

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

(21) **PI0707787-4 A2**

(22) Data de Depósito: 13/02/2007
(43) Data da Publicação: 10/05/2011
(RPI 2105)



(51) *Int.Cl.:*
C12Q 1/68
C40B 40/06

(54) Título: **POLIPEPTÍDEOS DE LIGAÇÃO QUIMÉRICOS DE PLANTA PARA RECONHECIMENTO MOLECULAR UNIVERSAL**

(30) Prioridade Unionista: 13/02/2006 US 60/773,086

(73) Titular(es): Divergence, INC

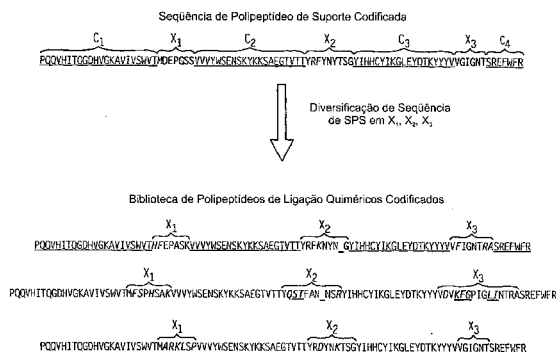
(72) Inventor(es): Jennifer Jones

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

(86) Pedido Internacional: PCT US2007003937 de 13/02/2007

(87) Publicação Internacional: WO 2007/095300 de 23/08/2007

(57) Resumo: POLIPEPTÍDEOS DE LIGAÇÃO QUIMÉRICOS DE PLANTA PARA RECONHECIMENTO MOLECULAR UNIVERSAL A presente invenção refere-se a bibliotecas de ácido nucléicos que codificam polipeptídeos de ligação quiméricos baseados em sequências de polipeptídeo de estrutura de planta. Também descritos são métodos para geração das bibliotecas.



Relatório Descritivo da Patente de Invenção para "**POLIPEPTÍ-
DEOS DE LIGAÇÃO QUIMÉRICOS DE PLANTA PARA RECONHECIMENTO
MOLECULAR UNIVERSAL**".

Antecedentes

5 A especificidade e afinidade de ligação de uma proteína para um alvo são determinadas principalmente pela seqüência de aminoácido de proteína dentro de uma ou mais regiões de ligação. Conseqüentemente, a variação da seqüência de aminoácido das regiões relevantes reconfigura as propriedades de ligação da proteína.

10 Na natureza, mudanças combinatoriais na ligação de proteína são melhores ilustradas pela vasta série de imunoglobulinas produzidas pelo sistema imune. Cada imunoglobulina inclui um conjunto de seqüências de aminoácido curtas, virtualmente únicas, conhecidas como regiões hipervariáveis (isto é, domínios de ligação de proteína), e outro conjunto de seqüên-
15 cias mais longas, não-variantes, conhecidas como regiões constantes. As regiões constantes formam folhas β que estabilizam as três estruturas tridimensionais da proteína apesar da enorme diversidade de seqüência entre regiões hipervariáveis na população de imunoglobulinas. Cada conjunto de regiões hipervariáveis confere especificidade e afinidade de ligação. O con-
20 junto de duas imunoglobulinas de cadeia pesada e duas imunoglobulinas de cadeia leve em um complexo de proteína grande (isto é, um anticorpo), aumentam adicionalmente o número de combinações com atividades de ligação diversas.

 A diversidade de ligação de anticorpos tem sido bem-sucedida-
25 mente explorada em muitas aplicações biomédicas e industriais. Por exemplo, bibliotecas foram construídas que expressam imunoglobulinas suportando regiões hipervariáveis artificialmente diversificadas. As bibliotecas de expressão de imunoglobulina são muito úteis para identificar anticorpos de alta afinidade a uma molécula-alvo (por exemplo, um receptor ou um ligante de
30 receptor). Um ácido nucléico que codifica a imunoglobulina identificada pode, então, ser isolado e expresso em células hospedeiras ou organismos.

 Contudo, apesar da utilidade de imunoglobulinas e anticorpos

em geral, sua expressão em plantas transgênicas pode ser problemática. As imunoglobulinas não podem se desdobrar corretamente em citoplasma de planta porque elas requerem a formação de ligações de dissulfeto múltiplas. Adicionalmente, o grande tamanho das imunoglobulinas impede seu enten-

5 dimento efetivo por algumas pestes de planta. Desse modo, as imunoglobulinas não são freqüentemente úteis como pesticidas de proteína ou moléculas que tem como alvo pesticidas. Finalmente, proteínas que expressam mamíferos, tais como imunoglobulinas (por exemplo), conforme assim de-

10 nominadas "anticorpos de planta" em plantas comestíveis também elevam a emissão potencial de aceitação do consumidor e é, desse modo, um impedimento à comercialização. Isto pode eficazmente impedir o uso de anticorpos de planta para muitas traços de entrada e saída em plantas transgênicas.

As desvantagens acima mencionadas de imunoglobulinas po-

15 dem ser evitadas pela geração de diversas bibliotecas de proteínas de ligação de outras classes de proteínas estruturalmente tolerantes, preferivelmente proteínas derivadas de planta. Estas bibliotecas podem ser classifica-

das para identificar proteínas individuais que se liga com especificidade e afinidade desejadas a um alvo de interesse. Em seguida, as proteínas de

20 ligação identificadas podem ser eficazmente expressas em plantas transgênicas.

SUMÁRIO

Diversas bibliotecas de ácidos nucléicos que codificam polipeptídeos de ligação quiméricos de planta, bem como métodos para gerá-las,

25 são aqui descritos. Os polipeptídeos de ligação quiméricos são conceitualmente análogos a imunoglobulinas em que elas caracterizam domínios de ligação altamente variados na estrutura de seqüências não-variantes que codificam uma proteína estruturalmente robusta. Contudo, os polipeptídeos de ligação quiméricos aqui descritos têm a vantagem considerável de serem

30 derivados de seqüências de proteína de planta, evitando, desse modo, muitos dos problemas associados com expressão de imunoglobulina em plantas. As seqüências de aminoácido das proteínas de ligação quiméricas de

planta codificada são derivadas de uma sequência de polipeptídeo de estrutura principal que inclui subsequências a serem variadas. As subsequências variadas correspondem aos domínios de ligação putativos dos polipeptídeos de ligação quiméricos de planta, e são altamente heterogêneos na biblioteca de proteínas de ligação quiméricas de planta codificada. Em contraste, a sequência das proteínas de ligação quiméricas codificadas fora das subsequências variadas é essencialmente a mesma conforme a sequência de polipeptídeo de estrutura principal de origem, e altamente homogênea através de toda a biblioteca de proteínas de ligação quiméricas de planta codificada.

10 Tais bibliotecas podem servir como uma plataforma de reconhecimento molecular universal para selecionar proteínas com ligação de alta seletividade e afinidade para expressão em plantas transgênicas.

Conseqüentemente, um aspecto aqui descrito é uma biblioteca de moléculas de ácido nucléico que codificam pelo menos dez (por exemplo, pelo menos 1.000, 10^5 ou 10^6) polipeptídeos de ligação quiméricos diferentes.

15 A sequência de aminoácido de cada polipeptídeo inclui $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, onde C_1-C_4 são subsequências de estrutura principal selecionadas de fosfatase ácida púrpura (isto é, SEQ ID NOs: 1-30, 31-60, 61-90 e 91-120, respectivamente) que podem incluir até 30 (por exemplo, 20, 10 ou 5) substituições, retiradas, inserções ou adições de aminoácido simples às sequências

20 de fosfatase ácida púrpura. As subsequências C_1-C_4 são homogêneas através de muitos dos polipeptídeos codificados na biblioteca. Em contraste às subsequências de estrutura principal C_1-C_4 , as subsequências de X_1-X_3 são subsequências independentes variáveis de 2-20 aminoácidos, e estas subsequências são heterogêneas através de muitos dos polipeptídeos na biblioteca. Por exemplo, a biblioteca de polipeptídeos quiméricos pode ter a sequência de aminoácidos de qualquer uma de SEQ ID Nos: 124-126, incluindo uma a dez substituições, retiradas, inserções ou adições de aminoácido simples para posições de aminoácido correspondendo a 23-39, 51-49 e 79-84

25 de SEQ ID Nos: 124-126.

30 Outro aspecto aqui descrito é um método para gerar a biblioteca aqui descrita. O método inclui provisão de um ácido nucléico parenteral que

codifica uma seqüência de polipeptídeo de estrutura de planta contendo $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, conforme definido acima. O método inclui adicionalmente replicação do ácido nucléico parenteral (por exemplo, pelo menos uma das subseqüências de X_1-X_3 é selecionada de SEQ ID Nos: 121-123) sob condições que introduzem até 10 substituições, retiradas, inserções ou adições de aminoácido simples às subseqüências X_1 , X_2 ou X_3 parenterais, pelo que uma população heteróloga de subseqüências aleatoriamente variadas que codificam X_1 , X_2 ou X_3 é gerada. A população de subseqüências variadas é, em seguida, substituída em uma população de ácidos nucléicos parenterais nas posições correspondentes àquelas que codificam X_1 , X_2 ou X_3 . As substituições, retiradas, inserções ou adições de aminoácido podem ser introduzidas nas subseqüências de ácido nucléico parenteral por replicação *in vitro* (por exemplo, usando-se uma polimerase mutagênica purificada ou análogos de nucleotídeo), ou *in vivo* (por exemplo, em uma cepa mutagênica de *E. coli*). A biblioteca antes descrita pode ser introduzida em um sistema de replicação biológico (por exemplo, *E. coli* ou bacteriófago), e amplificada.

Um aspecto relacionado aqui descrito é outro método para gerar a biblioteca acima descrita de ácidos nucléicos. O método inclui seleção de uma seqüência de aminoácido contendo $C_1-X_1-C_2-X_2-C_3-X_3-C_4$ conforme definido acima. O método inclui adicionalmente provisão de um primeiro e um segundo conjunto de oligonucleotídeos, tendo seqüências complementares de sobreposição. Os oligonucleotídeos do primeiro conjunto codificam as subseqüências C_1-C_4 e subseqüências X_1-X_3 heterogêneas múltiplas. Os oligonucleotídeos do segundo conjunto são complementares às seqüências de nucleotídeo que codificam as subseqüências C_1-C_4 e subseqüências X_1-X_3 heterogêneas múltiplas. Os dois conjuntos de oligonucleotídeos são combinados para formar uma primeira mistura e incubados sob condições que permitem hibridização das seqüências complementares de sobreposição. As seqüências hibridizadas resultantes são, em seguida, estendidas para formar uma segunda mistura contendo a biblioteca acima descrita.

Ainda outro aspecto da invenção é uma biblioteca de ácidos nu-

cléicos que codificam polipeptídeos de ligação quiméricos, cada um dos quais incluindo uma seqüência de aminoácidos pelo menos 70% (isto é, qualquer percentagem entre 70% e 100%) idêntica a qualquer de SEQ ID Nos: 127-129. A seqüência de aminoácidos de cada um dos polipeptídeos
5 codificados inclui aminoácidos que diferem daqueles de SEQ ID Nos:127-129 nas posições 14, 15, 33, 35-36, 38, 47-48, 66, 68-69, 71, 80, 81, 99, 101-102 e 104, e as diferenças de aminoácido são heterogêneas através de uma pluralidade dos polipeptídeos codificados. A seqüência de aminoácidos de cada um dos polipeptídeos codificados fora da posição acima listada é
10 homogênea através de uma pluralidade dos polipeptídeos quiméricos codificados.

Um aspecto relacionado aqui descrito é um método para gerar a biblioteca aqui descrita. O método inclui seleção de uma seqüência de aminoácido correspondente a qualquer uma de SEQ ID Nos:127-129, em que a
15 seqüência selecionada difere de SEQ ID Nos:127-129 em pelo menos uma das posições acima mencionadas. O método inclui adicionalmente provisão de um primeiro e um segundo conjunto de oligonucleotídeos tendo seqüências complementares de sobreposição. Os oligonucleotídeos do primeiro conjunto codificam subsequências da seqüência de aminoácido, as subsequências sendo heterogêneas nas posições acima mencionadas. Os oligo-
20 nucleotídeos do segundo conjunto são complementares às seqüências de nucleotídeo que codificam subsequências da seqüência de aminoácido selecionada, as subsequências sendo heterogêneas nas posições acima mencionadas. Os dois conjuntos de oligonucleotídeos são combinados para formar uma primeira mistura, e incubada sob condições que permitem hibridi-
25 zação das seqüências complementares de sobreposição. As seqüências hibridizadas resultantes são, em seguida, estendidas para formar uma segunda mistura contendo a biblioteca acima descrita.

Várias implementações da invenção podem incluir um ou mais
30 dos seguintes. Por exemplo, cada ácido nucléico em uma biblioteca pode incluir uma seqüência de vetor. Também caracterizado é qualquer ácido nucléico isolado de uma das bibliotecas acima descritas, bem como o polipep-

tídeo de ligação quimérico codificado por ela, na forma pura.

Em uma implementação, uma população de células (ou células individuais selecionadas a partir da população de células) é provida que expressa polipeptídeos de ligação quiméricos codificados por uma das bibliotecas. Outra implementação caracteriza uma biblioteca de polipeptídeos de ligação quiméricos codificados por uma das bibliotecas de ácido nucléico. Ainda outra implementação proporciona uma população de fago filamentoso revelando os polipeptídeos de ligação quiméricos codificados por uma das bibliotecas de ácido nucléico.

Em várias implementações de métodos para geração das bibliotecas de ácido nucléico acima descritas por conjunto de oligonucleotídeo, um ou mais dos seguintes podem ser incluídos. Por exemplo, o método pode adicionalmente incluir, após a segunda mistura que contém a biblioteca de ácido nucléico ser gerada, realização de um ciclo de desnaturação da população de ácidos nucléicos, seguida por uma hibridização e uma etapa de alongamento. Opcionalmente, este ciclo pode ser repetido (por exemplo, até 100 vezes). As bibliotecas de ácido nucléico podem ser uma reação de cadeia de polimerase que inclui um iniciador de avanço e reverso que hibridiza para as seqüências 5' e 3', respectivamente, de todos os ácidos nucléicos na biblioteca. Em uma implementação, aminoácidos a serem codificados em posições de seqüência variáveis são selecionados de um subconjunto (por exemplo, somente 4, 6, 8, 10, 12, 14 ou 16) de alanina, arginina, asparagina, aspartato, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina, cisteína e valina (os 20 aminoácidos ocorrem naturalmente). Em outros casos, 19 dos 20 aminoácidos (excluída cisteína). Em outros casos, todos os 20 são usados. Em outra implementação, o subconjunto de aminoácidos inclui pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático (por exemplo, alanina, aspartato, serina e tirosina).

É descrita aqui uma biblioteca de ácidos nucléicos que codificam pelo menos dez polipeptídeos diferentes, a seqüência de aminoácidos de

cada polipeptídeo compreendendo:

5 $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual (i) subsequência C_1 é selecionada de SEQ ID NOS:1-30, subsequência C_2 é selecionada de SEQ ID NOS:31-60, subsequência C_3 é selecionada de SEQ ID NOS:61-90, subsequência C_4 é selecionada de SEQ ID NOS:91-120, e cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; (ii) C_1-C_4 são homogêneas através de uma pluralidade dos polipeptídeos codificados; (iii) cada uma de X_1-X_3 é uma subsequência independentemente variável consistindo em 2-20 aminoácidos; e
10 cada uma de X_1-X_3 são heterogêneas através de uma pluralidade dos polipeptídeos codificados.

Também é descrita uma biblioteca de ácidos nucleicos que codificam pelo menos dez polipeptídeos diferentes, a sequência de aminoácidos de cada polipeptídeo compreendendo:

15 $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual: (i) subsequência C_1 é selecionada de figura 2 ou figura 4, subsequência C_2 é selecionada de figura 2 ou figura 4, subsequência C_3 é selecionada de figura 2 ou figura 4; subsequência C_4 é selecionada de figura 2 ou figura 4, e cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples
20 à subsequência selecionada; (ii) C_1-C_4 são homogêneas através de uma pluralidade dos polipeptídeos codificados; (iii) cada uma de X_1-X_3 é uma subsequência independentemente variável consistindo em 2-20 aminoácidos; e cada uma de X_1-X_3 são heterogêneas através de uma pluralidade dos polipeptídeos codificados.

25 Também descrita é uma biblioteca de ácidos nucleicos que codificam pelo menos dez polipeptídeos diferentes, a sequência de aminoácidos de cada polipeptídeo compreendendo:

30 $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual (i) subsequência C_1 é selecionada de figura 3 ou figura 5, subsequência C_2 é selecionada de figura 3 ou figura 5, subsequência C_3 é selecionada de figura 3 ou figura 5; subsequência C_4 é selecionada de figura 3 XX, e cada uma de C_1-C_4 compreende até 30 substituições, retiradas, inserções ou adições de aminoácido simples à subse-

qüência selecionada; (ii) C_1 - C_4 são homogêneas através de uma pluralidade dos polipeptídeos codificados; (iii) cada uma de X_1 - X_3 é uma subsequência independentemente variável consistindo em 2-20 aminoácidos; e cada uma de X_1 - X_3 são heterogêneas através de uma pluralidade dos polipeptídeos codificados.

Em várias concretizações: pelo menos 1.000 polipeptídeos diferentes são codificados; pelo menos 100.000 polipeptídeos diferentes são codificados; pelo menos 1.000.000 de polipeptídeos diferentes são codificados; cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; nenhuma de C_1 - C_4 compreende substituições, retiradas, inserções ou adições de aminoácido à subsequência selecionada; aminoácidos de X_1 - X_3 são selecionados de pouco menos do que 20 aminoácidos geneticamente codificados em plantas; aminoácidos de X_1 - X_3 são selecionados de todos 20 aminoácidos geneticamente codificados em plantas; os pouco menos do que 20 aminoácidos geneticamente codificados incluem pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático; pouco menos do que 20 aminoácidos geneticamente codificados compreendem alanina, aspartato, serina e tirosina.

Em alguns casos: a sequência de aminoácidos de cada polipeptídeo é selecionada de:

- (a) um polipeptídeo compreendendo C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 , no qual C_1 = SEQ ID NO:1, C_2 = SEQ ID NO:31, C_3 = SEQ ID NO:61, e C_4 = SEQ ID NO:91;
- (b) um polipeptídeo compreendendo C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 , no qual C_1 = SEQ ID NO:2, C_2 = SEQ ID NO:32, C_3 = SEQ ID NO:62, e C_4 = SEQ ID NO:92; e

(c) um polipeptídeo compreendendo $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual $C_1 = \text{SEQ ID NO:3}$, $C_2 = \text{SEQ ID NO:33}$, $C_3 = \text{SEQ ID NO:63}$, e $C_4 = \text{SEQ ID NO:93}$.

Em alguns casos: cada polipeptídeo codificado compreende
 5 $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual $C_1 = \text{SEQ ID NO: X1}$, $C_2 = \text{SEQ ID NO: X2}$, $C_3 = \text{SEQ ID NO: X3}$, e $C_4 = \text{SEQ ID NO: X4}$; SEQ designada. ID NO:130.

Em alguns casos, cada polipeptídeo codificado compreende
 $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual $C_1 = \text{SEQ ID NO: X1}$, $C_2 = \text{SEQ ID NO: X2}$, $C_3 = \text{SEQ ID NO: X3}$, e $C_4 = \text{SEQ ID NO: X4}$; designada SEQ. ID NO:130.

10 Em algumas concretizações: no qual cada um dos ácidos nucleicos compreende uma seqüência de vetor.

Também descritos: são um ácido nucléico isolado selecionado a partir da biblioteca e uma célula isolada que expressa o ácido nucléico, bem como uma biblioteca purificada de polipeptídeos purificados codificados pela
 15 biblioteca; e uma população de fago filamentoso que revela os polipeptídeos codificados pela biblioteca.

É descrito aqui um método de gerar uma biblioteca, compreendendo: (i) provisão de um ácido nucléico parenteral que codifica um polipeptídeo parenteral compreendendo a seqüência de aminoácido: $C_1-X_1-C_2-X_2-$
 20 $C_3-X_3-C_4$, no qual a subseqüência C_1 é selecionada de SEQ ID NOS:1-30, a subseqüência C_2 é selecionada de SEQ ID NOS:31-60, a subseqüência C_3 é selecionada de SEQ ID NOS:61-90, a subseqüência C_4 é selecionada de SEQ ID NOS:91-120, cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; e cada X_1-X_3 é uma subseqüência independente consistindo em 2-
 25 20 aminoácidos; (ii) replicação do ácido nucléico parenteral sob condições que introduzem até 10 substituições, retiradas, inserções ou adições de aminoácido simples às subseqüências X_1 , X_2 ou X_3 , pelo que uma população de seqüências aleatoriamente variadas que codificam X_1' , X_2' ou X_3' é gerada; e
 30 (iii) a população de seqüências aleatoriamente variadas X_1' , X_2' ou X_3' é substituída, em uma população de ácidos nucleicos parenterais nas posições correspondentes àquelas que codificam X_1' , X_2' ou X_3' .

Em vários exemplos: pelo menos uma das subsequências X_1 - X_3 é selecionada de SEQ ID Nos:121-123; cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; nenhuma de C_1 - C_4 compreende substituições, retiradas, inserções ou adições de aminoácido à subsequência selecionada; a replicação gera uma população heterogênea de subsequências aleatoriamente variadas pela introdução de até 5 substituições de aminoácido em cada uma de X_1 , X_2 ou X_3 ; o método compreende adicionalmente amplificar a biblioteca pela introdução da mesma em um sistema de replicação biológico e proliferação do sistema de replicação biológico; o sistema de replicação biológico é uma pluralidade de células de *E. coli*; o sistema de replicação biológico é uma pluralidade de bacteriófago; a replicação ocorre *in vitro*; a replicação é realizada com uma polimerase mutagênica purificada; a replicação é realizada na presença de um análogo de nucleotídeo; a replicação ocorre *in vivo*; a replicação *in vivo* ocorre em uma espécie mutagênica de *E. coli*.

Também descrito é um método de gerar a biblioteca de acordo com a reivindicação 1, compreendendo: (i) seleção de uma seqüência de aminoácido compreendendo a seqüência de aminoácido: C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 a ser codificada, no qual: (a) a subsequência C_1 é selecionada de SEQ ID NOS:1-30, a subsequência C_2 é selecionada de SEQ ID NOS:31-60, a subsequência C_3 é selecionada de SEQ ID NOS:61-90, a subsequência C_4 é selecionada de SEQ ID NOS:91-120; (b) cada uma de C_1 - C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; (c) cada X_1 , X_2 e X_3 consiste em uma subsequência de aminoácido de 2-20 aminoácidos de comprimento; (ii) provisão de uma primeira pluralidade e uma segunda pluralidade de oligonucleotídeos, no qual (a) oligonucleotídeos da primeira pluralidade codificam as sub-

seqüências C_1 - C_4 e subseqüências variantes X_1 - X_3 heterogêneas múltiplas X_1' - X_3' ; (b) oligonucleotídeos da segunda pluralidade são complementares às seqüências de nucleotídeo que codificam as subseqüências C_1 - C_4 e às seqüências de nucleotídeo que codificam as subseqüências X_1' - X_3' heterogêneas múltiplas; e (c) oligonucleotídeos das primeira e segunda pluralidades têm seqüências de sobreposição uma com a outra; (iii) combinação da população de oligonucleotídeos para formar uma primeira mistura; (iv) incubação da mistura sob condições eficazes para hibridização das seqüências complementares de sobreposição para formar uma pluralidade de seqüências complementares hibridizadas; e (v) alongamento da pluralidade de seqüências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

Em várias concretizações: cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; cada uma de C_1 - C_4 compreende independentemente de zero até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; o método compreende adicionalmente realizar um ciclo de etapas, o ciclo de etapas compreendendo desnaturação da biblioteca pelo aumento da temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento duplo, seguida pelas etapas (iv) e (v); o método compreende repetição do ciclo de etapas até 100 vezes; o método compreende adicionalmente amplificação da biblioteca por uma reação de cadeia de polimerase consistindo essencialmente da biblioteca, um iniciador de avanço, e um iniciador reverso, no qual os primers de avanço e reverso podem hibridizar para as seqüências terminais 5' e 3', respectivamente, de todos os ácidos nucleicos na biblioteca; o aminoácido a ser codificado em cada posição das subseqüências X_1 , X_2 ou X_3 , é selecionado de um subconjunto de alanina, arginina, asparagina, aspartato, cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, seri-

na, treonina, triptofan, tirosina e valina; aqui o aminoácido selecionado para cada substituição de aminoácido simples é selecionado de um grupo de aminoácidos consistindo em pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático; e o grupo de aminoácidos consiste em alanina, aspartato, serina e tirosina.

Também descrito aqui é um método de gerar uma biblioteca, compreendendo: (i) provisão de um ácido nucléico parenteral que codifica um polipeptídeo parenteral compreendendo a seqüência de aminoácido: $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual C_1 é selecionada de figura 2 ou figura 4, a subsequência C_2 é selecionada de figura 2 ou figura 4, a subsequência C_3 é selecionada de figura 2 ou figura 4, a subsequência C_4 é selecionada de figura 2 ou figura 4, cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; e cada uma de X_1-X_2 é uma subsequência independente consistindo em 2-10 aminoácidos; (ii) replicação do ácido nucléico parenteral sob condições que introduzem até 10 substituições, retiradas, inserções ou adições de aminoácido simples às subsequências de X_1 , X_2 , ou X_3 , pelo que uma população de subsequências aleatoriamente variadas que codificam X_1' , X_2' , ou X_3' é gerada; e (iii) a população de subsequências aleatoriamente variadas X_1' , X_2' , ou X_3' é substituída, em uma população de ácidos nucléicos parenterais nas posições correspondentes àquelas que codificam X_1 , X_2 , ou X_3 .

Em várias concretizações: pelo menos uma das subsequências X_1-X_2 é selecionada de SEQ ID Nos:121-123; cada uma de C_1-C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1-C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1-C_4 compreende independentemente até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; nenhuma de C_1-C_4 compreende substituições, retiradas, inserções ou

adições de aminoácido; a replicação gera uma população heterogênea de subsequências aleatoriamente variadas pela introdução de até 5 substituições de aminoácido em cada uma de X_1 , X_2 , ou X_3 , o método compreende adicionalmente amplificar a biblioteca pela introdução da mesma em um sistema de replicação biológico e proliferação do sistema de replicação biológico; o sistema de replicação biológico é uma pluralidade de células de *E. coli*; o sistema de replicação biológico é uma pluralidade de bacteriófago; a replicação ocorre *in vitro*; a replicação é realizada com uma polimerase mutagênica purificada; a replicação é realizada na presença de um análogo de nucleotídeo; a replicação ocorre *in vivo*; a replicação *in vivo* ocorre em uma espécie mutagênica de *E. coli*.

Também descrito é um método de gerar a biblioteca, compreendendo: (i) seleção de uma seqüência de aminoácido compreendendo C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 a ser codificada, no qual (a) a subsequência C_1 é selecionada de figura 2 ou figura 4, a subsequência C_2 é selecionada de figura 2 ou figura 4, a subsequência C_3 é selecionada de figura 2 ou figura 4, a subsequência C_4 é selecionada de figura 2 ou figura 4; (b) cada uma de C_1 - C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; (c) e cada uma de X_1 , X_2 , e X_3 consiste em uma seqüência de aminoácido de 2-20 aminoácidos de comprimento; (ii) provisão de uma primeira pluralidade e um segunda pluralidade de oligonucleotídeos, no qual (a) oligonucleotídeos da primeira pluralidade codificam as subsequências C_1 - C_4 e subsequências variantes X_1 - X_3 heterogêneas múltiplas X_1' - X_3' ; (b) oligonucleotídeos da segunda pluralidade são complementares às seqüências de nucleotídeo que codificam as subsequências C_1 - C_4 e às seqüências de nucleotídeo que codificam as subsequências X_1' - X_3' heterogêneas múltiplas; e (c) oligonucleotídeos das primeira e segunda pluralidades têm seqüências de sobreposição uma com a outra; (iii) combinação da população de oligonucleotídeos para formar uma primeira mistura; (iv) incubação da mistura sob condições eficazes para hibridização das seqüências complementares de sobreposição para formar uma pluralidade de seqüências complementares hibridizadas; e (v) alongamento da

pluralidade de seqüências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

Em vários casos: cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente de zero até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; o método compreende adicionalmente realizar um ciclo de etapas, o ciclo de etapas compreendendo desnaturação da biblioteca pelo aumento da temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento duplo, seguida pelas etapas (iv) e (v); o método compreende adicionalmente repetir o ciclo de etapas até 100 vezes; o método compreende adicionalmente amplificação da biblioteca por uma reação de cadeia de polimerase consistindo essencialmente da biblioteca, um iniciador de avanço, e um iniciador reverso, no qual os priméres de avanço e reverso podem hibridizar para as seqüências terminais 5' e 3', respectivamente, de todos os ácidos nucléicos na biblioteca; o aminoácido a ser codificado em cada posição das subsequências X_1 , X_2 ou X_3 , é selecionado de um subconjunto de alanina, arginina, asparagina, aspartato, cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina e valina; o aminoácido selecionado para cada substituição de aminoácido simples é selecionado de um grupo de aminoácidos consistindo em pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático; e o grupo de aminoácidos consiste em alanina, aspartato, serina e tirosina.

Também é descrito um método de gerar a biblioteca, compreendendo: (i) provisão de um ácido nucléico parenteral que codifica um polipeptídeo parenteral compreendendo a seqüência de aminoácido: C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 , no qual C_1 é selecionada de figura 3 ou figura 5, a subsequência

C_2 é selecionada de figura 3 ou figura 5, a subsequência C_3 é selecionada de figura 3 ou figura 5, a subsequência C_4 é selecionada de figura 3 ou figura 5, cada uma de C_1 - C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; e

5 cada uma de X_1 - X_3 é uma subsequência independente consistindo em 2-20 aminoácidos; (ii) replicação do ácido nucléico parenteral sob condições que introduzem até 10 substituições, retiradas, inserções ou adições de aminoácido simples às subsequências de X_1 , X_2 , ou X_3 , pelo que uma população de subsequências aleatoriamente variadas que codificam X_1' , X_2' , ou X_3' é gera-

10 da; e (iii) a população de subsequências aleatoriamente variadas X_1' , X_2' , ou X_3' é substituída, em uma população de ácidos nucléicos parenterais nas posições correspondentes àquelas que codificam X_1 , X_2 , ou X_3 .

Em vários exemplos: pelo menos uma das subsequências X_1 - X_3 é selecionada de SEQ ID Nos:121-123; cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de

15 aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; cada uma de C_1 - C_4 compreende independentemente até 5 substituições, retiradas, inser-

20 ções ou adições de aminoácido simples à subsequência selecionada; nenhuma de C_1 - C_4 compreende substituições, retiradas, inserções ou adições de aminoácido; a replicação gera uma população heterogênea de subsequências aleatoriamente variadas pela introdução de até 5 substituições de aminoácido em cada uma de X_1 , X_2 , ou X_3 ; o método compreende adicional-

25 mente amplificar a biblioteca pela introdução da mesma em um sistema de replicação biológico e proliferação do sistema de replicação biológico; o sistema de replicação biológico é uma pluralidade de células de *E. coli*; o sistema de replicação biológico é uma pluralidade de bacteriófago; a replicação ocorre *in vitro*; a replicação é realizada com uma polimerase mutagênica purificada;

30 a replicação é realizada na presença de um análogo de nucleotídeo; a replicação ocorre *in vivo*; a replicação *in vivo* ocorre em uma espécie mutagênica de *E. coli*.

Também descrito é um método de gerar a biblioteca compreendendo: (i) seleção de uma seqüência de aminoácido compreendendo C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 a ser codificada, no qual (a) a subseqüência C_1 é selecionada de figura 3 ou figura 5, a subseqüência C_2 é selecionada de figura 3 ou figura 5, a subseqüência C_3 é selecionada de figura 3 ou figura 5, a subseqüência C_4 é selecionada de figura 3 ou figura 5; (b) cada uma de C_1 - C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; (c) e cada uma de X_1 , X_2 , e X_3 consiste em uma seqüência de aminoácido de 2-20 aminoácidos de comprimento; (ii) provisão de uma primeira pluralidade e uma segunda pluralidade de oligonucleotídeos, no qual (a) oligonucleotídeos da primeira pluralidade codificam as subseqüências C_1 - C_4 e subseqüências variantes X_1 - X_3 heterogêneas múltiplas X_1' - X_3' ; (b) oligonucleotídeos da segunda pluralidade são complementares às seqüências de nucleotídeo que codificam as subseqüências C_1 - C_4 e às seqüências de nucleotídeo que codificam as subseqüências X_1' - X_3' heterogêneas múltiplas; e (c) oligonucleotídeos das primeira e segunda pluralidades têm seqüências de sobreposição uma com a outra; (iii) combinação da população de oligonucleotídeos para formar uma primeira mistura; (iv) incubação da mistura sob condições eficazes para hibridização das seqüências complementares de sobreposição para formar uma pluralidade de seqüências complementares hibridizadas; e (v) alongamento da pluralidade de seqüências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

Em várias concretizações: cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; cada uma de C_1 - C_4 compreende independentemente de zero até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; o método compreende adicionalmente realizar um ciclo de etapas, o ciclo de etapas compreendendo desnaturação da biblioteca pelo aumento da

temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento duplo, seguida pelas etapas (iv) e (v); o método compreende adicionalmente repetição do ciclo de etapas até 100 vezes; o método compreende adicionalmente amplificação da biblioteca por uma reação de cadeia de polimerase consistindo essencialmente da biblioteca, um iniciador de avanço, e um iniciador reverso, no qual os primers de avanço e reverso podem hibridizar para as seqüências terminais 5' e 3', respectivamente, de todos os ácido nucleicos na biblioteca; o aminoácido a ser codificado em cada posição das subseqüências X_1 , X_2 ou X_3 , é selecionado de um subconjunto de alanina, arginina, asparagina, aspartato, cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina e valina; o aminoácido selecionado para cada substituição de aminoácido simples é selecionado de um grupo de aminoácidos consistindo em pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático; e o grupo de aminoácidos consiste em alanina, aspartato, serina e tirosina.

Também descrita é uma biblioteca de ácidos nucleicos que codificam pelo menos dez polipeptídeos diferentes, no qual: (i) a seqüência de aminoácido de cada um dos polipeptídeos codificados compreende uma seqüência de aminoácido pelo menos 70% idêntica a qualquer de SEQ ID Nos:127-129; (ii) a seqüência de aminoácido de cada um dos polipeptídeos codificados inclui aminoácidos que diferem daquela de SEQ ID Nos:127-129 nas posições 14, 15, 33, 35-36, 38, 47-48, 66, 68-69, 71, 80, 81, 99, 101-102, e 104, e as diferenças de aminoácido são heterogênea através de uma pluralidade dos polipeptídeos codificados; e (iii) a seqüência de aminoácido de cada um dos polipeptídeos codificados fora dos resíduos correspondem às posições 14, 15, 33, 35-36, 38, 47-48, 66, 68-69, 71, 80, 81, 99, 101-102, e 104 de SEQ ID Nos:127-129 é homogênea através de uma pluralidade dos polipeptídeos codificados.

Em várias concretizações: a seqüência de aminoácido dos polipeptídeos tem pelo menos 75% de identidade para qualquer de SEQ ID

Nos:127-129; a seqüência de aminoácido dos polipeptídeos tem pelo menos 80% de identidade para qualquer de SEQ ID Nos:127-129; e a seqüência de aminoácido dos polipeptídeos tem pelo menos 85% de identidade para qualquer de SEQ ID Nos:127-129; cada um dos ácidos nucléicos compreende

5 uma seqüência de vetor. Também revelado: um ácido nucléico isolado que codifica um polipeptídeo selecionado a partir da biblioteca; um polipeptídeo purificado codificado pelo ácido nucléico; uma população de células que expressa os polipeptídeos codificados pela biblioteca; uma célula selecionada a partir da população de células; uma biblioteca purificada de polipeptídeos

10 codificados pela biblioteca; uma população de fago filamentosos revelando a biblioteca de polipeptídeos codificados pela biblioteca.

