



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

(11) PI 0809352-0 B1

(22) Data do Depósito: 26/03/2008

(45) Data de Concessão: 28/03/2017



(54) Título: PROTEÍNA INSETICIDA HÍBRIDA PROJETADA, MOLÉCULA DE ÁCIDO NUCLEICO CODIFICADORA DA DITA PROTEÍNA, CASSETE DE EXPRESSÃO, VETOR RECOMBINANTE, CÉLULA HOSPEDEIRA TRANSGÊNICA, COMPOSIÇÃO INSETICIDA, MÉTODOS DE PRODUÇÃO DA DITA PROTEÍNA E DE UMA PLANTA TRANSGÊNICA RESISTENTE A INSETOS E MÉTODO DE CONTROLE DE INSETOS".

(51) Int.Cl.: C07K 14/325; C12P 21/00

(30) Prioridade Unionista: 28/03/2007 US 60/920,493

(73) Titular(es): SYNGENTA PARTICIPATIONS AG

(72) Inventor(es): HOPE HART; JENG S. CHENG; CHERYL STACY; FREDERICK WALTERS

"PROTEÍNA INSETICIDA HÍBRIDA PROJETADA, MOLÉCULA DE ÁCIDO NUCLEICO CODIFICADORA DA DITA PROTEÍNA, CASSETE DE EXPRESSÃO, VETOR RECOMBINANTE, CÉLULA HOSPEDEIRA TRANSGÊNICA, COMPOSIÇÃO INSETICIDA, MÉTODOS DE PRODUÇÃO DA DITA PROTEÍNA E DE UMA PLANTA TRANSGÊNICA RESISTENTE A INSETOS E MÉTODO DE CONTROLE DE INSETOS"

PRÁTICA

A presente invenção refere-se a campos de engenharia de proteínas, biologia molecular de plantas e controle de pragas. Mais particularmente a invenção refere-se às recentes proteínas de engenharia híbrida com atividade inseticida, ácidos nucleicos cuja expressão resulta nas proteínas inseticidas, e métodos para fazer e métodos para usar as proteínas inseticidas e ácidos nucleicos correspondentes para controlar insetos.

As pragas de insetos são a maior causa das perdas de safras. Só nos EUA bilhões de dólares são perdidos a cada ano em função da infestação de vários gêneros de insetos. Além das perdas nos campos de safra, as pragas de insetos também são uma carga para plantadores de vegetais e frutas, produtores de flores ornamentais e são um incômodo para os jardineiros e proprietários de residências.

As espécies de larva do milho são as pragas de milho com maior poder de destruição. Nos EUA, as três espécies importantes são *Diabrotica virgifera virgifera*, a larva do oeste, *D. longicornis barberi*, larva do norte e *D. undecimpunctata howardi*, larva do sul. Apenas as larvas do oeste e do norte são consideradas como pragas primárias de milho na Faixa do Milho nos EUA. Uma larva importante no sul dos EUA é a larva mexicana, *Diabrotica virgifera zeae*. As larvas causam os danos mais substanciais nas plantas ao alimentar-se quase que exclusivamente das raízes do milho. Esta lesão foi comprovada de aumentar o alojamento da planta, reduzir o rendimento de grãos e rendimento vegetal assim como alterar o conteúdo e nutrientes do grão. A alimentação as larvas também causa efeitos indiretos no milho ao abrir avenidas nas raízes para infecções de bactérias e fungos que resultam em doenças de apodrecimento da raiz e do talo. As larvas adultas ficam ati-

vas nos milharais no final do verão quando se alimentam das espigas de milho, sedas e pólen, interferindo assim com a polinização normal.

