly Thr Ala Thr Arg Phe Leu Thr Ala Ala Ala Ala Leu Val Glu
 100 105 110

ggc aag gtc gtc gtt gat ggc gat gcc cat atg cgc aaa cgg ccg atc 384
 Gly Lys Val Val Val Asp Gly Asp Ala His Met Arg Lys Arg Pro Ile
 115 120 125

ggc ccg ctg gtc gac gcg ctc cgc tcg ctc ggc gtg gac gcc tcg acg 432
 Gly Pro Leu Val Asp Ala Leu Arg Ser Leu Gly Val Asp Ala Ser Thr
 130 135 140

gaa acc ggc tgc cca ccc gta aca atc aat ggc acc ggc cgt ttc gaa 480
 Glu Thr Gly Cys Pro Pro Val Thr Ile Asn Gly Thr Gly Arg Phe Glu
 145 150 155 160

gct agc cgc gtg cag atc gac 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 ctg ttg atg atg gcg gcc ggc ggc gac cgc gcc gtc gat atc gag	576
Ala Leu Leu Met Met Ala Ala Gly Gly Asp Arg Ala Val Asp Ile Glu	
180 185 190	
ctg ctc ggc gaa cat atc ggt gct ctg gga tac 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 gcc ttc ggt gcc aag gtc gag cgc gtg agc cca gtc	672
Ala Ala Met Arg Ala Phe Gly Ala Lys Val Glu Arg Val Ser Pro Val	
210 215 220	
gcc tgg cgt gtc gag ccc acc ggc tac cat gct gcc gat ttc ctg atc	720
Ala Trp Arg Val Glu Pro Thr Gly Tyr His Ala Ala Asp Phe Leu Ile	
225 230 235 240	
gag ccg gat gcc tct gct gcg acc tat ctc tgg gct gcc gaa gtg ctt	768
Glu Pro Asp Ala Ser Ala Ala Thr Tyr Leu Trp Ala Ala Glu Val Leu	
245 250 255	
ggt ggg ggc aag atc gat ctc ggg acg ccg gcg gaa caa ttc tcg cag	816
Gly Gly Gly Lys Ile Asp Leu Gly Thr Pro Ala Glu Gln Phe Ser Gln	
260 265 270	
ccg gat gcg aaa gca tat gat ctg atc tcg aaa ttc ccg cat ctg ccc	864
Pro Asp Ala Lys Ala Tyr Asp Leu Ile Ser Lys Phe Pro His Leu Pro	
275 280 285	
gcc gtc atc gac ggt tcg cag atg cag gac gcc atc ccg acg ctt gcg	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 gag acg ccg gtg cgc ttc gtc ggc atc gag	960
Val Leu Ala Ala Phe Asn Glu Thr Pro Val Arg Phe Val Gly Ile Glu	
305 310 315 320	
aat ctc cgc gtc aag gaa tgc gat cgt atc cgc gcg ctg tcg agc ggc	1008
Asn Leu Arg Val Lys Glu Cys Asp Arg Ile Arg Ala Leu Ser Ser Gly	
325 330 335	
ctg tcg cgc gtc gtt ccg aac ctc ggc acg gaa gag ggc gac gat ctg	1056
Leu Ser Arg Val Val Pro Asn Leu Gly Thr Glu Glu Gly Asp Asp Leu	
340 345 350	
atc gtc gca tcc gat ccg agc ctc gcc ggc aag acc ctg ccc gca gag	1104
Ile Val Ala Ser Asp Pro Ser Leu Ala Gly Lys Thr Leu Pro Ala Glu	
355 360 365	
atc gac agc ttt gcc gat cat cgc att gcc atg agc ttt gcg ctt gcc	1152
Ile Asp Ser Phe Ala Asp His Arg Ile Ala Met Ser Phe Ala Leu Ala	
370 375 380	
ggc ctg aag atc ggc ggc atc acc att ctc gac ccc gat 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 acc ttc ccg tcc tat tgg aac gtg ctg gct tcg ctt gga gtc aca	1248
Lys Thr Phe Pro Ser Tyr Trp Asn Val Leu Ala Ser Leu Gly Val Thr	
405 410 415	

tac gaa gac tga
Tyr Glu Asp *

1260

<210> SEQ ID NO: 22
<211> Comprimento: 419
<212> Tipo: PRT
<213> Organismo: *Rhizobium leguminosarum*

<400> Sequência: 22

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	Ala	Ile	Asp	Glu	Pro	Asp	Asp	Thr	Thr	Phe	Ile	Val	Thr
65					70					75				80	
Gly	Ser	Gly	Lys	Leu	Gln	Ala	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	Glu
			100					105					110		
Gly	Lys	Val	Val	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	Val	Asp	Ala	Ser	Thr
	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	Ile	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	Leu	Ile
225					230					235					240
Glu	Pro	Asp	Ala	Ser	Ala	Ala	Thr	Tyr	Leu	Trp	Ala	Ala	Glu	Val	Leu
				245					250					255	
Gly	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	Thr	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	Val	Val	Pro	Asn	Leu	Gly	Thr	Glu	Glu	Gly	Asp	Asp	Leu
			340					345					350		
Ile	Val	Ala	Ser	Asp	Pro	Ser	Leu	Ala	Gly	Lys	Thr	Leu	Pro	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 Ala Ser Leu Gly Val Thr
 405 410 415
 Tyr Glu Asp

<210> SEQ ID NO: 23
 <211> Comprimento: 1257
 <212> Tipo: DNA
 <213> Organismo: Sequência Artificial

<220> Características:
 <223> Outras Informações: syngrg50

<400> Sequência: 23
 atgatgatgg gccgcgccaa gctcaccatc atccccgccgg gcaagccgct caccggccgc 60
 gccatgccgc cgggcagcaa gagcatcacc aacagggcgc tgctgctggc cggcctcgcc 120
 aagggcacct caaggctcac cggcgcgctg aagagcgacg acaccgcta catggcggag 180
 gcgctgagag caatgggctg cgccatcgac gagccagatg acaccacctt catcgtcacc 240
 ggctcaggca agctgcaagc gccggcgggc cgcctcttcc tcggcaacgc cggcacggca 300
 acaagattct tgacggcggc ggcgcgcgct gtggaaggca aggtgggtgt ggatgggtgat 360
 gctcacatga ggaagaggcc aattgggccc ctgggtggac cgctgaggag cctcggcgtg 420
 gatgcttcaa cagaaactgg ctgcccgcgc gtgaccatca atggaactgg aagatttgaa 480
 gcatcaagag ttcagatcga cggcggcctc agctctcaat atgtctcggc gctgctgatg 540
 atggccgccc gcggcgaccg cgcctgggac attgagctgc tgggagagca catcggcgcg 600
 ctgggctaca tcgacctcac cgtcgccgcc atgagggcct tcggcgccaa ggtggagagg 660
 gtgtcaccag tggcatggag ggtggagcca actggctacc atgtctgctga ctctctcatc 720
 gagccagatg cttctgctgc cactacctc tgggcggcgg aggtgctggg cggcggcaag 780
 atagatcttg ggacgccggc ggagcagttc agccaacctg atgccaaggc ctacgacctc 840
 atctcaaaaat ttctcatct tcccgcgctc atcgacggca gccagatgca agatgccatc 900
 ccgacgctgg cgtgctggc ggcttcaat gagacgccgg tgaggttcgt cggcatcgag 960
 aaacctccgc tcaaggaatg tgacaggatc agggcgctga gcagcggcct ctcaaggggtg 1020
 gtgccccaac tcggcacaga agaaggagat gacctcatcg tggcttcaga tccaagcctc 1080
 gccggcaaga cgctgccggc ggagatcgac agctttgctg atcaccgcat cgccatgagc 1140
 ttgcgcgctc cgggcctcaa gatcggcggc atcaccatcc ttgatccaga ttgtgtggcc 1200
 aagaccttcc caagctactg gaatgtgctg gcaagcctcg gcgtcaccta tgaagat 1257