Também revelado é um método de gerar a biblioteca, compreendendo: (i) seleção de uma seqüência de aminoácido correspondente a qualquer uma de SEQ ID Nos:127-129 a ser codificada, no qual a seqüência

15 selecionada difere daquelas de SEQ ID Nos:127-129 em pelo menos uma de posições variáveis 14, 15, 33, 35-36, 38, 47-48, 66, 68-69, 71, 80, 81, 99, 101-102, e 104; (ii) provisão quimicamente de uma primeira e uma segunda pluralidade de oligonucleotídeos, no qual: (a) oligonucleotídeos da primeira pluralidade codificam subsequências de aminoácido da seqüência de ami-

20 noácido selecionada; as subsequências sendo heterogêneas nas posições variáveis codificadas; (b) oligonucleotídeos da segunda pluralidade são complementares às seqüências de nucleotídeo que codificam as subsequências da seqüência de aminoácido selecionada, as subsequências sendo heterogêneas nas posições variáveis codificadas; e (c) as primeira e segun-

25 da pluralidades compreendem oligonucleotídeos tendo seqüências de sobreposição uma com a outra; (iii) combinação da população de oligonucleotídeos para formar uma primeira mistura; (iv) incubação da mistura sob condições eficazes para hibridização das seqüências complementares de sobre-

30 zadas; e (v) alongamento da pluralidade de seqüências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

Em vários exemplos: o método compreende adicionalmente rea-

lizar um ciclo de desnaturação da biblioteca pelo aumento da temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento duplo, seguida pelas etapas (iv) e (v); o método compreende adicionalmente repetição do ciclo de etapas até 100 vezes; o método compre-

5 ende adicionalmente amplificação da biblioteca por uma reação de cadeia de polimerase consistindo essencialmente da biblioteca, um iniciador de avanço, e um iniciador reverso, no qual os prímeres de avanço e reverso podem hibridizar para as seqüências terminais 5' e 3', respectivamente, de todos os ácido nucléicos na biblioteca; os aminoácidos a serem codificados para as

10 posições variáveis são selecionados de um subconjunto de alanina, arginina, asparagina, aspartato, cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina e valina; os aminoácidos selecionados para as posições variáveis são selecionados de um grupo consistindo em pelo menos um aminoá-

15 cido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático; o grupo de aminoácidos consiste em alanina, aspartato, serina e tirosina.

Os detalhes de uma ou mais concretizações da invenção são colocados na descrição abaixo. Outras características, objetivos e vantagens

20 da invenção tornar-se-ão aparentes a partir da descrição e desenhos, e a partir das reivindicações.

DESCRIÇÃO DOS DESENHOS

A figura 1 é uma representação esquemática representando a geração de uma biblioteca de ácidos nucléicos que codifica polipeptídeos de

25 ligação quiméricos por diversificação de subsequências dentro de uma seqüência de estrutura principal de polipeptídeo codificado.

A figura 2 é um alinhamento das seqüências de um número de proteínas que têm regiões que podem ser usadas como um estrutura principal. Estas proteínas são homólogas a orizacistatin. As C1, C2, C3 e C4 são

30 colocadas em caixa e etiquetadas.

A figura 3 é um alinhamento das seqüências de um número de proteínas que têm regiões que podem ser usadas como um estrutura princi-

pal. Estas proteínas são homólogas a C2. Os C1, C2, C3 e C4 são colocados em caixa e etiquetados.

A figura 4 é um alinhamento das seqüências de um número de proteínas que têm regiões que podem ser usadas como um estrutura principal. Estas proteínas são homólogas a orizacistatin. As C1, C2, C3 e C4 são colocadas em caixa e etiquetadas.

A figura 5 é um alinhamento das seqüências de um número de proteínas que têm regiões que podem ser usadas como um estrutura principal. Estas proteínas são homólogas a C2. As C1, C2, C3 e C4 são colocadas em caixa e etiquetadas.

DESCRIÇÃO DETALHADA

Diversas bibliotecas de ácidos nucléicos (por exemplo, cDNA) que codificam polipeptídeos de ligação quiméricos de planta, bem como métodos para gerá-las, são aqui descritos. As seqüências de aminoácido da biblioteca de proteínas de ligação quiméricas codificadas são derivadas de uma seqüência de polipeptídeo de estrutura principal que inclui subseqüências a serem variadas. As subseqüências variadas correspondem a domínios de ligação putativos das proteínas de ligação quiméricas de planta, e são altamente heterogêneas na biblioteca de proteínas de ligação quiméricas de planta. Em contraste, a seqüência das proteínas de ligação quiméricas codificadas fora das subseqüências variadas é essencialmente a mesma conforme a seqüência de polipeptídeo de estrutura principal de origem, e altamente homogênea através de toda a biblioteca de proteínas de ligação quiméricas de planta codificada. Desse modo, bibliotecas de proteínas de ligação quiméricas de planta podem servir como uma plataforma de biblioteca de reconhecimento molecular universal para seleção de proteínas de ligação especializadas em plantas transgênicas. Bibliotecas de proteínas de ligação quiméricas de planta podem ser expressas por células transfectadas (isto é, como bibliotecas de expressão) e testadas para interação com um alvo molecular de interesse. Por exemplo, bibliotecas de expressão podem ser classificadas para identificar polipeptídeos que se ligam com alta especificidade e afinidade à polipeptídeos expressos por pestes de planta, incluindo

do nematóides. Por último, proteínas de ligação quiméricas individuais com propriedades de ligação de alvo desejadas podem ser expressas em uma planta transgênica.

I. Seqüências de Polipeptídeo de Estrutura de planta

5 Uma seqüência de polipeptídeo de estrutura de planta é uma seqüência de aminoácido baseada em uma proteína de planta que é estruturalmente tolerante de variação extrema de seqüência dentro de uma ou mais regiões. As regiões a serem variadas dentro da seqüência de polipeptídeo de estrutura principal são conceitualmente análogas às regiões hipervariáveis de imunoglobulinas, e formam domínios de ligação putativos em um polipeptídeo de ligação quimérico. Desse modo, uma biblioteca grande de seqüências de ácido nucléico que codifica polipeptídeos de ligação quiméricos de planta diversos é produzida por diversificação de seqüências específicas dentro de uma seqüência de polipeptídeo de estrutura principal, conforme é descrito em detalhe abaixo.

15 As seqüências de polipeptídeo de estrutura de planta são selecionadas para terem um número de propriedades, por exemplo, elas: (i) são derivadas de seqüências que são de origem de planta; (ii) codificam proteínas que toleram a introdução de seqüência estruturalmente diversa; (iii) somente contêm ligações de dissulfeto que não interferem com desdobramento do polipeptídeo quando expresso em uma planta; (iv) expressam em altos níveis em tecidos de planta diversos; e (v) podem ser alvo para localizações subcelulares diferentes (por exemplo, citoplasma, mitocôndria, plastídeo), ou secretadas a partir da célula. Baseado nestas propriedades, as seqüências de polipeptídeo de estrutura de planta permitem a regeneração de grandes quantidades de polipeptídeos de ligação quiméricos com atividades de ligação altamente diversas. Bibliotecas de polipeptídeos de ligação quiméricos podem ser classificadas para ligação a uma molécula-alvo. Proteínas de ligação quiméricas tendo a atividade de ligação desejada podem subseqüentemente serem expressas em plantas para conferir traços de entrada (por exemplo, resistência à peste ou patogenia, tolerância à secura), ou traços de saída (por exemplo, composição de lipídeo modificada, ligação de metal pe-

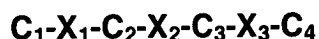
sado para fitoremediação, usos medicinais). Tais proteínas de ligação podem também serem usadas em várias especificações baseadas em afinidade, por exemplo, detecção de diagnóstico de um antígeno usando um sanduíche ELISA; detecção histoquímica de antígenos; geração de biochips de proteína; e purificação de afinidade de antígenos.

É proveitoso selecionar seqüência de polipeptídeo de estrutura principal baseada na seqüência de uma proteína de planta ou domínio de proteína de estruturas tridimensionais conhecidas (ver, por exemplo, Nygren et al. (2004) "Proteínas de Ligação de Estruturas principais Alternativas", *J. of Immun. Methods* 290:3-28). Contudo, mesmo sem dados estruturais experimentalmente determinados para uma seqüência de polipeptídeo de estrutura principal potencial, inferências valiosas podem ser providas de análise estrutural computacional de uma seqüência de aminoácido candidata. Programas úteis para prognóstico de estrutura de uma seqüência de aminoácido incluem, por exemplo, o conjunto "SCRATCH Protein Predictor" de programas disponíveis ao público na web mundial em ecs.uci.edu/~baldgi/scratch/index. É importante que a introdução de variação de seqüência não desestabilize a estrutura secundária conhecida ou prognosticada da seqüência de polipeptídeo de estrutura principal. Conseqüentemente, a estrutura secundária conhecida ou prognosticada da seqüência de polipeptídeo de estrutura principal informa a seleção de subsequências de aminoácido que pode ser variada dentro de uma seqüência de polipeptídeo de estrutura principal para formar domínios de ligação putativos. A adequação estrutural de uma seqüência de polipeptídeo de estrutura principal pode ser prontamente testada, por exemplo, por métodos de análise de expressão de revelação de fago que são comumente conhecidos na técnica. Por exemplo, uma seqüência de polipeptídeo de estrutura principal contendo 0, 1, 2, 3 ou mais ligações de dissulfeto pode ser testada para sua capacidade de se desdobrar em uma proteína estável. Desde que as proteínas que não se desdobram corretamente não serão incorporadas em um revestimento de fago, elas não serão reveladas. Desse modo, sem esforço indevido, muitas seqüências de polipeptídeo de estrutura principal candidatas podem

ser rapidamente classificadas por sua capacidade de se desdobrar em proteínas estáveis uma vez expressas.

As seqüências de polipeptídeo de estrutura de planta podem ser baseadas no domínio de acessório de fosfatase ácida púrpura (PAPs). A estrutura de cristal do domínio de acessório de feijão "kidney", *Phaseolus vulgaris*, foi determinada (Strater et al. (1995), *Science* 268 (5216):1489-1492). Três voltas expostas dentro da proteína são recordativas dos domínios hipervariáveis encontrados nas imunoglobulinas. As voltas são trazidos juntos pela estrutura de folha β anti-paralela rígida da proteína. As subsequências que formam cada volta a partir dos domínios de ligação putativos de uma proteína de ligação quimérica derivada de uma PAP. Estas subsequências são diversificadas pela substituição, retirada, inserção ou adição até 10 (por exemplo, até 3, 4, 6, 8) aminoácidos. As voltas que formam os domínios de ligação putativos são particularmente bem-adequadas para ligação de moléculas-alvo contendo bolsas ou fissuras.

As seqüências de polipeptídeo de estrutura principal baseadas em PAP tomam a fórmula geral:



onde C_1 , C_2 , C_3 e C_4 correspondem às seqüências de "estrutura principal" que podem incluir alguma variação introduzida, mas não são altamente diversificadas. Por outro lado, X_1 , X_2 e X_3 correspondem às subsequências altamente variadas que formam os domínios de ligação putativos de cada proteína de ligação quimérica baseada em PAP. A Tabela 1 mostra uma lista de subsequências de estrutura principal C_1 - C_4 adequadas derivadas das seqüências de aminoácido de 30 PAPs.

C_1 , C_2 , C_3 e C_4 correspondem a SEQ ID NOs: 1-30, 31-60, 61-90 e 91-120, respectivamente na Tabela 1.

X_1 , X_2 e X_3 podem ser baseados nas variantes que ocorrem naturalmente de seqüências de PAP correspondentes, por exemplo, aquelas mostradas na Tabela 2 como SEQ ID Nos:121-123. A Tabela 2 mostra a variação de faixa de cada posição de aminoácido em subsequências correspondendo, respectivamente, a X_1 , X_2 e X_3 , dentro de 30 seqüências de PAP

que ocorrem naturalmente. Alternativamente, as subsequências variáveis de origem X_1 - X_3 podem ser seqüências arbitrárias de 2-20 aminoácidos de comprimento.

- Em algumas concretizações, C_1 , C_2 , C_3 e C_4 de uma seqüência de polipeptídeo de estrutura principal podem ser selecionadas de seqüências de seqüência de polipeptídeo de estrutura principal baseado em PAP listadas na Tabela 1, em qualquer combinação, por exemplo, C_1 (SEQ ID NO:5), C_2 (SEQ ID NO:12), C_3 (SEQ ID NO:7) e C_4 (SEQ ID NO:19), C_1 (SEQ ID NO:5), C_2 (SEQ ID NO:12), C_3 (SEQ ID NO:5), e C_4 (SEQ ID NO:12), C_4 (SEQ ID NO:22), C_1 (SEQ ID NO:17), C_2 (SEQ ID NO:17), C_3 (SEQ ID NO:19), e C_4 (SEQ ID NO:19), e assim por diante.

Tabela 1: SPSs Baseadas no Domínio de Acessório de PAPs			
SEQ ID	C_1	SEQ ID	C_2
1	PQQVHITQGDHVGKAVIVSWVT	31	VVVYWSSENSKYKKSABGTVTT
2	PQQVHITQGDVLVGKAVIVSWVT	32	EVHYWSSENSDKKKIAEGKLV
3	PQQVHITQGDVLVGRAMIISWVT	33	AVRYWSEKNGRKRIAKGKMST
4	PQQVHITQGDVLVGKAVIVSWVT	34	EVHYWSSENSDKKKIAEGKLV
5	PQQVHITQGDHVGKAVIVSWVT	35	AVRYWSKNSKQKRLAKGKIVT
6	PQQVHITQGDHVGKAMIVSWVT	36	KVVYWSSENSQHKVKAGNIRT
7	PQQVHITQGDHVGKAMIVSWVT	37	KVVYWSSENSQHKVARGNIRT
8	PQQVHITQGDHEGKTVIVSWVT	38	TVLYWSEKSKQKNTAKGKVT
9	PQQVHITQGDVLVGQAMIIISWVT	39	QVIYWSDSLQNFATAEGEVFT
10	PQQVHITQGDVLVGQAMIIISWVT	40	QVIYWSDSLQNFATAEGEVFT
11	PQQVHITQGDHVGKAMIVSWVT	41	TVLYWSNNSKQKNTAKGAVTT
12	PQQVHITQGDLEGEAMIIISWVR	42	KVLYWIDGNSNQKHSANGKITK
13	PQQVHITQGDHVGKAVIVSWVT	43	TVVYWSEKSKLKNKANGKVTT
14	PQQVHITQGDHVGQAMIIISWVT	44	EVIYWSNSSLQNFATAEGEVFT
15	PQQVYITQGDHEGKGVIASWTT	45	SVLYWAENSNNVKSSAEGFVVS
16	PQQVHITQGDYEGKGVIIISWVT	46	TVVYWAENSNNVKRRADGVVVT
17	PQQVHITQGDVLVGRAMIISWVT	47	AVRYWSEKNGRKRIAKGKMST
18	PQQVHITQGDHVGKGVIVSWVT	48	KVLYWFEFNSKIQIAKGTVST
19	PQQVHITQGDVEGKAVIVSWVT	49	KVIYWKENSTKQKHAHGKTNT
20	PQQVHVTQGNHEGNGVIIISWVT	50	TVRYWCENKSRKQAEATVNT
21	PQQVHVTQGNHEGNGVIIISWVT	51	TVQYWCENKSRKQAEATVNT
22	PQQVHITQGDYDGKAVIVSWVT	52	KVQFGTSENKFTSAEGTVSN
23	PQQVHITQGDHEGRSIIISWIT	53	TVFYGTSENKLDQHAEGTVTM
24	PQQVHITLGDQGTGTAMTVSWVT	54	TVRYGSSPEKLDRAAEGSHTR
25	PQQVHITQGDYDGKAVIVSWVT	55	EVVYGTSPNSYDHSAGGKTIN
26	PQQVHITQGDYDGKAVIISWVT	56	HIQYGTSENKFTSEEGTVTN
27	PQQVHITQGDYDGEAVIISWVT	57	EVRYGLSEGKYDVTVEGTLNN
28	PQQVHITQGDYDGKAVIISWVT	58	QVHYGAVQGYEFVAQGTYHN
29	PQQVHITQGDYDGKAVIISWVT	59	QVHYGAVQGYEFVAQGTYHN
30	PQQVHITQGDYNGKAVIVSWVT	60	EVLYGKNEHQYDQRVEGTVTN

Tabela 1:-continuação							
SEQ ID	C ₃				SEQ ID	C ₄	
61	YIHHCYIKGLE YDTKYYYV				91	SREFWFR	
62	FIHHTTIRNLEYKTKYYYE				92	TRQFWFV	
63	FIHHTTIRKLYNTKYYYE				93	TRRFSPFI	
64	FIHHTTIRNLEYKTKYYYE				94	TRQFWFV	
65	FIHHTTIRNLEYNTKYYYE				95	TRQFWFV	
66	YIHHCTIRNLEYNTKYYYE				96	TRSFWFT	
67	YIHHCTIRNLEYNTKYYYE				97	TRSFWFT	
68	YIHHSTIRHLEFNTKYYYK				98	ARTFWFV	
69	FIHHTTITNLEFDTTYYYE				99	TRQFWFI	
70	FIHHTTITNLEFDTTYYYE				100	TRQFWFI	
71	YIHHCTIKHLEFNTKYYYE				101	PRTFWFV	
72	FIHHCTIRRLKHNTKYHYE				102	VRSFWFM	
73	YIHHCTIKNLEFNTKYYYK				103	ARTFWFT	
74	FIHHTNITNLEFNTTYPYV				104	TRQFWFI	
75	YIHHCTIKDLEFDTKYYYE				105	TRKFWFV	
76	YIHHCTIKDLE YDTKYYYE				106	KRQFWFV	
77	YIHHCTIKNLEYNTKYPYE				107	TRQFWFT	
78	YIHHCTIQNLKYNTKYYM				108	RRTFWFV	
79	FIHNCPIRNLEYDTKYYYV				109	ERKFWFF	
80	YIHHCLIDDLFDTKYYYE				110	SRRFWFF	
81	YIHHCLIDDLFDTKYYYE				111	SRRFWFF	
82	YVHHCLIEGLE YTKYYYR				112	SREFWFE	
83	YIHHCVLTDLKYDRKYFYK				113	ARLFWFK	
84	FIHHCTLGLTHATKYYA				114	VRTFSFT	
85	YIHHCLLDKLEYDTKYYYK				115	AREFWFH	
86	YIHHCLIEGLE YTKYYYR				116	SREFWFK	
87	YIHHCLVTGLQYDTKYYYE				117	ARKFWFE	
88	FIHHCLVSDLEHDTKYYYK				118	SREFWFV	
89	FIHHCLVSDLEHDTKYYYK				119	SREFWFV	
90	YIHHCLVDGLE YNTKYYYK				120	AREFWFE	

Tabela 2: Variação de Resíduo que Ocorre Naturalmente em Subseqüências de PAP Correspondentes a X₁, X₂ e X₃ (SEQ ID Nos:123-123)

X ₁ (SEQ ID NO:121)							X ₂ (SEQ ID NO:122)									X ₃ (SEQ ID NO:123)					
Posições							Posições									Posições					
a	b	c	d	e	f	g	a	b	c	d	e	f	g	h	i	A	b	c	d	e	f
M	D	E	P	G	S	S	Y	K	Y	Y	N	Y	T	S	G	V	G	L	R	N	T
V	E	A	K		P	N		R	F	F	T		S		P	I	E	I	G	H	
E	N	K	L		K	K		T		H	K		N			L		V	E	D	
P	V	D				T		F			D		K			M		E	D	Q	
Q	S				H							E		E				T		K	
T	I				T													S		S	
A	A																	E		E	
F																		F			
																		K			

Após diversificação das subsequências acima listadas da sequência de polipeptídeo de estrutura principal, as subsequências X_1' , X_2' e X_3' são altamente heterogêneas dentro da biblioteca de polipeptídeos de ligação quiméricos de planta codificados, e podem conter cada até 10 (por exemplo, 8, 6, 4, 3) substituições, retiradas, inserções e adições de aminoácidos simples com relação à SEQ ID Nos: 121-123 listados na Tabela 1, respectivamente (ver, por exemplo, Fig. 1). Por exemplo, o comprimento das sequências de aminoácido correspondente a X_1 , X_2 e X_3 pode ser inalterado, encurtado ou encompridado em relação à SEQ ID Nos: 121-123.

As regiões fora dos domínios de ligação putativos são referidas como regiões de "estrutura principal" (isto é, C_1 , C_2 , C_3 e C_4). Diferente das sequências de aminoácido para X_1 , X_2 e X_3 , as sequências de aminoácidos das regiões de estrutura principal são geralmente não diversificadas dentro da biblioteca de proteínas de ligação quiméricas codificadas, embora alguma variação de sequência nestas regiões dentro da biblioteca seja permissível. As regiões de estrutura principal de uma sequência de polipeptídeo de estrutura de planta podem ser pelo menos 70% (isto é, 80, 85, 90, 95 ou 100%) idênticas a qualquer de SEQ ID Nos: 1-120. Alternativamente, as regiões de estrutura principal podem conter até 30 (isto é, 28, 26, 24, 22, 20, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 ou 1) substituições, retiradas, inserções e adições de aminoácidos simples. Por exemplo, C_1 , C_2 , C_3 e C_4 podem cada uma incluir 0, 1, 2, 3, 4 ou 5 ou mais mudanças de aminoácido simples. Se substituições de aminoácidos são para serem introduzidas nas regiões de estrutura principal, é preferível produzir substituições conservativas. Uma substituição conservativa é uma que preserva nos substituintes um aminoácido com um que tem propriedades químicas similares (por exemplo, substituição de um aminoácido polar, tal como serina, com outro aminoácido polar, tal como treonina).

Em uma concretização, a sequência de polipeptídeo de estrutura de planta é uma de SEQ ID Nos: 124-126 mostrada abaixo. Sequências correspondentes a X_1 , X_2 e X_3 estão em negrito e sublinhadas.

SEQ ID NO:124

PPQVHITQGDHVGKAVIVSWVTMDEPGSSVVVYWSENSKYKKS AEGTVTT
YRFYNYTSGYIHHCYIKGLE YDTKYYYYVVGIGNTSREFWFR

SED ID NO: 125

- 5 PQQVHITQGD LVGKAVIVSWVTVDEPGSSEVHYWSENSDKKKIAEGKLVTY
RFFNYSSGFIHHTTIRNLEYKTKYYYYEVGLGNTTRQFWFV

SEQ ID NO: 126

PQQVHITQGD LVGRAMIISWVTMDEPGSSAVRYWSEKNGRKRIAKGKMST
YRFFNYSSGFIHHTTIEKLKYNTKYYYYEVGLRNTTRRFSFI

- 10 Em outras concretizações, uma seqüência de polipeptídeo de estrutura de planta é baseada na seqüência de aminoácido de proteínas de planta que tem repetições similares a ankyrin. Repetições similares a ankyrin são repetições de pequeno giro espiral-espiral (THT) consistindo em aproximadamente 33 aminoácidos. O número de repetições de THT dentro de uma
- 15 seqüência de polipeptídeo de estrutura principal pode variar de 2 a 20. Os locais de ligação putativos dentro das repetições de THT são tipicamente não-contíguos, mas agrupados no mesmo lado da proteína da qual eles são uma parte.

- Uma seqüência de polipeptídeo de estrutura principal contendo
- 20 repetição de THT de planta pode ter uma seqüência de aminoácido que é baseada em qualquer de SEQ ID NOs: 127-129 listadas abaixo. Altos níveis de variação de seqüência de aminoácido são introduzidos nos resíduos em negrito/sublinhados. As seqüências de polipeptídeo de estrutura principal contendo repetição de THT de planta pode conter substituições de até 3 a-
- 25 minoácidos, ou uma retirada no lugar dos aminoácidos correspondentes aos resíduos 12-13, 33, 35-36, 38, 46-47, 66, 68-69, 71, 79-80, 99, 101-102, 104 e 112-113 (resíduos em negrito e sublinhados) de SEQ ID Nos: 127-129.

SEQ ID NO: 127

- GDDL GKKLHLAASSRGHLEIVRVLVEAGADVNALDKFGRTALHIAASSRGHLE
- 30 VVKLLLEAGADVNALDKFGRTALHIAASSRGHLEVVKLLLEAGADVNALDKF
GDTALHVSIDNGNEDIAEILQ

SEQ ID NO: 128

GDDLGGKKLHLAA**SR**GHLEIVRVLVEAGADVNA**L****D****K****F****G****R****T****P****L****H****I****A****A****S****K****G****N****E****Q**
 VVKLLLEAGADPN**A****L****D****K****F****G****R****T****P****L****H****I****A****A****S****K****G****N****E****Q****V****V****K****L****L****L****E****A****G****A****D****P****N****A****Q****D****K****F**
 G**D****T****A****L****H****V****S****I****D****N****G****N****E****D****I****A****E****I****L****Q**

5 SEQ ID NO: 129

GSDLGKKLLEAA**R**AGQDDEVRLMANGADVNA**L****D****K****F****G****R****T****P****L****H****I****A****A****S****K****G****N****E**
 QVVKLLLEAGADPN**A****L****D****K****F****G****R****T****P****L****H****I****A****S****K****G****N****E****Q****V****V****K****L****L****L****E****A****G****A****D****P****N****A****Q****D****K****F**
 G**K****T****A****F****D****I****S****I****D****N****G****N****E****D****L****A****E****I****L****Q**

A seqüência das seqüências de polipeptídeo de estrutura principal pode ser pelo menos 70% (isto é, 80, 85, 90, 95, 98 ou 100%) idênticas à seqüência externa das posições de aminoácido precedentes (em negrito) de SEQ ID Nos: 127-129. Alternativamente, a seqüência das seqüências de polipeptídeo de estrutura principal externas das posições de aminoácidos precedentes (em negrito) de SEQ ID NOS: 127-129 pode conter até 30 (isto é, 28, 26, 24, 22, 20, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 ou 1) substituições, retiradas, inserções e adições de aminoácidos simples. Em alguns casos, pode ser desejável incluir umidades de repetição adicionais. SEQ ID NOS: 127-129 tem uma capa amino-terminal, duas repetições internas, e uma capa carbóxi-terminal. Pode ser desejável ter-se 1-6 repetições internas. A seqüência de capa amino-terminal é aa 1-33. A primeira repetição interna é 34-66, e a segunda repetição interna é 67-99. A seqüência de capa carbóxi-terminal é aa 100-123. As primeira ou segunda repetições internas, ou ambas, podem ser independentemente repetidas 1, 2, 3, 4, 5 ou 6 vezes.

Os locais de ligação putativos são formados por cadeias de aminoácido que se projetam a partir da estrutura secundária rígida formada pela seqüência de polipeptídeo de estrutura principal. Estas proteínas podem tipicamente formar uma superfície de ligação mais plana maior, e são particularmente úteis para ligação a alvos que não tem fissuras profundas ou bolsas.

Outro estrutura principal adequado pode estar baseado em orizacistatin (*J. Biol. Chem* 262:16793 (1987); *Biochemistry* 39: 14753 (2000)),

um membro da Família cistatin/Papaína (Pfam Identifier PF00031), que é identificado como um inibidor de cisteína proteinase de arroz. A sequência de orizacistatin é representada abaixo. Um estrutura principal tendo a sequência de aminoácido C1-X1-C2-X2-C3-X3-C4, onde cada uma de X1, X2, X3 e X4 é uma região variável, e C1, C2, C3 e C4 são as regiões de estrutura principal, podem ser criado baseado em orizacistatin.

MSSVGGPVLGGVEPVGNENDLHLVFLARFAVTEHNKKANSLLFEKLVSVK
QQVVAGTLYYFTLEVKEGDAKKLYEAKVWEKPWMDFKELQEFKPVDasAN
A

- 10 C1-MSS (aa 1-3)
 X1-VGGP (aa 4-7)
 C2-VLGGVEPVGNENDLHLVDLARFAVTEHNKKANSLLFEKLVSV (aa- 8-50)
 X2-KQQVVAGT (aa 51-58)
 15 C3-LYYFTLEVKEGDAKKLYEAKVWE (aa 59-81)
 X3-KPWM (aa 82-85)
 C4-DFKELQEFKPVDasANA (aa 86-102)

A figura 2 representa as seqüências de um grande número de proteínas de planta alinhadas com orizacistatin. Exemplos de regiões C1-C4 adequadas são indicados. A figura 4 representa as seqüências de um pequeno número de proteínas de planta alinhadas com orizacistatin. Exemplos de regiões C1-C4 adequadas são indicados. Em geral, X1 pode ser uma sequência de 2-20 aminoácidos aleatórios (por exemplo, 3 aminoácidos). X2 pode ser uma sequência de 2-20 aminoácidos aleatórios (por exemplo, 4 aminoácidos). X3 pode ser uma sequência de 2-20 aminoácidos aleatórios (por exemplo, 4 aminoácidos).

Ainda outra adequada pode ser baseada na proteína C2 de arroz (*Biochemistry* 42:11625 (2003)), um membro da família de domínio C2 (Pfam Identifier PF00168) que é pensada estar envolvida em sistemas de sinalização de defesa de planta. A sequência de C2 de arroz é representada abaixo. Um estrutura principal tendo a sequência de aminoácido C1-X1-C2-X2-C3-X3-C4, onde cada um de X1, X2, X3 e X4 é uma região variável e C1,

C2, C3 e C4 são as regiões de estrutura principal, pode ser criado baseado na C2 de arroz.

MAGSGVLEVHLVDAKGLTGNDFLGKIDPYVVVQYESQERKSSVARDQGKN
PSWNEVFKFQINSTAATGOHKLFLRLMDHDTFSRDDFLGEATINVTDLISLG
 5 MEHGTWEEMSESKHRVVLADKTYHGEIRVSLTFTASAKAQDHAEQVGGW
AHSFRQ

C1-MAGSGVLEVHLVDAKG (aa 1-16)

X1-LTGNDFLGKID (aa 17-27)

C2-PYVVVQYRSQERK (aa 28-40)

10 X2-SSVARDQGKNP (aa 41-51)

C3-SWNEVFKFQINSTAATGQHKLFLRL (aa 52-76)

X3-MDHDTFSRDDFL (aa 77-88)

C4-GEATINVTDLISLGMEHGTWEMSESKWRVVLADKTYHGEIRVSLTFRASAKAQ
 DHAEQVGGWAHSFRQ (aa 89-156)

15 A figura 3 representa as seqüências de um grande número de proteínas de planta alinhadas com C2 de arroz. Exemplos de regiões C1-C4 adequadas são indicados. A Figura 5 representa as seqüências de um pequeno número de proteínas de planta alinhadas com orizacistatin. Exemplos de regiões C1-C4 adequadas são indicados. Em geral, X1 pode ser uma se-
 20 qüência de 2-20 aminoácidos aleatórios (por exemplo, 11 aminoácidos). X2 pode ser uma seqüência de 2-20 aminoácidos aleatórios (por exemplo, 11 aminoácidos). X3 pode ser uma seqüência de 2-20 aminoácidos aleatórios (por exemplo, 12 aminoácidos).

25 As seguintes seções representam métodos para geração de bibliotecas de ácidos nucléicos que codificam proteínas de ligação quiméricas baseadas nas seqüências de polipeptídeo de estrutura de planta.

II. Geração de Bibliotecas de Ácido Nucléico baseada em uma Seqüência de polipeptídeo de estrutura de planta

30 Uma grande biblioteca de variantes de seqüência de ácido nucléico que codifica a seqüência de polipeptídeo de estrutura de planta é criada baseada em uma ou mais seqüências de polipeptídeo de estrutura de planta. A biblioteca de ácidos nucléicos codificam pelo menos 5 (por exem-

plo, 1.000 , 10^5 , 10^6 , 10^7 , 10^9 , 10^{12} , 10^{15} ou mais) seqüências de proteína de ligação quimérica diferentes. É reconhecido que nem todo membro de uma biblioteca gerado pelos métodos aqui descritos codificará uma seqüência de aminoácido única. Não obstante, é desejável que pelo menos 10% (por exemplo, 25%, 30%, 40%, 50%, 60%, 70% ou 90%) das proteínas de ligação quiméricas codificadas representadas na biblioteca sejam únicas.

Antes da diversificação de uma seqüência de polipeptídeo de estrutura de planta, pode ser útil estimar computacionalmente a diversidade de seqüência esperada a ser gerada com um dado conjunto de parâmetros de variação de seqüência. Um método de estimar diversidade de seqüência é descrito, por exemplo, em Volles et al., (2005), 33(11):3667-3677. Por exemplo, o número de seqüências diferentes esperadas em uma biblioteca de ácidos nucleicos gerada por PCR pode ser estimado baseado na freqüência de mutação da polimerase mutagênica usada para a amplificação. Algoritmos úteis para estimar diversidade de seqüência em bibliotecas de codificação de proteína randomizada podem também serem encontrados na web mundial, por exemplo, em guinevere.otago.ac.nz/mlrgd/STATS/index.

Bibliotecas de ácidos nucleicos que codificam proteínas de ligação quiméricas de planta podem ser geradas por um número de metodologias conhecidas. A diversidade de seqüência é introduzida em uma seqüência de polipeptídeo de estrutura de planta por substituição, retirada, inserção ou adição de aminoácidos em posições altamente variáveis de uma seqüência de polipeptídeo de estrutura principal, conforme descrito acima. Desde que o conjunto de 20 aminoácidos que são geralmente codificados em plantas tem propriedades químicas e estruturais um tanto redundantes, um subconjunto de aminoácidos (por exemplo, um subconjunto de 4 tipos de aminoácidos) que envolve esta diversidade estrutural pode ser adotado para substituições. Por exemplo, aminoácidos a serem usados para substituição ou inserção podem ser selecionados para incluir um aminoácido ácido, um aminoácido neutro, um aminoácido alifático, e um aminoácido aromático (ver Tabela 3). Por exemplo, os aminoácidos usados para substituição podem ser limitados a aspartato, serina, e tirosina. A limitação da redundância de substituições

de aminoácido aumentará a diversidade estrutural total e de ligação da biblioteca de proteínas de ligação quiméricas.

Tabela 3 Propriedades Químicas de Aminoácidos Geneticamente Codificados em Plantas				
Acídico	Neutro	Alifático	Aromático	Básico
Aspartato	Asparagina, Cisteína,	Alanina,	Histidina,	Arginina,
Glutamato	Glutamina, Metionina,	Glicina,	Fenilalanina,	Lisina
	Prolina, Serina, Treonina	Isoleucina,	Triptofan, Tirosina	
		Leucina, Valina		

A biblioteca de ácidos nucleicos pode ser gerada *in vitro* pela montagem de conjuntos de oligonucleotídeos com seqüências complementares de sobreposição. Primeiro, uma seqüência de seqüência de polipeptídeo de estrutura principal é selecionada que é para ser codificada por conjuntos de oligonucleotídeos montados. As seqüências a serem codificadas nas regiões variáveis de uma dada seqüência de polipeptídeo de estrutura principal incluirão uma multiplicidade de seqüências heterogêneas contendo substituições, inserções, retiradas e adições de acordo com a biblioteca de peptídeos de ligação quiméricos a ser gerada conforme descrito acima. As seqüências de polipeptídeo de estrutura principal a serem codificadas incluem subsequências C₁-C₄ correspondentes a qualquer de SEQ ID Nos:1-30, 31-60, 61-90 e 91-120, respectivamente.

Um conjunto de oligonucleotídeos codifica regiões da seqüência de polipeptídeo de estrutura de planta onde diversidade é para ser introduzida (por exemplo, em X₁, X₂ e X₃). Em contraste, regiões da seqüência de polipeptídeo de estrutura principal na qual pouco ou nenhuma variação é para ser introduzida (por exemplo, nos domínios de estrutura principal de seqüências de polipeptídeo de estrutura principal de PAP) são codificadas por um conjunto de oligonucleotídeos que codificam seqüências de aminoácido com não menos do que 70% (isto é, 75%, 80%, 85%, 90%, 95% ou 100%) de identidade para qualquer uma das seqüências de polipeptídeo de estrutura principal acima mencionadas. Os detalhes deste método são descritos, por exemplo, na Patente dos Estados Unidos N° 6.521.453, desse

modo incorporada por referência.

Oligonucleotídeos de sequência variada usados para gerar bibliotecas de ácidos nucléicos são tipicamente sintetizados quimicamente de acordo com o método de fosforamidite triéster de fase sólida descrito por
 5 Beaucage e Caruthers (1981), *Tetrahedron Metts.*, 22(20):1859-1862, por exemplo, usando-se um sintetizador automático, conforme descrito em Nee-
 dham-VanDevanter et al. (1984) *Nucleic Acids Res.*, 12:6159-6168. Uma
 ampla variedade de equipamento é comercialmente disponível para síntese
 automática de oligonucleotídeo. Aproximações de síntese de multinucleotí-
 10 deos (por exemplo, síntese de trinucleotídeos), conforme discutidas supra,
 são também úteis.

Ácidos nucléicos podem ser ordenados costumeiramente de uma variedade de fontes comerciais, tais como Sigma-Genosys (em sigma-genosys.com/oligo.asp); The Midland Certified Reagent Company
 15 (mcrc@aligos.com), The Great American Gene Company (em genco.com),
 ExpressGen Inc. (em expressgen.com), Operon Technologies Inc. (Alameda, Calif), e muitos outros.

Os oligonucleotídeos podem ter um uso de códon otimizado para expressão em um tipo de célula particular (por exemplo, em uma célula de
 20 planta, uma célula de mamífero, uma célula de levedura, ou uma célula bac-
 terial). Tabelas de frequência de uso de códon são publicamente disponí-
 veis, por exemplo, na web mundial em kazusa.or.jp/codon. Inclinação de có-
 don pode ser usada para otimizar expressão em uma célula ou na superfície
 de uma célula em que ligação de uma proteína de ligação quimérica de plan-
 25 ta é para ser avaliada, e pode também ser usada para otimizar expressão da
 proteína de ligação quimérica em um organismo transgênico de interesse
 comercial (por exemplo, uma planta transgênica). Em geral, códons com
 uma frequência de uso de menos do que 10% não são usados. Antes da
 síntese de seqüências de oligonucleotídeos serem verificadas para seqüên-
 30 cias potencialmente problemáticas, por exemplo, locais de restrição úteis
 para subclonagem, locais aceptores ou doadores de união de planta (ver,
 por exemplo, cbs.dtu.dk/services/FeatureExtract/), seqüências de desestabi-

lização de mRNA potenciais (por exemplo, "ATTTA"), e extensões de mais do que quatro ocorrências do mesmo nucleotídeo. Sequências potencialmente problemáticas são, conseqüentemente, mudadas.