As larvas do milho são controladas geralmente com aplicações intensivas de pesticidas químicos que são ativos pela inibição do crescimento de insetos, prevenção da alimentação ou reprodução de insetos, ou causar a morte. Um controle eficaz da larva do milho pode ser obtido, porém estes produtos químicos podem ocasionalmente afetar também outros organismos benéficos. Outro problema que resulta do amplo uso de pesticidas químicos é o aparecimento de variedades resistentes de insetos. Outro problema ainda se deve ao fato que a larva pode alimentar-se embaixo do solo dificultando assim a aplicação de tratamentos de resgate de inseticidas. Sendo assim, a maioria das aplicações de inseticida são realizadas de maneira profilática quando do plantio. Esta prática resulta em ampla carga ambiental. Isto foi parcialmente aliviado por várias práticas de gestão de agricultura, porém existe uma crescente necessidade de mecanismos alternativos de controle de pestes.

Agentes biológicos de controle de pestes, tais como *Bacillus thuringiensis* (Bt) grupo que expressa toxinas pesticidas como δ -endotoxinas (delta-endotoxinas; também denominadas toxinas de cristal ou proteínas Cry), também foram aplicados em plantas de safra com resultados satisfatórios contra pestes primárias de insetos lepidópteros. As δ -endotoxinas são proteínas contidas dentro de uma matriz cristalina que são conhecidas de possuir atividade inseticida quando ingeridas por certos insetos. As várias δ -endotoxinas foram classificadas com base em seu espectro de atividade e homologia de sequência. Antes de 1990, as classes maiores eram definidas por seu espectro de atividade com as proteínas Cry1 ativas contra Lepidópteros (traças e borboletas), proteínas Cry2 ativas contra Lepidópteros e Dípteros (moscas e mosquitos), proteínas Cry3 ativas contra Coleópteros (besouros) e proteínas Cry4 ativas contra Dípteros (Hofte & Whitely, 1989, Microbiol. Rev. 53:242-255). Uma nova nomenclatura foi desenvolvida em 1998 a qual sistematicamente classifica as proteínas Cry com base em homologia de sequência de aminoácidos ao invés de especificidades-alvo de

insetos (Crickmore *et al.* 1998, Microbiol. Molec. Biol. Rev. 62:807-813).

O espectro de atividade inseticida de uma δ -endotoxinas individual de Bt é bastante estrito, com dada δ -endotoxinas sendo ativa apenas contra algumas espécies dentro de uma Ordem. Por exemplo, uma toxina
5 Cry3A é conhecida de ser bastante tóxica para o besouro de batata do Colorado, *Leptinotarsa decemlineata*, mas de ter muito pouco efeito ou nenhuma toxicidade com besouros relacionados no gênero *Diabrotica* (Johnson *et al.*, 1993, J. Econ. Entomol. 86:330-333). De acordo com Slaney *et al.* (1992, Insect Biochem. Molec. Biol. 22:9-18) uma toxina Cry3A é pelo menos 2000
10 vezes menos tóxica para a larva do sul que para o besouro da batata do Colorado. Também é sabido que a Cry3A possui pouca ou nenhuma toxicidade para a larva do oeste ou do norte.

A especificidade das δ -endotoxinas é o resultado da eficiência dos vários passos envolvidos na produção de uma proteína tóxica ativa e a
15 sua subsequente interação com as células epiteliais no intestino médio dos insetos. Para ser inseticida, a maioria das δ -endotoxinas conhecidas devem ser primeiramente ingeridas pelo inseto e ativadas proteoliticamente para formar uma toxina ativa. A ativação das proteínas de cristal inseticida (Cry) é um processo com vários passos. Após a ingestão, os cristais devem ser pri-
20 meiramente solubilizados no intestino do inseto. Uma vez solubilizadas as δ -endotoxinas são ativadas por segmentações proteolíticas específicas. As proteases no intestino do inseto podem desempenhar um papel na especificidade ao determinar onde a δ -endotoxinas é processada. Assim que a δ -endotoxinas é solubilizada e processada, ela se junta com receptores espe-
25 cíficos na superfície do epitélio do intestino médio do inseto e subsequente-mente integra-se na dupla camada de lipídeo da membrana da borda em escova. Canais de íon formam-se interrompendo a função do intestino médio causando assim a morte do inseto.