<210> SEQ ID NO: 24
 <211> Comprimento: 438
 <212> Tipo: PRT
 <213> Organismo: *Streptomyces coelicolor* A3(2)

<400> Sequência: 24
 Met Thr Val Asn Pro Thr His Thr Ala Leu Trp Pro Ala Pro His Ala
 1 5 10 15
 Ser Gly Ala Val Asp Ala Thr Val His Val Pro Gly Ser Lys Ser Val
 20 25 30
 Thr Asn Arg Ala Leu Val Leu Ala Leu Ala Ser Glu Pro Gly Trp
 35 40 45
 Leu Arg Arg Pro Leu Arg Ser Arg Asp Thr Leu Leu Met Ala Glu Ala
 50 55 60
 Leu Arg Thr Leu Gly Val Glu Ile Glu Glu Gly Val Gly Pro Glu Gly
 65 70 75 80
 Thr Gly Glu Phe Trp Arg Val Ile Pro Ala Gly Leu Arg Gly Pro Ala
 85 90 95
 Thr Val Asp Val Gly Asn Ala Gly Thr Val Met Arg Phe Leu Pro Pro
 100 105 110
 Val Ala Thr Leu Ala Asp Gly Ala Val Arg Phe Asp Gly Asp Pro Arg
 115 120 125
 Ser Tyr Glu Arg Pro Leu His Gly Val Ile Asp Ala Leu Arg Val Leu
 130 135 140

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

<210> SEQ ID NO: 25

<211> Comprimento: 446

<212> Tipo: PRT

<213> Organismo: *Streptomyces avermitilis* MA-4680

<400> Sequência: 25

Met Thr Val Asn Pro Ala His Thr Ala Leu Trp Pro Ala Pro His Ala
 1 5 10 15
 Ser Gly Ala Val Asp Ala Thr Val His Val Pro Gly Ser Lys Ser Val
 20 25 30
 Thr Asn Arg Ala Leu Val Leu Ala Ala Leu Ala Ser Glu Pro Gly Trp
 35 40 45
 Leu Arg Arg Pro Leu Arg Ser Arg Asp Thr Leu Leu Met Ala Ala Ala
 50 55 60
 Leu Arg Glu Met Gly Val Gly Ile Glu Glu Thr Val Ser Ser Ser Ser
 65 70 75 80
 Ser Val Gly Gly Gly Ser Asp Gly Ser Gly Glu Ala Trp Arg Val Ile
 85 90 95
 Pro Ala Ala Leu His Gly Pro Ala Thr Val Asp Val Gly Asn Ala Gly
 100 105 110

```

Thr Val Met Arg Phe Leu Pro Pro Val Ala Ala Leu Ala Asp Gly Pro
      115                      120                      125
Ile Arg Phe Asp Gly Asp Pro Arg Ser Tyr Glu Arg Pro Leu Asn Gly
      130                      135                      140
Val Ile Asp Ala Leu Arg Ala Leu Gly Ala Arg Ile Asp Asp Asp Gly
145                      150                      155                      160
Arg Gly Ala Leu Pro Leu Thr Val His Gly Gly Gly Ala Leu Asp Gly
      165                      170                      175
Gly Pro Val Ala Ile Asp Ala Ser Ser Ser Ser Gln Phe Val Ser Ala
      180                      185                      190
Leu Leu Leu Ser Gly Pro Arg Phe Asn Gln Gly Val Glu Val Arg His
      195                      200                      205
Thr Gly Ser Thr Leu Pro Ser Met Pro His Ile Arg Met Thr Val Asp
      210                      215                      220
Met Leu Arg Ala Val Gly Ala Gln Val Asp Thr Pro Glu Ser Gly Gly
225                      230                      235                      240
Glu Ala Asn Val Trp Arg Val Thr Pro Gly Ala Leu Leu Gly Arg Asp
      245                      250                      255
Leu Thr Val Glu Pro Asp Leu Ser Asn Ala Gln Pro Phe Leu Ala Ala
      260                      265                      270
Ala Leu Val Thr Gly Gly Lys Val Val Ile Pro Asp Trp Pro Glu Arg
      275                      280                      285
Thr Thr Gln Pro Gly Asp Lys Leu Arg Glu Ile Phe Thr Glu Met Gly
      290                      295                      300
Gly Ser Cys Glu Leu Thr Glu Gln Gly Leu Glu Phe Thr Gly Ser Gly
305                      310                      315                      320
Ala Val His Gly Ile Asp Val Asp Leu Ser Glu Val Gly Glu Leu Thr
      325                      330                      335
Pro Gly Ile Ala Ala Val Ala Ala Leu Ala Asp Ser Pro Ser Thr Leu
      340                      345                      350
Arg Gly Val Ala His Leu Arg Leu His Glu Thr Asp Arg Leu Ala Ala
      355                      360                      365
Leu Thr Lys Glu Ile Asn Glu Leu Gly Gly Asp Val Thr Glu Thr Ala
      370                      375                      380
Asp Gly Leu Ser Ile Arg Pro Arg Arg Leu His Gly Gly Ile Phe His
385                      390                      395                      400
Thr Tyr Asp Asp His Arg Met Ala Thr Ala Gly Ala Ile Ile Gly Leu
      405                      410                      415
Ala Val Asp Gly Val Gln Ile Glu Asn Val Ala Thr Thr Ala Lys Thr
      420                      425                      430
Leu Pro Asp Phe Pro Asp Leu Trp Thr Gly Met Leu Gly Asn
      435                      440                      445