Populações de oligonucleotídeos são sintetizadas que codificam variações de aminoácido nas regiões de ligação putativas da sequência de polipeptídeo de estrutura principal selecionada (por exemplo, nas regiões X₁, X₂ e X₃ de uma sequência de polipeptídeo de estrutura principal de PAP).

Preferivelmente, todos os oligonucleotídeos de um comprimento selecionado (por exemplo, cerca de 10, 12, 15, 20, 30, 40, 50, 60, 70, 80, 90 ou 100 ou mais nucleotídeos) que correspondem a regiões onde diversidade de sequência é para ser introduzida na sequência de polipeptídeo de estrutura principal codificam todas as variações de aminoácido possíveis de um conjunto diverso de aminoácidos, conforme descrito acima. Isto inclui N oligonucleotídeos por N variações de sequência, onde N é o número de seqüências diferentes em um local. Os N oligonucleotídeos são idênticos em sequência, exceto para o(s) nucleotídeo(s) que codifica(m) o(s) aminoácido(s) variante(s). Na geração dos oligonucleotídeos de sequência variada, pode ser vantajoso utilizar estratégias de síntese paralelas ou reunidas nas quais reação de síntese simples ou conjunto de reagentes são usados para produzir porções comuns de cada oligonucleotídeo. Isto pode ser realizado, por exemplo, por técnicas de síntese de ácido nucléico de fase sólida bem conhecidas, ou, por exemplo, utilizando métodos sintéticos de oligonucleotídeo baseados em série (por exemplo, Fodor et al., (1991) *Science*, 251:767-777; Fodor (1997) "Genes, Chips and the Human Genome" *FASEB Journal*. 11:121-121; Fodor (1997) "Massively Paralell Genomics" *Science*. 277:393-395; e Chee et al., (1996) "Acessing Genetic Information with High-Density DNA Arrays" *Science* 274:610-614).

Em estratégias de síntese típicas, os oligonucleotídeos têm pelo menos cerca de 10 bases de identidade de sequência em qualquer lado de uma região de diferença para assegurar recombinação razoavelmente eficiente. Contudo, regiões de flaqueamento com bases idênticas podem ter poucas bases idênticas (por exemplo, 4, 4, 6, 7, 8 ou 9) e podem, natural-

mente, ter regiões maiores de identidade (por exemplo, 1, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 50, ou mais).

Os oligonucleotídeos a serem montados juntos são incubados para permitir hibridização entre oligonucleotídeos contendo seqüências complementares de sobreposição. Cada conjunto de hibridização de oligonucleotídeos de sobreposição forma, desse modo, um ácido nucléico contíguo interrompido por pequenas folgas. Estas folgas pequenas podem ser preenchidas para formar seqüências de comprimento total usando qualquer de uma variedade de métodos de remontagem mediados por polimerase, por exemplo, conforme descritos aqui, e conforme conhecidos a um técnico no assunto. A maior diversidade de freqüência é introduzida em oligonucleotídeos que codificam as regiões e resíduos de ligação putativos de seqüência de polipeptídeo de estrutura de planta. Contudo, oligonucleotídeos que codificam variações de seqüência específicas podem ser "bloqueados" na mistura de recombinação em qualquer concentração selecionada, causando, desse modo, incorporação preferencial de modificações desejáveis nas proteínas de ligação quiméricas de planta codificadas em regiões fora dos domínios de ligação putativos.

Por exemplo, durante alongamento do oligonucleotídeo, oligonucleotídeos hibridizados são incubados na presença de uma polimerase de ácido nucléico, por exemplo, Taq, Klenow, ou similares, e dNTPs (isto é, dATP, dCTP, dGTP e dTTP). Se regiões de identidade de seqüência são grandes, Taq ou outra polimerase de alta temperatura podem ser usados com uma temperatura de hibridização de entre cerca de temperatura ambiente (isto é, cerca de 25°C) e, por exemplo, cerca de 65°C. Se as áreas de identidade são pequenas, Klenow, Taq ou polimerases podem ser usados com uma temperatura de hibridização de abaixo da temperatura ambiente. A polimerase pode ser adicionada à reação de montagem antes de, simultaneamente com, ou após hibridização dos oligonucleotídeos. Em seguida, as seqüências de ácido nucléico de filamento duplo resultante são desnaturadas, hibridizadas e alongadas novamente. Este ciclo pode ser repetido por qualquer número desejado de vezes. O ciclo é repetido, por exemplo, de

cerca de 2 a cerca de 100 vezes.

Opcionalmente, após ciclos múltiplos de montagem de ácido nucléico combinatorial, os produtos resultantes podem ser amplificados, por exemplo, por reação de cadeia polimerase padrão (PCR). Uma porção do volume da reação de montagem acima descrita é incubada com priméres de avanço e reverso únicos que hibridizam universalmente para as extremidades dos ácidos nucléicos, bem como dNTPs e uma polimerase adequada (por exemplo, polimerase *pfu*). A reação de PCR é, em seguida, efetuada por cerca de 10 a 40 ciclos.

Para determinar a extensão de incorporação de oligonucleotídeo, qualquer aproximação que distingue ácidos nucléicos similares pode ser usada. Por exemplo, os ácidos nucléicos podem ser clonados e seqüenciados, ou amplificados (*in vitro*, ou por clonagem, por exemplo, em um vetor de clonagem ou expressão padrão), e clivados com uma enzima de restrição que reconhece especificamente uma variante de seqüência de oligonucleotídeo particular.

É útil incluir locais de restrição raros (por exemplo, Not I) nas extremidades 5' dos priméres 5' e 3' usados, ou na montagem, ou reações de PCR. A inclusão de locais de restrição nestes priméres facilita subclonagem dos ácidos nucléicos em um vetor por digestão de restrição e ligação subsequente. Alternativamente, a reação de montagem ou produtos de PCR podem também ser clonados, sem serem digeridas por restrição, usando-se métodos padrões, por exemplo, clonagem "TA".

Outros métodos para introdução de diversidade em uma seqüência de polipeptídeo de estrutura de planta podem também ser usados. Por exemplo, uma seqüência de polipeptídeo de estrutura principal pode ser codificada em um gabarito de ácido nucléico, por exemplo, uma construção de plasmídeo. Alternativamente, um produto de PCR, mRNA ou DNA genômico de uma espécie de planta apropriada, tal como feijão-soja, pode também servir como um gabarito que codifica uma seqüência de polipeptídeo de estrutura de planta. Uma ou mais subsequências de seqüência de polipeptídeo de estrutura principal a serem diversificadas (por exemplo, a

região X₂ de uma seqüência de polipeptídeo de estrutura principal de PAP) podem ser diversificadas durante ou após amplificação a partir do gabarito de ácido nucléico de seqüência de polipeptídeo de estrutura principal por qualquer de um número de métodos de PCR propensos a erro. Os métodos

5 de PCR propensos a erro podem ser divididos em (a) métodos que reduzem a fidelidade da polimerase por concentração de nucleotídeos não-equilibradas e/ou adição de composto quimérico, tais como cloreto de manganês (ver, por exemplo, Lin-Goerke et al. (1997) *Biotechniques*, 23, 409-412), (b) métodos que empregam análogos de nucleotídeo (ver, por exemplo,

10 plo, Patente dos Estados Unidos N° 6.153.745), (c) métodos que utilizam polimerases 'mutagênicas' (ver, por exemplo, Cline, J. e Hogrefe, H. H. (2000) *Strategies* (Stratagene Newsletter), 13, 157-161, e (d) métodos combinados (ver, por exemplo, Xu, H., Petersen, E.I., Petersen, S.B. e el-Gewely, M. R. (1999) *Biotechniques*, 27, 1102-1108. Outros métodos basea-

15 dos em PCR incluem aqueles, por exemplo, descritos por Osuna J, Yanes J, Soberon X, e Gaytan P. (2004), *Nucleic Acids Res.* 2004, 32(17):e136 e Wong TS, Tee KL, Hauer B, e Schwaneberg *Nucleic Acids Res.* 2004 Fev 10; 32(3):e26), e outros conhecidos na técnica.

Após gerar uma população de variantes de seqüência, estas

20 podem ser substituídas na região apropriada de um ácido de nucléico de seqüência de polipeptídeo de estrutura de planta escolhido (por exemplo, um plasmídeo contendo uma seqüência de polipeptídeo de estrutura principal) por subclonagem que, desse modo, age eficazmente como um vetor para a biblioteca de seqüências diversificadas.

25 Ainda outra aproximação para mutagênese de regiões de seqüência de polipeptídeo de estrutura de planta específicas é o uso de uma cepa mutagênica de *E. coli* (ver, por exemplo, Wu et al., (1999), *Plant. Mol. Biol.*, 39(2):381-386). Um vetor de ácido nucléico contendo uma seqüência alvo a ser mudado é introduzido na cepa de mudança, que é em seguida,

30 propagada. A replicação de DNA propensa a erro na cepa de mudança de *E. coli* introduz mutações na seqüência alvo introduzida. A população de seqüências alvo alteradas é, em seguida, recuperada e subclonada na posição

apropriada de um ácido nucléico que codifica a seqüência de polipeptídeo de estrutura de planta selecionada para gerar uma biblioteca diversa de ácidos nucléicos que codificam proteínas de ligação quiméricas de planta.

III. Expressão e Classificação de Proteínas de ligação quiméricas de planta

5 A biblioteca de ácidos nucléicos baseada em uma seqüência de polipeptídeo de estrutura de planta, e que codifica polipeptídeos de ligação quiméricos de planta são subclonados em um vetor de expressão e introduzidos em um sistema de replicação biológico para gerar uma biblioteca de expressão. A biblioteca de expressão pode ser propagada e classificada para identificar proteínas de ligação quiméricas de planta que se ligam a uma molécula-alvo (TM) de interesse (por exemplo, um nematóide, inseto, fungos, proteína viral ou de planta).

15 O sistema de replicação biológico em que classificação de proteínas de ligação quiméricas de planta será praticada deve ser capaz de crescer em um ambiente adequado, após seleção para ligação a um alvo. Alternativamente, o ácido nucléico que codifica a proteína de ligação quimérica de planta selecionada pode ser isolado por amplificação *in vitro*. Durante pelo menos parte do crescimento do sistema de replicação biológico, o aumento no número é preferivelmente e aproximadamente exponencial com relação ao tempo. A freqüência de membros da biblioteca que exibe as propriedades de ligação desejadas pode ser muito baixa, por exemplo, uma em 10^6 ou menos.

25 Os sistemas de replicação biológicos podem ser vírus de DNA bacterial, células bacteriais vegetativas, esporos bacteriais. Células eucarióticas (por exemplo, células de levedura) podem também ser usadas como um sistema de replicação biológico.

30 Em uma concretização particularmente útil, uma fusão de proteína de ligação quimérica-proteína de revestimento de fago é codificada em um constructo de fagos médios. Os constructos de fagos médios são transformados em bactérias hospedeiras, que são subsequenteiramente infectadas com um fago auxiliador que expressa proteínas de revestimento tipo selvagem. A progenia de fago resultante tem revestimentos de proteína que inclu-

em ambas proteína de fusão e proteínas de revestimento tipo selvagem. Esta aproximação tem a vantagem que a viabilidade de fago é maior comparada à viabilidade de fago que tem exclusivamente proteína de ligação química-proteína de fusão de revestimento. Construção de biblioteca de revelação baseada em fagos médios e kits de classificação são comercialmente disponíveis, por exemplo, the EZnet® Phage Display cDNA Library Construction Kit and Screening Kit (Maxim Biotech, Inc., San Francisco, CA).

Não obstante, uma cepa de qualquer célula viva ou vírus é potencialmente útil se a cepa pode ser: 1) geneticamente alterada com facilidade razoável para codificar uma proteína de ligação química de planta, 2) mantida e amplificada na cultura, 3) manipulada para mostrar o domínio de proteína de ligação potencial onde ela pode interagir com o material alvo, e 4) selecionada, enquanto retém a informação genética que codifica a proteína de ligação química de planta expressa na forma recuperável. Preferivelmente, o sistema de replicação biológico permanece visível após seleção baseada em afinidade.

Quando o sistema de replicação biológico é uma célula bacterial ou um fago que é montado no periplasma, o vetor de expressão para revelação da proteína de ligação química de planta codifica a própria proteína de ligação química fundida a dois componentes adicionais. O primeiro componente é um sinal de secreção que direciona o produto de expressão inicial para a membrana interna da célula (uma célula hospedeira quando o acondicionamento é um fago). Este sinal de secreção é clivado por uma peptidase de sinal para produzir uma proteína de ligação química de planta processada madura. O segundo componente é um sinal de transporte de superfície externa que direciona o sistema de replicação biológico para montar a proteína processada em sua superfície externa. Este sinal de transporte de superfície externa pode ser derivado de uma proteína de superfície nativa ao sistema de replicação biológico (por exemplo, a proteína gIII de revestimento de fago M13).

Por exemplo, o vetor de expressão compreende um DNA que codifica uma proteína de ligação química de planta operavelmente ligada a

uma seqüência de sinal (por exemplo, as seqüências de sinal do phoA bacteriana ou genes bla, ou a seqüência de sinal de gene III de fago M13), e ao DNA que codifica uma proteína de revestimento (por exemplo, o gene III M13 ou proteínas VIII de gene) de um fago filamentosos (por exemplo, M13).

- 5 O produto de expressão é transportado para a membrana interna (bicamada de lipídeo) da célula hospedeira, onde o peptídeo de sinal é clivado para deixar uma proteína híbrida processada. O terminal C do componente similar à proteína de revestimento desta proteína híbrida está preso na bicamada de lipídeo, de modo que a proteína híbrida não escapa no espaço periplásmico.
- 10 À medida que o DNA de filamento simples da partícula de fago nascente passa no espaço periplásmico, ele coleta ambas a proteína de revestimento tipo selvagem e a proteína híbrida a partir da bicamada de lipídeo. A proteína híbrida é, desse modo, acondicionada na blindagem da superfície do fago filamentosos, deixando a proteína de ligação quimérica de planta exposta em
- 15 sua superfície externa. Desse modo, o fago filamentosos, não a célula bacteriana hospedeira, é o sistema de replicação biológico nesta concretização. Se um sinal de secreção é necessário para a revelação da proteína de ligação quimérica de planta, uma cepa bacteriana de "secreção permissiva" pode ser usada para crescimento do sistema de replicação biológico de fago filamentosos.
- 20

É desnecessário usar um sinal de secreção de membrana interna quando o sistema de replicação biológico é um esporo bacteriana, ou um fago cujo revestimento é montado intracelularmente. Nestes casos, o meio de revelação é meramente o sinal de transporte de superfície externa, tipicamente de uma proteína de revestimento de esporo e de fago.

- 25 Os fagos filamentosos em geral são atrativos como sistemas de replicação biológicos para revelação de proteína de ligação quimérica de planta, e M13 em particular, é especialmente atrativo por que: 1) a estrutura 3D do virion é conhecida; 2) o processamento da proteína de revestimento é bem compreendida; 3) o genoma é expansível; 4) o genoma é pequeno; 5) a seqüência do genoma é conhecida; 6) o virion é tipicamente resistente a cisalhamento, calor, frio, uréia, guanidínio Cl, baixo pH, e alto sal; 7) o fago é
- 30

um vetor de seqüenciamento de modo que o seqüenciamento é especialmente fácil; 8) genes resistentes a antibióticos foram clonados no genoma; 9) É facilmente cultivado e armazenado, com requerimentos de meio não usuais e custosos para as células infectadas; 10) tem uma dimensão de ruptura alta, cada célula infectada produzindo 100 a 1000 progenias de M13 após infecção; e 11) é facilmente colhido e concentrado por métodos padrões.

Por exemplo, quando o sistema de replicação biológico é M13, as proteínas de gene III ou de gene VIII podem ser usadas como um sinal de alvo de superfície externa. Alternativamente, as proteínas de genes VI, VII e IX podem também serem usadas.

A proteína de ligação quimérica de planta codificada pode ser fundida ao sinal de alvo de superfície (por exemplo, a proteína de revestimento de gene III M13) em seu carboxi ou amino terminal. O limite de fusão entre a proteína de ligação quimérica de planta e o sinal de alvo pode também incluir uma seqüência ligadora curta (por exemplo, até 20 aminoácidos de comprimento) para evitar interações indesejáveis entre proteína de ligação quimérica e o sinal de alvo fundido. Em algumas concretizações, é vantajoso incluir dentro da seqüência ligadora um local de clivagem proteolítico específico. Em adição, o amino terminal ou carboxi terminal da proteína fundida podem incluir uma etiqueta de epítipo curta (por exemplo, etiqueta de hemaglutinina). A inclusão de um local de clivagem proteolítico ou uma etiqueta de epítipo curta é particularmente útil para purificação de uma biblioteca de proteínas de ligação quiméricas de uma população de células que expressa a biblioteca. Proteínas de ligação quiméricas etiquetadas de epítipo podem ser convenientemente purificadas pela clivagem proteolítica de seqüência ligadora, seguida por cromatografia de afinidade utilizando um anticorpo ou outro agente de ligação que reconhece a etiqueta de epítipo.

Muitos métodos existem para classificação de bibliotecas de revelação de fago (ver, por exemplo, Willats (2002), *Plant Mol. Biol.*, 50:837-854). Conforme comumente praticado, a molécula-alvo de interesse é adsorvida em um estrutura principal e, em seguida, exposta a soluções de proteí-

nas de ligação quiméricas de planta que revelam fago. A molécula-alvo pode ser imobilizada por adsorção passiva em um meio de estrutura principal, por exemplo, tubos, placas, colunas, ou gotas magnéticas. Geralmente, o meio de estrutura principal adsortivo é pré-bloqueado, por exemplo, com albumina de soro bovino, leite, ou gelatina, para reduzir ligação não-específica do fago durante classificação. Alternativamente, a molécula-alvo pode ser biotinizada, de modo que interação entre fago suportando proteína de ligação quimérica e a molécula-alvo pode ser efetuada em solução. O fago que se liga ao alvo pode então ser selecionado usando-se avidina ou estreptavidina ligado a um substrato sólido (por exemplo, gotas ou coluna).

Após os fagos serem permitidos interagirem com a molécula-alvo, os fagos não-interados são removidos por lavagem. Os fagos de ligação especificamente remanescentes são, então, eluídos por um de qualquer número de tratamento incluindo, por exemplo, abaixamento ou aumento do pH, aplicação de agentes de redução, ou uso de detergentes. Em uma concretização, um local de clivagem proteolítico específico é introduzido entre a sequência de proteína de ligação quimérica de planta e a sequência de proteína de revestimento de fago. Desse modo, eluição de fago pode ser efetuada simplesmente pela adição da protease apropriada.

Os fagos eluídos são, então, amplificados pela infecção de células hospedeiras, e podem substancialmente serem reclassificados pelo método já esboçado para reduzir o número de ligantes positivos falsos. Durante cada etapa de classificação de fago, cuidado deve ser tomado para incluir crescimento dos fagos em um meio sólido preferivelmente do que exclusivamente em um meio líquido à medida que este minimiza perda de clones de fagos que crescem subotimamente.

As proteínas de ligação quiméricas de planta podem também serem expressas e classificadas para ligação somente *in vitro* usando-se revelação ribossomal. Uma aproximação exclusivamente *in vitro* tira vantagem do requerimento de introduzir a biblioteca de ácido nucléicos que codifica proteínas de ligação quiméricas de planta em um sistema de replicação biológico. Métodos de classificação de polipeptídeos *in vitro* por revelação de

proteína ribossomal são descritos em detalhes, por exemplo, na Patente dos Estados Unidos Nº 6.589.741. Os ácidos nucleicos na seção acima são modificados por adição de uma seqüência promotora de fagos (por exemplo, um promotor T7) que capacita transcrição *in vitro*, uma seqüência de ligação

5 ribossomal a montante ao início de translação da proteína de ligação quimérica de planta, e uma seqüência de terminação de transcrição (por exemplo, de fagos T3). A biblioteca modificada de ácidos nucleicos é, então, transcrita *in vitro* para gerar uma população de mRNA correspondente que codifica proteínas de ligação quiméricas de planta. As proteínas de ligação quiméricas de planta são, então, expressas *in vitro* pela translação da população de

10 moléculas de mRNA destituídas de códons de parada na estrutura de leitura correta em um sistema de translação *in vitro*, sob condições que permitem a formação de polissomas. Os polissomas assim formados são então trazidos em contato com uma molécula-alvo sob condições que permitem a interação de proteínas de ligação quiméricas de planta com a molécula-alvo. Os polissomas que revelam proteínas de ligação quiméricas que interagem com a molécula-alvo são então separados dos polissomas não-interados que não revelam tais (poli)peptídeos; e o mRNA associado com o polissoma de interação é então amplificado (por exemplo, por PCR), e seqüenciado.

15 20 A interação de uma proteína de ligação quimérica de planta com uma proteína alvo pode também ser detectada em uma classificação genética. Na classificação, a proteína alvo funciona como uma "proteína de isca", e cada proteína de ligação quimérica de planta funciona como uma proteína "prey" potencial em um ensaio de ligação que utiliza um ensaio de dois híbridos, ou ensaio de três híbridos (ver, por exemplo, Patente dos Estados

25 Unidos Nº 5.283.317; Zervos et al., (1993) *Cell* 72:223-232; Madura et al., (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al., (1993) *Oncogene* 8:1693-1696; Hubsman et al., (2001) *Nuc. Acids Res.* Fev 15; 29(4): E18, e Brent WO94/10300).

30 Um ensaio de dois híbridos pode ser efetuado usando-se um polipeptídeo alvo como a proteína de isca. Em resumo, o polipeptídeo alvo é fundido ao domínio de ligação de LexA DNA, e usado como isca. O "prey" é

biblioteca de proteína de ligação quimérica de planta clonada na volta de local ativo de TrxA como uma proteína de fusão com um sinal de localização nuclear N-terminal, um domínio de ativação LexA, e uma etiqueta de epítopo (Colas et al., 1996 *Nature* 380:548; e Gyuris et al, *Cell* 1993 75:791). Células

5 de levedura são transformadas com genes de isca e "prey". Quando a proteína de fusão alvo se liga a uma proteína de fusão de proteína de ligação quimérica de planta, o domínio de ativação LexA é trazido em proximidade com o domínio de ligação de LexA DNA e expressão de genes relatores ou genes marcadores selecionáveis tendo um aumento de local de ligação LexA

10 xA apropriadamente posicionado. Genes relatores adequados incluem proteínas fluorescentes (por exemplo, EGFP), enzimas (por exemplo, luciferase, β -galactosidase, fosfatase alcalina, etc). Genes marcadores selecionáveis adequados incluem, por exemplo, o gene LEU2 de levedura.

Após identificação de uma ou mais proteínas de ligação quiméricas de ligação alvo, os ácidos nucleicos isolados que codificam as proteínas de ligação quiméricas podem ser mutagenizados pelos métodos aqui

15 descritos, para gerar bibliotecas de expressão pequenas que expressam proteínas de ligação quiméricas variantes. As proteínas de ligação quiméricas-bibliotecas de expressão variantes podem ser classificadas para identificar variantes de proteína de ligação quimérica com propriedades de ligação

20 alvo aperfeiçoadas (por exemplo, afinidade ou especificidade aumentadas).

Os seguintes exemplos específicos são para serem construídos como meramente ilustrativos, e não limitativos do restante da descrição. Sem elaboração adicional, acredita-se que um técnico no assunto pode, baseado na descrição aqui, utilizar a presente invenção na sua extensão mais

25 completa. Todas as publicações aqui citadas são, desse modo, incorporadas por referência em sua totalidade.

EXEMPLOS

Exemplo 1 Desenho e Expressão de Sequências de Polipeptídeo de Estrutura de planta

30

Várias famílias de domínio de proteína foram analisadas para seu uso potencial como estruturas principais. Uma pesquisa de domínios de

PFAM (pfam.wustl.edu; ver Bateman et al (2004)), restringindo a saída para Viridiplantae, foi conduzida para limitar domínios somente para aqueles presentes em plantas verdes. Quatro famílias de domínio de proteína foram selecionadas para desenvolver bibliotecas de reconhecimento molecular universal de planta; o domínio de acessório de fosfatase ácida púrpura (PAP),
 5 domínios C2 de planta, e o motivo de volta espiral-espiral (THT) encontrado em proteínas de repetição de ankyrin.

Três estruturas principais de fosfatase ácida púrpura foram designados tendo a seqüência de SEQ ID NOs:34-36. A seqüência de aminoácido do domínio de acessório de PAP de feijão "kidney" foi usada como uma
 10 seqüência de consulta para BLAST, à base de dados NCBI. Quando a saída é restringida às proteínas encontradas em *Viridiplanteae*, 62 seqüências únicas foram identificadas. A partir de um alinhamento destas seqüências, uma seqüência de PAP de planta de consenso foi gerada (SEQ ID NO:34) pela
 15 seleção do aminoácido mais freqüente em cada posição no alinhamento. O PAP de feijão "kidney" (*Phaseolus vulgaris*) foi selecionado como um estrutura principal parenteral (SEQ ID NO:35), devido à sua estrutura conhecida. Um PAP de feijão-soja, *Glycine max*, foi também escolhido (SEQ ID NO:36), conforme esta espécie representa um espécie de colheita comum em que
 20 produtos transgênicos são gerados.

Um conjunto de seqüências de polipeptídeo de estrutura principal que contém repetições similares a ankyrin de planta foi também designado. Repetições similares a ankyrin são motivos de volta espiral-espiral (THH) consistindo em aproximadamente 33 aminoácidos. Elas são elementos
 25 comuns de proteínas de todos os organismos, e são freqüentemente encontradas em séries em tandem de 2 a 20 repetições dentro de uma proteína.

Três estruturas principais de THH foram gerados. Estas proteínas são similares na estrutura a proteína de ligação de GA (GABP- β). Esta proteína consiste em capas amino e carbóxi terminal similares a THH com 3
 30 repetições internas de THH. Nesta proteína, é pensado que as capas ajudam a estabilizar a proteína por blindagem dos resíduos hidrofóbicos encontrados ns repetições internas.

Trezentas e doze repetições de ankyrin de Viridiplantae encontradas em PFMA foram alinhadas para auxiliar na designação de estruturas principais de THH específicos de planta. Uma seqüência de THH de consenso de planta foi gerada pela seleção do aminoácido que ocorre mais frequentemente em cada posição. Esta seqüência foi denominada a seqüência de repetição interna de consenso de planta. Esta seqüência foi usada para pesquisar as bases de dados NCBI por alinhamento de BLAST para encontrar a seqüência de THH natural mais próxima encontrada nas plantas. Uma seqüência de trigo (*Triticum aestivum*) foi encontrada. A repetição designada baseada em *T. aestivum* contém uma substituição de valina para a cisteína simples que ocorre na seqüência de *T. aestivum*. Um conjunto consiste em seqüências derivadas de GABP- β e o segundo conjunto foi derivado a partir da seqüência de consenso de THH de planta, e otimizado para assemelhar-se às estruturas de GABP- β . Em particular, a capa N-terminal tem uma estrutura alfa-espiral estendida, enquanto a capa C-terminal tem uma espiral truncada comparada à repetição de THH típica.

Três estruturas principais de THH foram designados, um consiste em capas N e C de consenso de planta e duas repetições de THH internas de consenso (SEQ ID NO:37). Outro consiste em capas N e C de consenso de planta e duas repetições internas de trigo (SEQ ID NO:38), e o terceiro consiste em capas N e C similares a ankyrin com duas repetições internas de trigo (SEQ ID NO:39).

Os genes que codificam as seqüências de polipeptídeo de estrutura de planta foram designados para teste de expressão em plantas, bactérias, e na superfície de fagos. Os códons foram selecionados para expressão de planta usando uma tabela de uso de códon *Glycine max* publicamente disponível (em kazusa.or.jp/codon, uso de códon tabulado das bases de dados de seqüência de DNA internacional: estado para o ano 2000. Nakamura, Y, Gojobori, T e Ikemura, T (2000) *Nucl. Acids Res.* 28:292). A seleção de códon foi feita manualmente com o auxílio da freqüência de códon final para refletir bruta a freqüência natural para *Glycine max*. Códon raramente usados (<10% de freqüência) não foram usados. As seqüências

5 finais foram verificadas para seqüências problemáticas potenciais, incluindo remoção de locais de restrição necessários para clonagem, locais receptores ou doadores de união de planta potenciais (ver website em [cbs.dtu.dk/ services/NetPgene/](http://cbs.dtu.dk/services/NetPgene/)), seqüências de desestabilização potenciais (ATTTA), e extensões de mais do que 4 ocorrências do mesmo nucleotídeo. Quaisquer seqüências problemáticas potenciais foram alteradas nos genes pela modificação de uso de códon. Desde que as seqüências de THH têm 4 seqüências de repetição similares dentro de cada proteína, às etapas foram tomadas para reduzir similaridade de nucleotídeo dentro das repetições; a identidade de repetição média foi reduzida 10-15% por estes meios.

10 Sete constructos foram produzidas usando-se montagem de gene (três baseadas nas seqüências de polipeptídeo de estrutura principal de THH, duas baseadas nas seqüências de polipeptídeo de estrutura principal de PAP, uma cistatina de planta, e uma proteína de domínio C2 de planta). As três seqüências de polipeptídeo de estrutura principal de THH foram colocadas em um vetor de fagos médios como seqüências de fusão com a proteína de revestimento III de gene (gIII) em seu terminal carbóxi (Phage 3.2, Maxim Biotech, Inc, South San Francisco, CA). Uma etiqueta 6-His foi incluída na extremidade 5' do gene, bem como uma etiqueta c-Myc entre o gene de estrutura principal e o terminal amino codificado de gIII. Os constructos de fagos médios foram então acondicionadas, e os fagos foram testados para expressão e revelação de superfície do estrutura principal de THH. Um fago ELISA usando ou anti-His e anti-Myc indicou que as proteínas de estrutura principal de THH foram expressas na superfície de fagos em fagos ELISA, sugerindo que todos os três constructos de seqüências de polipeptídeo de estrutura principal de THH se desdobraram e expressaram bem na superfície dos fagos. As seqüências de polipeptídeo de estrutura principal selecionadas foram então usadas para gerar vetores de expressão para avaliar sua expressão em plantas transgênicas por imunomanchamento.

30 Folhas de tabaco foram injetadas com *agrobacterium*, LB4404 transformadas com vetores de expressão de planta contendo THH. Dois dias depois, seções de folhas injetadas com *agrobacterium* foram colhidas, con-

geladas em gelo seco, em seguida moídas em um pó fino com um pilão. PBS contendo 0,2% de Tween-20 foi adicionado ao pó fino em uma razão de peso para volume de 1:1, e moagem adicional foi feita. O material insolúvel foi removido por centrifugação, e 10 ul do sobrenadante remanescente foi carregado em 4-12% de gel page SDS de acrilamida (Nupage, Invitrogen). As proteínas foram transferidas para membranas de PVDF. As proteínas foram detectadas usando-se um anticorpo HA-anti-rato (Rochet) e um anticorpo secundário conjugado HRP anti-rato (Chemicon). HPR foi detectado usando-se reagentes Amerham Lumigen.

- 10 Todos os três estruturas principais de THH foram encontrados para serem expressos, com o nível relativo de expressão dos três estruturas principais sendo TA-THH> CC-THH> TC-THH.

OUTRAS CONCRETIZAÇÕES

- 15 Todas as características reveladas neste relatório podem ser combinadas em qualquer combinação. Cada característica revelada neste relatório pode ser substituída por uma característica alternativa que serve para a mesma ou proposta equivalente. Desse modo, a menos que expressamente de outro modo citado, cada característica revelada é somente um exemplo de uma série genérica de características equivalentes ou similares.

- 20 A partir da descrição acima, um técnico no assunto pode facilmente determinar as características essenciais da presente invenção, e sem fugir do espírito e escopo da mesma, pode fazer várias mudanças e modificações da invenção para adaptá-la a vários usos e condições. Desse modo, outras concretizações estão também dentro do escopo das reivindicações
- 25 que se seguem.

REIVINDICAÇÕES

1. Biblioteca de ácidos nucléicos que codificam pelo menos dez polipeptídeos diferentes, a seqüência de aminoácidos de cada polipeptídeo compreendendo:

5 $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual

(i) subsequência C_1 é selecionada de SEQ ID NOS:1-30, subsequência C_2 é selecionada de SEQ ID NOS:31-60, subsequência C_3 é selecionada de SEQ ID NOS:61-90, subsequência C_4 é selecionada de SEQ ID NOS:91-120, e cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada;

(ii) C_1-C_4 são homogêneas através de uma pluralidade dos polipeptídeos codificados;

15 (iii) cada uma de X_1-X_3 é uma subsequência independentemente variável consistindo em 2-20 aminoácidos; e

(iv) cada uma de X_1-X_3 são heterogêneas através de uma pluralidade dos polipeptídeos codificados.

20 2. Biblioteca de ácidos nucléicos que codificam pelo menos dez polipeptídeos diferentes, a seqüência de aminoácidos de cada polipeptídeo compreendendo:

$C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual

25 (i) subsequência C_1 é selecionada de figura 2 ou figura 4, subsequência C_2 é selecionada de figura 2 ou figura 4, subsequência C_3 é selecionada de figura 2 ou figura 4; subsequência C_4 é selecionada de figura 2 ou figura 4, e cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada;

(ii) C_1-C_4 são homogêneas através de uma pluralidade dos polipeptídeos codificados;

30 (iii) cada uma de X_1-X_3 é uma subsequência independentemente variável consistindo em 2-20 aminoácidos; e

(iv) cada uma de X_1-X_3 são heterogêneas através de uma

pluralidade dos polipeptídeos codificados.

3. Biblioteca de ácidos nucleicos que codificam pelo menos dez polipeptídeos diferentes, a seqüência de aminoácidos de cada polipeptídeo compreendendo:

5 $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual

(i) subsequência C_1 é selecionada de figura 3 ou figura 5, subsequência C_2 é selecionada de figura 3 ou figura 5, subsequência C_3 é selecionada de figura 3 ou figura 5; subsequência C_4 é selecionada de figura 3 XX, e cada uma de C_1-C_4 compreende até 30 substituições, retiradas, in-
10 serções ou adições de aminoácido simples à subsequência selecionada;

(ii) C_1-C_4 são homogêneas através de uma pluralidade dos polipeptídeos codificados;

(iii) cada uma de X_1-X_3 é uma subsequência independentemente variável consistindo em 2-20 aminoácidos; e

15 (iv) cada uma de X_1-X_3 são heterogêneas através de uma pluralidade dos polipeptídeos codificados.

4. Biblioteca, de acordo com qualquer uma das reivindicações 1 a 3, no qual pelo menos 1.000 polipeptídeos diferentes são codificados.

5. Biblioteca, de acordo com qualquer uma das reivindicações 1
20 a 3, no qual pelo menos 100.000 polipeptídeos diferentes são codificados.

6. Biblioteca, de acordo com qualquer uma das reivindicações 1 a 3, no qual pelo menos 1.000.000 de polipeptídeos diferentes são codificados.

7. Biblioteca, de acordo com qualquer uma das reivindicações 1
25 a 3, no qual cada uma de C_1-C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

8. Biblioteca, de acordo com a reivindicação 7, no qual cada uma de C_1-C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.
30

9. Biblioteca, de acordo com a reivindicação 8, no qual cada

uma de C₁-C₄ compreende independentemente até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

5 10. Biblioteca, de acordo com a reivindicação 7, no qual nenhuma de C₁-C₄ compreende substituições, retiradas, inserções ou adições de aminoácido à subsequência selecionada.

11. Biblioteca, de acordo com qualquer uma das reivindicações 1 a 3, no qual os aminoácidos de X₁-X₃ são selecionados de pouco menos do que 20 aminoácidos geneticamente codificados em plantas.

10 12. Biblioteca, de acordo com a reivindicação 11, no qual os pouco menos do que 20 aminoácidos geneticamente codificados incluem pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático.

15 13. Biblioteca, de acordo com a reivindicação 12, no qual os pouco menos do que 20 aminoácidos geneticamente codificados compreendem alanina, aspartato, serina e tirosina.

14. Biblioteca, de acordo com a reivindicação 1, no qual a sequência de aminoácidos de cada polipeptídeo é selecionada de:

20 (a) um polipeptídeo compreendendo C₁-X₁-C₂-X₂-C₃-X₃-C₄, no qual C₁ = SEQ ID NO:1, C₂ = SEQ ID NO:31, C₃ = SEQ ID NO:61, e C₄ = SEQ ID NO:91;

(b) um polipeptídeo compreendendo C₁-X₁-C₂-X₂-C₃-X₃-C₄, no qual C₁ = SEQ ID NO:2, C₂ = SEQ ID NO:32, C₃ = SEQ ID NO:62, e C₄ = SEQ ID NO:92; e

25 (c) um polipeptídeo compreendendo C₁-X₁-C₂-X₂-C₃-X₃-C₄, no qual C₁ = SEQ ID NO:3, C₂ = SEQ ID NO:33, C₃ = SEQ ID NO:63, e C₄ = SEQ ID NO:93.