Nos lepidópteros, que possuem intestino com pH alcalino, as
30 proteases do intestino processam δ -endotoxinas, por exemplo, Cry1Aa, Cry1Ab, Cry1Ac, Cry1B e Cry1F, de protoxinas 130-140 kDa para proteínas tóxicas de aproximadamente 60-70 kDa. O processamento da protoxina em

toxina foi relatado de proceder por meio da remoção de ambos os aminoácidos terminais N- e C- com a localização exata de processamento dependendo da δ -endotoxinas específica e os fluídos específicos de intestino de inseto envolvidos (Ogiwara *et al.*, 1992, J. Invert. Pathol. 60:121-126). Assim a ativação requer que toda a região da cauda protoxina C-terminal- seja transpassada. Esta ativação proteolítica de uma δ -endotoxinas pode ter um papel significativo na determinação da sua especificidade.

Os insetos coleópteros possuem intestinos que são mais neutros a acidíferos e as δ -endotoxinas específicas de coleópteros são similares ao tamanho das toxinas ativadas específicas de lepidópteros. Portanto, o processamento das δ -endotoxinas específicas de coleópteros era previamente considerado desnecessário para toxicidade. Não obstante, dados sugerem que as δ -endotoxinas ativas de coleópteros são solubilizadas e proteolizadas para polipeptídios tóxicos menores. Uma proteína 73 kDa Cry3A δ -endotoxinas produzida por *B. thuringiensis* var. *tenebrionis* é na verdade processada pela bactéria no término N-, perdendo 49-57 resíduos durante ou após a formação de cristais para produzir a forma 67 kDa comumente isolado (Carroll *et al.*, 1989, Biochem. J. 261:99-105). McPherson *et al.*, (1988, Biotechnology 6:61-66) também demonstrou que uma sequência de código nativa *cry3A* contém dois códons funcionais translacionais de iniciação no mesmo quadro de leitura, uma codificação para proteína 73 kDa e a outra codificação para uma proteína 67 kDa começando no Met-1 e Met-48 respectivamente, da sequência de aminoácido deduzida. Ambas as proteínas podem ser consideradas como proteínas Cry3A de extensão completa ocorrendo naturalmente.

Conforme mais conhecimentos são obtidos sobre a função das δ -endotoxinas, as tentativas de engenharia das δ -endotoxinas para ter novas atividades têm aumentado. A engenharia das δ -endotoxinas tornou-se mais possível ao resolver a estrutura tridimensional de uma Cry3A em 1991 (Li *et al.*, 1991, Nature 353:815-821). Li *et al.* Determinou que uma proteína Cry3A possui três domínios estruturais: O domínio I do N-terminal-, de resíduos 58-290, consiste em 7 α -hélices, domínio II, de resíduos 291-500, contém três

folhas- β em uma chamada formação chave grega, e o domínio III de C-terminal, de resíduos 501-644, é um sanduíche β -em conformação denominada rocambole. A estrutura tridimensional da toxina ativa Cry1Aa de lepidóptero também foi resolvida (Grochulski *et al.*, 1995, J. Mol. Biol. 254:447-464). A toxina Cry1Aa possui também três domínios: O domínio I N-terminal, de resíduos 33-253, domínio II de resíduos 265-461, e domínio III de resíduos 463-609 com fita adicional externo é uma das folhas- β de resíduos 254-264. Se as estruturas Cry3A e Cry1Aa são projetadas em outras sequências Cry1, o domínio I corre aproximadamente de resíduo de aminoácido de 28 a 260, o domínio II de aproximadamente 260 a 460 e o domínio III de aproximadamente 460 a 600. Ver, Nakamura *et al.*, Agric. Biol. Chem. 54(3): 715-724 (1990); Li *et al.*, Nature 353: 815-821 (1991); Ge *et al.*, J. Biol. Chem. 266(27): 17954-17958 (1991); e Honee *et al.*, Mol. Microbiol. 5(11): 2799-2806 (1991); cada um dos quais aqui incorporados por referência. Assim, agora se sabe que com base na homologia de sequência de aminoácidos, as δ -endotoxinas Bt possuem uma estrutura tridimensional similar compreendendo os três domínios.