```

<210> SEQ ID NO: 26

<211> Comprimento: 447

<212> Tipo: PRT

<213> Organismo: *Bacillus halodurans* C-125

<400> Sequência: 26

```

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

```

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

<210> SEQ ID NO: 27

<211> Comprimento: 446

<212> Tipo: PRT

<213> Organismo: *Bacillus clausii* KSM-K16

<400> Sequência: 27

Met	Val	Gln	Phe	Asp	Ser	Gln	Ala	Arg	Ser	Pro	Trp	Thr	Pro	Leu	Ala
1				5					10					15	
Gly	Val	Glu	Arg	Leu	Arg	Leu	Thr	Pro	Ser	Gln	Lys	Arg	Ile	Asn	Ala
			20					25					30		
Thr	Leu	Glu	Val	Pro	Gly	Ser	Lys	Ser	Ala	Thr	Asn	Arg	Ala	Leu	Leu
		35					40					45			

```

Leu Ala Ala Val Ala Ser Gly Thr Ser Thr Leu Arg Asn Ala Leu Lys
 50          55          60
Ser Asp Asp Thr Tyr Trp Cys Ile Glu Ala Leu Lys Lys Thr Gly Val
65          70          75          80
Glu Ile Ala Val Asp Gly Ser Asn Val Thr Val Tyr Gly Arg Gly Gly
          85          90          95
Val Phe His Ser Gly Ser Leu Tyr Ile Gly Ser Ala Gly Thr Ala Gly
          100          105          110
Arg Phe Leu Pro Gly Met Leu Ala Ala Ala Thr Gly Asn Trp His Val
          115          120          125
Glu Ala Ser His Ser Met Asn Lys Arg Pro Ile Ala Pro Leu Val Lys
          130          135          140
Thr Leu Gln Ala Leu Gly Ala Asn Ile Gln Tyr Gly Ser Arg Arg Gly
145          150          155          160
His Tyr Pro Leu Ser Ile Ser Gly Glu Gly Leu Asn Gly Gly Lys Val
          165          170          175
Asn Met Ser Gly Gln Leu Ser Ser Gln Phe Ile Ser Gly Cys Leu Leu
          180          185          190
Ala Ala Pro Leu Ala Lys Asn Pro Val Ser Ile Thr Val Lys Asp Gly
          195          200          205
Ile Val Gln Gln Ala Tyr Val Arg Ile Thr Ile Asp Leu Met Ala Ala
          210          215          220
Phe Gly Val Glu Val Lys Ala Ala Pro Asp Trp Ser Leu Leu Glu Val
225          230          235          240
Asn Pro Ser Pro Tyr Val Ala Asn Asp Ile Ala Ile Glu Ala Asp Ala
          245          250          255
Ser Thr Ala Cys Tyr Phe Leu Ala Leu Ala Ala Ile Thr Ala Gly Lys
          260          265          270
Ile Arg Ile Arg His Phe Ser Thr Lys Thr Ser Gln Pro Asp Ile Leu
          275          280          285
Phe Val Ser Ile Leu Lys Arg Met Gly Cys Asn Phe Glu Ile Gly Pro
          290          295          300
Ser Phe Val Glu Gly Glu Gly Pro Thr Arg Leu Arg Gly Gly Phe Thr
305          310          315          320
Val Asn Met Asn Glu Leu Ser Asp Gln Ala Leu Thr Leu Ala Ala Ile
          325          330          335
Ser Pro Phe Ala Asp Gly Pro Ile Ala Ile Glu Gly Val Gly His Ile
          340          345          350
Arg His His Glu Cys Asp Arg Ile Arg Ala Ile Cys Thr Glu Leu Ser
          355          360          365
Arg Leu Gly Ile Arg Val Glu Glu Arg His Asp Gly Leu Thr Val Tyr
          370          375          380
Pro Gly Gln Pro Lys Pro Thr Val Val Asn Thr Tyr Asp Asp His Arg
385          390          395          400
Met Ala Met Ala Leu Ala Leu Ile Gly Ala Lys Val Asp Gly Ile Glu
          405          410          415
Leu Asp Asp Pro Gly Cys Val Ala Lys Thr Cys Pro Ser Tyr Phe Ser
          420          425          430
Met Leu Ala Gln Thr Gly Ile Gly Val Lys Ala Val Ser Pro
          435          440          445

```

<210> SEQ ID NO: 28

<211> Comprimento: 460

<212> Tipo: PRT

<213> Organismo: *Arthrobacter* sp. FB24

<400> Sequência: 28

```

Met Thr Gly Thr Ala Pro Thr Glu Ser Ala Thr Ser Gly Pro Val Ala
 1          5          10          15

```

Asp	Val	Pro	His	Trp	Pro	Ala	Pro	Phe	Ala	Glu	Ala	Pro	Val	Asp	Ala		
			20					25					30				
Thr	Val	Thr	Val	Pro	Gly	Ser	Lys	Ser	Leu	Thr	Asn	Arg	Tyr	Leu	Val		
		35					40					45					
Leu	Ala	Ala	Leu	Ala	Asp	Gly	Pro	Ser	Arg	Leu	Arg	Ala	Pro	Leu	His		
	50					55					60						
Ser	Arg	Asp	Ser	Ala	Leu	Met	Ile	Glu	Ala	Leu	Arg	Gln	Leu	Gly	Ala		
65					70					75					80		
Gly	Ile	Arg	Glu	Val	His	Ser	Asp	Gly	Ala	Phe	Gly	Pro	Asp	Leu	Glu		
				85					90					95			
Val	Thr	Pro	Leu	Arg	Ala	Asp	Ala	Ala	Ala	Thr	Asp	Ala	Ala	Ile	Asp		
			100					105					110				
Cys	Gly	Leu	Ala	Gly	Thr	Val	Met	Arg	Phe	Val	Pro	Pro	Val	Ala	Ala		
		115					120					125					
Leu	Arg	Asn	Gly	Ala	Thr	Val	Phe	Asp	Gly	Asp	Pro	His	Ala	Arg	Lys		
	130					135					140						
Arg	Pro	Met	Gly	Thr	Ile	Ile	Glu	Ala	Leu	Ala	Ala	Leu	Gly	Val	Asp		
145					150					155					160		
Val	Arg	Ala	Ala	Asp	Gly	Thr	Pro	Pro	Ser	Ala	Leu	Pro	Phe	Thr	Val		
				165					170					175			
Ala	Gly	Ser	Gly	His	Val	Arg	Gly	Gly	His	Leu	Val	Ile	Asp	Ala	Ser		
			180					185					190				
Ala	Ser	Ser	Gln	Phe	Val	Ser	Ala	Leu	Leu	Leu	Val	Gly	Ala	Arg	Phe		
		195					200					205					
Thr	Glu	Gly	Leu	His	Leu	Glu	His	Val	Gly	Lys	Pro	Val	Pro	Ser	Leu		
	210					215					220						
Asp	His	Ile	Asn	Met	Thr	Val	Ala	Val	Leu	Arg	Glu	Val	Gly	Val	Ser		
225					230					235					240		
Val	Asp	Asp	Ser	Val	Pro	Asn	His	Trp	Val	Val	Ala	Pro	Gly	Arg	Ile		
				245					250					255			
Arg	Ala	Phe	Asp	Arg	Arg	Ile	Glu	Gln	Asp	Leu	Ser	Asn	Ala	Gly	Pro		
			260					265					270				
Phe	Leu	Ala	Ala	Ala	Leu	Ala	Thr	Arg	Gly	Thr	Val	Arg	Ile	Pro	Asn		
		275					280					285					
Trp	Pro	Ser	Pro	Thr	Thr	Gln	Val	Gly	Asp	Leu	Trp	Arg	Ser	Ile	Leu		
	290					295					300						
Thr	Ala	Met	Gly	Ala	Thr	Val	Thr	Leu	Asp	Asn	Gly	Thr	Leu	Thr	Val		
305					310					315					320		
Thr	Gly	Gly	Pro	Glu	Ile	Thr	Gly	Ala	Asp	Phe	Ala	Asp	Thr	Ser	Glu		
				325					330					335			
Leu	Ala	Pro	Thr	Val	Ala	Ala	Leu	Cys	Ala	Leu	Ala	Thr	Gly	Pro	Ser		
			340					345					350				
Arg	Leu	Thr	Gly	Ile	Ala	His	Leu	Arg	Gly	His	Glu	Thr	Asp	Arg	Leu		
		355					360					365					
Ala	Ala	Leu	Val	Thr	Glu	Ile	Asn	Arg	Leu	Gly	Gly	Asp	Ala	Glu	Glu		
		370				375					380						
Thr	Ser	Asp	Gly	Leu	Val	Ile	Arg	Pro	Ala	Lys	Leu	His	Gly	Gly	Val		
385					390					395					400		
Val	His	Ser	Tyr	Ala	Asp	His	Arg	Met	Ala	Thr	Ala	Gly	Ala	Ile	Leu		
				405					410					415			
Gly	Leu	Ala	Val	Pro	Gly	Val	Glu	Val	Glu	Asp	Ile	Gly	Thr	Thr	Ser		
			420					425					430				
Lys	Thr	Met	Pro	Asp	Phe	Pro	Gln	Leu	Trp	Glu	Ser	Met	Leu	Thr	Gln		
		435					440					445					
Gln	Pro	Gly	Arg	Gln	Thr	Glu	Gln	Ala	Arg	Gly	Ala						
	450					455					460						