30 15. Biblioteca, de acordo com a reivindicação 2, no qual cada polipeptídeo codificado compreende C₁-X₁-C₂-X₂-C₃-X₃-C₄, no qual C₁ = SEQ ID NO: X1, C₂ = SEQ ID NO: X2, C₃ = SEQ ID NO: X3, e C₄ = SEQ ID NO: X4; SEQ designada. ID NO:130.

16. Biblioteca, de acordo com a reivindicação 3, no qual cada

polipeptídeo codificado compreende $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual $C_1 =$ SEQ ID NO: X1, $C_2 =$ SEQ ID NO: X2, $C_3 =$ SEQ ID NO: X3, e $C_4 =$ SEQ ID NO: X4; designada SEQ ID NO:130.

17. Biblioteca, de acordo com qualquer uma das reivindicações 1 a 3, no qual cada um dos ácidos nucleicos compreende uma sequência de vetor.

18. Ácido nucleico isolado selecionado a partir da biblioteca como definida em qualquer uma das reivindicações 1 a 3.

19. Ácido nucleico isolado selecionado a partir da biblioteca como definida em qualquer uma das reivindicações 14 a 16.

20. Célula que expressa o ácido nucleico isolado como definido na reivindicação 19.

21. Ácido nucleico isolado selecionado a partir da biblioteca como definida em reivindicação 17.

22. Polipeptídeo purificado codificado pelo ácido nucleico isolado como definido na reivindicação 18.

23. População de células que expressam os polipeptídeos codificados pela biblioteca como definido em qualquer uma das reivindicações 1 a 3.

24. Célula selecionada a partir da população como definida na reivindicação 23.

25. Biblioteca de polipeptídeos purificados codificados pela biblioteca como definida em qualquer uma das reivindicações 1 a 3.

26. População de fago filamentoso que revela os polipeptídeos codificados pela biblioteca como definida em qualquer uma das reivindicações 1 a 3.

27. Método de gerar a biblioteca como definida na reivindicação 1, compreendendo:

(i) provisão de um ácido nucleico parenteral que codifica um polipeptídeo parenteral compreendendo a sequência de aminoácido: $C_1-X_1-C_2-X_2-C_3-X_3-C_4$, no qual a subsequência C_1 é selecionada de SEQ ID NOS:1-30, a subsequência C_2 é selecionada de SEQ ID NOS:31-60, a sub-

seqüência C₃ é selecionada de SEQ ID NOS:61-90, a subsequência C₄ é selecionada de SEQ ID NOS:91-120, cada uma de C₁-C₄ compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada; e cada X₁-X₃ é uma subsequência independente consistindo em 2-20 aminoácidos;

(ii) replicação do ácido nucléico parenteral sob condições que introduzem até 10 substituições, retiradas, inserções ou adições de aminoácido simples às subsequências X₁, X₂ ou X₃, pelo que uma população de seqüências aleatoriamente variadas que codificam X₁', X₂' ou X₃' é gerada; e

(iii) a população de seqüências aleatoriamente variadas X₁', X₂' ou X₃' é substituída, em uma população de ácidos nucléicos parenterais nas posições correspondentes àquelas que codificam X₁', X₂' ou X₃'.

28. Método, de acordo com a reivindicação 27, no qual pelo menos uma das subsequências X₁-X₃ é selecionada de SEQ ID Nos:121-123.

29. Método, de acordo com a reivindicação 27, no qual cada uma de C₁-C₄ compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

30. Método, de acordo com a reivindicação 29, no qual cada uma de C₁-C₄ compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

31. Método, de acordo com a reivindicação 30, no qual cada uma de C₁-C₄ compreende independentemente até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

32. Método, de acordo com a reivindicação 31, no qual cada uma de C₁-C₄ compreende substituições, retiradas, inserções ou adições de aminoácido à subsequência selecionada.

33. Método, de acordo com a reivindicação 27, no qual a replicação gera uma população heterogênea de subsequências aleatoriamente variadas pela introdução de até 5 substituições de aminoácido em cada uma

de X_1 , X_2 ou X_3 .

34. Método, de acordo com a reivindicação 27, compreendendo adicionalmente amplificar a biblioteca pela introdução da mesma em um sistema de replicação biológico e proliferação do sistema de replicação biológico.

5 35. Método, de acordo com a reivindicação 34, no qual o sistema de replicação biológico é uma pluralidade de células de *E. coli*.

36. Método, de acordo com a reivindicação 34, no qual o sistema de replicação biológico é uma pluralidade de bacteriófago.

10 37. Método de acordo com a reivindicação 27, em que ocorreu a replicação *in vitro*.

38. Método, de acordo com a reivindicação 37, no qual a replicação é realizada com uma polimerase mutagênica purificada.

39. Método, de acordo com a reivindicação 37, no qual a replicação é realizada na presença de um análogo de nucleotídeo.

15 40. Método, de acordo com a reivindicação 27, no qual a replicação ocorre *in vivo*.

41. Método, de acordo com a reivindicação 40, no qual a replicação *in vivo* ocorre em uma espécie mutagênica de *E. coli*.

20 42. Método de gerar a biblioteca como definida na reivindicação 1, compreendendo:

(i) seleção de uma seqüência de aminoácido compreendendo a seqüência de aminoácido: C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 a ser codificada, no qual

25 (a) a subsequência C_1 é selecionada de SEQ ID NOS:1-30, a subsequência C_2 é selecionada de SEQ ID NOS:31-60, a subsequência C_3 é selecionada de SEQ ID NOS:61-90, a subsequência C_4 é selecionada de SEQ ID NOS:91-120;

30 (b) cada uma de C_1 - C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada;

(c) cada X_1 , X_2 e X_3 consiste em uma subsequência de aminoácido de 2-20 aminoácidos de comprimento;

(ii) provisão de uma primeira pluralidade e uma segunda pluralidade de oligonucleotídeos, no qual

(a) oligonucleotídeos da primeira pluralidade codificam as subseqüências C_1 - C_4 e subseqüências variantes X_1 - X_3 heterogêneas múltiplas
5 X_1' - X_3' ;

(b) oligonucleotídeos da segunda pluralidade são complementares às seqüências de nucleotídeo que codificam as subseqüências C_1 - C_4 e às seqüências de nucleotídeo que codificam as subseqüências X_1' - X_3' heterogêneas múltiplas; e

10 (c) oligonucleotídeos das primeira e segunda pluralidades têm seqüências de sobreposição uma com a outra;

(iii) combinação da população de oligonucleotídeos para formar uma primeira mistura;

(iv) incubação da mistura sob condições eficazes para hibridização das seqüências complementares de sobreposição para formar uma
15 pluralidade de seqüências complementares hibridizadas; e

(v) alongamento da pluralidade de seqüências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

43. Método, de acordo com a reivindicação 42, no qual cada
20 uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada.

44. Método, de acordo com a reivindicação 43, no qual cada
25 uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada.

45. Método, de acordo com a reivindicação 44, no qual cada
30 uma de C_1 - C_4 compreende independentemente de zero até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada.

46. Método, de acordo com a reivindicação 42, compreendendo adicionalmente realizar um ciclo de etapas, o ciclo de etapas compreenden-

do desnaturação da biblioteca pelo aumento da temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento duplo, seguida pelas etapas (iv) e (v).

47. Método, de acordo com a reivindicação 46, compreendendo
5 repetição do ciclo de etapas até 100 vezes.

48. Método, de acordo com a reivindicação 42, compreendendo
adicionalmente amplificação da biblioteca por uma reação de cadeia de po-
limerase consistindo essencialmente da biblioteca, um iniciador de avanço, e
um iniciador reverso, no qual os prímeres de avanço e reverso podem hibri-
10 dizar para as seqüências terminais 5' e 3', respectivamente, de todos os áci-
do nucléicos na biblioteca.

49. Método, de acordo com a reivindicação 42, no qual o amino-
ácido a ser codificado em cada posição das subseqüências X_1 , X_2 ou X_3 , é
selecionado de um subconjunto de alanina, arginina, asparagina, aspartato,
15 cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina,
metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina e valina.

50. Método, de acordo com a reivindicação 49, no qual o amino-
ácido selecionado para cada substituição de aminoácido simples é selecio-
nado de um grupo de aminoácidos consistindo em pelo menos um aminoá-
cido alifático, pelo menos um aminoácido ácido, pelo menos um aminoáci-
do neutro, e pelo menos um aminoácido aromático.
20

51. Método, de acordo com a reivindicação 50, no qual o grupo
de aminoácidos consiste em alanina, aspartato, serina e tirosina.

52. Método de gerar a biblioteca como definida na reivindicação
25 2, compreendendo:

(i) provisão de um ácido nucléico parenteral que codifica
um polipeptídeo parenteral compreendendo a seqüência de aminoácido: C_1 -
 X_1 - C_2 - X_2 - C_3 - X_3 - C_4 , no qual C_1 é selecionada de figura 2 ou figura 4, a sub-
seqüência C_2 é selecionada de figura 2 ou figura 4, a subseqüência C_3 é
30 selecionada de figura 2 ou figura 4, a subseqüência C_4 é selecionada de
figura 2 ou figura 4, cada uma de C_1 - C_4 compreende até 10 substituições,
retiradas, inserções ou adições de aminoácido simples à subseqüência sele-

cionada; e cada uma de X_1 - X_2 é uma subsequência independente consistindo em 2-10 aminoácidos;

(ii) replicação do ácido nucléico parenteral sob condições que introduzem até 10 substituições, retiradas, inserções ou adições de aminoácido simples às subsequências de X_1 , X_2 , ou X_3 , pelo que uma população de subsequências aleatoriamente variadas que codificam X_1' , X_2' , ou X_3' é gerada; e

(iii) a população de subsequências aleatoriamente variadas X_1' , X_2' , ou X_3' é substituída, em uma população de ácidos nucléicos parenterais nas posições correspondentes àquelas que codificam X_1 , X_2 , ou X_3 .

53. Método, de acordo com a reivindicação 52, no qual pelo menos uma das subsequências X_1 - X_2 é selecionada de SEQ ID Nos:121-123.

54. Método, de acordo com a reivindicação 52, no qual cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

55. Método, de acordo com a reivindicação 54, no qual cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

56. Método, de acordo com a reivindicação 55, no qual cada uma de C_1 - C_4 compreende independentemente até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

57. Método, de acordo com a reivindicação 56, no qual nenhuma de C_1 - C_4 compreende substituições, retiradas, inserções ou adições de aminoácido.

58. Método, de acordo com a reivindicação 52, no qual a replicação gera uma população heterogênea de subsequências aleatoriamente variadas pela introdução de até 5 substituições de aminoácido em cada uma de X_1 , X_2 , ou X_3 .

59. Método, de acordo com a reivindicação 52, compreendendo

adicionalmente amplificar a biblioteca pela introdução da mesma em um sistema de replicação biológico e proliferação do sistema de replicação biológico.

5 60. Método, de acordo com a reivindicação 59, no qual o sistema de replicação biológico é uma pluralidade de células de *E. coli*.

61. Método, de acordo com a reivindicação 59, no qual o sistema de replicação biológico é uma pluralidade de bacteriófago.

62. Método, de acordo com a reivindicação 52, no qual a replicação ocorre *in vitro*.

10 63. Método, de acordo com a reivindicação 62, no qual a replicação é realizada com uma polimerase mutagênica purificada.

64. Método, de acordo com a reivindicação 62, no qual a replicação é realizada na presença de um análogo de nucleotídeo.

15 65. Método, de acordo com a reivindicação 52, no qual a replicação ocorre *in vivo*.

66. Método, de acordo com a reivindicação 65, no qual a replicação *in vivo* ocorre em uma espécie mutagênica de *E. coli*.

67. Método de gerar a biblioteca como definida na reivindicação 2, compreendendo:

20 (i) seleção de uma seqüência de aminoácido compreendendo $C_1-X_1-C_2-X_2-C_3-X_3-C_4$ a ser codificada, no qual

(a) a subsequência C_1 é selecionada de figura 2 ou figura 4, a subsequência C_2 é selecionada de figura 2 ou figura 4, a subsequência C_3 é selecionada de figura 2 ou figura 4, a subsequência C_4 é selecionada de
25 figura 2 ou figura 4;

(b) cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada;

(c) e cada uma de X_1 , X_2 , e X_3 consiste em uma seqüência de
30 aminoácido de 2-20 aminoácidos de comprimento;

(ii) provisão de uma primeira pluralidade e um segunda pluralidade de oligonucleotídeos, no qual

(a) oligonucleotídeos da primeira pluralidade codificam as subsequências C_1 - C_4 e subsequências variantes X_1 - X_3 heterogêneas múltiplas X_1' - X_3' ;

5 (b) oligonucleotídeos da segunda pluralidade são complementares às seqüências de nucleotídeo que codificam as subsequências C_1 - C_4 e às seqüências de nucleotídeo que codificam as subsequências X_1' - X_3' heterogêneas múltiplas; e

(c) oligonucleotídeos das primeira e segunda pluralidades têm seqüências de sobreposição uma com a outra;

10 (iii) combinação da população de oligonucleotídeos para formar uma primeira mistura;

(iv) incubação da mistura sob condições eficazes para hibridização das seqüências complementares de sobreposição para formar uma pluralidade de seqüências complementares hibridizadas; e

15 (v) alongamento da pluralidade de seqüências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

68. Método, de acordo com a reivindicação 67, no qual cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

69. Método, de acordo com a reivindicação 68, no qual cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

25 70. Método, de acordo com a reivindicação 44, no qual cada uma de C_1 - C_4 compreende independentemente de zero até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

71. Método, de acordo com a reivindicação 67, compreendendo
30 adicionalmente realizar um ciclo de etapas, o ciclo de etapas compreendendo desnaturação da biblioteca pelo aumento da temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento

duplo, seguida pelas etapas (iv) e (v).

72. Método, de acordo com a reivindicação 71, compreendendo repetição do ciclo de etapas até 100 vezes.

73. Método, de acordo com a reivindicação 67, compreendendo adicionalmente amplificação da biblioteca por uma reação de cadeia de polimerase consistindo essencialmente da biblioteca, um iniciador de avanço, e um iniciador reverso, no qual os primers de avanço e reverso podem hibridizar para as seqüências terminais 5' e 3', respectivamente, de todos os ácidos nucleicos na biblioteca.

74. Método, de acordo com a reivindicação 67, no qual o aminoácido a ser codificado em cada posição das subseqüências X_1 , X_2 ou X_3 , é selecionado de um subconjunto de alanina, arginina, asparagina, aspartato, cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina e valina.

75. Método, de acordo com a reivindicação 74, no qual o aminoácido selecionado para cada substituição de aminoácido simples é selecionado de um grupo de aminoácidos consistindo em pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático.

76. Método, de acordo com a reivindicação 75, no qual o grupo de aminoácidos consiste em alanina, aspartato, serina e tirosina.

77. Método de gerar a biblioteca de acordo com a reivindicação 3, compreendendo:

(i) provisão de um ácido nucleico parenteral que codifica um polipeptídeo parenteral compreendendo a seqüência de aminoácido: C_1 - X_1 - C_2 - X_2 - C_3 - X_3 - C_4 , no qual C_1 é selecionada de figura 3 ou figura 5, a subseqüência C_2 é selecionada de figura 3 ou figura 5, a subseqüência C_3 é selecionada de figura 3 ou figura 5, a subseqüência C_4 é selecionada de figura 3 ou figura 5, cada uma de C_1 - C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada; e cada uma de X_1 - X_3 é uma subseqüência independente consistindo em 2-20 aminoácidos;

(ii) replicação do ácido nucléico parenteral sob condições que introduzem até 10 substituições, retiradas, inserções ou adições de aminoácido simples às subsequências de X_1 , X_2 , ou X_3 , pelo que uma população de subsequências aleatoriamente variadas que codificam X_1' , X_2' , ou X_3' é gerada; e

(iii) a população de subsequências aleatoriamente variadas X_1' , X_2' , ou X_3' é substituída, em uma população de ácidos nucléicos parenterais nas posições correspondentes àquelas que codificam X_1 , X_2 , ou X_3 .

78. Método, de acordo com a reivindicação 77, no qual pelo menos uma das subsequências X_1 - X_3 é selecionada de SEQ ID Nos:121-123.

79. Método, de acordo com a reivindicação 27, no qual cada uma de C_1 - C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

80. Método, de acordo com a reivindicação 79, no qual cada uma de C_1 - C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

81. Método, de acordo com a reivindicação 80, no qual cada uma de C_1 - C_4 compreende independentemente até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada.

82. Método, de acordo com a reivindicação 81, no qual nenhuma de C_1 - C_4 compreende substituições, retiradas, inserções ou adições de aminoácido.

83. Método, de acordo com a reivindicação 77, no qual a replicação gera uma população heterogênea de subsequências aleatoriamente variadas pela introdução de até 5 substituições de aminoácido em cada uma de X_1 , X_2 , ou X_3 .

84. Método, de acordo com a reivindicação 77, compreendendo adicionalmente amplificar a biblioteca pela introdução da mesma em um sistema de replicação biológico e proliferação do sistema de replicação biológico.

co.

85. Método, de acordo com a reivindicação 84, no qual o sistema de replicação biológico é uma pluralidade de células de *E. coli*.

86. Método, de acordo com a reivindicação 84, no qual o sistema de replicação biológico é uma pluralidade de bacteriófago.

87. Método, de acordo com a reivindicação 77, no qual a replicação ocorre *in vitro*.

88. Método, de acordo com a reivindicação 87, no qual a replicação é realizada com uma polimerase mutagênica purificada.

89. Método, de acordo com a reivindicação 87, no qual a replicação é realizada na presença de um análogo de nucleotídeo.

90. Método, de acordo com a reivindicação 77, no qual a replicação ocorre *in vivo*.

91. Método, de acordo com a reivindicação 90, no qual a replicação *in vivo* ocorre em uma espécie mutagênica de *E. coli*.

92. Método de gerar a biblioteca de acordo com a reivindicação 3, compreendendo:

(i) seleção de uma seqüência de aminoácido compreendendo $C_1-X_1-C_2-X_2-C_3-X_3-C_4$ a ser codificada, no qual

(a) a subsequência C_1 é selecionada de figura 3 ou figura 5, a subsequência C_2 é selecionada de figura 3 ou figura 5, a subsequência C_3 é selecionada de figura 3 ou figura 5, a subsequência C_4 é selecionada de figura 3 ou figura 5;

(b) cada uma de C_1-C_4 compreende até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subsequência selecionada;

(c) e cada uma de X_1 , X_2 , e X_3 consiste em uma seqüência de aminoácido de 2-20 aminoácidos de comprimento;

(ii) provisão de uma primeira pluralidade e um segunda pluralidade de oligonucleotídeos, no qual

(a) oligonucleotídeos da primeira pluralidade codificam as subsequências C_1-C_4 e subsequências variantes X_1-X_3 heterogêneas múltiplas

$X_1'-X_3'$;

(b) oligonucleotídeos da segunda pluralidade são complementares às seqüências de nucleotídeo que codificam as subseqüências C_1-C_4 e às seqüências de nucleotídeo que codificam as subseqüências $X_1'-X_3'$ heterogêneas múltiplas; e

(c) oligonucleotídeos das primeira e segunda pluralidades têm seqüências de sobreposição uma com a outra;

(iii) combinação da população de oligonucleotídeos para formar uma primeira mistura;

(iv) incubação da mistura sob condições eficazes para hibridização das seqüências complementares de sobreposição para formar uma pluralidade de seqüências complementares hibridizadas; e

(v) alongamento da pluralidade de seqüências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

93. Método, de acordo com a reivindicação 92, no qual cada uma de C_1-C_4 compreende independentemente até 20 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada.

94. Método, de acordo com a reivindicação 93, no qual cada uma de C_1-C_4 compreende independentemente até 10 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada.

95. Método, de acordo com a reivindicação 94, no qual cada uma de C_1-C_4 compreende independentemente de zero até 5 substituições, retiradas, inserções ou adições de aminoácido simples à subseqüência selecionada.

96. Método, de acordo com a reivindicação 92, compreendendo adicionalmente realizar um ciclo de etapas, o ciclo de etapas compreendendo desnaturação da biblioteca pelo aumento da temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento duplo, seguida pelas etapas (iv) e (v).

97. Método, de acordo com a reivindicação 96, compreendendo

repetição do ciclo de etapas até 100 vezes.

98. Método, de acordo com a reivindicação 92, compreendendo adicionalmente amplificação da biblioteca por uma reação de cadeia de polimerase consistindo essencialmente da biblioteca, um iniciador de avanço, e um iniciador reverso, no qual os primers de avanço e reverso podem hibridizar para as seqüências terminais 5' e 3', respectivamente, de todos os ácidos nucleicos na biblioteca.

99. Método, de acordo com a reivindicação 92, no qual o aminoácido a ser codificado em cada posição das subsequências X_1 , X_2 ou X_3 , é selecionado de um subconjunto de alanina, arginina, asparagina, aspartato, cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina e valina.

100. Método, de acordo com a reivindicação 99, no qual o aminoácido selecionado para cada substituição de aminoácido simples é selecionado de um grupo de aminoácidos consistindo em pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático.

101. Método, de acordo com a reivindicação 100, no qual o grupo de aminoácidos consiste em alanina, aspartato, serina e tirosina.

102. Biblioteca de ácidos nucleicos que codificam pelo menos dez polipeptídeos diferentes, no qual

(i) a seqüência de aminoácido de cada um dos polipeptídeos codificados compreende uma seqüência de aminoácido pelo menos 70% idêntica a qualquer de SEQ ID Nos:127-129;

(ii) a seqüência de aminoácido de cada um dos polipeptídeos codificados inclui aminoácidos que diferem daquela de SEQ ID Nos:127-129 nas posições 14, 15, 33, 35-36, 38, 47-48, 66, 68-69, 71, 80, 81, 99, 101-102, e 104, e as diferenças de aminoácido são heterogênea através de uma pluralidade dos polipeptídeos codificados; e

(iii) a seqüência de aminoácido de cada um dos polipeptídeos codificados fora dos resíduos correspondem às posições 14, 15, 33, 35-36, 38, 47-48, 66, 68-69, 71, 80, 81, 99, 101-102, e 104 de SEQ ID

Nos:127-129 é homogênea através de uma pluralidade dos polipeptídeos codificados.

5 103. Biblioteca, de acordo com a reivindicação 102, no qual a seqüência de aminoácido dos polipeptídeos tem pelo menos 75% de identidade para qualquer de SEQ ID Nos:127-129.

 104. Biblioteca, de acordo com a reivindicação 102, no qual a seqüência de aminoácido dos polipeptídeos tem pelo menos 80% de identidade para qualquer de SEQ ID Nos:127-129.

10 105. Biblioteca, de acordo com a reivindicação 102, no qual a seqüência de aminoácido dos polipeptídeos tem pelo menos 85% de identidade para qualquer de SEQ ID Nos:127-129.

 106. Biblioteca, de acordo com a reivindicação 102, no qual cada um dos ácidos nucléicos compreende uma seqüência de vetor.

15 107. Ácido nucléico isolado que codifica um polipeptídeo selecionado a partir da biblioteca como definido em qualquer uma das reivindicações 102-106.

 108. Polipeptídeo purificado codificado pelo ácido nucléico como definido na reivindicação 107.

20 109. População de células que expressam os polipeptídeos codificados pela biblioteca como definida na reivindicação 102.

 110. Célula selecionada a partir da população de células como definida na reivindicação 109.

 111. Biblioteca purificada de polipeptídeos codificados pela biblioteca como definida na reivindicação 102.

25 112. População de fago filamentoso revelando a biblioteca de polipeptídeos codificados pela biblioteca como definida na reivindicação 102.

 113. Método de gerar a biblioteca como definida na reivindicação 102, compreendendo:

30 (i) seleção de uma seqüência de aminoácido correspondente a qualquer uma de SEQ ID Nos:127-129 a ser codificada, no qual a seqüência selecionada difere daquelas de SEQ ID Nos:127-129 em pelo menos uma de posições variáveis 14, 15, 33, 35-36, 38, 47-48, 66, 68-69,

71, 80, 81, 99, 101-102, e 104;

(ii) provisão quimicamente de uma primeira e uma segunda pluralidade de oligonucleotídeos, no qual

5 (a) oligonucleotídeos da primeira pluralidade codificam subsequências de aminoácido da sequência de aminoácido selecionada; as subsequências sendo heterogêneas nas posições variáveis codificadas;

(b) oligonucleotídeos da segunda pluralidade são complementares às sequências de nucleotídeo que codificam as subsequências da sequência de aminoácido selecionada, as subsequências sendo heterogêneas
10 nas posições variáveis codificadas; e

(c) a primeira e segunda pluralidades compreendem oligonucleotídeos tendo sequências de sobreposição uma com a outra;

(iii) combinação da população de oligonucleotídeos para formar uma primeira mistura;

15 (iv) incubação da mistura sob condições eficazes para hibridização das sequências complementares de sobreposição para formar uma pluralidade de sequências complementares hibridizadas; e

(v) alongamento da pluralidade de sequências complementares hibridizadas para formar uma segunda mistura contendo a biblioteca.

20 114. Método, de acordo com a reivindicação 113, compreendendo adicionalmente realizar um ciclo de desnaturação da biblioteca pelo aumento da temperatura da segunda mistura a uma temperatura eficaz para desnaturação de DNA de filamento duplo, seguida pelas etapas (iv) e (v).

25 115. Método, de acordo com a reivindicação 114, compreendendo repetição do ciclo de etapas até 100 vezes.

116. Método, de acordo com a reivindicação 115, compreendendo adicionalmente amplificação da biblioteca por uma reação de cadeia de polimerase consistindo essencialmente da biblioteca, um iniciador de avanço, e um iniciador reverso, no qual os primers de avanço e reverso podem hibridizar para as sequências terminais 5' e 3', respectivamente, de todos os ácidos nucleicos na biblioteca.

30 117. Método, de acordo com a reivindicação 113, no qual os

aminoácidos a serem codificados para as posições variáveis são selecionados de um subconjunto de alanina, arginina, asparagina, aspartato, cisteína, glutamina, glutamato, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofan, tirosina e valina.

5 118. Método, de acordo com a reivindicação 117, no qual os aminoácidos selecionados para as posições variáveis são selecionados de um grupo consistindo em pelo menos um aminoácido alifático, pelo menos um aminoácido ácido, pelo menos um aminoácido neutro, e pelo menos um aminoácido aromático.

10 119. Método, de acordo com a reivindicação 118, no qual o grupo de aminoácidos consiste em alanina, aspartato, serina e tirosina.

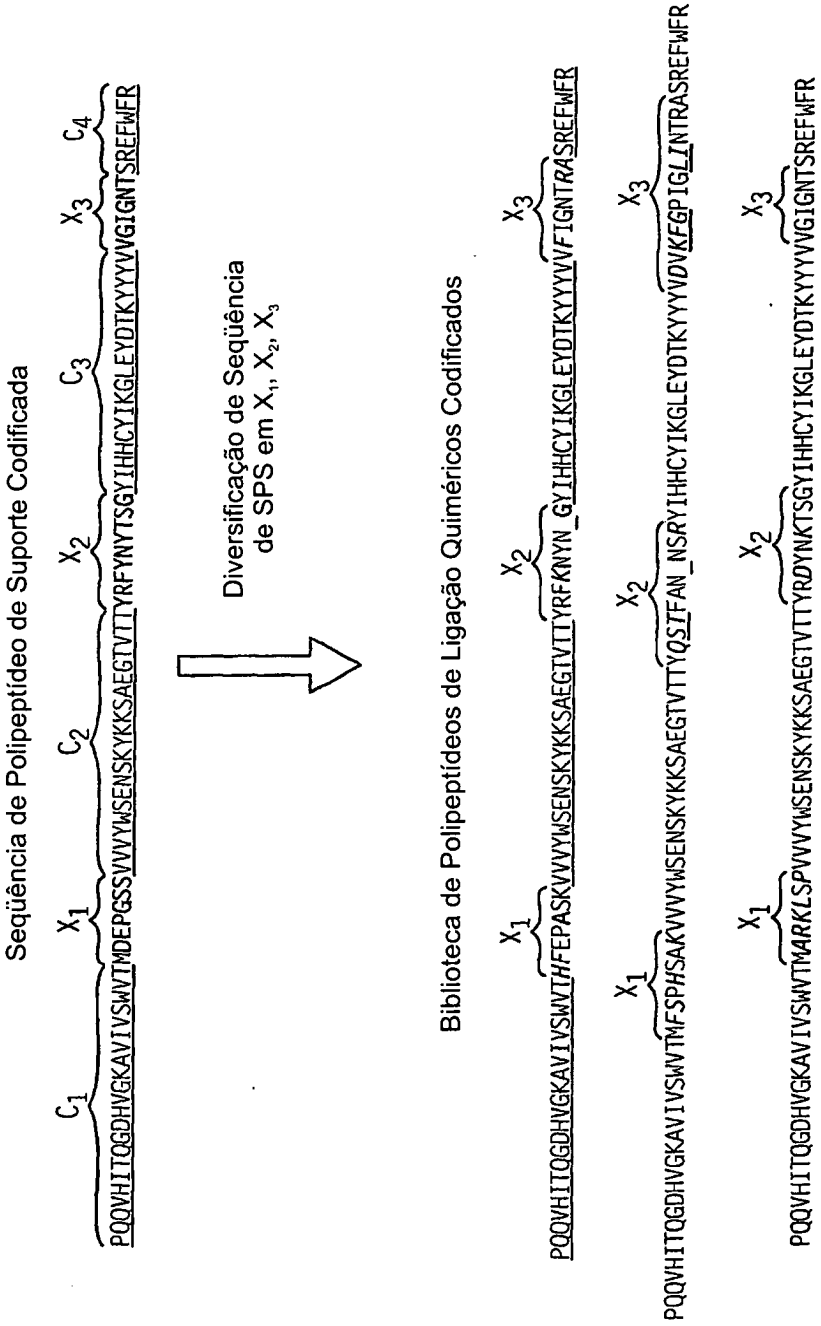


FIG. 1

C1	C2	C3	C4
Q2V816_CUCHA_1441/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q2V817_CUCHA_1441/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q2V818_CUCHA_1441/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q2V8J5_BRVDI_431/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q6DL_C9_CUCHA_1441/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q2V815_CUCHE_431/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q81131_CUCHA_633/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q81132_CUCHA_633/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q2V814_CUCHO_734/1-28	GEVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q2V8H9_LAGLE_431/1-28	GGVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q6DKJ9_CUCHA_1441/1-28	GGVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q6DL_C8_CUCHA_1441/1-28	GGVIGIS--VNDPRVKEIAEFAL--KQHAECN		
Q8C389_CUCSA_795/1-89	GGVPGCKD--PNDPHWKDI AEIAV--AEYNKSGGH	HL--TLVSLKCESQ--VWAGWNRVLKCKD--	ENNGEGWETVWKEKIVENFRQLITFDHLL
Q1RVK3_MEDTR_2578/1-54	GGYSP1KN--LNQHLHLEIANFAV--TEYKQOOG	IKOI--KLEKITKGETQ--WOGTNYRL	
Q1T1R3_MEDTR_2579/1-55	GGYSP1KN--LNQPHVTEIANFAV--TEYKQOOR	FYS--KLDKITKCESQ--AVDGTNYRFTL	
Q2HUA2_MEDTR_2698/1-73	GGWSP1KO--INDPHVTVIANFAV--TEYNKHTG	ANL--KLDKITKCESQ--VTSGIYVDLILSAGO	GSASNITYKALVWIK
Q1SRH3_MEDTR_26102/1-77	GGWEP1KN--INDPHVIDIANFAV--TEHDKQAG	LNH--KLEKVISGETR--VWDGJTYCLNITASD	GSASNITYKALVLELECH
Q1S343_MEDTR_2688/1-63	GNLSP1NN--INDPKVIDIANFAV--KEYNNRRKP	EEKL--RLWKVTKGESQIVADGWYRLTISATK	
Q9SL01_ARATH_87117/1-31		L--QFSRWSAQKQ--WAGLKYLLRIEVTQ	PNGS
Q8L5T9_ARATH_84114/1-31		L--QFSRWSAQKQ--WAGLKYLLRIEVTQ	PNGS
Q4F747_MAIZE_53118/1-66	GGARTEVPDVEGGEVQELGRFSV--AEYNRQL--REGGGGGG	RL--EFGRWVAAQRO--VWGLKYLLRWAVE	
Q4ALJ1_MAIZE_1983/1-65	GGARTEVPDVEGGEVQELGRFSV--AEYNRQL--RGGGGGGG	RL--EFGRWVAAQRO--VWGLKYLLRWAVE	
Q5N806_ORYSA_51116/1-66	GGARTEVPDVEGGEVQELGRFSV--EENHRRR--RSROCGDV	RL--EFGRWVAAQRO--VWGLKYLLRWAAE	
Q7XKS0_ORYSA_40109/1-70	GGARTEVPDVEGGEVQELGRFSV--AEHRRRL--RHGGSGGPADVPVKL	AFARVWEAQKQ--VWGGIAYLLKVA	
Q8AXY6_MALDO_2692/1-67	GGARKEIENWKTKEVQELGRFSV--EENRORGTOKMDGGG	EL--QFLEWEASQ--VWGLKYLLKSAVR	
CYTB_HELAN_366/1-64	GGARTEVPDVEGGEVQELGRFSV--DEYNRQL--RTKKTG	AGDL--KFSQVIAEETQ--WAGLKYLLKTEA	
Q04430_DALCA_33132/1-100	GGARTEIPDVEGGEVQELGRFSV--EENYQOQ--HNGDGGSDTS	AGDL--KFKVWAAEQKQ--VWAGLKYLLKVA	GGHKKKEDAEIWDAAKTKQLMSFAPSH
Q8C2V2_CHEM_340/1-38		FSV--DEYMSKN	
Reference/1-102	HSSGGGPTLGGVEPV--GNENDLHLVQELARFV--TEHNKKA	NSLL--EFEKLVSVKQO--WAGLKYLLLEWKE	GOAKKLYEAKWKEKPMDFKELQEFKPDASANA
CYTL_ORYSA_1097/1-88	GGVEPV--GNENDLHLVQELARFV--TEHNKKA	NSLL--EFEKLVSVKQO--WAGLKYLLLEWKE	GOAKKLYEAKWKEKPMDFKELQEFKPD
Q9ANT9_COTILA_42130/1-89	GGIKOVPANENDLHLQELARFV--DEHNKKA	NALL--GYEKLWAKTKQ--WAGLKYLLTIEVKD	GEWKLYEAKWKEKPMDFKELQEFKPD
Q41294_SORBE_37125/1-89	GGIKOVPANENDLHLQELARFV--DEHNKKA	NALL--GYEKLWAKTKQ--WAGLKYLLTIEVKD	GEWKLYEAKWKEKPMDFKELQEFKPD
P93627_MAIZE_41125/1-89	GGIKOVPGNENDLHLQELARFV--DEHNKKA	NALL--GYEKLWAKTKQ--WAGLKYLLTIEVKD	GEWKLYEAKWKEKPMDFKELQEFKPD

FIG. 2 (Folha 1 de 6)

	C2	C3	C4
Q30K40_ZEAMP_41129/1-89	GGIKOVPGENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q41897_MAIZE_41129/1-89	GGIKOVPGENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q30K48_ZEAMP_41129/1-89	GGIQDVPSENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q30K47_ZEAMP_41129/1-89	GGIQDVPSENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q30K46_ZEAMP_41129/1-89	GGIQDVPSENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q41825_MAIZE_41129/1-89	GGIQDVPSENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q41K00_MAIZE_42130/1-89	GGIKOVPGENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
CYT1_MAIZE_42130/1-89	GGIKOVPGENDLHLOELARFV - DEHNKKA	NALL - GFEKLVKAKTO - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q45KJ0_WHEAT_167/1-67	V - SEHNKKT	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q9FS03_WHEAT_172/1-72	LARFV - SEHNKKT	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GEVKKLYEAKVWEKPMENFKELQEFKPV
Q2KNE8_WHEAT_48136/1-89	GGI VDS - LGRENDPYTIVOLARFV - SEHNKKG	NTQL - ELEKVKVKEQ - AVAGR YTTITQVDE	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q9FS05_WHEAT_48136/1-89	GGI VDS - LGRENDPYTIVOLARFV - SEHNKKG	NTQL - ELEKVKVKEQ - AVAGR YTTITQVDE	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q9FS07_WHEAT_391/1-89	GGI VDS - PGRENDPYTIVOLARFV - SEHNKKA	NTQL - ELEKVKVKEQ - AVAGR YTTITQVDE	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q2KMF0_WHEAT_48137/1-90	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q9FS06_WHEAT_48137/1-90	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q9FS04_WHEAT_31120/1-90	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q4ALJ8_MAIZE_1395/1-83	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q4F753_MAIZE_14105/1-92	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q7Y009_SACOF_24104/1-81	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q9LE17_HORVU_18107/1-90	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q4W4C8_HORVU_18107/1-90	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
CYT2_ORYSA_14104/1-91	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q687J0_AMACO_44133/1-90	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q42380_BRACH_189/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q710E1_BRACH_1199/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q48589_ARATH_1199/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q94501_ARATH_1199/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q5ZF46_POPTN_40128/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q5ZF44_POPTN_40128/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q5ZF48_POPTN_40128/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q5ZF46_POPTN_40128/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q5ZF40_POPTN_40128/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV
Q5ZFV2_POPTN_40128/1-89	GGI SDSPMGQENDL DVIATLARFV - SEHNKKA	NALL - EFERVRLKQ - VWAGTHYLTIEVKD	GGAKKLYEAKVWEKPMENFKELQEFKPV