As porções de toxina das proteínas Cry Bt são também caracterizadas por ter cinco blocos conservados atravessando sua sequência de aminoácidos numerados CB1 a CB5 de terminais N a terminais C, respectivamente (Hofte & Whiteley, *supra*). O bloco conservado 1 (CB1) compõe aproximadamente 29 aminoácidos, o bloco conservado 2 (CB2) compõe aproximadamente 67 aminoácidos, o bloco conservado 3 (CB3) compõe aproximadamente 48 aminoácidos, o bloco conservado 4 (CB4) compõe aproximadamente 10 aminoácidos e o bloco conservado 5 (CB5) compõe aproximadamente 12 aminoácidos. As sequências antes e depois destes cinco blocos conservados são altamente variáveis e, portanto designadas com as "regiões variáveis" V1-V6. O domínio I de uma δ -endotoxinas Bt tipicamente compõe a região variável 1, bloco conservado 1, região variável 2, e os 52 aminoácidos N-terminal do bloco conservado 2. O domínio II tipicamente compõe aproximadamente os 15 aminoácidos C-terminal do bloco conservado 2, região variável 3, e aproximadamente os 10 aminoácidos do

N-terminal do bloco conservado 3. O domínio III tipicamente compõe aproximadamente os 38 aminoácidos C-terminal do bloco conservado 3, região variável 4, o bloco conservado 4, região variável 5, e o bloco conservado 5. As toxinas ativas Cry1 lepidópteras, entre outras delta-toxinas, possuem região variável 6 com aproximadamente 1-3 aminoácidos dentro do domínio III.

Muitas fitas Bt e δ -endotoxinas são ativas contra diferentes espécies de insetos e nematódeos. Não obstante, relativamente poucas destas fitas e toxinas possuem atividade contra insetos coleópteros. Adicionalmente, a maioria das agora conhecidas δ -endotoxinas nativas ativas de coleópteros, por exemplo, Cry3A, Cry3B, Cry3C, Cry7A, Cry8A, Cry8B, e Cry8C, possuem toxicidade insuficiente oral contra a larva do milho para fornecer controle de campo adequado caso aplicadas, por exemplo, por meio de micróbios ou plantas transgênicas. Assim sendo, outras abordagens para a produção de toxinas recentes ativas contra a larva do milho precisam ser exploradas.

As δ -endotoxinas lepidópteras ativas passaram por engenharia na tentativa de melhorar a atividade específica ou para ampliar o espectro da atividade inseticida. Por exemplo, o domínio de especificidade do bicho-da-seda (*Bombyx mori*) de proteína Cry1Aa foi movida para uma proteína Cry1Ac, impondo assim nova atividade inseticida na proteína híbrida Bt resultante (Ge *et al.* 1989, PNAS 86: 4037-4041). Também, Bosch *et al.* 1998 (Patente US 5,736,131, incorporada aqui por referência) descreve toxinas híbridas de *Bacillus thuringiensis* compreendendo no seu C-terminal domínio III de uma proteína de primeiro Cry e seus terminais N domínios I e II de uma segunda proteína Cry. Tais toxinas híbridas mostraram possuir especificidade inseticida alterada contra insetos lepidópteros. Por exemplo, a toxina híbrida H04, que também é descrita em De Maagd *et al.*, *Appl. Environ. Microbiol.* 62(5): 1537-1543 (1996) compõe os seus domínios I e II de N-terminal de Cry1Ab e seu domínio III de C-terminal de Cry1C. H04 é relatado como sendo altamente tóxico ao inseto lepidóptero *Spodoptera exigua* comparado com a toxina Cry1Ab paterna e significativamente mais tóxico do que a toxina Cry1C paterna. Também foi demonstrado que a substituição do domínio