<210> SEQ ID NO: 29

<211> Comprimento: 419

<212> Tipo: PRT

<213> Organismo: *Brevundomonas vesicularis*

<400> Sequência: 29

```

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> SEQ ID NO: 30

<211> Comprimento: 444

<212> Tipo: PRT

<213> Organismo: *Ochrobactrum/Brucella cepa C58*

<400> Sequência: 30

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	
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	
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	
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	
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	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	
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> SEQ ID NO: 31

<211> Comprimento: 425

<212> Tipo: PRT

<213> Organismo: *Agrobacterium tumefaciens* cepa C58

<400> Sequência: 31

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	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> SEQ ID NO: 32

<211> Comprimento: 418

<212> Tipo: PRT

<213> Organismo: *Pseudomonas syringae* cepa 1448a

<400> Sequência: 32

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> SEQ ID NO: 33

<211> Comprimento: 418

<212> Tipo: PRT

<213> Organismo: *Pseudomonas syringae* cepa DC3000

<400> Sequência: 33

```

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
Leu Arg Val Lys Glu Cys Asp Arg Ile Ser Ala Leu Cys Asp 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 His Ala Asn Pro Ala Leu Ala Gly Thr Thr Val Asn 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 Lys Gly Ile His Ile Gln 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> SEQ ID NO: 34

<211> Comprimento: 418

<212> Tipo: PRT

<213> Organismo: *Pseudomonas syringae* cepa B728

<400> Sequência: 34

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> SEQ ID NO: 35

<211> Comprimento: 425

<212> Tipo: PRT

<213> Organismo: *Agrobacterium tumefaciens* cepa C58

<400> Sequência: 35

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 Ala His Glu Glu Gly Asp Asp Leu Ile Val His
 340 345 350
 Ala Asp Pro Ser Leu Ala Gly Gln Thr Val Asp 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	Ala	Lys	Thr	Tyr
385					390					395					400
Pro	Gly	Tyr	Trp	Lys	Ala	Leu	Ala	Ser	Leu	Gly	Val	Asp	Tyr	Thr	Glu
				405					410						415
Lys	Glu	Ser	Ala	Ala	Glu	Pro	Gln	His							
			420					425							

<210> SEQ ID NO: 36

<211> Comprimento: 444

<212> Tipo: PRT

<213> Organismo: Desconhecido

<220> Características:

<223> Outras Informações: Isolado de amostra de solo

<400> Sequência: 36

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	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	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 Trp Pro Ser Leu Gly Val His Leu Asn Asp
      435      440

```

<210> SEQ ID NO: 37

<211> Comprimento: 427

<212> Tipo: PRT

<213> Organismo: *Escherichia coli*

<400> Sequência: 37

```

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 Ser Val Ser Asn Arg Ala Leu 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 Thr 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 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> SEQ ID NO: 38

<211> Comprimento: 444

<212> Tipo: PRT

<213> Organismo: *Zea mays*

<400> Sequência: 38

```

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 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> SEQ ID NO: 39

<211> Comprimento: 455

<212> Tipo: PRT

<213> Organismo: *Agrobacterium* sp. CP4

<400> Sequência: 39

```

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 Gly 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
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> SEQ ID NO: 40