FIG. 2 (Folha 2 de 5)

	C2	C3	C4
052FV8_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FX4_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FX1_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKESHEFKDAG--
052FU7_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FU3_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FV6_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
052FX9_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
052FV7_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FN3_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FT9_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FH8_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FU0_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FX2_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
Q40123_9RUST_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAG--
052FT8_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
052FM0_POPTN_40128/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q43635_RICCO_695/1-90	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q84LB7_MALDO_48136/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q8L5J8_COLES_593/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q8H0X6_ARATH_593/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q8VZA2_ARATH_593/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q9LHF9_ARATH_38126/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q9FU80_ARATH_13101/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q11007_BOEDR_38126/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q39268_BRACH_592/1-88	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q39270_BRACH_593/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q8VYX5_BRAOL_593/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q22202_ARATH_36124/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q41906_ARATH_35123/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q8RXS7_ARATH_27115/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q9FXN6_ARATH_391/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q1SNM2_MEDTR_43131/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--
Q1S5N8_MEDTR_43131/1-89	GGVHDS-QSSONSASAEIDGLARFV--DEHNKKE--	NAIL--EFARVVKAKEQ-VFAGTHHLTIEAIE--	AGKKLYEAKVWVKPMLNFKELHEFKDAS--

FIG. 2 (Folha 3 de 6)

	C2	C3	C4
004720_S0YBN_49137/1-89	GGLRDS-QSSQNSVQTEALARFV--DEHNKQ	NSLL--EFSRVRTLEQ-VVAGT-HHLTLEATE	AGEKKLYEAKVWPKPMLNFKELQEFKPAG
04F752_MAIZE_45133/1-89	GGWKEN-PAANSAOSDGLGRFV--DEHNRE	NALL--EFVRVKEAQ-VVAGT-HHLTLEAVE	AGRKLYEAKVWPKPMLNFKELQEFKSHG
04A1J7_MAIZE_36124/1-89	GGWKEN-PAANSAESDGLGRFV--DEHNRE	NALL--EFVRVKEAQ-VVAGT-HHLTLEAVE	AGRKLYEAKVWPKPMLNFKELQEFKSHG
Q1YHC6_WHEAT_42130/1-89	GGKSEN-PAANSLLETDGLARFV--DEHNRE	NALL--EFVRVKEAQ-VVAGT-HHLTLEAVE	AGRKLYEAKVWPKPMLNFKELQEFKSHG
Q5N811_ORYSA_795/1-89	GGGAHDA-PSAANSVETDALARFV--DEHNRE	NALL--EFVRVKEAQ-VVAGT-HHLTLEAVE	AGRKLYEAKVWPKPMLNFKELQEFKSHG
Q3HVL8_SOLTU_38127/1-90	GGVHDSHGSSONSDETHSLAKFV--DEHNRE	NAMI--ELARVVKAEQ-VVAGT-HHLTLEWD	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9SE09_LYCES_39128/1-90	GGVHDSHGSSONSDETHSLAKFV--DEHNRE	NAMI--ELARVVKAEQ-VVAGT-HHLTLEWD	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q5Y7Z8_PETHY_55144/1-90	GGIRDSPESONSDETHSLAKFV--DEHNRE	NAMF--ELARVVKAEQ-VVAGT-HHLTLEWD	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9H537_IPDBA_58146/1-89	GGI SDS-ASAENSIVETESLARFV--EHNKKE	NAMI--ELVRVKAQEQ-VVAGT-HHLTLEVID	AGRKLYEAKVWPKPMLNFKELQEFKHW
Q9ZSC2_IPDBA_58146/1-89	GGI SDS-ASAENSIVETESLARFV--EHNKKE	NAMI--ELVRVKAQEQ-VVAGT-HHLTLEVID	AGRKLYEAKVWPKPMLNFKELQEFKHW
Q8LC76_ARATH_44133/1-90	GGFSDSKNDWNGKKEIDDI AL FAV--DEHNRE	NAVL--ELARVKAQEQ-VVAGT-HHLTLEVID	AGEKKLYEAKVWPKPMLNFKELQEFKHW
Q9FF63_ARATH_46135/1-90	GGFSDSKNDWNGKKEIDDI AL FAV--DEHNRE	NAVL--ELARVKAQEQ-VVAGT-HHLTLEVID	AGEKKLYEAKVWPKPMLNFKELQEFKHW
Q9AUC9_SESIN_590/1-86	GGVHDS--NSNPOTHSLARFV--DEHNKKE	NGLL--ELVRVKAQEQ-VVAGT-HHLTLEWD	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q4A1J6_MAIZE_166/1-66	GYLAALR-ERAEADAARFV--AHYNKQ	GAAL--EFTRVLSKQEQ-VVAGT-HHLTLEWD	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q4F751_MAIZE_36121/1-86	GYLAALR-ERAEADAARFV--AHYNKQ	GAAL--EFTRVLSKQEQ-VVAGT-HHLTLEWD	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q61570_ORYSA_48134/1-87	GNLAATRREDAEADAARFV--AEYNKQ	GAEL--EFARVKAQEQ-VVAGT-HHLTLEWD	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q8SA65_SANAU_594/1-90	GAPROVPAGGSENSADEELARFV--AEHNKKE	NALL--EFGRVVKAEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q4G2T8_FRAAN_33122/1-90	GGIRDSPAGSENSLETEALGRFV--DEHNKQ	NGML--EFVRVKAQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q24462_PYRCO_593/1-89	GAYRON-OGVANSIVETESLARFV--DEHNKKE	NDLL--EFVRVLDKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9HB08_HELAN_195255/1-61	GGITEWKDF-ANSLVTDLALARFV--DEYKQ	NTLL--EFERVLDAQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9HB08_HELAN_593/1-89	GGITEWKDF-ANSLVTDLALARFV--DEYKQ	NTLL--EFERVLDAQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9HB08_HELAN_100188/1-89	GGITEWKDF-ANSLVTDLALARFV--DEYKQ	NTLL--EFERVLDAQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q38678_AMBAR_592/1-88	GGITEWKDF-ANSLVTDLALARFV--DEYKQ	NTLL--EFERVLDAQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
CYTA_HELAN_178/1-78	GGITEWKDF-ANSLVTDLALARFV--DEYKQ	NTLL--EFERVLDAQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q8YX72_RUMOB_593/1-89	GGIKOVEGS-ANSLVTDLALARFV--DEHNKQ	NAML--EFSKVVMTKEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9SE08_LYCES_189/1-89	GGITREAGGS-ENSLVTDLALARFV--DEHNKQ	NALL--EFGRVVKAEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q2Y67_9ERIC_593/1-89	GGIHEKEGS-ANSVDELARFV--DEHNKQ	NALL--EFGRVVKAEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9ZTB4_DIACA_1194/1-84	GGSSAISLEIDELARFV--DEHNKKE	NALL--EFGRVVKAEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9ARH0_DIACA_594/1-90	GGIKOSGSSANSLEIDELARFV--DEHNKKE	NTLL--QFKKVVMAKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q8S315_9R0ST_272/1-71	GGVSDVKGH-ENSLQIDDLARFV--DEHNKKA	NTLL--QFKKVVMAKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q8S374_9R0ST_272/1-71	GGVSDVKGH-ENSLQIDDLARFV--DEHNKKA	NTLL--QFKKVVMAKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q9ZRP6_CASSA_694/1-89	GGVSDVKGH-ENSLQIDDLARFV--DEHNKKA	NTLL--QFKKVVMAKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q39561_CARPA_896/1-89	GGLODVEGD-ANSLVTDLALARFV--DEHNKKT	NAML--QFKKVVMAKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
Q8A184_9FABA_684/1-79	GGVSDVKGH-ENSLQIDDLARFV--DEHNKKT	NALL--QFKKVVMAKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW
CYTI_VIGUN_593/1-89	GGVSDVKGH-ENSLQIDDLARFV--DEHNKKT	NALL--QFKKVVMAKQEQ-VVAGT-HHLTLEVID	AGKKLYEAKVWPKPMLNFKELQEFKHW

FIG. 2 (Folha 4 de 6)

	C2	C3	C4
Q39840_SOYBN_1088/1-79	NSVEIDALARFV--EEHNKKQ--	NALL--EFFECTVTKQ--VWSGL YTTTLEAKD--	GGQKKVYEAKWESLNFKEVQEFKLVG--
Q9H404_MANES_593/1-89	GGTKEVES--ANSVEIDNARFV--DDYNKKQ--	NALL--EFRRVSTKQ--VWAGIYYITLVAID--	GGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q35368_MEDSA_593/1-89	GGVWDVPGH--QNSLAIDGLARFV--EEHNKKQ--	NALL--EFSRVTSKQ--VWAGIHHITLWKD--	GNKKVYEAKWESLNFKEVQEFKLVG--
Q64766_ARAHY_593/1-89	GAPREVAGN--ENSLEIDSLARFV--DEHNKKQ--	NGLL--EFRRVTSKQ--VWAGIHHITLWAS--	GDSKNVYEAKWEPKLVNFKEVQEFKLAG--
Q71S08_VTGUN_997/1-89	GGITDVPGA--ANSVEIANLARFV--DDHNKKQ--	NGVL--EFRRVTSKQ--VWSGL YTTLEAKD--	GETKKVYTKWPRELNPKEVQEFNLVT--
Q6V4X1_VIGUN_141229/1-89	GGITDVPGA--ANSVEIANLARFV--DDHNKKQ--	NGVL--EFRRVTSKQ--VWSGL YTTLEAKD--	GETKKVYTKWPRELNPKEVQEFNLVT--
Q6V4X1_VIGUN_12100/1-89	GGITDVSQS--ANCLFALLARFV--DDHNKKQ--	NGVL--EFRRVTSKQ--VWSGL YTTLEAKD--	GETKKVYTKWPRELNPKEVQEFNLVT--
Q39841_SOYBN_1098/1-89	GGITDVGGA--ANSVEINNLARFV--EEDNKKQ--	NSVL--EFRRVTSKQ--VWAGIYYITLVAID--	GLIKNEYEAKWPRELNPKEVQEFNLVT--
Q71S08_VIGUN_106192/1-87	GGVGDVP--SDTLHIENLARFV--DDYNKKQ--	NALL--EFRRVTSKQ--VWAGIYYITLVAID--	GESKNVYEAKWESLNFKEVQEFKPIG--
Q6V4X1_VIGUN_238324/1-87	GGVGDVP--SDTLHIENLARFV--DDYNKKQ--	NALL--EFRRVTSKQ--VWAGIYYITLVAID--	GESKNVYEAKWESLNFKEVQEFKPIG--
Q39842_SOYBN_290/1-89	GGFTDITGA--QNSDIDENLARFV--DEHNKKQ--	NAVJ--EFRRVTSKQ--VWSGL YTTLEAKD--	GVTKKVYTKWPRELNPKEVQEFKPIG--
CYT_SOYBN_148/1-48	GGFTDITGA--QNSDIDENLARFV--DEHNKKQ--	NAVJ--EFRRVTSKQ--VWSGL YTTLEAKD--	GVTKKVYTKWPRELNPKEVQEFKPIG--
Q9H52_ARTUJ_590/1-86	GGVTECKNF--ENNVEIETAKFV--EENKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	G--KTYEAKLWPKWEPKLVNFKEVQEFKPIG--
Q7H224_PEARL_491/1-88	GGVWDV--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q9SE07_LYCES_258/1-57	GGVWDV--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q81693_LYCES_69156/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_571658/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_194281/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_99187/1-89	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q9S809_SOLTU_259/1-58	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_288375/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_477564/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q81693_LYCES_163250/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q9SE07_LYCES_65152/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_665752/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q9S808_SOLTU_259/1-58	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_383470/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q81693_LYCES_662/1-57	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYT_SOLTU_462/1-59	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
CYTH_SOLTU_592/1-88	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q84RKL_ABAH_101135/1-35	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q441J4_MAZE_464/1-61	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q4FZ49_MAZE_3696/1-61	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q4FZ50_MAZE_2891/1-64	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--
Q4A1J3_MAZE_4098/1-59	GGITDVP--PDHNSAETELARFV--DEHNKKQ--	NATL--EFSKVVSKEQ--VWSGL YTTTLEAKD--	AGQTKVYEAKWEPKLVNFKEVQEFKPIG--

FIG. 2 (Folha 5 de 6)

	C2	C3	C4
Q4FZ46_MATZE_4098/1-59	GGKKPI--KNVNDPHWQEIQRMAV--SEHITKA--	NDGL--GFGVWSGEQ--TVAGTYRRLRTOATK--	
Q84252_WHEAT_4198/1-58	GGKPI--GNINDQHIQELGRMAV--LEFGKHV--	NCVL--KFNKVVSGRQD--LVSGMYELITEASD--	
Q2A9P9_BRAOL_2790/1-64	GGNSPI--SDAKDPHWEIGFVAV--SEYDKOS--	KSLG--KFVTVWSESQ--VAGTYRRLIVTV--DG--SIGVA--	
Q2A9T7_BRAOL_24115/1-92	GGNSPI--SNWKDPHWEVGKFAV--SEYDMES--	KSEL--KFVTVWSESK--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPW--	
Q2A9T6_BRAOL_27120/1-94	GGRRPL--SNVNDPHWQEIQRMAV--SEYDKOS--	KAGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q8LB6_ARATH_29117/1-89	GGNSPI--SNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q9FGK2_ARATH_29117/1-89	GGNSPI--SNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q41916_ARATH_29117/1-89	GGNSPI--SNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q9FQ13_CITPA_30117/1-88	GGKPI--EDPKKHMWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q84AT8_ARATH_30115/1-86	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	KSEL--KFVTVWSESK--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPW--	
Q23494_ARATH_30115/1-86	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q4FZ48_MATZE_88143/1-56	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q6TPK4_ACTDE_3089/1-60	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q6TPK3_ACTDE_2988/1-60	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q15WH2_MEDTR_152194/1-43	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q15WH8_MEDTR_152194/1-43	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q84LB7_MALDO_157199/1-43	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q4GZT8_FRAAN_147192/1-46	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q1XHC6_WHEAT_149195/1-47	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q5NB11_ORYSA_114160/1-47	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q4ALJ7_MATZE_143189/1-47	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q4FZ52_MATZE_152198/1-47	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q9FUB0_ARATH_120208/1-89	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q9VZA2_ARATH_112182/1-71	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q9LHF9_ARATH_145215/1-71	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q84X66_ARATH_112182/1-71	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q39270_BRACH_109197/1-89	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q39268_BRACH_115203/1-89	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q8VYX5_BRAOL_118206/1-89	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q1L007_BOEDR_142230/1-89	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q84UC9_SESIN_109155/1-47	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q43635_RICCO_121167/1-47	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q8LC76_ARATH_150212/1-63	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	
Q9F63_ARATH_152214/1-63	GGNSPI--KNVNDPHWQEIQRMAV--SEYDKOS--	ESGL--KFVTVWSESQ--VAGTYRRLIVAVNDG--VAGPGASKNYEATWCKQMLKSHMLTSFRPAI--	

FIG. 2 (Folha 6 de 6)

	C1										
	10	20	30	40	50	60	70				
Q9H366_ARATH_43120/1-78	-----LYVQVTDAINSVVNPSAR-----										TCC
Q9LZE5_ARATH_42107/1-66	-----LYVDVIRAIKNSDVPGP-----										CD
Q6K6B6_ORYSA_263333/1-71	-----LYVRVVRARGVAAVG-----										E
Q64492_ARATH_281355/1-75	-----LYVNIVKAKDLVLG-----										E
Q6EUH5_ORYSA_49126/1-78	-----LYVRVVKARGLKWS-----										GEFD
Q7XR21_ORYSA_51139/1-89	-----LYVRVVRARGTAAASTVAGG-----										GGCN
Q84TJ7_ARATH_280360/1-81	-----LYVRVVKARELPIMDIT-----										GSVD
Q9T0C8_ARATH_126206/1-81	-----LYVRVVKARELPIMDIT-----										GSVD
Q259Q9_ORYSA_276356/1-81	-----LFVRVVKARCLPMDVT-----										GSLD
Q7XPV3_ORYSA_276356/1-81	-----LFVRVVKARCLPMDVT-----										GSLD
Q8H2Q5_ORYSA_277357/1-81	-----LFVRVVKARCLPHMDIT-----										GSLD
Q65279_ARATH_84164/1-81	-----LYVRVVKARCLPNKDLT-----										GSLD
Q8RXU9_ARATH_270350/1-81	-----LYVRVVKARCLPNKDLT-----										GSLD
Q2HRE0_MEDTR_283363/1-81	-----LFIRVVKARCLPRMDLT-----										GSLD
Q9SKA3_ARATH_296377/1-82	-----LYVSVVKARCLPVMDVS-----										GSLD
Q80558_ARATH_50131/1-82	-----LYVSVVKARCLPVMDVS-----										GSLD
Q7XKA3_ORYSA_291371/1-81	-----LYVSVVKARCLPNMDIT-----										GALC
Q9M2D4_ARATH_251331/1-81	-----LFIKIVKARCLPSMDLT-----										GSLD
Q1RSQ4_MEDTR_457538/1-82	-----LYVRVVKAKCLPPGTIT-----										SSCD
Q1S9Y9_MEDTR_80161/1-82	-----LYVRVVKAKCLTNSLT-----										STCD
Q9FL59_ARATH_56137/1-82	-----LYVRVVKAKCLPPNPVT-----										SNCD
Q9F132_ARATH_296378/1-83	-----LYVRVVKAKCLPPGSIT-----										GGCD
Q94JQ8_ARATH_41121/1-81	-----LYVRVVKAKCLPGKDMT-----										GSCD
Q9M2R0_ARATH_41121/1-81	-----LYVRVVKAKCLPGKDMT-----										GSCD
Q9C8H3_ARATH_41121/1-81	-----LYVRVVKAKCLPGKDLT-----										GSCD
Q9LXU2_ARATH_42122/1-81	-----LYVRVVKAKCLPGKDLT-----										GSCD
Q60EW9_ORYSA_42122/1-81	-----LYVRVVKAKCLPSKDLT-----										GSCD
Q5TKJ0_ORYSA_73154/1-82	-----LYVHVVKAKCLPAVSAA-----										GTID
Q9F1Z1_ARATH_40112/1-73	-----LYIRIVKARCLPSN-----										D
Q48584_ARATH_40112/1-73	-----LYIRIVKARCLPSN-----										D
Q49435_ARATH_48124/1-77	-----LYARIVKARCLPVN-----										D
Q2QWP5_ORYSA_200283/1-84	-----LFVRVVKARCLPMDAN-----										GSLD
Q9CA47_ARATH_337419/1-83	-----LFVRVVKARCLPPNE-----										S
Q9SS7_ARATH_337419/1-83	-----LFVRVVKARCLPPNE-----										S
Q8S1F8_ORYSA_344423/1-80	-----LFVRVVKARCLTRACE-----										G
Q7XID7_ORYSA_66152/1-87	-----LFVRVVRARGPAGA-----										H
Q9FJG3_ARATH_325405/1-81	-----VFIRVVKARCLPTSG-----										S
Reference/1-156	MAGSGVLEVHLVDAKGLT-----										GKID
ERG1_ORYSA_795/1-89	-----LEVHLVDAKGLT-----										GKID
Q259F2_ORYSA_163/1-63	-----LEVLLVCAKGLE-----										MD
Q7XPW6_ORYSA_687/1-82	-----LEVLLVCAKGLE-----										DMD
Q24582_MAIZE_686/1-81	-----LEVLLVSAKGLE-----										NMD
Q9SWH6_MAIZE_686/1-81	-----LEVLLVSAKGLE-----										NMD
Y1322_ARATH_687/1-82	-----LEVLLVSAKGLE-----										NMD
ERG3_ORYSA_687/1-82	-----LEVLLVSAKGLE-----										NMD
Q25AG5_ORYSA_687/1-82	-----LEVLLVSAKGLE-----										NMD
Q6H7E3_ORYSA_687/1-82	-----LEVLLVSAKGLE-----										NMD
Q9ZR6_CICAR_687/1-82	-----LEVLLVSAKGLE-----										NMD
PP16A_CUCMA_591/1-87	-----MEVHLISGKGLQ-----										KPID
Q4JH17_PROSI_692/1-87	-----MEVHLISGKGLQ-----										KPID
Q4JHJ0_CUCMA_692/1-87	-----MEVHLISGKGLQ-----										KPID
Q4JH15_PROSI_692/1-87	-----MEVHLISGKGLQ-----										KPID
Q4JH19_CUCMA_692/1-87	-----MEVHLISGKGLQ-----										KPID
PP16B_CUCMA_591/1-87	-----MEVHLISGKGLQ-----										KPID
Q4JH18_CUCMA_692/1-87	-----MEVHLISGKGLQ-----										KPID

FIG. 3 (Folha 1 de 27)

C2				C3			
80	90	100	110	120	130	140	150
PVVEITL	---GNY-KSSTKNL---	PMGPN	---MDWNO---	VFAFD	---KSKGD---	VLSVTL	
PVVEITL	---GNY-KSSTKDL---	PVGPN	---MDWNO---	VFAFD	---KTKGD---	VLSVTL	
TVAEVKL	---GNY-RGVTPAT---	AAH	---HWDQ---	VFAFS	---KETIQSS---	FVEVTV	
VVSEVKL	---GNY-RGVTKKV---	SSNSSN	---PEWNO---	VVFVS	---KERIQSS---	VVELFV	
PFaelRL	---GGY-SQITRHH---	EKTAS	---PEWDD---	VFAFS	---RERIHAP---	FLDLV	
PVVEVRL	---GNY-RGTTRHH---	ERKAA	---PEWNO---	VFAFS	---RERVOAS---	VLEVTV	
PFVEVRV	---GNY-KGTTRHF---	EKRQH	---PEWNO---	VFAFA	---KERMOAS---	VLEVTV	
PFVEVRV	---GNY-KGTTRHF---	EKRQH	---PEWNO---	VFAFA	---KERMOAS---	VLEVTV	
PFVEVRV	---GNY-RGTTRHF---	EKQKN	---PEWNA---	VFAFS	---RDRMOAT---	ILEVTV	
PVVEVRV	---GNY-RGTTRHF---	EKQKN	---PEWNA---	VFAFS	---RDRMOAT---	ILEVTV	
PVVEVHL	---GNY-KMTRHF---	EKNQR	---PEWDE---	VFAFP	---REVMOST---	SLEVIV	
PVVVKI	---GNF-KGVTHF---	NKNTD	---PEWNO---	VFAFA	---KDNLQSN---	FLEVTV	
PVVVKI	---GNF-KGVTHF---	NKNTD	---PEWNO---	VFAFA	---KDNLQSN---	FLEVTV	
PVIVVKI	---GNF-KGTTRHF---	EKNNS	---PEWNL---	VFAFA	---KENQOAT---	TLEVVI	
PVVEVKL	---GNY-KGLTKHL---	EKNSN	---PIWKO---	IFAFS	---KERLOSN---	LLEVTV	
PVVEVKL	---GNY-KGLTKHL---	EKNSN	---PIWKO---	IFAFS	---KERLOSN---	LLEVTV	
PVVEVRL	---GNF-KGVTRHL---	EKNPN	---PVWRO---	VFAFS	---RDHLQSS---	QLEVTV	
PVIEVKL	---GNY-TGKTKHF---	EKNQN	---PVWNE---	VFAFS	---KSNQOQS---	VLEVTV	
PVVEVKL	---GNY-RGTTRHL---	EKKLN	---PEWNO---	VFAFS	---KDRIOQS---	VLEVTV	
PVVEVRL	---GNY-KGTTRHL---	DKRSN	---PEWNO---	VYAFS	---KDOIQSS---	ILEVTV	
PVVEVKI	---GNY-KGTTRHF---	EKRTH	---PEWNO---	VFAFS	---KDKVOSS---	TVEVTV	
PVVEVKL	---GNY-KGTTRHF---	EKKSN	---PEWNO---	VFAFT	---KERIQSS---	VLEVTV	
PVVEVKL	---GNY-KGTTRHF---	EKKSN	---PEWNO---	VFAFS	---KDRIOAS---	FLEATV	
PVVEVKL	---GNY-KGTTRHF---	EKKSN	---PEWNO---	VFAFS	---KDRIOAS---	FLEATV	
PVVEVKL	---GNY-KGTTRHF---	EKKSN	---PEWNO---	VFAFS	---KDRVOAS---	YLEATV	
PVVEVKL	---GNY-KGTTRHF---	EKKSN	---PEWNO---	VFAFS	---KDRVOAS---	YLEATV	
PVVEVKL	---GNY-KGTTRHF---	EKKTN	---PEWNO---	VFAFS	---KERIQSS---	VVEIIV	
PFVEVKL	---GNF-KGTTPVL---	GGNHN	---PSWKQ---	VFAFS	---ATHLOAH---	VLEVTV	
LFVEVTI	---GRY-KGRTKRS---	TNPYPN	---LEFDE---	VFAFN	---SDRLQGN---	MLEVMK	
LFVEVTI	---GRY-KGRTKRS---	TNPYPN	---LEFDE---	VFAFN	---SDRLQGN---	MLEVMK	
SFVAVKI	---GSY-KGRTKQI---	LNSNPN	---PEFHE---	TFAFT	---KTRLQGD---	ILEVTV	
PVVEVKF	---GAYNRQVTRCF---	KRNKN	---PEWNE---	TFAFS	---FQHDKIPSP---	TVDIIV	
AYVAVRT	---SNHFVRSPAVNRPGESVDS---	PEWNO	---VFAFGHNRSDSAVTG---	ATLEISA			
AYVAVRT	---SNHFVRSPAVNRPGESVDS---	PEWNO	---VFAFGHNRSDSAVTG---	ATLEISA			
PVVKIQA	---GPHTLRSPGRDVS---	TGN	---PEWNO---	VFAINHAKE	---PTLEISV		
PHVRVAA	---GGRHASTREAR---	RGAF	---FEWDO---	TFAFVRDPGATDSPG---	PTLEISV		
PVTKISL	---SGTMIQSKPAR---	KTSC	---FEWDO---	TFAFLRDPDLSSS---	PILEISV		
PVVVVQY	---RSQERKSSVARD---	QGKN	---PSWNE---	VFKFOINSTAATGQ---	HKLFLRL		
PVVVVQY	---RSQERKSSVARD---	QGKN	---PSWNE---	VFKFOINSTAATGQ---	HKLFLRL		
PVILTC	---RTQEQSSVAKG---	AGSE	---PEWNE---	TFVFTVSDDVP---	QLNVKI		
PVILTC	---RTQEQSSVAKG---	AGSE	---PEWNE---	TFVFTVSDDVP---	QLNVKI		
PFVILTC	---RTQEQSSVANG---	AGSE	---PEWNE---	TFVFTVSDDTP---	QLHLKI		
PFVILTC	---RTQEQSSVANG---	AGSE	---PEWNE---	TFVFTVSDDTP---	QLHLKI		
PVQLTC	---RTQEQSSVANG---	AGSE	---PEWNE---	TFVFTVSDDTP---	QLHLKI		
PVQLTC	---RTQEQSSVANG---	AGSE	---PEWNE---	TFVFTVSDDTP---	QLHLKI		
PVAVLKC	---RSQEQSSVASG---	KGSD	---PEWNE---	TFMFVTHNAT---	ELIKL		
PVAVLKC	---RSQEQSSVASG---	KGSD	---PEWNE---	TFMFVTHNAT---	ELIKL		
PVAILKC	---RSQEQSSVASG---	KGSD	---PEWNE---	TFMFVTHNAT---	ELIKL		
PVILSY	---RAQEHKSTVQEG---	AGSN	---PEWNE---	TFLFTVSDSAS---	ELNLRI		
PVAILNF	---KGOERMSKVAKN---	AGPN	---PLWDE---	KFKFLAEYPGSGGD---	FHILFKV		
PVAILNF	---KGOERMSKVAKN---	AGPN	---PLWDE---	KFKFLAEYPGSGGD---	FHILFKV		
PVAILNF	---KGOERMSKVAKN---	AGPN	---PLWDE---	KFKFLAEYPGSGGD---	FHILFKV		
PVAILNF	---KGOERMSKVAKN---	AGPD	---PIWNE---	KFKFLVEYPGSGGD---	FHILFKV		
PVAILNF	---KGOERMSKVAKN---	AGPD	---PIWNE---	KFKFLVEYPGSGGD---	FHILFKV		
PVAILNF	---KGOERMSKVAKN---	AGPD	---PIWNE---	KFKFLVEYPGSGGD---	FHILFKV		
PVAILNF	---KGOERMSKVAKN---	AGPD	---PIWNE---	KFKFLVEYPGSGGD---	FHILFKV		
PVAILNF	---KGOERMSKVAKN---	AGPD	---PIWNE---	KFKFLVEYPGSGGD---	FHILFKV		

FIG. 3 (Folha 2 de 27)

10/36

C4

	160	170	180	190	200	210	220	230	240
KDG	P		TNTV	IKRN					
KDR									
RAR	G		SDDHY	GRVN					
KEG	N		KDEY	GRVL					
RGR	GFA		KDDY	GSTR					
RDKDA	VAAVA		RDGY	GRVA					
KDK	DLL		KDDY	GFVR					
KDK	DLL		KDDY	GFVR					
KDK	DLL		KDDF	GLVR					
KDK	DLL		KDDF	GLVR					
KDK	DFI		RDDY	GRVS					
KDK	DIL		LDDF	GIVK					
KDK	DIL		LDDF	GIVK					
KDK	DTI		HDDF	GTVR					
KDK	DLLT		KDDF	GRVH					
KDK	DLLT		KDDF	GRVH					
KDK	DVL		KDDF	GRVV					
KDK	DMV		KDDF	GLIR					
KDK	EMVG		RDDY	GRVI					
KDK	ETVG		RDDY	GRVA					
RDK	EMVT		RDEY	GKVV					
KDK	ETLG		RDDI	GKVV					
KDK	DFV		KDOL	GRVV					
KDK	DFV		KDOL	GRVV					
KDK	DLV		KDOL	GRVV					
KDK	DVV		LDDL	GRIM					
KDK	DFV		KDDF	GRVL					
KAK	DLAG		GDOL	GRVG					
K	MN		EEEI	GQCR					
K	MN		EEEI	GQCR					
RNR	DNPN		EODI	GKCK					
NDK	DLV		RDDF	GKLI					
WD	ASS		ESFL	GGVC					
WD	ASS		ESFL	GGVC					
WDG	GAPSI		EAFI	GGVC					
MDLPP	DADVSDAD		DRHFL	GGLC					
WDSST	GJETS		OFLGGI	C					
MDHDT			FSRDDFL	GEAT	INVTDLISLGM	HGTWEMSESKHRVVLADKTYHGEIRVSLTFTASAKAQDHAEQVGGWAHSFRQ			
MDHDT			FSRDDFL	GEAT					
MDSDA			FSADDF	GEAN					
MDSDA			FSADDF	GEAN					
MDSD			LTNDDF	VGERT					
MDSD			LTNDDF	VGERT					
FDKDV			GTEDDA	GEAT					
MDSDS			GTDDDF	GEAT					
MDSDS			GTDDDF	GEAT					
LDSDT			GSADDF	GEAT					
MEKDN			FNNDNL	GEAT					
MDHDA			IDGDDY	GDVK					
MDHDA			IDGDDY	GDVK					
MDHDA			IDGDDY	GDVK					
MDHDA			IDGDDY	GDVK					
MDHDA			IDGDDY	GDVK					
MDHDA			IDGDDY	GDVK					
MDHDA			IDGDDY	GDVI					

FIG. 3 (Folha 3 de 27)

	10	20	30	40	50	60	70
Q4JHJ3_CUCMO_692/1-87	-----LEVHLISGKGLQ-----	-----AHDPLN-----	-----KPID-----				
Q4JH16_9ROS1_692/1-87	-----LEVHLISGKGLQ-----	-----AHDPLN-----	-----KPID-----				
Q4JHJ2_CUCMO_692/1-87	-----LEVHLISGKGLQ-----	-----AHDPLN-----	-----KPID-----				
Q4JH14_9ROS1_692/1-87	-----LEVHLISGKGLQ-----	-----AHDPLN-----	-----KPID-----				
Q4JHJ1_CUCMO_692/1-87	-----LEVHLISGKGLR-----	-----AHDPLN-----	-----KPID-----				
Q9M2T2_ARATH_692/1-87	-----LEVSLISGKGLK-----	-----RSDFLG-----	-----K-ID-----				
Q49490_ARATH_39115/1-77	-----VGCKLK-----	-----DTEWFS-----	-----ROD-----				
Q945K9_ARATH_1293/1-82	-----LEVTVVGCKLK-----	-----DTEWFS-----	-----ROD-----				
Q9SDM4_DUNTE_485/1-82	-----VDCTLVARGIK-----	-----DVEIVG-----	-----KOS-----				
Q67UD3_ORYSA_502605/1-104	-----LRASVIEAHLRVP-----	-----APSPGLPF-----	-----D-----				
Q8S1F8_ORYSA_499586/1-88	-----LRASVIEAHLRVP-----	-----APSPGLPF-----	-----D-----				
Q9CA47_ARATH_499585/1-87	-----LRVTVLEAQLHIA-----	-----PNLPPLTAP-----	-----E-----				
Q9SSF7_ARATH_499585/1-87	-----LRVTVLEAQLHIA-----	-----PNLPPLTAP-----	-----E-----				
Q49435_ARATH_205287/1-83	-----LRNVNIEAQLVLL-----	-----H-PNRINP-----	-----E-----				
Q7XZ74_ORYSA_473566/1-94	-----LRLSVIAQQLRLP-----	-----APPDAAKAP-----	-----M-GPAFPE-----				
Q9SS68_ARATH_441533/1-93	-----LRLTVIQTQQLLG-----	-----LGSEAKSK-----	-----IPTTE-----				
Q9FJG3_ARATH_479562/1-84	-----LRATVIEAQLLPP-----	-----QLTAFKEA-----	-----S-----				
Q7XR21_ORYSA_228303/1-7	-----LRISVLEAQQVPG-----	-----AVAGAGGDK-----	-----GRHGEAF-----				
Q64492_ARATH_436531/1-96	-----LRISVIEAQQV-----	-----AIMDKGSSL-----	-----MRFPE-----				
Q8H205_ORYSA_438519/1-82	-----VRNVNIEAQLI-----	-----FPMEN-----	-----HIPD-----				
Q9M2D4_ARATH_406486/1-81	-----LRNVNIEAQLVI-----	-----VPDRT-----	-----RLPN-----				
Q2HRE0_MEDTR_448519/1-72	-----LRVKVIEAHLVS-----	-----HDNKS-----	-----RAPD-----				
Q1S9Y9_MEDTR_243325/1-83	-----LRNVNIEAQLVI-----	-----SSDRN-----	-----RVPE-----				
Q1RS04_MEDTR_620702/1-83	-----LRNVNIEAQLVI-----	-----PSDRN-----	-----RLPE-----				
Q9F132_ARATH_461543/1-83	-----LRNVNIEAQLVI-----	-----PSDRN-----	-----RLPD-----				
Q93ZA2_ARATH_94176/1-83	-----LRNVNIEAQLVI-----	-----PSDRS-----	-----QPPQ-----				
Q9FL59_ARATH_219301/1-83	-----LRNVNIEAQLVI-----	-----PSDRS-----	-----QPPQ-----				
Q69T22_ORYSA_46128/1-83	-----LRNVNIEAQLVI-----	-----POARG-----	-----RAPE-----				
Q7XPV3_ORYSA_438520/1-83	-----LRVNIEAQLIA-----	-----ITDKT-----	-----RYPD-----				
Q25909_ORYSA_438520/1-83	-----LRVNIEAQLIA-----	-----ITDKT-----	-----RYPD-----				
Q94JQ8_ARATH_203285/1-83	-----LRNVNIEAQLI-----	-----PTDKQ-----	-----RYPE-----				
Q9M2R0_ARATH_203285/1-83	-----LRNVNIEAQLI-----	-----PTDKQ-----	-----RYPE-----				
Q9C8H3_ARATH_203285/1-83	-----LRNVNIEAQLI-----	-----PSDKG-----	-----RYPE-----				
Q9LXU2_ARATH_203285/1-83	-----VRNVNIEAQLI-----	-----PHDKT-----	-----KFPE-----				
Q60EW9_ORYSA_203285/1-83	-----LRNVNIEAQLI-----	-----PNDRT-----	-----RFPD-----				
Q9T0C8_ARATH_290372/1-83	-----VRNVNIEAQLI-----	-----PTDKT-----	-----RFPD-----				
Q84TJ7_ARATH_444526/1-83	-----VRNVNIEAQLI-----	-----PTDKT-----	-----RFPD-----				
Q6K6B6_ORYSA_415506/1-92	-----LRVSVIEAQLI-----	-----PHDKGPMA-----	-----IGRYPE-----				
Q65279_ARATH_209276/1-68	-----D-----	-----DKS-----	-----RVPE-----				
Q8RXU9_ARATH_434516/1-83	-----LRVQILEAQLVII-----	-----VSDKS-----	-----RVPE-----				
Q9SKA3_ARATH_459541/1-83	-----LRIVHVEAQLV-----	-----PSDKG-----	-----RVPD-----				
Q80558_ARATH_213295/1-83	-----LRIVHVEAQLV-----	-----PSDKG-----	-----RVPD-----				
Q7XKA3_ORYSA_453534/1-82	-----LKVVATIAAQLI-----	-----PAEKG-----	-----RPLAP-----				
Q5TKJ0_ORYSA_236319/1-84	-----LRVAATGAQLV-----	-----PLDAS-----	-----RP-AN-----				
Q9LZE5_ARATH_195276/1-82	-----VRVTIVSGHOLI-----	-----STDNR-----	-----RTPS-----				
Q9M366_ARATH_201281/1-81	-----VRVTIVSGHOLI-----	-----SKDKN-----	-----KTPS-----				
Q93ZM0_ARATH_483561/1-79	-----IIVTVLAGKNL-----	-----VSKDKSGKC-----	-----DAS-----				
Q9L5S3_ARATH_453531/1-79	-----IIVTVLAGKNL-----	-----VSKDKSGKC-----	-----DAS-----				
Q93ZM0_ARATH_293392/1-100	-----IYVTVVGNNLNRRLRGSPSKSSEIGEGSSGNS-----	-----SSKPVO-----	-----SSKPVO-----				
Q9L5S3_ARATH_263362/1-100	-----IYVTVVGNNLNRRLRGSPSKSSEIGEGSSGNS-----	-----SSKPVO-----	-----SSKPVO-----				
Q5NA77_ORYSA_287368/1-82	-----VKLEILEGSDMKP-----	-----SDMNG-----	-----LSD-----				
Q6Z6R6_ORYSA_287368/1-82	-----AKVEILEGSDMKP-----	-----SDPNG-----	-----LAD-----				
Q9LUD5_ARATH_208289/1-82	-----ALVEVVEACQVKP-----	-----SDLNG-----	-----LAD-----				
Q67XP8_ARATH_240321/1-82	-----ALVEVVEACQVKP-----	-----SDLNG-----	-----LAD-----				
Q93XX4_ARATH_284365/1-82	-----VLVEVFEASDLKP-----	-----SDLNG-----	-----LAD-----				
Q9C8L5_ARATH_239320/1-82	-----VLVEVFEASDLKP-----	-----SDLNG-----	-----LAD-----				