III das toxinas, que não são ativas contra o *Spodoptera exigua* tais como Cry1E e Cry1Ab, pelo domínio III de Cry1C, que é ativo contra o *Spodoptera exigua*, pode produzir toxinas híbridas que são ativas contra referido inseto. Todos os híbridos descritos em Bosch *et al.* utilizam domínios de proteínas ativas Cry lepidópteras para fazer novas toxinas com atividade lepidóptera. Os resultados sugerem que o domínio III de Cry1C é um importante determinante da especificidade para o *Spodoptera exigua*. Ver também, Bosch *et al.*, *FEMS Microbiology Letters* 118: 129-134 (1994); Bosch *et al.*, *Bio/Technology* 12: 915-918 (1994); De Maagd *et al.*, *Appl. Environ. Microbiol.* 62(8): 2753-2757 (1996); e De Maagd *et al.*, *Mol. Microbiol.* 31(2): 463-471 (1999); cada um dos quais aqui incorporados por referência.

Várias tentativas de engenharia de δ -endotoxinas ativas coleópteras foram relatadas. Chen e Stacy (Patente US 7,030,295, aqui incorporada por referência) criaram com sucesso uma toxina ativa de larva do milho ao inserir um sítio de reconhecimento de protease de ocorrência não-natural no domínio I, domínio III, ou ambos os domínios I e III de uma proteína Cry3A. Uma das proteínas Cry3A modificadas resultantes, designada Cry3A055, com sítio de reconhecimento de protease inserido no domínio I, foi ativa contra várias espécies de *Diabrotica*. Van Rie *et al.*, 1997, (Patente US Nº 5.659.123) realizou engenharia de Cry3A ao repor aleatoriamente aminoácidos, acreditado de ser importante na acessibilidade de solvente, no domínio II com o aminoácido alanino. Várias destas substituições aleatórias confinadas ao domínio II foram reportadamente envolvidas na toxicidade incrementada da larva do milho do oeste. Não obstante, outras mostraram que algumas substituições de alanino no domínio II de Cry3A resultam na interrupção da junção de receptor ou instabilidade estrutural (Wu e Dean, 1996, *J. Mol. Biol.* 255: 628-640). English *et al.*, 1999, (Pedido de Patente Internacional Publicado Nº WO 99/31248) relatou substituições de aminoácidos em Cry3Bb que causaram aumentos na toxicidade para as larvas do milho do sul e do oeste. Porém, dos 35 mutantes de Cry3Bb relatados, apenas três, com mutações primariamente no domínio II e na interface domínio I-domínio II, foram ativos contra a larva do milho do oeste. Adicionalmente, a

variação em toxicidade do tipo selvagem de Cry3Bb contra a larva do milho do oeste nos mesmos estudos aparenta ser maior contra quaisquer das diferenças entre as toxinas Cry3Bb com mutação e Cry3Bb tipo selvagem. Shadenkov *et al.* (1993, Mol. Biol. 27:586-591), fez uma proteína híbrida ao fundir aminoácidos 48-565 de uma proteína Cry3A em aminoácidos 526-725 de uma proteína Cry1Aa. Assim sendo, o cruzamento entre a sequência Cry3A e Cry1Aa ocorreu no bloco conservado 4 localizado no domínio III. A Cry3A é muito ativa contra o besouro da batata do Colorado (*Leptinotarsa decemlineata*). Não obstante, a proteína híbrida descrita por Shadenkov *et al.* não foi ativa contra o besouro da batata do Colorado apesar de que mais de 75% da proteína híbrida ser composta por uma sequência Cry3A. Assim, a adição de apenas 25% de uma sequência Cry1Aa destruiu a atividade contra inseto coleóptero que a Cry3A paterna era ativa contra. Isto sugere que proteínas híbridas feitas com fusão de porções de proteína Cry ativa coleóptera, por exemplo Cry3A, e uma proteína ativa lepidóptera, por exemplo Cry1A, não teriam atividade contra insetos coleópteros, particularmente um inseto coleóptero que normalmente não é suscetível a Cry3A como a larva do milho.