<211> Comprimento: 428

<212> Tipo: PRT

<213> Organismo: *Bacillus subtilis*

<400> Sequência: 40

```

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> SEQ ID NO: 41

<211> Comprimento: 431

<212> Tipo: PRT

<213> Organismo: *Enterobacteriaceae*

<400> Sequência: 41

Met Lys Val Thr Ile Gln Pro Gly Asp Leu Thr Gly Ile Leu 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		

REIVINDICAÇÕES

1. Molécula de ácido nucléico isolada caracterizada por compreender uma seqüência de nucleotídeos selecionada do grupo consistindo em:

a) seqüência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, ou um complemento dessas;

b) seqüência de nucleotídeos tendo pelo menos 90% de identidade de seqüência em relação à seqüência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, ou 13, ou um complemento dessas;

c) seqüência de nucleotídeos tendo pelo menos 97% de identidade de seqüência em relação à seqüência de nucleotídeos de SEQ ID NO: 15, 17, 18, 20, 21, ou 23, ou um complemento dessas;

d) seqüência de nucleotídeos de resistência a herbicidas do inserto de DNA do plasmídeo depositado como Nos. de Acesso NRRL B-30934, B-30932, B-30933, B-30945, B-30946, B-30947, ou B-30948, ou um complemento dessas;

e) seqüência de nucleotídeos que codifica um polipeptídeo compreendendo a seqüência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, 14, 16, 19, ou 22;

f) sequência de nucleotídeos codificando um polipeptídeo tendo pelo menos 90% identidade de sequência de aminoácidos em relação à sequência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, ou 14; e,

5

g) sequência de nucleotídeos codificando um polipeptídeo tendo pelo menos 97% identidade de sequência de aminoácidos em relação à sequência de aminoácidos de SEQ ID NO: 16, 19, ou 22.

10

2. Molécula de ácido nucléico isolada de acordo com a reivindicação 1, caracterizada pelo fato de que dita sequência de nucleotídeos é uma sequência sintética que foi projetada para expressão em uma planta.

15

3. Vetor caracterizado por compreender a molécula de ácido nucléico da reivindicação 1.

4. Vetor de acordo com a reivindicação 3, caracterizado por compreender ainda uma molécula de ácido nucléico codificando um polipeptídeo heterólogo.

20

5. Célula hospedeira caracterizada por conter o vetor da reivindicação 3.

25

6. Célula hospedeira de acordo com a reivindicação 5, caracterizada por ser uma célula hospedeira bacteriana.

7. Célula hospedeira de acordo com a reivindicação 5, caracterizada por ser uma célula vegetal.

8. Planta transgênica caracterizada por compreender a
5 célula hospedeira da reivindicação 7.

9. Planta de acordo com a reivindicação 8, caracterizada pelo fato de que dita planta é selecionada do grupo consistindo de milho, sorgo, trigo, girassol, tomate,
10 crucíferas, pimentas, batata, algodão, arroz, soja, beterraba, cana-de-açúcar, tabaco, cevada e colza.

10. Semente transgênica caracterizada por compreender a molécula de ácido nucléico da reivindicação 1.

15

11. Polipeptídeo isolado caracterizado por ser selecionado do grupo consistindo em:

a) polipeptídeo compreendendo a sequência de aminoácidos
20 de SEQ ID NO: 8, 2, 5, 11, 14, 16, 19, ou 22;

b) polipeptídeo codificado pela sequência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23;

25

c) polipeptídeo compreendendo uma sequência de aminoácidos tendo pelo menos 90% de identidade de sequência em relação à sequência de aminoácidos de SEQ ID NO: 8, 2,

5, 11, ou 14, em que dito polipeptídeo possui atividade de resistência a herbicidas;

d) polipeptídeo compreendendo uma sequência de aminoácidos tendo pelo menos 97% de identidade de sequência em relação à sequência de aminoácidos de SEQ ID NO: 16, 19, ou 22, em que dito polipeptídeo possui atividade de resistência a herbicidas;

10 e) polipeptídeo que é codificado por uma sequência de nucleotídeos que é pelo menos 90% idêntica à sequência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, ou 13, em que dito polipeptídeo possui atividade de resistência a herbicidas;

15

f) polipeptídeo que é codificado por uma sequência de nucleotídeos que é pelo menos 97% idêntica à sequência de nucleotídeos de SEQ ID NO: 15, 17, 18, 20, 21, ou 23, em que dito polipeptídeo possui atividade de resistência a herbicidas; e,

20

g) polipeptídeo que é codificado pela sequência de nucleotídeos de resistência a herbicidas do inserto de DNA do plasmídeo depositado como Nos. de Acesso NRRL B-30934, B-30932, B-30933, B-30945, B-30946, B-30947, ou B-30948.

25

12. Polipeptídeo de acordo com a reivindicação 11, caracterizado por compreender ainda uma sequência de aminoácidos heteróloga.

13. Método para produzir um polipeptídeo com atividade de resistência a herbicidas, caracterizado por compreender cultivar a célula hospedeira da reivindicação 5 sob condições nas quais uma molécula de ácido nucléico
5 codificando o polipeptídeo é expressa, dito polipeptídeo sendo selecionado do grupo consistindo em:

a) polipeptídeo compreendendo a seqüência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, 14, 16, 19, ou 22;

10

b) polipeptídeo codificado pela seqüência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23;

15 c) polipeptídeo compreendendo uma seqüência de aminoácidos tendo pelo menos 90% de identidade de seqüência em relação à seqüência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, ou 14, em que dito polipeptídeo possui atividade de resistência a herbicidas;

20

d) polipeptídeo compreendendo uma seqüência de aminoácidos tendo pelo menos 97% de identidade de seqüência em relação à seqüência de aminoácidos de SEQ ID NO: 16, 19, ou 22, em que dito polipeptídeo possui atividade de
25 resistência a herbicidas;

e) polipeptídeo que é codificado por uma seqüência de nucleotídeos que é pelo menos 90% idêntica à seqüência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, ou 13,

caracterizado pelo fato de que dito polipeptídeo possui atividade de resistência a herbicidas;

- 5 f) polipeptídeo que é codificado por uma sequência de nucleotídeos que é pelo menos 97% idêntica à sequência de nucleotídeos de SEQ ID NO: 15, 17, 18, 20, 21, ou 23, em que dito polipeptídeo possui atividade de resistência a herbicidas; e,
- 10 g) a polipeptídeo que é codificado pela sequência de nucleotídeos de resistência a herbicidas do inserto de DNA do plasmídeo depositado como Nos. de Acesso NRRL B-30934, B-30932, B-30933, B-30945, B-30946, B-30947, ou B-30948.
- 15 14. Método para conferir resistência a um herbicida em uma planta, caracterizado por compreender transformar dita planta com uma construção de DNA, dita construção compreendendo um promotor que dirige a expressão em uma célula vegetal operacionalmente ligada com uma sequência de
- 20 nucleotídeos, e regenerar uma planta transformada, em que dita sequência de nucleotídeos é selecionada do grupo consistindo em:
- a) sequência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6,
- 25 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, ou um complemento dessas;
- b) sequência de nucleotídeos tendo pelo menos 90% de identidade de sequência em relação à sequência de

nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, ou 13, ou um complemento dessas, em que dito polipeptídeo possui atividade de resistência a herbicidas;

- 5 c) sequência de nucleotídeos tendo pelo menos 97% de identidade de sequência em relação à sequência de nucleotídeos de SEQ ID NO: 15, 17, 18, 20, 21, ou 23, ou um complemento dessas, em que dito polipeptídeo possui atividade de resistência a herbicidas;

10

d) sequência de nucleotídeos de resistência a herbicidas do inserto de DNA do plasmídeo depositado com os Nos. de Acesso NRRL B-30934, B-30932, B-30933, B-30945, B-30946, B-30947, ou B-30948, ou um complemento desses;

15

e) sequência de nucleotídeos que codificada um polipeptídeo compreendendo a sequência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, 14, 16, 19, ou 22;

- 20 f) sequência de nucleotídeos codificando o polipeptídeo tendo pelo menos 90% identidade de sequência de aminoácidos em relação à sequência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, ou 14, em que dito polipeptídeo possui atividade de resistência a herbicidas; e,

25

g) sequência de nucleotídeos codificando um polipeptídeo tendo pelo menos 97% identidade de sequência de aminoácidos em relação à sequência de aminoácidos de SEQ ID NO: 16, 19,

ou 22, em que dito polipeptídeo possui atividade de resistência a herbicidas.

15. Método de acordo com a reivindicação 14, caracterizado
5 pelo fato de que dita planta produz um polipeptídeo tendo atividade de resistência a herbicidas.

16. Método de acordo com a reivindicação 14, caracterizado pelo fato de que dito herbicida é um glifosato.

10

17. Planta caracterizada por ter estavelmente incorporada em seu genoma uma construção de DNA compreendendo uma sequência de nucleotídeos que codifica uma proteína tendo atividade de resistência a herbicidas, em que dita
15 sequência de nucleotídeos é selecionada do grupo consistindo em:

a) sequência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, 13, 15, 17, 18, 20, 21, ou 23, ou um complemento
20 dessas;

b) sequência de nucleotídeos tendo pelo menos 90% de identidade de sequência em relação à sequência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9, 10, 12, ou 13,
25 ou um complemento dessas, em que dito polipeptídeo possui atividade de resistência a herbicidas;

c) sequência de nucleotídeos tendo pelo menos 97% de identidade de sequência em relação à sequência de

nucleotídeos de SEQ ID NO: 15, 17, 18, 20, 21, ou 23, ou um complemento dessas, em que dito polipeptídeo possui atividade de resistência a herbicidas;

5 d) sequência de nucleotídeos de resistência a herbicidas do inserto de DNA do plasmídeo depositado como Nos. de Acesso NRRL B-30934, B-30932, B-30933, B- 30945, B-30946, B-30947, ou B-30948, ou um complemento desses;

10 e) sequência de nucleotídeos que codifica um polipeptídeo compreendendo a sequência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, 14, 16, 19, ou 22;

f) sequência de nucleotídeos codificando um polipeptídeo
15 tendo pelo menos 90% identidade de sequência de aminoácidos em relação à sequência de aminoácidos de SEQ ID NO: 8, 2, 5, 11, ou 14, em que dito polipeptídeo possui atividade de resistência a herbicidas; e,

20 g) sequência de nucleotídeos codificando a polipeptídeo tendo pelo menos 97% identidade de sequência de aminoácidos em relação à sequência de aminoácidos de SEQ ID NO: 16, 19, ou 22, em que dito polipeptídeo possui atividade de resistência a herbicidas;

25

em que dita sequência de nucleotídeos é operacionalmente ligada a um promotor que dirige a expressão de uma sequência codificadora em uma célula vegetal.

18. Planta de acordo com a reivindicação 17, caracterizada pelo fato de que dita planta é uma célula vegetal.

19. Método para gerar um polinucleotídeo variante de um
5 polinucleotídeo parental *grg33*, *grg35*, *grg36*, *grg37*, *grg38*,
grg39, ou *grg50*, caracterizado por compreender usar a
seqüência de nucleotídeos de SEQ ID NO: 7, 1, 3, 4, 6, 9,
10, 12, 13, 15, 17, 18, 20, 21, ou 23, ou um fragmento
dessas, em um procedimento recombinogênico, e testar um
10 polipeptídeo codificado pela seqüência de polinucleotídeos
resultante para uma atividade de interesse.

20. Método de acordo com a reivindicação 19, caracterizado
pelo fato de que a atividade de interesse é atividade de
15 resistência a glifosato.

21. Método de acordo com a reivindicação 20, caracterizado
pelo fato de que dito polipeptídeo codificado pela
seqüência de polinucleotídeos resultante possui uma
20 atividade de resistência a glifosato maior do que um
polipeptídeo codificado por dito polinucleotídeo parental.

22. Método de acordo com a reivindicação 19, caracterizado
pelo fato de que dito procedimento recombinogênico é
25 embaralhamento de DNA.

FIGURA 1A

GRG33	MTDSNPSTDPSTGR	Y A A	+	20	+	40	+	60	52
GRG35	MTAASPAHPDPSQPAVPAGPDL	Y B P	+	20	+	40	+	60	60
S_avermitilis_MA-468	MTVNPAHTAL	H S A	+	20	+	40	+	60	48
S_coelicolor_A3(2)	MTVNPTHTAL	H S A	+	20	+	40	+	60	48
Zea_maiZe	AGAEIVLOPIKE	SG K K P	+	20	+	40	+	60	42
E_coli	MESLTIOPIAR	G N	+	20	+	40	+	60	40
GRG33	EG	K D	+	80	+	100	+	120	104
GRG35	EG	K D	+	80	+	100	+	120	112
S_avermitilis_MA-468	A A E	C L T	+	80	+	100	+	120	108
S_coelicolor_A3(2)	A A E	C L T	+	80	+	100	+	120	100
Zea_maiZe	DNT M E	V H Y L G	+	80	+	100	+	120	97
E_coli	ETNT D D	V R H L N T A	+	80	+	100	+	120	92
GRG33	AK	Q E	+	140	+	160	+	180	161
GRG35	AK	Q E	+	140	+	160	+	180	169
S_avermitilis_MA-468	AK	Q E	+	140	+	160	+	180	165
S_coelicolor_A3(2)	AK	Q E	+	140	+	160	+	180	157
Zea_maiZe	AK	Q E	+	140	+	160	+	180	157
E_coli	AK	Q E	+	140	+	160	+	180	150
GRG33	AK	Q E	+	200	+	220	+	240	221
GRG35	AK	Q E	+	200	+	220	+	240	229
S_avermitilis_MA-468	AK	Q E	+	200	+	220	+	240	225
S_coelicolor_A3(2)	AK	Q E	+	200	+	220	+	240	217
Zea_maiZe	AK	Q E	+	200	+	220	+	240	216
E_coli	AK	Q E	+	200	+	220	+	240	207

FIGURA 1B