FIG. 3 (Folha 4 de 27)

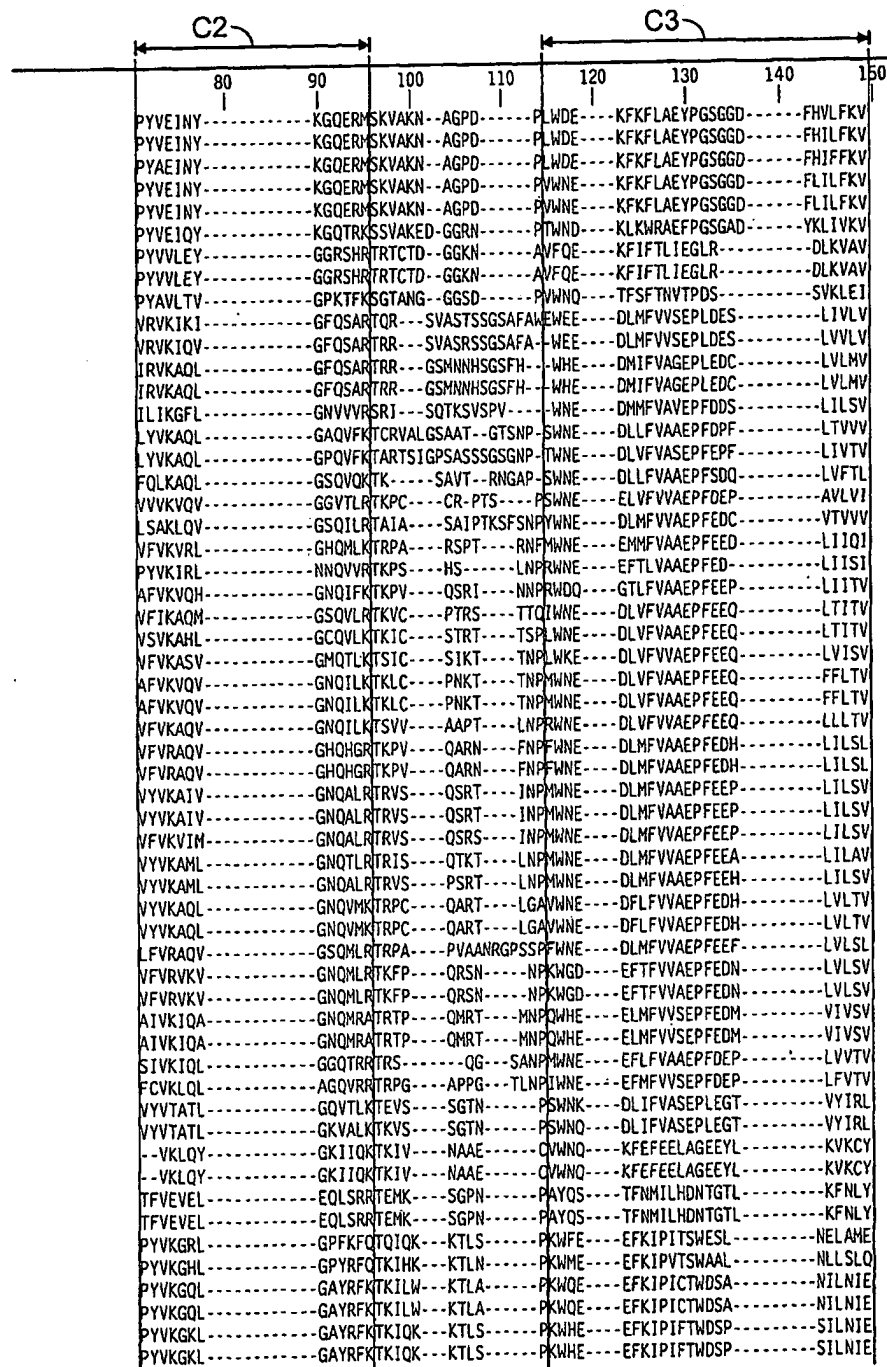
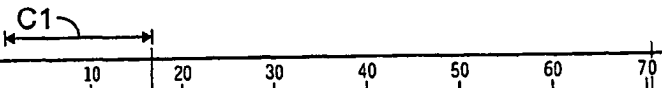


FIG. 3 (Folha 5 de 27)

C4

	160	170	180	190	200	210	220	230	240
MDHDA			IDGDDYIGDVK						
MDHDA			IDGDDYIGDVK						
MDHDA			IDDDDLIGEVK						
MDHDV			IDGDDYIGDVS						
MDHDV			IDGDDYIGDVS						
MDHDT			FSSDDFIGEAT						
WNSNT			LSTDDFIGNAT						
WNSNT			LSTDDFIGNAT						
FNSNV			VLRDVAJGGCK						
KORT			MIKEPARRGARPTSALLPAKEAAHVC						
EDRS			MIKEPALLGHAT						
EDR			TTKEATLLGHAM						
EDR			TTKEATLLGHAM						
EDKV			GPRE-ECLGRCE						
EDIF			S-G-OPVGOAR						
EDIT			N-G-QSIGOTK						
EYRT			SKGP-VTVGMAR						
E									
EDRLNG			GAIGGQNDVAVGRVQ						
EDRV			AQNKDEVIGETH						
EDRV			APNREETLGEVH						
ED			KD						
EDRV			HGSKDEVLGKIM						
EDHV			OPSKDEVLGRIIS						
EDRV			HTSKDEVIGKIT						
ENKV			TPAKDEVNGRLI						
ENKV			TPAKDEVNGRLI						
EDRV			TPRKDDLGRAA						
EDRV			APNKDEVIGRVI						
EDRV			APNKDEVIGRVI						
EDRV			APNKDEVIGRCA						
EDRV			APNKDEVIGRCA						
EDRV			APNKDEVIGRCA						
EDRV			APNKDETLGRCA						
EDRT			APGKDDVLGRTI						
EDRV			APGKDEIVGRTY						
EDRV			APGKDEIVGRTY						
EDHV			SPGRDDVLGRLV						
EDHT			APNRDEPVGKAV						
EDHT			APNRDEPVGKAV						
DDRI			GPGKDEILGRVF						
DDRI			GPGKDEILGRVF						
EERV			AAGROEPVGRVI						
EDRV			GPGRDEPLGRIM						
IDRVD			DQHEERTIGKLE						
IDRED			EQHEG-CIGTLK						
REE			MLGTDNIGTAT						
REE			MLGTDNIGTAT						
ENNP			GSVRYDSLASCE						
ENNP			GSVRYDSLASCE						
VCD	KDHMF		DDSLGTCT						
VRD	KDPIF		DDTLGDCS						
VQD	KDRFS		DDSLGDCS						
VQD	KDRFS		DDSLGDCS						
VGD	KDRFV		DDTLGDCS						
VGD	KDRFV		DDTLGDCS						

FIG. 3 (Folha 6 de 27)



	10	20	30	40	50	60	70
Q9LPH4_ARATH_284365/1-82	VLVEVFEASDLP	SDLNG					LAD
Q6Z7A3_ORYSA_23103/1-81	LCVHVL EARGLOA	AYLTG					HSD
Q66VB0_ORYSA_24104/1-81	LQVRVVEARGLPA	VRVDG					TSD
Q9ZVT9_ARATH_383/1-81	LQVRVVEARMLPA	MDLNG					FSD
Q6Z8U1_ORYSA_383/1-81	LNVVVEIARMLRA	MDSNG					FSD
Q5Z6I4_ORYSA_384/1-82	LLVOVSEARMLPA	IDGGGG					LSD
Q6SR86_CHLRE_569/1-65	LCVSVIEARMLQA	DDFAG					KNF
Q8W4D4_ARATH_83163/1-81	VKVELLAAKNLIG	ANLNG					TSD
Q9M1A2_ARATH_83163/1-81	VKVELLAAKNLIG	ANLNG					TSD
Q69MT7_ORYSA_91171/1-81	VKLELLAAKNLIA	ANLNG					TSD
Q6K8R7_ORYSA_117207/1-91	TKLELLCAKYLIG	ANLNG					SSD
Q6Y3R3_PTEVI_2858/1-31	LEVHLL EAHGLD	TDTFG					KSD
Q9LS53_ARATH_577655/1-79	TELVLVEAROLVA	ADIRG					TSD
Q93ZM0_ARATH_607685/1-79	TELVLVEAROLVA	ADIRG					TSD
Q1T680_MEDTR_420502/1-83	LVVIVHEAQDVEG	KH					HTN
Q6ETC4_ORYSA_421503/1-83	LVVIVHEAQDVEG	KH					HTN
Q9LNT5_ARATH_418499/1-82	LVVIVHEAQDLEG	KY					HTN
Q1PF28_ARATH_1882/1-65	FVVI VHSACDVEG	KH					HTN
Q9SKQ7_ARATH_1882/1-65	FVVI VHSACDVEG	KH					HTN
Q9SKR0_ARATH_137201/1-65	LVVIVHSACDVEG	KH					HTN
Q5MD16_BRANA_158241/1-84	LAVIVHSACDVEG	KH					HTN
Q9SKR2_ARATH_422505/1-84	LVVIVHSACDVEG	KH					HTN
Q69JE2_ORYSA_421504/1-84	LYVVVHEAQDLEG	KH					HTN
Q5QLZ9_ORYSA_398461/1-64	LYVVVHEAKLEG	KC					NTN
Q7XA06_ARATH_422505/1-84	LSVAVQSADVEG	KKK					HSN
Q3E9M4_ARATH_200283/1-84	LSVAVQSADVEG	KKK					HSN
Q9FYD9_ARATH_465548/1-84	LSVAVQSADVEG	KKK					HSN
Q655F0_ORYSA_421505/1-85	LLSVVENAEDVEG	KR					HTN
Q7XPV3_ORYSA_692/1-87	LGVEVASAHLMP	KDGGG					SAS
Q259Q9_ORYSA_692/1-87	LGVEVASAHLMP	KDGGG					SAS
Q9T0C9_ARATH_893/1-86	LGVDVIGAHNLP	KDGGG					TSN
Q84TJ7_ARATH_893/1-86	LGVDVIGAHNLP	KDGGG					TSN
Q8H2Q5_ORYSA_791/1-85	LGVEVTSADLMP	KE-QG					TCN
Q2HRE0_MEDTR_1398/1-86	LCVEVVGANDLVA	KDGGG					SST
Q65279_ARATH_664/1-59	LGVEVISAQGLQ	RDKHN					SCS
Q8RXU9_ARATH_691/1-86	LGVEVISAQGLQ	RDKHN					SCS
Q9M2D4_ARATH_790/1-84	LGVEVISAR-LKP	REDYG					GVN
Q2QWP5_ORYSA_689/1-84	LVVEVISAD-LPS	SSNTS					QTN
Q7XZZ4_ORYSA_10103/1-94	VVVEVCNARNLMP	KDGGG					TAS
Q9SS68_ARATH_994/1-86	LIVE ICSARNLMP	KDGGG					TAS
Q9CA47_ARATH_19108/1-90	LVVEVVEARNLIP	KDGGG					SSS
Q9SSF7_ARATH_19108/1-90	LVVEVVEARNLIP	KDGGG					SSS
Q8S1F8_ORYSA_27121/1-95	LAVEVVDARDLVP	KDGLG					TSS
Q1RSQ4_MEDTR_389/1-87	LIVEVINADLMP	KDGGG					SAS
Q9F132_ARATH_793/1-87	LVVHVVDADLMP	RDGGG					SAS
Q6K6B6_ORYSA_8104/1-97	LVVEVVAARNLMP	KDGGG					SSS
Q64492_ARATH_1398/1-86	LVVEIVGAHNLMP	KDGED					SSS
Q9SKA3_ARATH_489/1-86	LVVEIVDASDLP	KDGGG					SAS
Q7XKA3_ORYSA_393/1-91	VGVEILDASELAP	KDGAG					ACN
Q9FJG3_ARATH_796/1-90	LVVEVVDADLTP	KDGGG					TSS
Q04823_SPOST_1394/1-82	LNVVRVVRGSLA	ICDPL					THTSD
Q69RN2_ORYSA_1596/1-82	LNVVRVVRGSLA	IADPL					THTSD
Q6Z117_ORYSA_889/1-82	LKVRVMRGLMLA	ICDPL					THSSD
Q7XIV3_ORYSA_887/1-80	LKIRRVVRGIMLA	YRD					TRGSD
Q7XIU9_ORYSA_887/1-80	LSVRVLRGVMLV	SRD					AGGSD
Q6YWF1_ORYSA_987/1-79	VKVRVVRGVMLA	VRD					LRSSD

FIG. 3 (Folha 7 de 27)

C2				C3			
80	90	100	110	120	130	140	150
PYVKGKL	GAYRFKTKIQK	KTLS	PKWHE	EFKIPIFTWOSP	SILNIE		
PYVRLQM	GRRRAKTTVVK	RCLS	PLWDE	EFGFVGDAAEE	ELVVS		
PFVKLQL	GKRRAKTAVAR	RTLA	PAWDE	EFSFLVGDIAE	ELVVS		
PYVRLQL	GKORSRTKVVK	KNLN	PKWTE	DFSFGVDDLND	ELVVS		
PYVKLQL	GKORFKTKVVK	KNLN	PANDQ	EFSFSGVDVRD	VLKLYV		
PYAKLQL	GRORGKTRVAK	RTLS	PTWDE	EFAFRVVDLKD	ELVVVV		
FYVKLRV	GSNEVKTDVVK	GSLA	PKFLK	DCRLAVPTPES	DYLRIE		
PYAIVNC	GSEKRFSSMVP	GSRN	PMWGE	EFNFPDDEL P	AKINVTI		
PYAIVNC	GSEKRFSSMVP	GSRN	PMWGE	EFNFPDDEL P	AKINVTI		
PYALITC	GEEKRFSSMVP	GSRN	PMWGE	EFNFFVDSL P	VKINVTI		
PYAVISC	GEORRFSSMVP	SSRN	PLWGE	EFNFLVREL PVECTAPVND	SKVTITM		
PYAIVYC	QKE						
PYVRVQY	GEKKORTKVIY	KTLO	PKWNO	THEFPDDGSS	LELHV		
PYVRVQY	GEKKORTKVIY	KTLO	PKWNO	THEFPDDGSS	LELHV		
PQARLIF	RGEKKTKRIK	KNRD	PRWED	EFQFIAEPPTN	DKLHVEV		
PYVRIVF	RGEERKTKHIK	KNRD	PRWEQ	EFQFVCEPPIN	DKMQIEV		
PSVRLLF	RGEERKTKRVK	KNRE	PRWDE	DFQFPLOEPPIN	DKLHVEV		
PYVHIYF	KGEERKTKHVK	KNKD	PKWNE	EFSFMLEEPPIH	EKMHVKV		
PYVHIYF	KGEERKTKHVK	KNKD	PKWNE	EFSFMLEEPPIH	EKMHVKV		
PYVHIYF	KGEERKTKNVK	KNKD	PKWNE	EFSFMLEEPPVH	EKLHVEV		
PYVRIYF	KGEERKTKHVK	KNRD	PRWEE	EFTFMLEEPPVR	XKLHVEV		
PYVRIYF	KGEERKTKHVK	KNRD	PRWNE	EFTFMLEEPPVR	EKLHVEV		
PYAKIIF	KGEKKTKVIK	KNRD	PRWED	EFEFVCEEPPVN	DKLHIEV		
PYVKLTF	KGVEKTKVVK	ENRN	PRWKE	EFEFECEETPAN	DKLHVEV		
PYAVVLF	RGEKKTKMLK	KTRD	PRWNE	EFQFTLEEPPVK	ESIRVEV		
PYAVVLF	RGEKKTKMLK	KTRD	PRWNE	EFQFTLEEPPVK	ESIRVEV		
PYAVVLF	RGEKKTKMLK	KTRD	PRWNE	EFQFTLEEPPVK	ESIRVEV		
PYAVVHF	RGERKETIK	KTRD	PRWNE	EFQFVDEAPVD	DKIHIEV		
ACVELTF	DGORRRTAID	KDLN	PVWNE	RFYFVNSD-PSNL	P-ELALEAYV		
ACVELTF	DGORRRTAID	KDLN	PVWNE	RFYFVNSD-PSNL	P-ELALEAYV		
AYVELYF	DGOKHRTTIK	RDLN	PVWNE	SFFFNISD-PSRL	H-YLNLEAQA		
AYVELYF	DGOKHRTTIK	RDLN	PVWNE	SFFFNISD-PSRL	H-YLNLEAQA		
PYVEIEF	DDOKFRTAIKE	RDLN	PVWNE	QFYFNISD-PSRL	T-EKDLEAYV		
TFVELEF	DDOKFRTTTKD	KDLS	PVWNE	IFYFNITD-PSKL	S-NLNLEACI		
PFVELKF	DNQIFRATTKH	NDPN	PVWHE	CFYFVNSD-PSV			
PFVELKF	DNQIFRATTKH	NDPN	PVWHE	CFYFVNSD-PSV			
AYVELRF	DDQKVITMTKI	DDSS	PVWNE	KFFFNISD-TEDL	S-NOFLDAYV		
YSVELHF	NSOSKSTTIKE	N-V	AVWNE	RFSFDMRO-REDP	SGNLILEAAV		
AYAVVDF	DGORRRRTATRP	RDLN	PQWGE	RLEFLVHD-PDA	MCAETLELNL		
AYAVVDF	DGORRRRTATRP	RDLN	PQWDE	KLEFFVHD-VAT	MGEETLELNL		
AYVVVDF	DAQKRTSTKF	RDLN	PIWNE	MLDFAVSD-PKN	MDYDELDIEV		
AYVVVDF	DAQKRTSTKF	RDLN	PIWNE	MLDFAVSD-PKN	MDYDELDIEV		
AFVAVDF	DGOKRRTRTVP	RDLN	PQWHE	RLEFAVHD-PAA	MHAALDVSL		
TFVEVDF	ENQLSRTRTVP	KNLN	PTWNO	KLVFNLDT-TKP	YHHKTIEVS		
PFVEVDF	LNQLSRTRTVP	KSLN	PVWNO	KLVFNDQ-SVIN	QHNQHTIEVS		
AYVEVEF	EHORRRRTARP	KELN	PVWNE	RLVFVAD-PDD	LPYRAIDVGV		
PFVEVDF	ENQLSRTTKVP	KDLN	PIWNE	KLVFHVID-VND	LRHKAETINV		
PFVEVEF	DEORRRTQTRF	KDLN	PQWNE	KLVFVND-LKR	LNNKTVDVTV		
AFVEVEF	DGOKRRTPTKP	ADRS	PQWNE	TLVFDVRD-PSR	LPSLPVDVSV		
PYVLDY	YGORRRRTITV	RDLN	PVWNE	TLEFSLAKRPSHO	LFTDVLLEDM		
PYVVLHY	GAOKVKTSVOK	KPNP	PVWNE	VLQSVTNP	TKPVHLEV		
PYVVLHY	GAOKVKTSVOK	KPNP	PVWNE	VLQSVTNP	TKPVHLEV		
PYVVLRY	GSOKVKTSTY	HSIN	PEWNE	ELTSLITNM	MLPVKTEV		
PYVVLRL	GKOKVKTSVKK	KSVN	PIWHE	ELTSLITNM	IAPIKLGV		
PYVVLHL	DNOKLKTGVVK	KTTN	PVWNE	ELTSLITNM	ETPIQLEV		
PYVIVRM	GKOKLKTGVVK	KTTN	PEWNE	ELTSLITNM	AVPVRLLEV		

FIG. 3 (Folha 8 de 27)

	160	170	180	190	200	210	220	230	240
VGD	---KDRFV		---DDTLGFAP						
LNE	---EGYFG		---GGFLGRVK						
LNE	---DKYFS		---NDLLGKVR						
LDE	---DKYFN		---DDFVGQVR						
YDE	---DMIGI		---DDFLGQVK						
VDE	---DRYFS		---DDFLGQVR						
L									
HDW	---DIIWK		---STVLGSSVT						
HDW	---DIIWK		---STVLGSSVT						
YDW	---DIVWK		---STVLGSSVI						
YDW	---DTVCK		---CKVJGSSVT						
KDY	---NTLLP		---TSSJGNCV						
KDY	---NTLLP		---TSSJGNCV						
VSS	---SSRT		---LLHOKESLGYVD						
ISR	---PPSI		---GIHSENLGYVV						
ISS	---SSR		---LIHPKETLGYVV						
F									
F									
F									
LST	---SSRI		---GLLHPKETLGYVD						
LST	---SSRI		---GLLHPKETLGYVD						
LSK	---ASKK		---GLIHKETLGYID						
MSK	---GTG		---FHFRSKEELGHVD						
MSK	---GTG		---FHFRSKEELGHVD						
MSK	---GTG		---FHFRSKEELGHVD						
VSK	---RRGLR		---LPFRNKESLGHVD						
YNI	---NRSID		---GSRSLGKVR						
YNI	---NRSVD		---GSRSLGKVR						
YSH	---NRSTN		---G-RSFLGKVS						
YSH	---NRSTN		---G-RSFLGKVS						
YHA	---NRASN		---S-KTCLGKVR						
NHY	---NK-TN		---GSKIPLGKVK						
YSY	---QNEFD		---A-KPFLGKVR						
Y	---NKTSS		---ITKSCLGKIR						
YCF	---DQMSN		---S-KSLLGKVL						
YND	---KKAJ		---AATGGGRRGGTFLGKVK						
CND	---KK		---TG-KRS-TFLGKVK						
YND	---KRF		---GNGGGRKNHFLGRVK						
YND	---KRF		---GNGGGRKNHFLGRVK						
YHD	---RRFNP		---SGGGGGGGKKNHFLGRVR						
YND	---RRQPN		---PGRNFLGRVR						
YHE	---RR-PI		---PGRSFLGRVK						
YND	---RAASGGV		---AGGGGAAPHGRNFLGKVR						
YNE	---KRSS		---NSRNFLGKVR						
YDD	---RRDN		---QPGKFLGRVK						
HHD	---RSLTD		---HHATRLHTFLGRVR						
YHD	---KNFG		---QTRRNFLGRIR						
FDE	---DKFT		---ADDSMGVAE						
FDE	---DKFT		---ADDSMGVAE						
FDK	---DTFT		---KDDSMGDAE						
FDK	---DTFS		---RDDPMGDAE						
FDK	---DTFS		---KDDMGDAE						
YDK	---DTF		---IDDMGNAE						

FIG. 3 (Folha 9 de 27)

	10	20	30	40	50	60	70
Q9C5M6_ARATH_887/1-80	-----LRIRVKRGINLA-----	-----	-----	-----QRD-----	-----	-----	-----TLGSD
Q9SSL1_ARATH_887/1-80	-----LRIRVKRGINLA-----	-----	-----	-----QRD-----	-----	-----	-----TLSSD
Q9CAC6_ARATH_839/1-32	-----LRIRVKRGINLA-----	-----	-----	-----QRD-----	-----	-----	-----TLSSD
Q9S764_ARATH_1796/1-80	-----VRILVKRGIDLA-----	-----	-----	-----RRD-----	-----	-----	-----ALSSD
Q49303_ARATH_887/1-80	-----LRIRVKRGINLV-----	-----	-----	-----SRD-----	-----	-----	-----SNTSD
Q9S7J9_ARATH_988/1-80	-----LRVHVKRGINLA-----	-----	-----	-----IRD-----	-----	-----	-----ATTSO
Q9ZVF1_ARATH_988/1-80	-----LTIHVKRGINLA-----	-----	-----	-----IRD-----	-----	-----	-----HRSSD
Q1S855_MEDTR_989/1-81	-----LKLRIKRGINLA-----	-----	-----	-----IRD-----	-----	-----	-----SNSSD
Q9LVH4_ARATH_2099/1-80	-----LRIRIKRGVILA-----	-----	-----	-----VRD-----	-----	-----	-----TSSSD
Q9LP65_ARATH_44123/1-80	-----LRIRIKRGVILA-----	-----	-----	-----VRD-----	-----	-----	-----LNSSD
Q1SSA3_MEDTR_24103/1-80	-----LRIRIKRGVILA-----	-----	-----	-----VRD-----	-----	-----	-----VNTSD
Q9FHP6_ARATH_887/1-80	-----LRHVKRGVILA-----	-----	-----	-----IRD-----	-----	-----	-----TSSSD
Q9C8Y2_ARATH_887/1-80	-----LRLHVIRGVILA-----	-----	-----	-----IRD-----	-----	-----	-----SQSSD
Q9C6B7_ARATH_1190/1-80	-----LRVRVQGVILA-----	-----	-----	-----VRD-----	-----	-----	-----VSSSD
Q8LFN9_ARATH_182261/1-80	-----LKVTIKKGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MMSSD
Q9M0W2_ARATH_180259/1-80	-----LKVTIKKGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MMSSD
Q49557_ARATH_215294/1-80	-----LKVTIKKGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MMSSD
Q8L9H2_ARATH_183262/1-80	-----LKVTIKKGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MMSSD
Q9FVJ3_ARATH_183262/1-80	-----LKVTIKKGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MMSSD
Q1RV25_MEDTR_191270/1-80	-----LKVVVKGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MRTSD
Q49U73_ORYSA_166245/1-80	-----LNVKVKGGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MSSSD
Q6H738_ORYSA_166245/1-80	-----LNVKVKGGTILA-----	-----	-----	-----IRD-----	-----	-----	-----MSSSD
Q69V47_ORYSA_167246/1-80	-----IKVKVIRGTILA-----	-----	-----	-----VRD-----	-----	-----	-----TLSSD
Q6Z653_ORYSA_179258/1-80	-----LNTVVVGIDILA-----	-----	-----	-----VRD-----	-----	-----	-----MLTSD
Q6L4C8_ORYSA_240319/1-80	-----IKVDIRGTILA-----	-----	-----	-----VRD-----	-----	-----	-----VMSSD
Q8RZA2_ORYSA_227306/1-80	-----IKVNVIRGTILA-----	-----	-----	-----VRD-----	-----	-----	-----MMSSD
Q8L7A4_ARATH_231310/1-80	-----IKVNVVKGTILA-----	-----	-----	-----VRD-----	-----	-----	-----VMTSD
Q9SFC0_ARATH_219298/1-80	-----IKVNVVKGTILA-----	-----	-----	-----VRD-----	-----	-----	-----VMTSD
Q1RU67_MEDTR_887/1-80	-----IKVNVVKGTILA-----	-----	-----	-----IRD-----	-----	-----	-----VVTSD
Q9F1K8_ARATH_887/1-80	-----LQVTIIGKKLV-----	-----	-----	-----IRD-----	-----	-----	-----FKSSD
Q2A9R2_BRAOL_887/1-80	-----LQVTIIRGKILA-----	-----	-----	-----IRD-----	-----	-----	-----FKSSD
Q2HV28_MEDTR_887/1-80	-----LKVIIVOGKRLV-----	-----	-----	-----IRD-----	-----	-----	-----FKTSD
Q6K295_ORYSA_1594/1-80	-----LKVVVASGTILA-----	-----	-----	-----VRD-----	-----	-----	-----FTSSD
Q8L173_ORYSA_1594/1-80	-----VKVKVVRGTILA-----	-----	-----	-----VRD-----	-----	-----	-----VFSSD
Q7XPV3_ORYSA_600693/1-94	-----LELGILGAQGIVP-----	-----	-----	-----MK-TRDGK-----	-----	-----	-----GSSD
Q25909_ORYSA_600693/1-94	-----LELGILGAQGIVP-----	-----	-----	-----MK-TRDGK-----	-----	-----	-----GSSD
Q84TJ7_ARATH_606693/1-88	-----LELGILNAVGLHP-----	-----	-----	-----MK-TREGR-----	-----	-----	-----GTSD
Q9T0C8_ARATH_452539/1-88	-----LELGILNAVGLHP-----	-----	-----	-----MK-TREGR-----	-----	-----	-----GTSD
Q43085_PEA_150/1-50	-----	-----	-----	-----	-----	-----	-----
Q2HRE0_MEDTR_630702/1-73	-----	-----	-----	-----RDGR-----	-----	-----	-----GAAD
Q64492_ARATH_604694/1-91	-----LEIGILSATGLMP-----	-----	-----	-----MK-VRDGK-----	-----	-----	-----CGGIAD
Q48584_ARATH_329399/1-71	-----LVLGVISASGISP-----	-----	-----	-----MK-SRDG-----	-----	-----	-----RGTTD
Q49435_ARATH_363455/1-93	-----LELGVLNATGLMP-----	-----	-----	-----MK-SRGG-----	-----	-----	-----RGTTD
Q94JQ8_ARATH_364455/1-92	-----LELGVLNATGLMP-----	-----	-----	-----MK-TKDG-----	-----	-----	-----RGTTD
Q9M2R0_ARATH_364455/1-92	-----LELGVLNATGLMP-----	-----	-----	-----MK-TKDG-----	-----	-----	-----RGTTD
Q9C8H3_ARATH_365458/1-94	-----LELGVLNATGLMP-----	-----	-----	-----MK-AKEGG-----	-----	-----	-----RGTTD
Q6DEW9_ORYSA_365456/1-92	-----LELGILTAOGLLP-----	-----	-----	-----MK-TKDG-----	-----	-----	-----RGTTD
Q93ZA2_ARATH_258351/1-94	-----LEVGIISAQGLSP-----	-----	-----	-----MK-TKDG-----	-----	-----	-----KATTD
Q9FL59_ARATH_383476/1-94	-----LEVGIISAQGLSP-----	-----	-----	-----MK-TKDG-----	-----	-----	-----KATTD
Q9LXU2_ARATH_364451/1-88	-----LEVGIISAQGLMP-----	-----	-----	-----MK-SKDG-----	-----	-----	-----KGTTD
Q1RS04_MEDTR_784876/1-93	-----LEMGIIGAKGLLP-----	-----	-----	-----MK-MKDG-----	-----	-----	-----HGSTD
Q1S9Y9_MEDTR_407504/1-98	-----LEVGIIGAKGLLP-----	-----	-----	-----MK-MNNS-----	-----	-----	-----RGSTD
Q9F132_ARATH_626718/1-93	-----LEIGILGANGLVP-----	-----	-----	-----MK-LKDG-----	-----	-----	-----RGSTN
Q69T22_ORYSA_212314/1-103	-----LEVGIIGAAQLP-----	-----	-----	-----MK-NRDG-----	-----	-----	-----RGTTD
Q765H8_FLATR_33118/1-86	-----LEMGIIGAHGLPP-----	-----	-----	-----MK-SKDG-----	-----	-----	-----HTTDD
Q7XR21_ORYSA_405496/1-92	-----LEVGVLAGAQLPP-----	-----	-----	-----MKTAADGG-----	-----	-----	-----RGTTD

FIG. 3 (Folha 10 de 27)

[illegible]

FIG. 3 (Folha 11 de 27)

C4

	160	170	180	190	200	210	220	230	240
YDK---DTFT---			SHDKMGDAK						
YDK---DTFT---			SHDKMGDAK						
YDK---DTFT---									
YDK---DTFT---			ADDDMGDAQ						
YDK---DTFT---			SHDPMGDAE						
FDK---DRFS---			GDDKMGDAE						
FDW---DKFT---			GDDKMGDAN						
FDK---DTFF---			VDDKMGDAE						
YDH---DMFS---			KDDKMGDAE						
YDY---DTFT---			KDDKMGDAE						
YDY---DTFS---			KDDKMGDAE						
YDY---DMLS---			ADDDMGDAE						
YDR---DRFS---			RDDKMGDAI						
YDH---DFFS---			KDDKMGDAE						
YDY---DTFS---			ADDIMGDAE						
YDY---DTFS---			ADDIMGDAE						
FDY---DTFS---			ADDIMGDAE						
FDY---DTFS---			ADDIMGDAE						
FDY---DTFS---			ADDIMGDAE						
FDH---DLFS---			ADDIMGDAE						
FDH---DMLS---			KDDLMEGAE						
FDH---DMLS---			KDDLMEGAE						
YDH---DVLS---			RDDIMGDAE						
YDH---DTFS---			ADDIMGDAE						
FDK---DTFS---			SDDRMGDAE						
YDK---DTFT---			TDDRMGDAE						
YDK---DTFS---			TDDFMGDAE						
YDK---DTFS---			TDDFMGDAE						
YDK---DTFS---			TDDFMGDAE						
FDK---DRFK---			ADDDMGHAS						
FDK---DRFK---			ADDDMGHAT						
FDK---DLLK---			ADDDMGNAF						
FDW---DRFK---			YDDKMGHAF						
YDR---DPF---			VDDPMGAFF						
FDN---GQLG---			EKGGEKTSSSKDAKIGKVR						
FDN---GQLG---			EKGGEKTSSSKDAKIGKVR						
FDN---GQLG---			EKG---NRDVKIGKIR						
FDN---GQLG---			EKG---NRDVKIGKIR						
FDN---CQVN---			GPD---NKDLLIGKVR						
FNN---GQLN---			DSN---DSNDSKIGKVR						
FDN---ARVN---			ENN---NSRDVRIGKVR						
FDN---									
FDN---LKL---			GAGNENRLIN-DSRIGKIR						
FDN---CHLH---			GGEK-IGGAK-DSRIGKVR						
FDN---CHLH---			GGEK-IGGAK-DSRIGKVR						
FDN---CHLH---			GGDKNNGGGK-DSRIGKVR						
FDN---CHLN---			GGEK-ANGAR-DTRIGKVR						
FDN---CHLG---			GSEKNSGAKVDSRIGKVR						
FDN---CHLG---			GSEKNSGAKVDSRIGKVR						
FDN---GHIP---			GGSG---K-DLRIGKVR						
FDN---CHLG---			EKAPSGSS-1KDSRIGKVR						
FDN---CHLGGG---			EKAPSGGSNAARDSRIGKVR						
FDN---SHLG---			SAQSGTAD-SRDARIGKVR						
FDN---NHLGNGNGNGNNAGGGGGSPPARDAVIGKIR									
FDNN---FHLQ---			GGDK---RIGKVR						
FDN---CNLG---			NGGGGG---KDQRTIGKVR						

FIG. 3 (Folha 12 de 27)