Em vista da discussão acima, resta a necessidade de desenhar novos e eficazes agentes de controle de pestes que forneçam um benefício econômico para fazendeiros e que seja ambientalmente aceitável. São de particular necessidade as toxinas que são tóxicas para espécies de *Diabrotica*, uma peste principal do milho, que tenham diferente modo de ação do que os produtos de controle de pestes existentes como forma de mitigar o desenvolvimento da resistência. Adicionalmente, as entregas de agentes de controle em produtos que minimizam a carga ao meio ambiente, como por plantas transgênicas, são desejáveis.

SUMÁRIO

Em vista destas necessidades, é um dos objetivos da presente invenção fornecer proteínas híbridas inseticidas recentes projetadas (eHIPs). Referidas eHIPs recentes são desenvolvidas ao fundir combinações únicas de regiões variáveis e blocos conservados de pelo menos duas proteínas

Cry diferentes e opcionalmente incluir uma região de cauda de protoxina de uma proteína Cry Bt no C-terminal ou fragmento terminal peptidila N-terminal ou ambos. Por exemplo, sem-limitação, ao combinar regiões variáveis completas ou parciais e blocos conservados de uma primeira proteína Cry que possui atividade coleóptera com regiões variáveis, completas ou parciais, e blocos conservados de uma segunda proteína Cry com atividade lepidóptera, diferente da primeira proteína Cry, e opcionalmente incluindo uma região de cauda de protoxina de proteína ativa lepidóptera Cry Bt, ou fragmento de peptidila N-terminal ou ambos, novas proteínas híbridas inseticidas projetadas são criadas possuindo atividade contra um espectro de insetos diferente do que a primeira ou segunda proteína Cry paterna ou de ambas. Referidas eHIPs recentes podem compor regiões variáveis completas ou parciais, blocos conservados ou domínios de uma proteína Cry3A modificada e uma proteína Cry diferente da proteína Cry3A modificada. O fragmento de peptidila pode conferir atividade inseticida ao eHIP, ou pode aumentar a atividade inseticida do eHIP sobre um eHIP sem fragmento de peptidila, ou tornar o eHIP mais estável do que um eHIP sem o fragmento de peptidila. Os eHIPs da invenção possuem surpreendente e inesperada toxicidade contra a larva do milho (*Diabrotica sp.*). A invenção é ainda mais puxada para ácidos nucleicos que codificam o eHIPs ou que é complementar a um que se torna híbrido sob condições estritas com os ácidos nucleicos híbridos recombinantes de acordo com a invenção.

Também incluídos na invenção estão vetores contendo referidos aminoácidos recombinantes (ou complementares a estes); uma planta ou micro-organismo que inclui e permite a expressão destes ácidos nucleicos; plantas transformadas com estes ácidos nucleicos, por exemplo, plantas de milho transgênico; a progênie destas plantas que contém os ácidos nucleicos estavelmente incorporados e hereditários de modo Mendelian, e/ou as sementes destas plantas e progênie.

A invenção também inclui composições e formulações contendo os eHIPs, que são capazes de inibir a habilidade das pestes de insetos de sobreviver, crescer e reproduzir, ou de limitar danos ou perdas a plantas de

safras relacionadas com insetos, por exemplo aplicando os eHIPs ou composições ou formulações em áreas infestadas com insetos, ou tratar de modo profilático áreas ou plantas suscetíveis a insetos para oferecer proteção contra as pestes de insetos.

5 A invenção e adicionalmente puxada para um método de fazer os eHIPs e os métodos de usar os ácidos nucleicos, por exemplo em microorganismos para controlar insetos ou em plantas transgênicas para oferecer proteção contra danos causados por insetos.