```

GRG33      : 260      +      280      +      300      : 281
GRG35      : 289
S_avermitilis_MA-468 : 285
S_coelicolor_A3(2)   : 277
Zea_maize      : 275
E_coli      : 265

GRG33      : 320      +      340      +      360      : 334
GRG35      : 342
S_avermitilis_MA-468 : 338
S_coelicolor_A3(2)   : 330
Zea_maize      : 335
E_coli      : 317

GRG33      : 380      +      400      +      420      : 393
GRG35      : 401
S_avermitilis_MA-468 : 397
S_coelicolor_A3(2)   : 389
Zea_maize      : 395
E_coli      : 377

GRG33      : 440      +      460      : 442
GRG35      : 450
S_avermitilis_MA-468 : 446
S_coelicolor_A3(2)   : 438
Zea_maize      : 444
E_coli      : 427

```

```

GRG36      : -KQNN[REDACTED]E[REDACTED]R[REDACTED]ET[REDACTED]L[REDACTED]T[REDACTED]S[REDACTED]Y[REDACTED]N[REDACTED]D[REDACTED]E[REDACTED]K[REDACTED]N[REDACTED]E[REDACTED]H[REDACTED]S[REDACTED]L[REDACTED]L[REDACTED]E[REDACTED]H[REDACTED]K[REDACTED]E[REDACTED]A[REDACTED] : 2300
H_halodura : -N[REDACTED]A[REDACTED]T[REDACTED]R[REDACTED]Q[REDACTED]L[REDACTED]L[REDACTED]L[REDACTED]A[REDACTED]Y[REDACTED]D[REDACTED]T[REDACTED]T[REDACTED]Q[REDACTED]S[REDACTED]L[REDACTED]H[REDACTED]A[REDACTED]F[REDACTED]E[REDACTED]L[REDACTED]H[REDACTED]M[REDACTED]S[REDACTED]E[REDACTED]A[REDACTED] : 2280
H_clausii  : -E[REDACTED]N[REDACTED]E[REDACTED]K[REDACTED]N[REDACTED]Q[REDACTED]Q[REDACTED]E[REDACTED]I[REDACTED]C[REDACTED]E[REDACTED]I[REDACTED]G[REDACTED]A[REDACTED]K[REDACTED]N[REDACTED]S[REDACTED]T[REDACTED]G[REDACTED]O[REDACTED]A[REDACTED]R[REDACTED]R[REDACTED]D[REDACTED]M[REDACTED]A[REDACTED]E[REDACTED] : 2270
E_coli     : -G[REDACTED]E[REDACTED]T[REDACTED]N[REDACTED]D[REDACTED]A[REDACTED]D[REDACTED]B[REDACTED]E[REDACTED]E[REDACTED]A[REDACTED]E[REDACTED]T[REDACTED]A[REDACTED]P[REDACTED]E[REDACTED]D[REDACTED]T[REDACTED]V[REDACTED]R[REDACTED]G[REDACTED]D[REDACTED]L[REDACTED]S[REDACTED]K[REDACTED]P[REDACTED]D[REDACTED]C[REDACTED]T[REDACTED]N[REDACTED]A[REDACTED]T[REDACTED] : 2113
Zea_mai ze : I[REDACTED]G[REDACTED]E[REDACTED]P[REDACTED]E[REDACTED]K[REDACTED]K[REDACTED]S[REDACTED]S[REDACTED]E[REDACTED]E[REDACTED]L[REDACTED]A[REDACTED]E[REDACTED]E[REDACTED]E[REDACTED]L[REDACTED]G[REDACTED]D[REDACTED]E[REDACTED]E[REDACTED]I[REDACTED]K[REDACTED]I[REDACTED]S[REDACTED]I[REDACTED]P[REDACTED]E[REDACTED]E[REDACTED]E[REDACTED]R[REDACTED]I[REDACTED]E[REDACTED]R[REDACTED]E[REDACTED] : 2222

```

FIGURA 2B

```

GRG36      : KKYDSSLKE  Y--  K  Q  N  V  T  N  Q  D  N  Y  E  I  K  : 288
B_halodura : YDQQLQT  H--  P  C  Q  D  P  K  N  N  A  S  T  : 286
B_clausii  : KAAPD  S  L  E  N--  P  V  A  N  A  R  H  F  T  K  S  : 285
E_coli     : N-QH  Q  Q  F  K  G  Q  Y  Q  G  T  Y  L  G  S  A  K  G  T  T  G  G  R  N  M  G  : 272
Zea_maize  : KAHSDS  D  R  F  Y  K  G  Q  K  Y  R  K  A  Y  G  C  T  T  E  G  C  G  T  T  L  G  : 282
              6                               6E Das3A YFLA AA6t G 6 6 3 Q
              *                               *