C1

	10	20	30	40	50	60	70
Q6K6B6_ORYSA_587682/1-96	-----LELGVLGATQIP	-----MK-ARDGR	-----GATSD				
Q9SKA3_ARATH_624711/1-88	-----LELGILSARNMP	-----MK-GKDG	-----RMTD				
Q80558_ARATH_378465/1-88	-----LELGILSARNMP	-----MK-GKDG	-----RMTD				
Q5TKJ0_ORYSA_399486/1-88	-----LELGILGARNLP	-----MK-GKDG	-----RTTD				
Q7XKA3_ORYSA_621701/1-81	-----LELGILGARNLA	-----G-GK	-----S				
Q65279_ARATH_285357/1-73	-----R	-----R	-----KGTSD				
Q8RXU9_ARATH_598688/1-91	-----LELGILNANVFHS	-----MK-TREG	-----KGTSD				
Q9M2D4_ARATH_565653/1-89	-----LELGILRIEQLN	-----LS-QEGK	-----KETVD				
Q8H2Q5_ORYSA_599693/1-95	-----LEVGLSANDLNP	-----TK-TKHE	-----RGSCD				
Q6EUH5_ORYSA_383461/1-79	-----RDG	-----RDG	-----RGSCD				
Q9FJG3_ARATH_636729/1-94	-----VELGITGCKNLLP	-----MKT-VNG	-----KGSTQ				
Q7XID7_ORYSA_400497/1-98	-----VELGITVGCKLLP	-----MKT-ADG	-----KGCTD				
Q9SSF7_ARATH_671762/1-92	-----LELGILGARGLLP	-----MKA-KNGG	-----KGSTQ				
Q9CA47_ARATH_671762/1-92	-----LELGILGARGLLP	-----MKA-KNGG	-----KGSTQ				
Q8S1F8_ORYSA_670764/1-95	-----LELGITGACQLP	-----MKT-KGGA	-----KGSTQ				
Q7XZZ4_ORYSA_639736/1-98	-----LEVGIKGAANLP	-----MKIADGA	-----SGSTD				
Q9SS68_ARATH_606698/1-93	-----LEVGIKGAANLP	-----MKT-RDGT	-----RGTTD				
Q9LZE5_ARATH_351437/1-87	-----LEIGILGATELKG	-----SDEKQK	-----ID				
Q9M366_ARATH_360447/1-88	-----LEIGILGATELKG	-----SDEKQK	-----ID				
Q2QWP5_ORYSA_478560/1-83	-----VHLGILRATGLP	-----LRMG	-----KSTVN				
Q7XTM4_ORYSA_443525/1-83	-----LSVTVISGEDLP	-----MDHNGK	-----SD				
Q25AB2_ORYSA_443525/1-83	-----LSVTVISGEDLP	-----MDHNGK	-----SD				
Q23994_HORVU_146/1-46	-----	-----	-----				
Q8L706_ARATH_437519/1-83	-----LSVTVISAEIPI	-----QDLHGK	-----AD				
Q9ZVY8_ARATH_405487/1-83	-----LSVTVISAEIPI	-----QDLHGK	-----AD				
Q1S211_MEDTR_434516/1-83	-----LSVTVISAEIPI	-----VDFHGK	-----AD				
Q1SF66_MEDTR_443525/1-83	-----LSVTVISAEIPI	-----VDFHGK	-----SD				
Q6UU05_ORYSA_295377/1-83	-----LSVTVISAEIPI	-----MDVHGK	-----AD				
Q69UK6_ORYSA_435517/1-83	-----LSVTVISAEIPI	-----MDVHGK	-----AD				
Q9FY55_ARATH_450532/1-83	-----LSVTVISAEIPI	-----VDFHGK	-----AD				
P92940_ARATH_248329/1-82	-----LIVTVVKATNLKN	-----KELIGK	-----SD				
Q9LEX1_ARATH_265346/1-82	-----LIVTVVKATNLKN	-----KELIGK	-----SD				
Q9LDM1_ARATH_216259/1-44	-----	-----	-----				
Q5MD17_BRANA_179/1-79	-----TVVKATNLKN	-----KEFIGK	-----SD				
Q48645_LYCES_264346/1-83	-----LTVTVVKATNLKN	-----HEMIGK	-----SD				
Q7XAL6_ORYSA_264346/1-83	-----LTVTVVKATNLKN	-----KELIGK	-----SD				
Q6UU05_ORYSA_124208/1-85	-----LEVKLVEARDLTN	-----KDLVGK	-----SD				
Q69UK6_ORYSA_264348/1-85	-----LEVKLVEARDLTN	-----KDLVGK	-----SD				
Q9ZVY8_ARATH_232316/1-85	-----LEVKLVOAKNLTN	-----KDLVGK	-----SD				
Q8L706_ARATH_264348/1-85	-----LEVKLVOAKNLTN	-----KDLVGK	-----SD				
Q7XTM4_ORYSA_264348/1-85	-----LEVKLVOAKNLTN	-----KDLIGK	-----SD				
Q25AB2_ORYSA_264348/1-85	-----LEVKLVOAKNLTN	-----KDLIGK	-----SD				
Q1SF66_MEDTR_264348/1-85	-----LEVKLVOAKNLTN	-----KDLIGK	-----SD				
Q1S211_MEDTR_251335/1-85	-----LDVKLVQAKNLKN	-----KDLIGK	-----SD				
Q9FY55_ARATH_273357/1-85	-----LDVKLVQAKNLKN	-----KDLIGK	-----SD				
Q5MD16_BRANA_182/1-82	-----KVVRAVGLRK	-----KDMHGG	-----AD				
Q9SKR2_ARATH_262346/1-85	-----VHVKVVRVGLRK	-----KDLHGG	-----AD				
Q9SKR0_ARATH_269/1-68	-----	-----GM	-----IN				
Q17680_MEDTR_262346/1-85	-----LHVKVLHAMKLLK	-----KDLIGA	-----SD				
Q9LNT5_ARATH_260344/1-85	-----LSVKVIAKIKLLK	-----KDLGG	-----SD				
Q6ETC4_ORYSA_262346/1-85	-----LHVNJVRVAVKLT	-----KDFLGK	-----SD				
Q3E9M4_ARATH_41125/1-85	-----LHVSILRARNLLK	-----KDLGT	-----SD				
Q7XA06_ARATH_263347/1-85	-----LHVSILRARNLLK	-----KDLGT	-----SD				
Q9FYD9_ARATH_306390/1-85	-----LHVSILRARNLLK	-----KDLGT	-----SD				
Q655F0_ORYSA_262346/1-85	-----LHVKVIRAHNLLK	-----MDLLGK	-----SD				
Q5QLZ9_ORYSA_262340/1-79	-----LLVKVLRAQNLRE	-----KGPLGK	-----RD				

FIG. 3 (Folha 13 de 27)

C2				C3			
80	90	100	110	120	130	140	150
AYCVAKY	-----GOKWIRTRTVV	-----DSVC	-----PRWNE	-----QYTWEVFDP	-----CTVITVGV		
PYCVAKY	-----GNKWIRTRTLL	-----DALA	-----PKWNE	-----QYTWEVHDP	-----CTVITIGV		
PYCVAKY	-----GNKWIRTRTLL	-----DALA	-----PKWNE	-----QYTWEVHDP	-----CTVITIGV		
AYCVAKY	-----GPKWIRTRTIL	-----NTLN	-----PKWNE	-----QYTWEVFDP	-----CTVITVVV		
PYCVAKY	-----GAKWIRTRTLV	-----GTAA	-----PRWNE	-----QYTWEVFDL	-----CTVVTAVV		
TYVVAKY	-----GHKWIRSRTVI	-----NSMN	-----PKYNE	-----QYTWEVFDP	-----ATVLTICV		
TYVVAKY	-----GHKWIRSRTVI	-----NSMN	-----PKYNE	-----QYTWEVFDP	-----ATVLTICV		
AYCVAKY	-----GTKWIRTRTVT	-----NCLN	-----PRFNE	-----QYTWEVYEP	-----ATVITIGV		
AYCVAKY	-----GOKWIRTRTIV	-----DNLN	-----PRFNE	-----QYTWDVFDH	-----GTVLTIGL		
AYCVAKY	-----GVKWIRTRTVT	-----DSIS	-----PRFHQ	-----QYHWEVHGH	-----CTVLTAVV		
AYTVAKY	-----GSKWIRTRTVS	-----DSL	-----PKWNE	-----QYTWKVYDP	-----CTVLTIGV		
AYAVAKY	-----GPKWIRTRTIS	-----DSFD	-----PAWNE	-----QYTWPVYDP	-----CTVLTGVG		
AYCVAKY	-----GKKWIRTRTIT	-----DSFD	-----PRWHE	-----QYTQVYDP	-----CTVLTGVG		
AYCVAKY	-----GKKWIRTRTIT	-----DSFD	-----PRWHE	-----QYTQVYDP	-----CTVLTGVG		
AYCVAKY	-----GKKWIRTRTVT	-----DSL	-----PRWNE	-----QYTQVYDP	-----CTVLTAVV		
AYVVLKY	-----GPKWIRTRTIL	-----DOFN	-----PRWNE	-----QYAWDVFDH	-----CTVLTIGV		
AYVVLKY	-----GPKWIRTRTIL	-----DRFN	-----PRWNE	-----QYTWDVYDP	-----CTVLTIGV		
SYVVAKY	-----GNKWIRTRTVV	-----NSVT	-----PKWNE	-----QYSWDDYEP	-----CTVLTIGI		
SYVVAKY	-----GNKWIRTRTVV	-----NSVS	-----PKWNE	-----QYSHOVYEP	-----CTVLTIGI		
PYCVAKY	-----GOKWIRTRTIL	-----DGPE	-----HVFNE	-----OHTWSVYDI	-----ATVLTAGV		
PYVVLSLKK	-----SKTK-YKTRVVS	-----ESLN	-----PWNQ	-----TFDFVVEDGLHD	-----MLMLEV		
PYVVLSLKK	-----SKTK-YKTRVVS	-----ESLN	-----PWNQ	-----TFDFVVEDGLHD	-----MLMLEV		
	-----RVVN	-----ESLN	-----PWNQ	-----TFDFVVEDGLHD	-----MLMLEV		
PYVVL SMK	-----SGAK-SKTRVVN	-----DSL	-----PWNQ	-----TFDFVVEDGLHD	-----MLMLEV		
PYVVL SMK	-----SGAK-SKTRVVN	-----DSL	-----PWNQ	-----TFDFVVEDGLHD	-----MLMLEV		
PFVVLALKK	-----SEKK-QKTRVVN	-----ETLN	-----PWNQ	-----TFDFVVEDGLHD	-----MLLVEL		
PFVVLTKK	-----AETK-NKTRVVN	-----NSLN	-----PWNQ	-----TFDFVVEDGLHD	-----MLLVEL		
PFVVLTKK	-----GETK-KKTRVVN	-----ETLN	-----PWNQ	-----TFDFVVEDGLHD	-----MLLVEL		
PFVVLTKK	-----GETK-KKTRVVN	-----ETLN	-----PWNQ	-----TFDFVVEDGLHD	-----MLLVEL		
AFVVI LKK	-----SETK-SKTRVVP	-----DSL	-----PWNQ	-----TFDFVVEDGLHD	-----LLTLEV		
PYATIYIRP	-----VFK--YKTNATD	-----NNLN	-----PWNQ	-----TFELIAEDKETQ	-----SLTVEV		
PYATIYIRP	-----VFK--YKTKAIE	-----NNLN	-----PWNQ	-----TFELIAEDKETQ	-----SLTVEV		
	-----IE	-----NNLN	-----PWNQ	-----TFELIAEDKETQ	-----SLTVEV		
PYATIHIRP	-----VFK--YNTKAIE	-----NNLN	-----PWNQ	-----TFDLIAEDKETQ	-----SLTIEV		
PYAVVHIRP	-----LFK--VKTKTID	-----NNLN	-----PWNQ	-----TFELIAEDKETQ	-----SLFIEV		
PYVILYVRP	-----MFK--VKTKVID	-----DNLN	-----PEWNE	-----TFPLIVEDKETQ	-----SVIFEV		
PFAVLYIRP	-----LQDKMKSKTIN	-----NDLN	-----PIWNE	-----HYEFVVEDTSTQ	-----RLTVKI		
PFAVLYIRP	-----LQDKMKSKTIN	-----NDLN	-----PIWNE	-----HYEFVVEDTSTQ	-----RLTVKI		
PFAKMFIRP	-----LREKTKRSKTIN	-----NDLN	-----PIWNE	-----HFEFVVEDASTQ	-----HLVVRI		
PFAKMFIRP	-----LREKTKRSKTIN	-----NDLN	-----PIWNE	-----HFEFVVEDASTQ	-----HLVVRI		
PFAIVYVRP	-----LPDKMKRSKTIN	-----NDLN	-----PIWNE	-----HFEFIVEDADTQ	-----TVTVKI		
PFAIVYVRP	-----LPDKMKRSKTIN	-----NDLN	-----PIWNE	-----HFEFIVEDADTQ	-----TVTVKI		
PYAVLYIRP	-----LRNRTKSKTIN	-----NDLN	-----PIWNE	-----HFEFIVEDASTQ	-----HLFVKV		
PFAVVFVRP	-----LRDKTKTSKIIN	-----NOLN	-----PIWNE	-----HFEFIEEDSTQ	-----HLTIRI		
PYAIVFIRP	-----LPDRTKTKTIS	-----NSLN	-----PIWNE	-----HFEFIVEDVSTQ	-----HLTVRV		
PYVKIKLSE	-----DKIPSKKTTVKH	-----KNLN	-----PEWNE	-----EHKFSVRDPQTO	-----VLEFSV		
PFVKIKLSE	-----DKIPSKKTTVKH	-----KNLN	-----PEWNE	-----EHKFSVRDPQTO	-----VLEFSV		
PYVQIELSE	-----DKISSKKTIVKH	-----KNLN	-----PEWNE	-----EFKFSVRDPKTO	-----VLEFNV		
PYVKLKLTD	-----DKMPSKKTIVKH	-----KNLN	-----PEWNE	-----EFNLVVKDPETO	-----VLQLNV		
PYVKLTLTG	-----DKVPGKKTIVKH	-----SNLN	-----PEWNE	-----EFDLVVKEPESQ	-----ELQLIV		
PYVKLKLTE	-----EKLPSKKTIVKR	-----SNLN	-----PEWNE	-----DFKLIVKDPESQ	-----ALELTV		
PYVKLSLTG	-----EKLPAKKTITIKK	-----RNLN	-----PEWNE	-----HFKLIVKDPNSQ	-----VLQLEV		
PYVKLSLTG	-----EKLPAKKTITIKK	-----RNLN	-----PEWNE	-----HFKLIVKDPNSQ	-----VLQLEV		
PYVKLSLTG	-----EKLPAKKTITIKK	-----RNLN	-----PEWNE	-----HFKLIVKDPNSQ	-----VLQLEV		
PYVKLRLSG	-----EKLPSKKTISJKM	-----SNLN	-----PEWNE	-----HFRFIVKDPETO	-----ILELRM		
PYVKLKMSG	-----SKLPSKKTIVKH	-----SNLN	-----PEWNE	-----EFKFIIVKDPETO	-----ELDIN		

FIG. 3 (Folha 14 de 27)

	160	170	180	190	200	210	220	230	240
FDN---CHVD-----	KPASGNTTLAVRDNCIGKVR								
FDN---SHVN-----	DGGD-----	FKDQRIJKVR							
FDN---SHVN-----	DGGD-----	FKDQRIJKVR							
FDN---NQIG-----	KNGD-----	ARDESIJKVR							
FDN---CHLT-----	GCGD-----	AKDQRIJKVR							
FDN---AHFA-----	AGDGG-----	NKRQPIJKVR							
FDN---AHFA-----	AGDGG-----	NKRQPIJKVR							
FDN---NQIN-----	SGNG-----	NKGDKIJKIR							
FDN---CHIS-----	ADSNHSSSPGHMOKPIJKVR								
FHN---SQIG-----	DKGGLVAGDPVKDVLIGKVR								
FDS---WGVY-----	EVOGGKEATRDRIJKVR								
FDDPPPPSPSQ-----	LPDGAKDAAAFSRPIJKVR								
FDN---WRMF-----	S---DASDD-RPDTRIJKIR								
FDN---WRMF-----	S---DASDD-RPDTRIJKIR								
FDN---WRMF-----	AFAGAGDEQRQDYRIJKVR								
FDNVR--YRSA-----	EASGDAGKLPKDARIJKLR								
FDNGR--YKRD-----	ES---GKQGRDVRVGIKIR								
YDN---RQIFKE-----	DOANDVPIJKVR								
YDN---RQILEDK-----	NKANDVPIJKVR								
FDH---FPHTR-----	KAHREIJKVQ								
YDH---DTFS-----	RD-YMGRCI								
YDH---DTFS-----	RD-YMGRCI								
YDH---DTFS-----	RD-YMGRCI								
WDH---DTFG-----	KD-YIGRCI								
WDH---DTFG-----	KD-YIGRCI								
WDH---DTFG-----	KE-KMGKVI								
YDH---DTFG-----	KD-YMGKVI								
WDH---DTFG-----	KD-YIGRCI								
WDH---DTFG-----	KD-YIGRCI								
WDH---DKFG-----	KD-KIGRVI								
FDK---D-VG-----	QDERLGLVK								
FDK---D-VG-----	QDERLGLVK								
FDK---D-VG-----	QDERLGLVK								
FDK---D-VG-----	QDERLGLVK								
FDK---DNIG-----	QDORMGVAK								
YDE---DRLQ-----	QDKKLGVAK								
YDD---EGLQ-----	ASELISCAR								
YDD---EGLQ-----	ASELISCAR								
YDD---EGVQ-----	ASELISCAQ								
YDD---EGVQ-----	ASELISCAQ								
YDD---DGIO-----	ESELISCAQ								
YDD---DGIO-----	ESELISCAQ								
YDD---EGLQ-----	SSELISCTD								
FDD---EGIO-----	AAELISCAQ								
FDD---EGVG-----	SSQLISAAQ								
YDW---GQLG-----	KHDKMGMN								
YDW---EQVG-----	NPEKMGMN								
YDW---EKIG-----	KHDKMGMN								
YDW---EQVG-----	KHDKMGMN								
YDW---EQVG-----	KHDKMGMN								
YDW---EQVG-----	KHDKIGMSV								
FDW---DKVG-----	GHDRLGMOM								
FDW---DKVG-----	GHDRLGMOM								
FDW---DKVG-----	GHDRLGMOM								
FDW---EKVK-----	MHDKLSMOV								
FG-----	KDEKLGCK								

FIG. 3 (Folha 15 de 27)

C1

	10	20	30	40	50	60	70
Q69JE2_ORYSA_262346/1-85	-----LLVKVLRQAQRLRK-----	-----KDLLGK-----	-----SD				
Q8L626_ARATH_403488/1-86	-----LSVTLVDAQKRLRY-----	-----M-FFGK-----	-----TD				
Q9SX44_ARATH_391476/1-86	-----LSVTLVDAQKRLRY-----	-----M-FFGK-----	-----TD				
Q9LT26_ARATH_388473/1-86	-----LSVTLVNAQKLPY-----	-----M-FSGR-----	-----TD				
Q1T4J1_MEDTR_78119/1-42	-----LSVTLVDAQKLPY-----	-----FFGK-----	-----TD				
Q6K9U1_ORYSA_403488/1-86	-----LSVTLVDAQKLSF-----	-----V-LFGK-----	-----TD				
Q04042_ARATH_262345/1-84	-----FRCVNL DN-----	-----KDLFSK-----	-----SD				
Q5XQC7_ARATH_202285/1-84	-----FRCVNL DN-----	-----KDLFSK-----	-----SD				
Q1KS96_ARATH_202285/1-84	-----FRCVNL DN-----	-----KDLFSK-----	-----SD				
Q5S1W2_ARATH_199283/1-85	-----VFRGLNLES-----	-----KDTFSK-----	-----SD				
Q94EW4_ARATH_199283/1-85	-----VFRGLNLES-----	-----KDTFSK-----	-----SD				
Q9LY30_ARATH_217301/1-85	-----VFRGLNLES-----	-----KDTFSK-----	-----SD				
Q9FH53_ARATH_199283/1-85	-----VFRCSNLES-----	-----KDLFSK-----	-----SD				
Q941L3_ARATH_199283/1-85	-----VFRCSNLES-----	-----KDLFSK-----	-----SD				
Q6H563_ORYSA_198286/1-89	-----IMENVFRCSNLEI-----	-----KDLFSK-----	-----SD				
Q5S1W2_ARATH_53146/1-94	-----FSASNL RD-----	-----RDVISK-----	-----SD				
Q94EW4_ARATH_53146/1-94	-----FSASNL RD-----	-----RDVISK-----	-----SD				
Q9LY30_ARATH_71164/1-94	-----FSASNL RD-----	-----RDVISK-----	-----SD				
Q9FH53_ARATH_54147/1-94	-----FSASNL RD-----	-----RDVISK-----	-----SD				
Q941L3_ARATH_54147/1-94	-----FSASNL RD-----	-----RDVISK-----	-----SD				
Q6H563_ORYSA_58151/1-94	-----LSASNL GD-----	-----QEFTTK-----	-----SN				
Q1KS96_ARATH_58151/1-94	-----LSASNL LD-----	-----CDITSK-----	-----SD				
Q5XQC7_ARATH_58151/1-94	-----LSASNL LD-----	-----CDITSK-----	-----SD				
Q04042_ARATH_58145/1-88	-----VNQLTSLASNL LD-----	-----CDITSK-----	-----SD				
Q6T5P8_ORYSA_65157/1-93	-----FSASKLRN-----	-----MDAFSK-----	-----SD				
Q66VB0_ORYSA_606687/1-82	-----MTVALIEGT-----	-----GITNSNK-----	-----ELFDM-				
Q6Z7A3_ORYSA_632713/1-82	-----LTVALIEGS-----	-----GVVGSOTP-----	-----GLPDP-				
Q6Z8U1_ORYSA_590672/1-83	-----LTVALIEGT-----	-----KLAPVDAT-----	-----GFSDP-				
Q5Z614_ORYSA_558640/1-83	-----LTVALIDGT-----	-----NLAATKSS-----	-----GYSOP-				
Q9ZVT9_ARATH_537619/1-83	-----LTVALIEGV-----	-----DLAAVDPS-----	-----GHCDP-				
Q9FGS8_ARATH_541623/1-83	-----LTIALIKGT-----	-----NLASVEAT-----	-----ELFDP-				
Q7XK16_ORYSA_549630/1-82	-----LTVALLEAT-----	-----SLPPVS-S-----	-----GSVDP-				
Q9T0H5_ARATH_90154/1-65	-----	-----	-----GTSOP-				
Q56W08_ARATH_430523/1-94	-----LKVKIYTGEQNDL-----	-----DFHHTHFD-----	-----QYSPPD				
Q56W08_ARATH_430523/1-94	-----LKVKIYTGEQNDL-----	-----DFHHTHFD-----	-----QYSPPD				
Q9SZN3_ARATH_392485/1-94	-----LKVKIYTGEQNDL-----	-----DFHHTHFD-----	-----QYSPPD				
Q38811_ARATH_392485/1-94	-----LKVKIYTGEQNDL-----	-----DFHHTHFD-----	-----QYSPPD				
Q39032_ARATH_432524/1-93	-----LKVKIYTGEQNMH-----	-----DFPLDHFD-----	-----RYSPPD				
Q49970_ARATH_432524/1-93	-----LKVKIYTGEQNMH-----	-----DFPLDHFD-----	-----RYSPPD				
Q2V2X4_ARATH_362454/1-93	-----LKVKVCMGDGJLL-----	-----DFKKTTHFD-----	-----SYSPPD				
Q940R9_ARATH_464556/1-93	-----LKVKVCMGDGJLL-----	-----DFKKTTHFD-----	-----SYSPPD				
Q944C1_ARATH_470562/1-93	-----LKVKVCMGDGJLL-----	-----DFKKTTHFD-----	-----SYSPPD				
Q75IL8_ORYSA_471563/1-93	-----LKVTVYMGEQNYF-----	-----DFRHTHFD-----	-----KCSPPD				
Q39033_ARATH_454546/1-93	-----LRVTVYMGEQNYF-----	-----DFRHTHFD-----	-----QYSPPD				
Q8LG47_ARATH_57149/1-93	-----LRVTVYMGEQNYF-----	-----DFRHTHFD-----	-----QYSPPD				
Q56XL3_ARATH_136228/1-93	-----LRVTVYMGEQNYF-----	-----DFRHTHFD-----	-----QYSPPD				
Q9XEK4_BRANA_454546/1-93	-----LRVTIYMGEQNYF-----	-----DFPHTHFD-----	-----QYSPPD				
Q9LY51_ARATH_457549/1-93	-----LRVTIYMGEQNYF-----	-----DFPHTHFD-----	-----RYSPPD				
Q43443_SOYBN_471563/1-93	-----LKVTVYMGEQNYF-----	-----DFKHTHFD-----	-----QYSPPD				
Q43439_SOYBN_471563/1-93	-----LKVTVYMGEQNYF-----	-----DFKHTHFD-----	-----QYSPPD				
Q6SA76_9FABA_36128/1-93	-----LKVTVYMGEQNYF-----	-----DFKHTHFD-----	-----QYSPPD				
Q2PEW5_TRIPR_476568/1-93	-----LKVTVYMGEQNYF-----	-----DFKHTHFD-----	-----QFSPPD				
Q93YX8_MEDTR_466558/1-93	-----LKVTVYMGEQNYF-----	-----DFKHTHFD-----	-----QFSPPD				
Q24297_PEA_467559/1-93	-----LKVTVYMGEQNYF-----	-----DFKHTHFD-----	-----QFSPPD				
Q8LLW1_PEA_467559/1-93	-----LKVTVYMGEQNYF-----	-----DFKHTHFD-----	-----QFSPPD				
Q43444_SOYBN_424516/1-93	-----LKVTIYMGEQNYF-----	-----DFKHTHFD-----	-----KFSPPD				

FIG. 3 (Folha 16 de 27)

[illegible]

FIG. 3 (Folha 17 de 27)

C4-

[illegible]

FIG. 3 (Folha 18 de 27)

C1

	10	20	30	40	50	60	70
Q6TM10_9FABA_464556/1-93	-----LKVTIYMGECHFH-----	-----DFKHTHFD-----					QYSPPD
Q9M5Q5_TOBAC_459551/1-93	-----LKVTVMGEGNYI-----	-----DFKHTHFD-----					AYTPPD
Q9M5Q2_TOBAC_461553/1-93	-----LKVTVMGEGNYI-----	-----DFNHTHFD-----					AYAPPD
Q49952_SOLTU_458550/1-93	-----LKVTVMGEGNYI-----	-----DFEHTHFD-----					AYSPPD
Q49950_SOLTU_468560/1-93	-----LKVKVYMGDGMHL-----	-----DFKRTHFD-----					AYSPPD
Q2QNK2_ORYSA_463555/1-93	-----LKVKVYMGDGMHL-----	-----DFSHTHFD-----					TFSPPD
Q8RW31_ORYSA_153245/1-93	-----LKVKVYMGDGMHL-----	-----DFSHTHFD-----					TFSPPD
Q6QJ78_MAIZE_459551/1-93	-----LKVKVYMGDGMHL-----	-----DFSHTHFD-----					AFSPPD
Q6Z3Y9_ORYSA_471563/1-93	-----LKVKVYMGDGMHL-----	-----DFTQTHFD-----					QYSPPD
Q9AXF1_ORYSA_472564/1-93	-----LKVKVYMGDGMHL-----	-----DFTQTHFD-----					QYSPPD
Q2YGU1_9LILI_406498/1-93	-----LKVKVYMGDGMHL-----	-----DFKSTHFD-----					TYSPPD
Q2YGU0_9LILI_420512/1-93	-----LKVKVYMGDGMHL-----	-----DFKQTYFD-----					AYSPPD
Q49902_NICRU_461553/1-93	-----LKVKVYMGDGMHL-----	-----DFKQTHFD-----					LYSPPD
P93341_NICRU_461553/1-93	-----LKVKVYMGDGMHL-----	-----DFKQTHFD-----					LYSPPD
Q5GA62_LYCES_449539/1-91	-----LQVRVYMGDGMHL-----	-----DFSHTHFD-----					AYSPPD
Q8GV43_ARATH_417509/1-93	-----LKVKVYMGDGMHL-----	-----DFSHTHFD-----					AYSPPD
Q9FSW1_9POAL_504595/1-92	-----LKVKVYMGEGNHH-----	-----DFKQTHFD-----					TYSPPD
Q49951_SOLTU_438530/1-93	-----LKVTVMGEGNHH-----	-----DFOQTHFD-----					TYSPPD
Q5GA63_LYCES_456548/1-93	-----LKVTVMGEGNHH-----	-----DFOQTHFD-----					TYSPPD
P93620_VIGUN_422514/1-93	-----LKVKVYLGKQNSL-----	-----DFSPSDFD-----					SYSPPD
Q6TM09_9FABA_321413/1-93	-----LKVKVYLGKQNSL-----	-----DFSPSDFD-----					SYSPPD
Q43442_SOYBN_426518/1-93	-----LKVKVYMGEGNHH-----	-----DFSHTHFD-----					SFSPPD
Q1SD14_MEDTR_462553/1-92	-----LKVKVYKGVGNSS-----	-----DFSPTHFD-----					RFSPPD
Q1SDG3_MEDTR_461552/1-92	-----LKVKVYKGVGNSS-----	-----DFSPTHFD-----					RFSPPD
Q944C2_ARATH_451543/1-93	-----LKVKVYMGEGNHH-----	-----GFQRTCFN-----					TWSSPN
Q9LUZ0_ARATH_451543/1-93	-----LKVKVYMGEGNHH-----	-----GFQRTCFN-----					TWSSPN
Q76604_PHYPA_521593/1-73	-----LKVKIYMGDGMIV-----	-----DFKKR-IG-----					RLSKPD
Q6NMA7_ARATH_405496/1-92	-----LKVKIYMGDGMIV-----	-----DFKKR-IG-----					RLSKPD
Q9SD51_ARATH_387478/1-92	-----LKVKIYMGDGMIV-----	-----DFKKR-IG-----					RLSKPD
Q42582_ARATH_405496/1-92	-----LKVKIYMGDGMIV-----	-----DFKKR-IG-----					RLSKPD
Q9STZ3_ARATH_405496/1-92	-----LKVKIYMGDGMIV-----	-----DFKKR-IG-----					RLSKPD
Q762E2_PHYPA_515603/1-89	-----LKVTLLGTDHKK-----	-----N-----YD-----					VFKKPG
Q8RUY6_ARATH_271/1-70	-----LVVTIKRGN-----	-----NMKQ-----					SVGNPS
Q6NPD6_ARATH_191271/1-81	-----LVVTIKRGN-----	-----NMKQ-----					SVGNPS
Q700A9_CICAR_120200/1-81	-----LVVIVKRGH-----	-----NMRO-----					SVGIPS
Q1S353_MEDTR_20862166/1-81	-----LVVIVKRGH-----	-----NLRO-----					SVGNPS
Q67U15_ORYSA_864944/1-81	-----LTVTIKRGH-----	-----NLRO-----					SMATTN
Q9CAQ9_ARATH_19842064/1-81	-----LTVNVMRAN-----	-----NLKQ-----					SMATTN
Q8GXS1_ARATH_308388/1-81	-----LTVNVMRAN-----	-----NLKQ-----					TMGSTN
Q2R9P0_ORYSA_20022082/1-81	-----LTVTIKRGH-----	-----NLKQ-----					S-----N
Q9C6Y4_ARATH_19952071/1-77	-----LTVAIKRGD-----	-----NLKR-----					
Q5HZ03_ARATH_53140/1-88	-----LEVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q9FJ58_ARATH_87174/1-88	-----LEVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q84W25_ARATH_53140/1-88	-----LEVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q9LPS7_ARATH_41128/1-88	-----LEVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q9C6Q0_ARATH_328415/1-88	-----LEVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q8LAD1_ARATH_41128/1-88	-----LEVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q8H7W2_ORYSA_41128/1-88	-----LEVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q1SS74_MEDTR_58145/1-88	-----VDVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q8H5M0_ORYSA_29118/1-90	-----VDVYVHQARDIHNIC-----	-----IYHKQDV-----					
Q94CL2_ARATH_23113/1-91	-----LDIYVHGARGIHNIC-----	-----IYAAQDV-----					
Q941Z2_ORYSA_27120/1-94	-----LEIEVISAEGKVDR-----	-----KPLKKK-----					
Q80843_ARATH_1094/1-85	-----LEIEVISAEGKVDR-----	-----KPLKKK-----					
Q58FX0_ARATH_1094/1-85	-----LEIEVISAEGKVDR-----	-----KPLKKK-----					
Q84N41_ARATH_1094/1-85	-----LEIEVISAEGKVDR-----	-----KPLKKK-----					
Q941L2_ARATH_20101/1-82	-----LEIDLISAEGKLN-----	-----RPIKKK-----					

FIG. 3 (Folha 19 de 27)

C2				C3			
80	90	100	110	120	130	140	150
FYARVGIAG	VPYDTVMKKT	KSVE	DNWS	PSWNE	EFKFP	LSVPELAL	LRVEV
FYAKIGIAG	VPADNVMMKKT	LTLE	DMOT	PTWDE	KFEFPL	TVPELAL	LRVEV
FYAKIGIAG	VPADNVMMKKT	RTLE	NNWJ	PTWDE	KFEFPL	TVPELAL	LRVEV
FYAKIGIAG	VDADIYMMKKT	LTLE	DNWJ	PTWDE	QFEFPL	TVPELAL	LRVEV
FYVKIGIAG	VAADSRVKK	TKAIE	DNWJ	PTWND	EFEFPL	TVPELAL	LRVEV
FYTRVGIAG	VRADCVMMKKT	RTIE	DQWV	PMWDE	EFTFPL	TVPELAV	LRIEV
FYTRVGIAG	VRADCVMMKKT	RTIE	DQWV	PMWDE	EFTFPL	TVPELAV	LRIEV
FYTKVGIAG	VKADSVMMKKT	RVIE	DQWV	PMWDE	EFTFLL	TVPELAL	LRVEV
FYARVGIAG	VPADSVMMKKT	TRAIE	DNWV	PMWEE	DFTFKL	TVPELAL	LRVEV
FYARVGIAG	VPADSVMMKKT	TRAIE	DNWV	PMWEE	DFTFKL	TVPELAL	LRVEV
FYTRVGIAG	VPADCTMMKKT	RTIE	DDWT	PMWDE	ELVFP	TVPELAL	LRIEV
FYTRIGIAG	VPADXYMMKKT	KAIE	DDWT	PMWNE	EFVFP	TVPELAL	LRIEV
FYTRVGIAG	VPADIMKKKT	KTKE	DKWT	PMWDE	AFTFPL	TVPELAL	LRIEV
FYTRVGIAG	VPADIMKKKT	KTKE	DKWT	PMWDE	EFTFPL	TVPELAL	LRIEV
FYTK - VTG	VPADSRKKKT	TRILE	DDWC	PMWDE	EFNFPL	TVPELAL	LRIEV
FYTKMFTVG	VPADNAKKKT	KTIE	DNWV	PMWDE	EFSPFL	TVPELAL	LRIEV
FYVEVGIAG	VPLDSVMRK	TKAVE	DNWV	PMWEE	EFAFPL	TVPELAV	LRVEV
FYAKLGIAG	VPADVKKKT	TKTMD	DNWJ	PSWDE	QFEFPL	TVPELAL	LRJVK
FYAKLGIAG	VPADVKKKT	TETJD	DNWJ	PSWNE	QFEFPL	TVPELAL	LRJVK
FYVKVCIAG	VPADMMKKKT	SVIS	NNWF	PMWNE	EFDFPL	TVPELAL	LRIEV
FYVKVCIAG	VPADMMKKKT	SVIS	NNWF	PMWNE	EFDFPL	TVPELAL	LRIEV
FYTKVCIAG	VPADKANKKT	TKVQ	DNWF	PMWDE	EFEFPL	TVPELAL	LRIEV
FYTVKCIAG	VGADSVMMKKT	SVKM	DNWV	PMWDE	EFEFQL	TVPELAL	LRIEV
FYTKVSIAG	VRADCAKKKT	SVKM	DNWV	PMWDE	EFEFRL	TVPELAL	LRIEV
FYTRVGIAG	VRGDKVMKKKT	KEEQ	KTWE	PMWNE	EFEFQL	TVPELAL	LRIEV
FYTRVGIAG	VRGDKVMKKKT	KEEQ	KTWE	PMWNE	EFEFQL	TVPELAL	LRIEV
FFTRLLVTG	VPADVAKMK	TSVID	DVWE	PMWNE	DHEFYL	KCPALAL	LRIEV
LYVRISIAG	VPHDEKIMKT	TVKN	NEWK	PTWGE	EFTFPL	TYPDAL	ISFEV
LYVRISIAG	VPHDEKIMKT	TVKN	NEWK	PTWGE	EFTFPL	TYPDAL	ISFEV
LYVRISIAG	VPHDENIMKT	TVKN	NEWT	PTWGE	EFTFPL	TYPDAL	ISFEV
LYVRISIAG	VPHDENIMKT	TVKN	NEWT	PTWGE	EFTFPL	TYPDAL	ISFEV
LYVRISIHG	MHDDEQKF	KTIVCK	RSRE	PMWEVE	EFVQIR	VPKLAI	LRLEV
VFCKITLGN	NPP	RQTKVIS	TGPN	PEWDE	SFSWSF	SPPKGQK	LHISC
VFCKITLGN	NPP	RQTKVIS	TGPN	PEWDE	SFSWSF	SPPKGQK	LHISC
VYCKITLGN	NPP	RLTKVIS	TGPN	PEWDE	SFSWSF	SPPKGQK	LHISC
VYCKITLGN	SP	KLTKVVS	TGPN	PEWEE	SFSWSF	SPPKGQK	LHISC
AFCKLTLGN	NPP	RLTKVIS	TGAT	PEWDE	AFAMAF	DSPPKGQK	LHISC
AFCOLTIGN	CPP	RQTKVVS	NSTT	PEWKE	GFTWAF	DVPPKGQK	LHIC
AFCOLTIGN	CPP	RQTKVVS	NSTT	PEWKE	GFTWAF	DVPPKGQK	LHIC
AFCCLOTIGN	GPP	RQTKVVN	NSIC	PMWNE	GFTWFL	DIPPKGQK	LYILC
AFCLRLION	CPT	KKTKVVK	RSSS	PMWKE	SFTWDF	FAAPRGQF	LEIVC
-YAKLCLTSD	P-DKS	-VSTKIJNG	-GGRN	-PVFDDNV	-K-LDVRV	-LDTSLKC	-EYMHMSR
-YAKLCLTSD	P-DKS	-VSTKIJNG	-GGRN	-PVFDDNV	-K-LDVRV	-LDTSLKC	-EYMHMSR
-YAKLCLTSD	P-DKS	-VSTKIJNG	-GGRN	-PVFDDNV	-K-LDVRV	-LDTSLKC	-EYMHMSR
-YAKLCLTND	P-ENS	-LSTKIJNG	-GGQN	-PVFDDTL	-Q-FDVKN	-LDCSLKC	-EYMHMSR
-YAKLCLTND	P-ENS	-LSTKIJNG	-GGQN	-PVFDDTL	-Q-FDVKN	-LDCSLKC	-EYMHMSR
-YAKLCLTND	P-ENS	-LSTKIJNG	-GGQN	-PVFDDTL	-Q-FDVKN	-LDCSLKC	-EYMHMSR
-YAKLCLTSD	P-DVS	-CSTKVIJNG	-GGRN	-PVFDDGL	-R-LDVRT	-VDASLKC	-EYMHMSR
-YAKLSLSD	P-ENS	-VNTKIJNG	-GGRN	-PVFNDL	-R-LSVRT	-VDSSLKC	-EYMHMSR
-YARLSLPG	G-APA	-ASTQVJNG	-GGRN	-PVFDQSL	-R-LGVRAG	-DVDGALRC	-EYMHMSR
-YAKFSLTYN	P-DDT	-ISTRI THR	-AGKN	-PEFNQKL	-M-IDVTO	-IDAHAAYLKC	-EYMHMSR
-YARLALTS	P-DDAPAL	-DTRVAAG	-GGAN	-PRFDERL	-PPLVRARR	-LGTVDLKC	-EYMHMSR
TSYVVRIDEK	-SWASKVDEL	-GGSY	-PTWKD	-RFDMEHPT	-NASV	-RFISIEV	-RFISIEV
TSYVVRIDEK	-SWASKVDEL	-GGSY	-PTWKD	-RFDMEHPT	-NASV	-RFISIEV	-RFISIEV
TSYVXRIDEK	-SWASKVDEL	-GGSY	-PTWKD	-RFDMEHPT	-NASV	-RFISIEV	-RFISIEV
TFAVVXIDEK	-CRKSNLDES	-RRSN	-PTWN	-YKSEHPT	-NGNE	-QFIFIEV	-QFIFIEV