Os eHIPs recentes aqui descritos são altamente ativos contra os insetos. Por exemplo, os eHIPs da presente invenção podem ser utilizados para controlar pestes de insetos de importância econômica tais como a larva do milho do oeste (*Diabrotica virgifera virgifera*) a larva do milho do norte (*D. longicornis barberi*) e a larva do milho mexicana (*D. virgifera zea*). Determinados eHIPs podem também ser utilizados para controlar a broca do milho da Europa (*Ostrinia nubilalis*) e outros insetos lepidópteros. Os eHIPs podem ser utilizados de modo individual ou em combinação com outras estratégias de controle de pestes de insetos para oferecer maximização na eficiência de controle de pestes com impacto ambiental minimizado.

Outros aspectos e vantagens da presente invenção tornar-se-ão aparentes para aqueles versados na técnica de estudar a seguinte descrição da invenção e exemplos sem-limitação.

BREVE DESCRIÇÃO DOS DESENHOS

Os seguintes desenhos formam parte da presente especificação e são incluídos para adicionalmente demonstrar determinados aspectos da presente invenção. A invenção pode ser melhor entendida por meio de referência a um ou mais destes desenhos em combinação com a descrição detalhada de incorporações específicas aqui apresentadas.

A figura 1A-1E mostra um alinhamento de sequência de algumas incorporações de eHIP com proteínas Cry paternas ou Cry 3A modificado utilizadas para construir os eHIPs, incluindo, um Cry3A, Cry1Ab, e Cry3A055, e indica identidade de porcentagem. Os fragmentos de peptídila N-terminal estão sublinhados. Os 5 blocos conservados estão etiquetados

CB1-CB5. A localização das junções entre os domínios I, II e III, está designada por uma linha vertical pontilhada. Uma sequência de reconhecimento de protease de catepsina G, AAPF, está em negrito.

A figura 2A-2E ilustra um alinhamento de incorporações de eHIP
 5 que são ativas contra pelo menos a larva do milho do oeste e indicam identidade de porcentagem comparada com o eHIP 8AF. Os fragmentos de peptídila N-terminal estão sublinhados. As regiões de cauda de protoxinas C-terminal estão duplamente sublinhadas. Os 5 blocos conservados estão etiquetados CB1-CB5. As localizações das junções entre os domínios I, II e III
 10 estão indicadas pelo símbolo " " e etiquetadas de acordo. As localizações das posições dos cruzamentos estão indicadas pelo símbolo " ". Uma sequência de reconhecimento de protease de catepsina G, AAPF, está em negrito.

A figura 3 mostra um mapa do vetor recombinante 12207, utilizado para transformar milho compreendendo um cassette de expressão
 15 com promotor ubiquitina de milho operável relacionado com uma sequência de codificação *FRCG* operável relacionada a um terminador NOS.

A figura 4 mostra um mapa do vetor recombinante 12161, utilizado para transformar milho compreendendo um cassette de expressão
 20 com promotor ubiquitina de milho operável relacionado com uma sequência de codificação *FR8a* operável relacionada a um terminador NOS.

A figura 5 mostra um mapa do vetor recombinante 12208, utilizado para transformar milho compreendendo um cassette de expressão
 25 com um vírus amarelo de cestrum enrolador de folhas promotor (cmp) operável relacionado com uma sequência de codificação *FRCG* operável relacionada a um terminador NOS.

A figura 6 mostra um mapa do vetor recombinante 12274, utilizado para transformar milho compreendendo um cassette de expressão
 com um vírus amarelo de "Cestrum" enrolador de folhas promotor (cmp)
 30 operável relacionado com uma sequência de codificação *FR8a* operável relacionada a um terminador NOS.