GRG36      : D  M  D  L  E  A  N  K  R  S  S  I  R  V  V  Q  S  R  V  V  : 342
B_halodura : H  M  T  F  E  R  T  R  S  S  E  L  V  V  Q  S  R  V  V  : 340
B_clausii  : L  L  S  L  K  T  N  F  E  I  P  E  R  G  P  R  T  S  : 339
E_coli     : A  D  L  E  A  T  C  W  D  D  S  C  T  R  G  E  L  N  A  I  D  H  H  P  A  T  A  L  : 324
Zea_maize  : A  A  E  L  E  M  A  K  M  T  R  S  T  T  P  P  R  E  P  E  G  R  K  H  L  K  A  I  D  H  H  P  A  T  A  L  : 342
              6le MG          g 6 6 g          6 Mn 6 D a6T6A a F
              *                               *

GRG36      : A  I  E  T  E  N  D  B  T  H  H  L  H  I  S  V  V  D  S  E  R  I  E  F  K  L  K  N  N  A  L  H  : 400
B_halodura : A  I  E  T  E  N  D  B  T  H  H  L  H  I  S  V  V  D  S  E  R  I  E  F  K  L  K  N  N  A  L  H  : 400
B_clausii  : A  I  E  T  E  N  D  B  T  H  H  L  H  I  S  V  V  D  S  E  R  I  E  F  K  L  K  N  N  A  L  H  : 397
E_coli     : K  T  T  R  I  S  Y  N  W  V  K  E  T  E  F  A  R  A  E  C  H  Y  R  T  P  K  L  N  F  A  E  I  A  : 383
Zea_maize  : A  I  E  T  E  N  D  B  T  H  H  L  H  I  S  V  V  D  S  E  R  I  E  F  K  L  K  N  N  A  L  H  : 401
              Adgp 6 6 R E R6 6 L 46G VeE D 6 P          6 Ty1
              *                               *

GRG36      : T  H  R  A  I  N  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  : 447
B_halodura : T  H  R  A  I  N  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  : 447
B_clausii  : T  H  R  A  I  N  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  : 446
E_coli     : T  H  R  A  I  N  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  : 427
Zea_maize  : T  H  R  A  I  N  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  T  I  Q  T  E  P  S  : 444
              DHR6AM sL          6 6 DpgC Kt P YF L

```

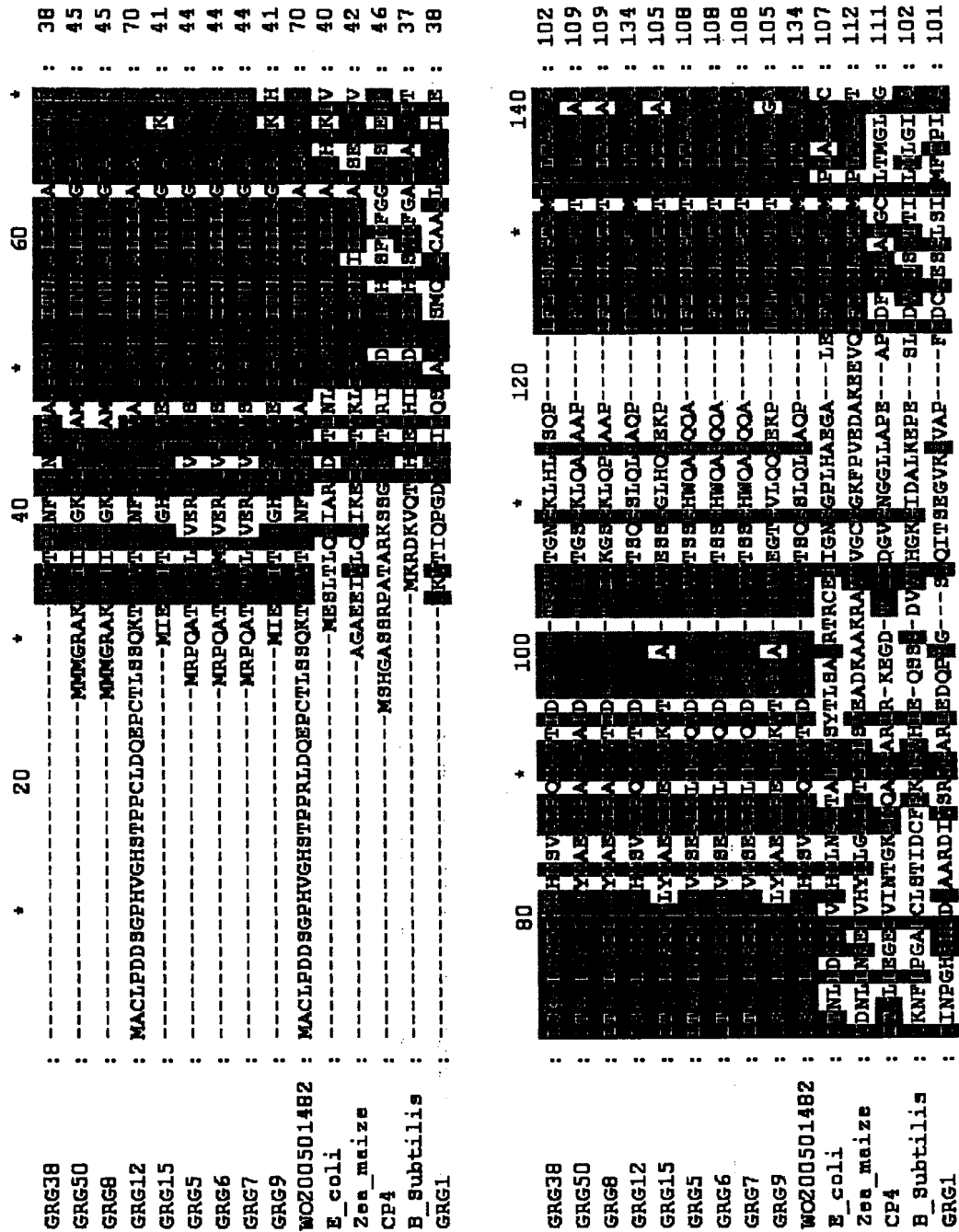


FIGURA 3B

```

GRG38      : TVE--T--T--DY--Q--T--G--AT--G--G--AT--DBP--* 160 200 240 280
GRG50      : LVE--K--V--A--H--E--G--D--A--R--S--D--A--S--T--E--* 167
GRG8       : LVD--K--V--A--H--E--G--D--A--R--S--D--A--S--E--* 174
GRG12      : TVC--T--V--E--Y--Q--H--E--G--D--A--R--S--D--A--S--E--* 174
GRG15      : LVD--A--I--E--H--E--G--D--A--R--S--D--A--S--E--* 199
GRG5       : NFE--D--F--V--E--Y--H--E--G--D--A--R--S--D--A--S--E--* 171
GRG6       : NFE--D--F--V--E--Y--H--E--G--D--A--R--S--D--A--S--E--* 173
GRG7       : NFE--D--F--V--E--Y--H--E--G--D--A--R--S--D--A--S--E--* 173
GRG9       : LVD--A--I--E--H--E--G--D--A--R--S--D--A--S--E--* 171
WO200501482 : TVC--T--V--E--Y--Q--H--E--G--AT--G--G--AT--DBP--* 199
E_coli      : LGSN--D--T--E--P--R--E--H--E--D--A--R--L--G--A--K--T--Y--L--E--O--E--N--Y--E--R--Q--* 172
Zea_maize   : AAG--MAT--Y--L--V--P--R--E--H--E--D--A--R--L--G--A--K--T--Y--L--E--O--E--N--Y--E--R--Q--* 181
CP4         : VYDF--D--B--T--F--I--A--S--T--E--P--R--E--H--E--D--A--R--L--G--A--K--T--Y--L--E--O--E--N--Y--E--R--Q--* 176
B_subtilis  : GRPF--Y--S--A--A--E--S--A--A--K--R--T--E--P--K--K--A--R--D--G--R--A--G--E--P--T--E--S--A--* 167
GRG1       : LSKS--E--T--K--S--G--S--V--T--E--D--F--F--D--E--I--P--H--K--K--S--N--O--* 164

```