FIG. 3 (Folha 20 de 27)

	160	170	180	190	200	210	220	230	240
HEYD			MSEKDDF	GGQTC					
HEYD			MSEKDDF	AGQTC					
HVYD			MSEKDDF	AGQTC					
HEYD			MSEKDDF	AGQTC					
HEYD			MSEIDDF	GGQTC					
HEYD			MSEKHDF	GGQTC					
HEYD			MSEKHDF	GGQTC					
QEYD			MSEKHDF	GGQTV					
HEYD			MSEKDDF	GGQTV					
HEYD			MSEKDDF	GGQTV					
HEYD			MSDKDDF	GGQTC					
HEYD			MSEKDDF	GGQTC					
HEYD			MSEKDDF	AGQTC					
HEYD			MSEKDDF	AGQTC					
REYD			MSEKDDF	GGQTC					
REYD			MSEKDDF	GGQTC					
HEQD			VSE-DDF	GGQTA					
LDYN			LSDKDEF	AGQTC					
LDYN			LSDKDEF	AGQTC					
REDD			KHQKDDF	GGQTC					
REED			KHQKDDF	GGQTC					
REYD			KHEKDDF	GGQTC					
KOKD			KG-SDDF	AGQTC					
KOKD			QT-KDDF	AGQTC					
HDYN			MPEKDDF	GGQTC					
HDYN			MPEKDDF	GGQTC					
RDHD			EGSQDEF	EGQAC					
YDYE			VSTPDYF	CGQTC					
YDYE			VSTPDYF	CGQTC					
YDYE			VSTADAF	CGQTC					
YDYE			VSTADAF	CGQTC					
REYD			RIVRDDM	VGQSC					
KNKS			KMGKSSF	GKVT					
KNKS			KMGKSSF	GKVT					
KNKS			KVGKSKF	GKVT					
KNKS			KVGKSKF	GKVT					
KNNS			KFGKKSF	GKVT					
KSKS			TFGKTTL	GRVT					
KSKS			TFGKTTL	GRVT					
KSKN			TFGKSTL	GRVT					
KSNN			IFRNKNL	GKVR					
VKN			YLEDQL	LGFTL					
VKN			YLEDQL	LGFTL					
VKN			YLEDQL	LGFTL					
VKN			YLEDQL	LGFSL					
VKN			YLEDQL	LGFSL					
VKN			YLEDQL	LGFSL					
VRN			YLEDQL	LGFAL					
VKN			YLEDQL	LGFAL					
VKN			YLEDQL	LGFAL					
ARR			YLEDQL	LGFAL					
ARR			LLDDQL	LGFAL					
YYRTS			GSGROKN	VGyak					
YYRTS			GSGROKN	VGyak					
YYRTS			GSGROKN	VGyak					
FYRT			GSGHDKK	IGeak					

FIG. 3 (Folha 21 de 27)

	10	20	30	40	50	60	70
Q8LBT3_ARATH_586/1-82	-----LEIDLRSAGL	KLNR-----	-----RPIKKK-----				
Q9M2E5_ARATH_687/1-82	-----LEIDLRSAGL	KLNR-----	-----RPIKKK-----				
Q8SA25_ORYSA_21115/1-95	-----LEVTLISAQGL	KPPSG-----	-----LRRRLLOA-----				
Q6K8N6_ORYSA_30117/1-88	-----LEVTVISAQGL	HR--R-----	-----LGRRVRAA-----				
Q22783_ARATH_1094/1-85	-----LELNIIISAQDL	APVS-----	-----RKMKT-----				
Q23030_ARATH_892/1-85	-----LELNIIISAQDL	APVA-----	-----RKTKT-----				
Q9S142_ARATH_1094/1-85	-----LELNIIISAQDL	APVA-----	-----RCHKT-----				
Q1SSZ2_MEDTR_1195/1-85	-----LELNIIISAQDL	AEVG-----	-----RSMRT-----				
Q9M8Z2_ARATH_1197/1-87	-----LELNIIISAQDL	APVS-----	-----RNMKT-----				
Q8L8S6_ARATH_1197/1-87	-----LELNIIISAQDL	APVS-----	-----RNMKT-----				
Q9M148_ARATH_794/1-88	-----LELNIIISAQGL	KEPTG-----	-----KLRRLOT-----				
Q8W0F9_ORYSA_50140/1-91	-----LEVTVVSGKH	KNVNW-----	-----RRGDLRA-----				
Q9LNV0_ARATH_995/1-87	-----LVVTVVSAKH	KNVNW-----	-----RNGDLKP-----				
Q6ZBX9_ORYSA_795/1-89	-----VEVTVSSARD	KNVNW-----	-----RNGDLKP-----				
Q651B1_ORYSA_896/1-89	-----VEVTVASARD	KNVNW-----	-----RNGDLKP-----				
Q5DVL6_HORVD_795/1-89	-----VEVTVGAARD	KNVNW-----	-----RNGDLKP-----				
Q5PNU6_ARATH_895/1-88	-----VEVTISSAKD	KNVNW-----	-----RNGPNKP-----				
Q9FF85_ARATH_24111/1-88	-----VEVTISSAKD	KNVNW-----	-----RNGPNKP-----				
Q60EX0_ORYSA_899/1-92	-----LELTLLSASD	RGVN-----	-----LVSKMEV-----				
Q94Z23_ORYSA_691/1-86	-----LELTLLSASD	KDVN-----	-----LLSKMEV-----				
Q8GRM0_ORYSA_691/1-86	-----LEVTLISARN	KKVN-----	-----LITPMEV-----				
Q8GS52_ORYSA_693/1-88	-----LEVTLHSARD	KNVN-----	-----FISRMHV-----				
Q94172_ORYSA_48138/1-91	-----LELTIVSASD	KKVT-----	-----LFSRMHV-----				
Q93W71_ORYSA_696/1-91	-----LELTIVSASD	KKVT-----	-----LFSRMHV-----				
Q65XE8_ORYSA_696/1-91	-----LEVTLVSAKN	KKVT-----	-----MFSKMNV-----				
Q9LK74_ARATH_697/1-92	-----LELNIVSAKD	LENVN-----	-----LITKMDV-----				
Q23425_ARATH_41134/1-94	-----LELKIVSASD	VNHID-----	-----ATDKMDV-----				
Q23427_ARATH_388479/1-92	-----LELNINSARN	LLNVN-----	-----LITKMDV-----				
Q04133_SOYBN_898/1-91	-----LELNIIISAQDL	KNVN-----	-----LFSKMDV-----				
Q23425_ARATH_232323/1-92	-----LELVVKFAXN	JEDVN-----	-----AFSSMDV-----				
Q04023_ARATH_695/1-90	-----LDLTIIISAED	KDVQ-----	-----LIGKQDL-----				
Q81814_ARATH_695/1-90	-----LDLTIIISAED	KDVQ-----	-----LIGKQDL-----				
Q69P64_ORYSA_15113/1-99	-----LELTVYEADD	LHN-----	-----AINGRIIKAES-----	-----LKESLG-----			-----VHRLAHR
Q8SAG6_ORYSA_15113/1-99	-----LELTVYEADD	LHN-----	-----AINGRIIKAES-----	-----LKESLG-----			-----VHRLAHR
Q1RZP8_MEDTR_84121/1-38							
Q1RZP4_MEDTR_59118/1-60							
PLDA1_PIMBR_10109/1-100	-----LHVTIYFEVDH	LK-----	-----AGSVVVFSES	LRRT-----	-----LRKPLV-----		-----LAKGTPK
Q1T525_MEDTR_10109/1-100	-----LHATIFEVDK	LK-----	-----NIGGGNILSK	IRQN-----	-----FEETVG-----		-----FGKGTTK
Q2HUA3_MEDTR_10109/1-100	-----LHATIFEVDK	LK-----	-----NIGGGNILSK	IRQN-----	-----FEETVG-----		-----FGKGTTK
PLDA1_VIGUN_10109/1-100	-----LHATIYEVDEL	LH-----	-----GGGGGNFFSK	LKQN-----	-----FEETVG-----		-----IGKGVTK
Q2Q0A8_CUCME_10109/1-100	-----LHATIYEIDRL	LH-----	-----TGGSSNVFSL	MRQN-----	-----FEEAVG-----		-----IGKGVTK
Q9XF8_CRAPL_10109/1-100	-----LHVTIYEVDDL	LH-----	-----SGGGGNFFTK	LKAN-----	-----FEETVG-----		-----FGKGTTPK
Q9XF7_CRAPL_10109/1-100	-----LHVTIYEVDDL	LH-----	-----AGGGGNIFSK	LKAN-----	-----FEETVG-----		-----FGKGTTPK
Q70EW5_CYNCA_10109/1-100	-----LHVTIYEVDDL	LR-----	-----EGGGPNVFGK	LMAN-----	-----FEETVG-----		-----FGKGTTPK
PLDA1_RICCO_10109/1-100	-----LHVTIYEVDDL	LH-----	-----SGGGPHFFRK	LVEN-----	-----FEETVG-----		-----FGKGVSK
Q2HWT8_ARAHY_10108/1-99	-----LHVTIYEVDDL	LK-----	-----TSGG-NVFTK	LQVN-----	-----FEETVG-----		-----FGKGVTK
PLDA1_BRAOC_10110/1-101	-----LHATIYEVDDL	HT-----	-----GGLRSGFFGK	ILAN-----	-----VEETIG-----		-----VGKGETQ
PLDA1_ARATH_51110/1-60							
PLDA2_BRAOC_52111/1-60							
PLDA2_ARATH_10110/1-101	-----LHATIYEVDDL	HA-----	-----EGGRSGFLGS	ILAN-----	-----VEETIG-----		-----VGKGETQ
Q9AWC0_LYCES_10110/1-101	-----LHVTIYEVDDL	LG-----	-----EEEGGHFFSK	IKQH-----	-----FEETVG-----		-----IGKGTTPK
Q9SDZ6_LYCES_10110/1-101	-----LHVTIYEVDDL	LG-----	-----EEEGGHFFSK	IKQH-----	-----FEETVG-----		-----IGKGTTPK
PLDA1_TOBAC_10109/1-100	-----LHVTIYEVDDL	LG-----	-----KEGGGHFFSK	IKKEH-----	-----VEETIG-----		-----FGKGTTPA
Q533V0_FRAAN_13111/1-99	-----LHATIYEVDDL	LH-----	-----GSSGNFLRK	ITGK-----	-----LEETVG-----		-----LGKGVSK
PLDA1_ORYSA_10114/1-105	-----LHATIFEASL	SNPHRASGSAPK	IRKFEV-----	-----IEDTVG-----			-----VGKGATK
PLDA1_MAIZE_10114/1-105	-----LHATIFEASL	SNPHRATGAPK	IRKLEVE-----	-----IEDTVG-----			-----VGKGATK

FIG. 3 (Folha 22 de 27)

[illegible]

FIG. 3 (Folha 23 de 27)

FIG. 3 (Folha 24 de 27)

	10	20	30	40	50	60	70
Q8VWE9_PAPSO_10114/1-105	-----LHVTIFEANSTSHPDRKTGGAPKFFRKLVEN-----	IEETVG-----	-----	-----	-----	-----	FGKGAS
Q8W1B2_PAPSO_10114/1-105	-----LHVTIFEANSTSHPDRKTGGAPKFFRKLVEN-----	IEETVG-----	-----	-----	-----	-----	FGKGAS
Q9AWB8_LYCES_1775/1-59	-----LHVTIFEVDKLR-----	TNFGREIFNKVVQG---	IEGAIG-----	-----	-----	-----	FNKTAST
Q9FR61_LYCES_10108/1-99	-----LHVTIFEVDRLH-----	TNFGRODFNKVVQG---	IEGAIG-----	-----	-----	-----	FNKAASR
Q75KP6_ORYSA_72130/1-59	-----MHVTIFEAESESNPSRPSSQAPQFLRKLVEG-----	IETDVG-----	-----	-----	-----	-----	VGKGTSK
Q65XR9_ORYSA_10114/1-105	-----LEATILEADHLSNPTRATGAAPGIFRKVEG-----	FEDSLG-----	-----	-----	-----	-----	LGKGATR
PLDA2_ORYSA_10100/1-91	-----LEATILEADHLSNPTRATGAAPGIFRKVEG-----	FEDSLG-----	-----	-----	-----	-----	LGKGATR
Q69X20_ORYSA_10114/1-105	-----LDATIFEATNLNPTRLTGNAPGFRKWNLENGLEKTTG-----	-----	-----	-----	-----	-----	LGPGGTR
Q69X21_ORYSA_10118/1-109	-----LDATIFEATNLNPTRLTGNAPGFRKWNLENGLEKTTG-----	-----	-----	-----	-----	-----	LGPGGTR
Q69X22_ORYSA_11115/1-105	-----LDATIFEATNLNPTRLTGSAPGIRKWNLENG-----	VEKTTG-----	-----	-----	-----	-----	VGOGGTR
Q9LKM3_ORYSA_11115/1-105	-----LDATIFEATNLNPTRLTGSAPGIRKWNLENG-----	VEKTTG-----	-----	-----	-----	-----	VGOGGTR
Q8H1U0_GOSHI_15129/1-115	-----LDLTIVEARRLPNMDFMVNHLSRCLT-CEPCKSPAQTAAKE-----	GDS-KIRGHRKIITSD	-----	-----	-----	-----	-----
Q8H6B9_GOSHI_15129/1-115	-----LDLTIVEARRLPNMDFMVNHLSRCLT-CEPCKSPAQTAAKE-----	GDS-KIRGHRKIITSD	-----	-----	-----	-----	-----
Q8H1T9_GOSHI_16133/1-118	-----LDLTIEARKLPNMDIVSNHLRKLCT-CETCKAPSOAAAQEP-----	GEVGKVHHHKIITSD	-----	-----	-----	-----	-----
Q3E9Q5_ARATH_16138/1-123	-----LDLKIIVKARRLPNMDMFEHLRRLFTACNACARPTDDVDPRDKGEFGDK-NIRSHRKVIITSD	-----	-----	-----	-----	-----	-----
Q8L891_ARATH_7129/1-123	-----LDLKIIVKARRLPNMDMFEHLRRLFTACNACARPTDDVDPRDKGEFGDK-NIRSHRKVIITSD	-----	-----	-----	-----	-----	-----
PLDD1_ARATH_16138/1-123	-----LDLKIIVKARRLPNMDMFEHLRRLFTACNACARPTDDVDPRDKGEFGDK-NIRSHRKVIITSD	-----	-----	-----	-----	-----	-----
Q7XJ06_ORYSA_25139/1-115	-----LDLWVVEARLPNMDMFEHVRRCFA-ACKPPTSCATARP-----	RHARGHHRKIITSD	-----	-----	-----	-----	-----
Q9LKM1_ORYSA_22137/1-116	-----LDIWIIEAKLPNMDIMSERMRFFFTGYGACGSSCAG-DNA-----	RRGGVGVPRPKIITSD	-----	-----	-----	-----	-----
Q6AVR2_ORYSA_22139/1-118	-----LDIWIIEAKLPNMDIMSERMRFFFTGYGACGSSCGTGDN-----	RRAGGVPRPKIITSD	-----	-----	-----	-----	-----
Q8LGN5_ORYSA_18128/1-111	-----LDIQUIEAKLPNMDLMTMRKCTGYGACSTECGK-----	SOPHTOVR-KIITSD	-----	-----	-----	-----	-----
Q1RUP7_MEDTR_19130/1-112	-----LDLFIIEAKLPNLDLSTEARKCLTMGNSCTPPFYKG-----	LKTHSGKD-KIITSD	-----	-----	-----	-----	-----
PLDG2_ARATH_29113/1-85	-----LMVELLH-----	GRRIRKVD-----	-----	-----	-----	-----	GEKSSKFTSD
Q3EA52_ARATH_38145/1-108	-----LDIWKIEAKLPNMIICYRNKLVGGISFSELGRRIRKVD-----	GEKSSKFTSD	-----	-----	-----	-----	-----
PLDG1_ARATH_44147/1-104	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
PLDG3_ARATH_48154/1-107	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
Q9ANB7_LYCES_39142/1-104	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
PLDB2_ARATH_119221/1-103	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
PLDB1_ARATH_158260/1-103	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
Q9ANB5_LYCES_99193/1-95	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
Q9XGT0_GOSHI_20126/1-107	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
Q8H1U1_GOSHI_353459/1-107	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
Q8H1U2_GOSHI_275381/1-107	-----LDIWKIEAKLPNMDGFHN-RLGGMLSG-LGRK-KVE-----	GEKSSKFTSD	-----	-----	-----	-----	-----
Q710M6_ORYSA_31135/1-105	-----LDVWVYDARNLPNKDLFSK-RVGDLLG-PRLIGAVGS-----	KMSSANMTSD	-----	-----	-----	-----	-----
Q8H093_ORYSA_243347/1-105	-----LDVWVYDARNLPNKDLFSK-RVGDLLG-PRLIGAVGS-----	KMSSANMTSD	-----	-----	-----	-----	-----
Q8H048_ORYSA_102202/1-101	-----LDIWIIEAKLPNMDIVSK-TVVDILG-TKKK-K-----	KAANGAMTSD	-----	-----	-----	-----	-----
Q8SAG7_ORYSA_102202/1-101	-----LDIWIIEAKLPNMDIVSK-TVVDILG-TKKK-K-----	KAANGAMTSD	-----	-----	-----	-----	-----
Q6YU55_ORYSA_38161/1-124	-----LDLTIIEAKLPNMDLSTLRLRLCLCLRPARRSPGOSRGVPADEGRRQPHGHLLPTSD	-----	-----	-----	-----	-----	-----
Q6ES10_ORYSA_58121/1-64	-----LDVTLKSASDLRE-----	DMSVK-----	-----	-----	-----	-----	LD
Q39485_CHLEU_1073/1-64	-----LDVTLKSASDLRE-----	DMSVK-----	-----	-----	-----	-----	LD
Q9LZ17_ARATH_697/1-92	-----LEINVTSAGLKK-----	VSKMD-----	-----	-----	-----	-----	-----
Q9MA57_ARATH_796/1-90	-----VEINVLSAQLNS-----	INLLFR-----	-----	-----	-----	-----	PT
Q9FGS8_ARATH_384/1-82	-----LYVYILOAKDLPA-----	-----	-----	-----	-----	-----	KE
Q9SS68_ARATH_283366/1-84	-----LYIRVAKAKRAKN-----	DGS-----	-----	-----	-----	-----	NP
Q9C7Z6_ARATH_26602763/1-104	-----MTVOILEAKDLHI-----	IDDGN-----	-----	-----	-----	-----	SHS
Q2HRV4_MEDTR_7107/1-101	-----IDLKISCKDINA-----	FNFFOK-----	-----	-----	-----	-----	LT
Q2R360_ORYSA_1396/1-84	-----VTIRISIRGKVA-----	FVPFOK-----	-----	-----	-----	-----	PP
Q1T1V3_MEDTR_14104/1-91	-----FELRIIOARNIES-----	VKS-----	-----	-----	-----	-----	TKN

FIG. 3 (Folha 25 de 27)

C2				C3			
80	90	100	110	120	130	140	150
LYASVOLD-----	KARVGRTRI1KD--	EPVN-----	PKWYE-----	SFHIYCAHMAAN--	-----	VIFTV-----	
LYASVOLD-----	KARVGRTRI1KD--	EPVN-----	PKWYE-----	SFHIYCAHMAAN--	-----	VIFTV-----	
-YATIDLG-----	KARVGRTRLLD--	EHKN-----	PRWYE-----	SFHIYCAHMASD--	-----	VVFTV-----	
LYATIDLG-----	KARVGRTRLLD--	EHKN-----	PRWYE-----	SFHIYCAHMASD--	-----	VVFTV-----	
LYATIDLG-----	KARVGRTRLLD--	DHKN-----	PRWYE-----	SFHIYCAHMAAN--	-----	VITTV-----	
-YATVOID-----	KARVARTRTVE--	PTGT-----	PRWKE-----	SFHIYCAHYAGD--	-----	VIFTV-----	
VYATIGLD-----	KARVGRTRTLAD--	DTAA-----	PRWYE-----	SFHVYCAHLATH--	-----	VAFTL-----	
LYATIDLG-----	RARVGRTRVDD--	EPVN-----	PRWYE-----	SFHIYCAHFAAD--	-----	VVFSV-----	
LYATIDLG-----	RARVGRTRVDD--	EPVN-----	PRWYE-----	SFHIYCAHFAAD--	-----	VVFSV-----	
LYATVDLG-----	RARLGRTRVDD--	EPVS-----	PRWDE-----	SFHIYCAHFAEN--	-----	VVFSV-----	
LYATVDLG-----	RARLGRTRVDD--	EPVS-----	PRWDE-----	SFHIYCAHFAEN--	-----	VVFSV-----	
LYATVDLG-----	RARLGRTRVDD--	EPVN-----	PRWDE-----	SFHIYCAHFAEN--	-----	VVFSV-----	
LYATVDLG-----	KARLGRTRVDD--	EPVN-----	PRWDE-----	SFHIYCAHFAEN--	-----	VVFSV-----	
PYVTCLP-----	QATVARTRVLK--	NSQN-----	PKWNE-----	HFI1PLAHPVTE--	-----	LDINV-----	
PYVTCLP-----	QATVARTRVLK--	NSQN-----	PKWNE-----	HFI1PLAHPVTE--	-----	LDINV-----	
PYVTITVP-----	QSTLARTPVLK--	SADN-----	REWNE-----	RFI1PLAHPVTE--	-----	LDINV-----	
PYVTVVVP-----	QATLARTRVLK--	NSQE-----	PLWDE-----	KFNISIAHPFAY--	-----	LEFOV-----	
PYVTVVVP-----	QATLARTRVLK--	NSQE-----	PLWDE-----	KFNISIAHPFAY--	-----	LEFOV-----	
PYVTVVVP-----	QATLARTRVLK--	NSQE-----	PLWDE-----	KFNISIAHPFAY--	-----	LEFOV-----	
PYVTVSV-----	GAVVARTRVIP--	NDQD-----	PRWDE-----	RFVPLAHPYAAA--	-----	LEFHV-----	
PYVSVCLA-----	GATVARTRVIP--	NSEN-----	PRWEE-----	RFRVEVAHVSRS--	-----	LEFHV-----	
PYVSVCLA-----	GATVARTRVIP--	NSEN-----	PRWEE-----	RFRVEVAHVSRS--	-----	LEFHV-----	
PYVSVCLA-----	GATVARTRVIA--	NSEN-----	PRWDE-----	HFYVQVAHSVSR--	-----	VEFHV-----	
PYVSICLA-----	GATVARTRVIP--	NCEN-----	PLWDE-----	HFLVPVAHPAHK--	-----	TEFLV-----	
PYVTVSIS-----	GAVIGRTFVIS--	NSEN-----	PRWMO-----	HFDVPVAHSAAE--	-----	VHFVV-----	
PYVTVSIS-----	GAVIGRTFVIS--	NSEN-----	PRWMO-----	HFDVPVAHSAAE--	-----	VHFVV-----	
PYVTVSIS-----	GAVIGRTFVIS--	NSEN-----	PRWMO-----	HFDVPVAHSAAE--	-----	VHFVV-----	
PYVTVSIS-----	GAVIGRTFVIS--	NSEN-----	PRWMO-----	HFDVPVAHSAAK--	-----	VHFVV-----	
PYVTVSVS-----	NAVARTYVIN--	NSEN-----	PRWMO-----	HFYVPVAHYASE--	-----	VHFVV-----	
PYVTVSVS-----	GAVIGRTYVIS--	NSEN-----	PRWMO-----	HFYVPVAHHAEE--	-----	VHFVV-----	
PYVTVSVS-----	GAVIGRTYVMS--	NSEN-----	PRWMO-----	HFYVPVAHHAEE--	-----	VHFVV-----	
PYVTVSVS-----	DATIGRTYVIN--	NNEN-----	PRWMO-----	HFNVVPVAHYAAE--	-----	VQFLV-----	
PYVTJAVS-----	GAVIGRTFVIS--	NNEN-----	PRWMO-----	HFNVVPVAHHAEE--	-----	VQFVV-----	
PYVTJAVS-----	GAVIGRTFVIS--	NNEN-----	PRWMO-----	HFNVVPVAHHAEE--	-----	VQFVV-----	
PYVTJAVS-----	GAVIGRTFVIN--	NDEN-----	PRWRO-----	HFYVPVAHHAEE--	-----	VQFVV-----	
PYVTIQVS-----	YATVARTYVVP--	NNEN-----	PRWTO-----	NFLVPVGHDAEE--	-----	VEFVV-----	
PYVTIQVS-----	YATVARTYVVP--	NNEN-----	PRWTO-----	NFLVPVGHDAEE--	-----	VEFVV-----	
PYVTIQVS-----	SATVARTYVVN--	DDEN-----	PRWAO-----	HFLIPVAHEAPA--	-----	VHFLV-----	
PYVTIQVS-----	SATVARTYVVN--	DDEN-----	PRWAO-----	HFLIPVAHEAPA--	-----	VHFLV-----	
PYAAVVVA-----	GNTLARTHVVV--	DSED-----	REWST-----	HVLHLAHATG--	-----	VAFHV-----	
-YVNIQF-----	GDQIFTSKITQG--	KGKK-----	VWNE-----	KFRFPLSSDECKEL--	-----	AKVTLKI-----	
AYCVVSC-----	ASTAHSNTVTD--	AGKT-----	MWNE-----	TFHFDKVASTS--	-----	VKLLE-----	
VFAVAVKSGDP-----	KCSOHREQRTQAARD--	GGTS-----	PKWSN-----	DVMKFILQNLAEAN--	-----	RLVITFKI-----	
VYVSVSVTR-----	GSRDKQVTPAAAWDKKFL--	-----	RWNY-----	RMKFYIEDDKVRRN--	-----	ESVVFQOI-----	
TFAKLHV-----	GRHKSRTVAR--	DTSS-----	PIWNE-----	EFVFRISDVDEGD--	-----	DVVVSI-----	
VYAKLVI-----	GTNGVMTSQT--	GKO-----	WDQ-----	VFAFEKESLNSTS--	-----	LEVSF-----	
FFCTLRVLVDSQGAEPQKLF--	PQSARTKCVKPSTTIVNDLMECTSKWNE--	-----	LFIFEIPRKGVAR--	-----	LEVEF-----		
LYAQVSIISTN-----	PKTKLTKQQQRTPTHRDODDGTN--	-----	PEWNHQTFRNLSFLSSHPDSE--	-----	FFLSFER-----		
LYAAVSLA-----	GRREKTSQDPD--	GGEN-----	PDWDA-----	VFAFDLPAAADG--	-----	MLQFEP-----	
LFARLYLP-----	TGNKRILNSKSVSTKS--	-----	VPFWD-----	SFNLDCSCQOEFLN--	-----	LNQQSLEVEL-----	

FIG. 3 (Folha 26 de 27)

FIG. 3 (Folha 27 de 27)

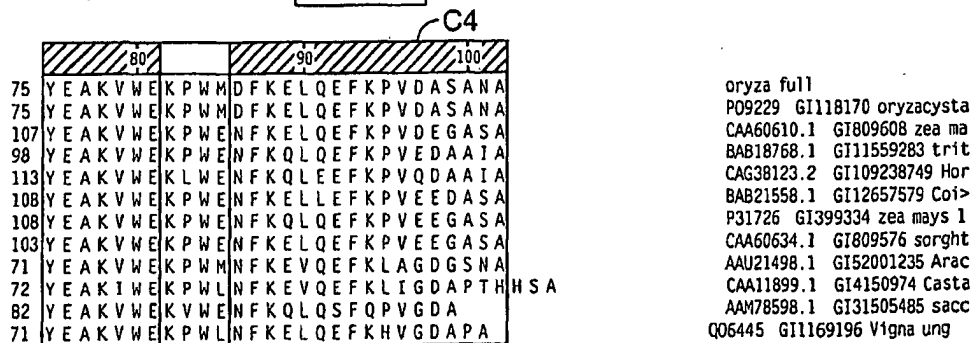
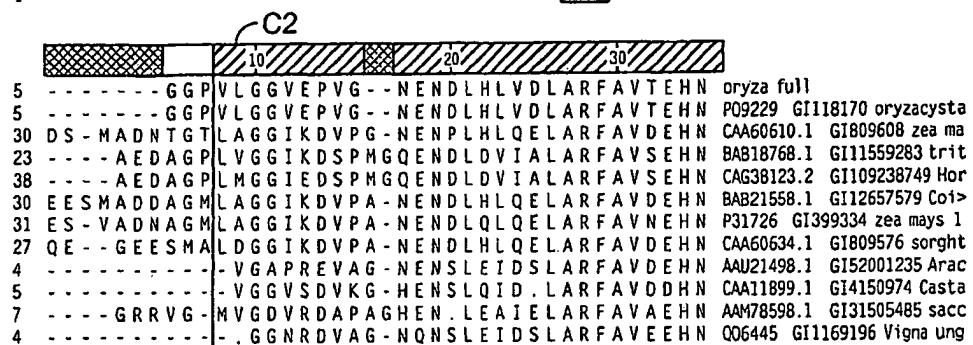


FIG. 4

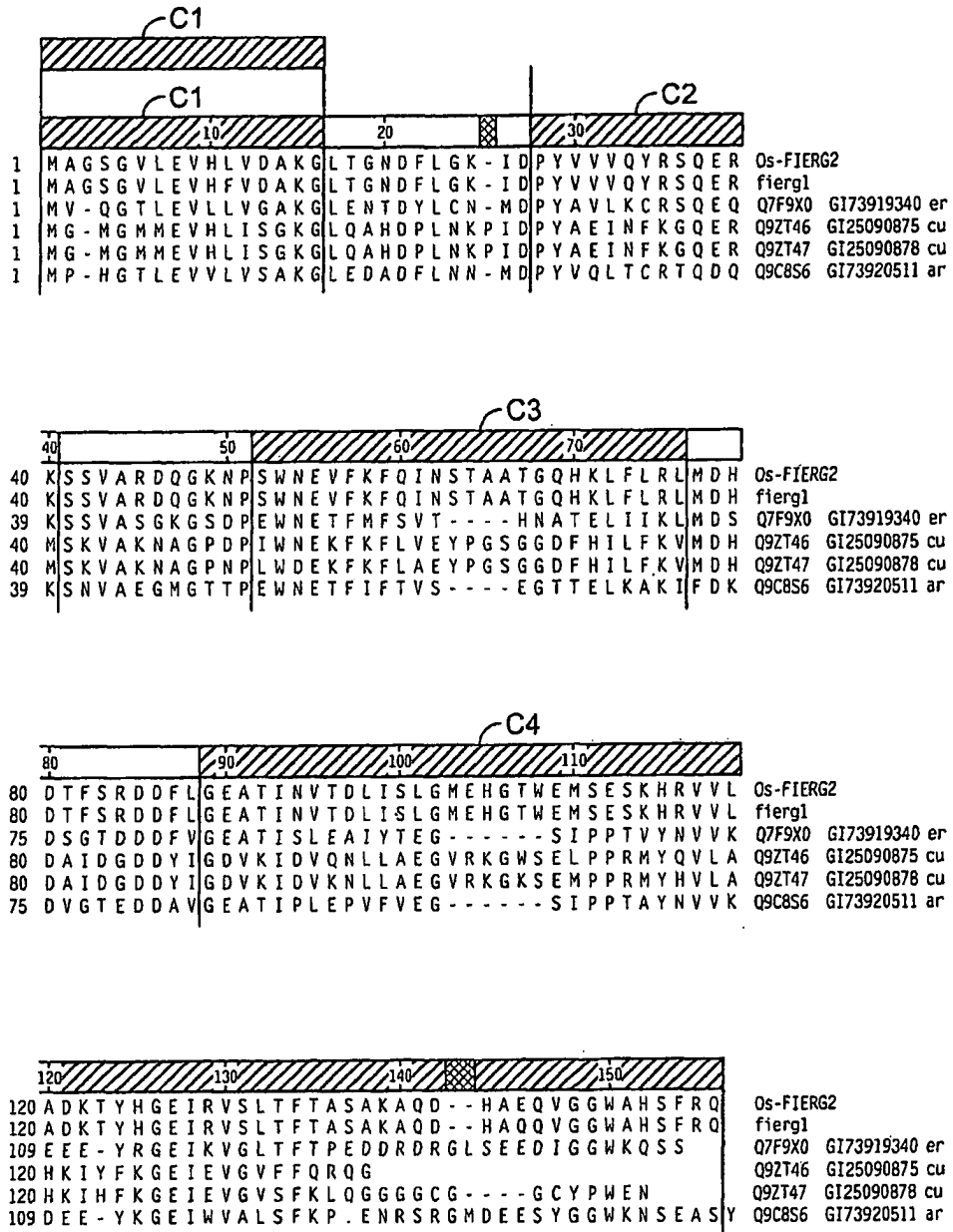


FIG. 5

RESUMO

Patente de Invenção: **"POLIPEPTÍDEOS DE LIGAÇÃO QUIMÉRICOS DE PLANTA PARA RECONHECIMENTO MOLECULAR UNIVERSAL"**.

5 A presente invenção refere-se a bibliotecas de ácido nucléicos que codificam polipeptídeos de ligação quiméricos baseados em seqüências de polipeptídeo de estrutura de planta. Também descritos são métodos para geração das bibliotecas.