```

GRG38      : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 220 260 300 340
GRG50      : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 232
GRG8       : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 239
GRG12      : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 239
GRG15      : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 264
GRG5       : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 236
GRG6       : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 238
GRG7       : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 238
GRG9       : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 236
WO200501482 : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 264
E_coli      : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 238
Zea_maize   : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 248
CP4         : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 240
B_subtilis  : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 226
GRG1       : VAS--D--L--C--E--A--P--E--A--T--K--D--E--H--E--G--D--A--R--S--D--A--S--E--* 233

```

FIGURA 3C

	300	320	340
GRG38	...V...R-...VAAQD...QALAQ...H...Q		281
GRG50	...V...K-...TEAEQ...YDSR...H...P		288
GRG8	...V...K-...TEAEQ...YDSR...H...P		288
GRG12	...V...R-...VAAQD...QALAQ...N...Q		313
GRG15	...V...R-...TPADK...HEHAQ...H...P		285
GRG5	...A...D-...VASDA...ILSQ...IAS...N...P		287
GRG6	...A...N-...VASDA...ILSQ...IDS...N...P		287
GRG7	...A...E-...VASNA...ILSQ...IAS...N...P		287
GRG9	...V...R-...TPADK...YDSR...H...P		285
WO2000501482	...V...R-...VAAQD...QALAQ...H...Q		313
E_coli	...K...T-...KTEGIERNS...G...LIFAD...EREGATICWGDDYIS---	CTRG---	299
Zea_maze	...T-...TEGCGTTSLS...GVFAF...EMNGAKVTMTTETSVT---	VTGPPRPF---	313
CP4	...V...S-...TILNVLMNPTRTGILIT...	QEMGADIEVINPALAGSEDVADLRVS	308
B_subtilis	...V...S-...RVLVKNVGLNPTRTGILIT...	QNMGALEIKPSADSGAEFYGDLLIET	294
GRG1	...V...S-...RGGLDIAS...A...RAIVQAMBANAGIAIDAKEIKLHP---		292

	360	*	380	*	400	*	420													
GRG38	---	TV	HT	ET	NT	ET	HT	QD	HD	MA	RP	GL	AT	ES	: 343					
GRG50	---	V	C	P	BT	V	ET	VT	ET	BT	SS	SR	VP	GT	ES	: 350				
GRG8	---	V	C	M	ET	V	EM	VP	ET	VT	ET	SS	SR	VP	GT	ES	: 350			
GRG12	---	TV	P	ET	NT	ET	NT	ET	NT	ET	QD	HD	NE	RP	GL	AT	ES	: 375		
GRG15	---	ET	HT	ET	BT	ET	HT	ET	BT	ET	ST	NE	RD	GL	AT	ES	: 347			
GRG5	---	V	C	M	ET	V	EM	VP	ET	VT	ET	SS	SR	CA	AP	GL	AT	ES	: 349	
GRG6	---	V	C	M	ET	V	EM	VP	ET	VT	ET	SS	CD	CA	AP	GL	AT	ES	: 349	
GRG7	---	V	C	M	ET	V	EM	VP	ET	VT	ET	SS	SN	CA	AP	GL	AT	ES	: 349	
GRG9	---	ET	HT	ET	BT	ET	HT	ET	BT	ET	ST	NE	RD	GL	AT	ES	: 347			
WQ200501482	---	TV	HT	ET	NT	ET	NT	ET	NT	ET	QD	HD	NE	RP	GL	AT	ES	: 375		
E_coli	---	---	ET	MA	ET	TA	ET	AG	TT	ET	MA	ET	TA	ET	AG	TT	ET	MA	ET	: 363
Zea_maze	:	---	ET	MA	ET	TA	ET	AG	TT	ET	MA	ET	TA	ET	AG	TT	ET	MA	ET	: 381
Cp4	:	FG	RR	HE	KA	ID	MA	KE	VA	LA	AD	TA	IR	SM	TE	MA	IR	TE	MA	: 376
B_subtilis	:	ST	KG	TV	PD	RA	PS	ET	ET	ET	ET	ET	ET	ET	ET	ET	ET	ET	ET	: 362
GRG1	:	SS	KA	VE	IG	DI	PR	ET	ET	ET	ET	ET	ET	ET	ET	ET	ET	ET	ET	: 357
	:	---	AD	NA	PE	A	DC	P	ET	VA	SS	CK	GE	ET	IR	SR	LA	HE	SS	: 357

FIGURA 3D

```

GRG38      : 440 460 480
GRG50      : 408
GRG8        : 415
GRG12       : 415
GRG15       : 440
GRG5        : 412
GRG6        : 414
GRG7        : 414
GRG9        : 412
WO200501482 : 440
E_coli      : 440
Zea_maiZe   : 425
CP4         : 443
B_subtilis  : 445
GRG1        : 426
              : 420

```

```

GRG38      : 500
GRG50      : 412
GRG8        : 419
GRG12       : 419
GRG15       : 444
GRG5        : 425
GRG6        : 418
GRG7        : 418
GRG9        : 418
WO200501482 : 425
E_coli      : 444
Zea_maiZe   : 427
CP4         : 444
B_subtilis  : 455
GRG1        : 428
              : 431

```

RESUMO

MOLÉCULA DE ÁCIDO NUCLÉICO E POLIPEPTÍDEO ISOLADOS, VETOR, CÉLULA HOSPEDEIRA, PLANTA E SEMENTE TRANSGÊNICAS E MÉTODOS PARA PRODUIR UM POLIPEPTÍDEO COM ATIVIDADE DE RESISTÊNCIA A HERBICIDAS, PARA CONFERIR RESISTÊNCIA A UM HERBICIDA EM UMA PLANTA E PARA GERAR UM POLINUCLEOTÍDEO VARIANTE DE UM POLINUCLEOTÍDEO PARENTAL GRG33, GRG35, GRG36, GRG37, GRG38, GRG39, OU GRG50

10 São providos composições e métodos para conferir resistência a herbicidas para bactérias, plantas, células, tecidos e sementes vegetais. Composições incluem moléculas de ácidos nucleicos codificando polipeptídeos de resistência ou tolerância a herbicidas, vetores
15 compreendendo tais moléculas de ácidos nucleicos e células hospedeiras compreendendo os vetores. As seqüências de nucleotídeos da invenção podem ser usadas em construções de DNA ou cassetes de expressão para transformação e expressão em organismos, incluindo microrganismos e plantas.
20 Composições também compreendem bactérias, plantas, células tecidos e semente vegetais transformados. Em particular, a presente invenção provê moléculas de ácidos nucleicos isoladas compreendendo a seqüência de nucleotídeos apresentada em SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15,
25 17, 18, 20, 21, ou 23, uma seqüência de nucleotídeos codificando a seqüência de aminoácidos mostrada em SEQ ID NO: 2, 5, 8, 11, 14, 16, 19, ou 22, a seqüência de nucleotídeos de resistência a herbicidas depositada em um

hospedeiro bacteriano como Nos. de Acesso NRRL B-30932, B-30933, B-30934, B-30945, B-30946, B-30947, ou B-30948, assim como variantes e fragmentos dessas.