A figura 7 mostra um mapa do vetor recombinante 12473, utili-

zado para transformar milho compreendendo umum cassete de expressão com promotor ubiquitina (ubi) de milho operável relacionado com uma sequência de codificação *FRD3* operável relacionada a um terminador NOS.

5 A figura 8 mostra um mapa do vetor recombinante 12474, utilizado para transformar milho compreendendo umum cassete de expressão com um vírus amarelo de cestrum enrolador de folhas promotor (cmp) operável relacionado com uma sequência de codificação *FRD3* operável relacionada a um terminador NOS.

BREVE DESCRIÇÃO DAS SEQUÊNCIAS NA LISTAGEM DE SEQUÊNCIAS

10 CIAS

SEQ ID NO: 1 é a sequência de nucleotídeo *2OL-8a*.

SEQ ID NO: 2 é o *2OL-8a* codificado pelo SEQ ID NO: 1.

SEQ ID NO: 3 é a sequência de nucleotídeo *FR8a*.

SEQ ID NO: 4 é o *FR8a* codificado pelo SEQ ID NO: 3.

15 SEQ ID NO: 5 é a sequência de nucleotídeo *FRCG*.

SEQ ID NO: 6 é o *FRCG* codificado pelo SEQ ID NO: 5.

SEQ ID NO: 7 é a sequência de nucleotídeo *FR8a-9F*.

SEQ ID NO: 8 é o *FR8a-9F* codificado pelo SEQ ID NO: 7.

SEQ ID NO: 9 é a sequência de nucleotídeo *FR-9F-catg*.

20 SEQ ID NO: 10 é a cat. *FR-9F* codificada pelo SEQ ID NO: 9.

SEQ ID NO: 11 é a sequência de nucleotídeo *FR8a-12AA*.

SEQ ID NO: 12 é o *FR8a-12AA* codificado pelo SEQ ID NO: 11.

SEQ ID NO: 13 é a sequência de nucleotídeo *WR-9mut*.

SEQ ID NO: 14 é o *WR-9mut* codificado pelo SEQ ID NO: 13.

25 SEQ ID NO: 15 é a sequência de nucleotídeo *FRD3*.

SEQ ID NO: 16 é o *FRD3* codificado pelo SEQ ID NO: 15.

SEQ ID NO: 17 é a sequência de nucleotídeo *FR-12-cg-dm3*.

SEQ ID NO: 18 é a *FR-12-cg-dm3* codificada pelo SEQ ID NO: 17.

SEQ ID NO: 19 é a sequência de nucleotídeo *9F-cg-de16*.

30 SEQ ID NO: 20 é a *9F-cg-de16* codificada pelo SEQ ID NO: 19.

SEQ ID NO: 21 é a sequência de nucleotídeo *FR-cg-dm3*.

SEQ ID NO: 22 é a *FR-cg-dm3* codificada pelo SEQ ID NO: 21.

- SEQ ID NO: 23 é a sequência de nucleotídeo *9F-cg-dm3*.
 SEQ ID NO: 24 é a *9F-cg-dm3* codificada pelo SEQ ID NO: 23.
 SEQ ID NO: 25 é a sequência de nucleotídeo *B8a*.
 SEQ ID NO: 26 é o *B8a* codificado pelo SEQ ID NO: 25.
- 5 SEQ ID NO: 27 é a sequência de nucleotídeo *5*B8a*.
 SEQ ID NO: 28 é a *5*B8a* codificado pelo SEQ ID NO: 27.
 SEQ ID NO: 29 é a sequência de nucleotídeo *V3A*.
 SEQ ID NO: 30 é o *V3A* codificado pelo SEQ ID NO: 29.
 SEQ ID NO: 31 é a sequência de nucleotídeo *V4F*.
- 10 SEQ ID NO: 32 é o *V4F* codificado pelo SEQ ID NO: 31.
 SEQ ID NO: 33 é a sequência de nucleotídeo *5*V4F*.
 SEQ ID NO: 34 é a *5*V4F* codificado pelo SEQ ID NO: 33.
 SEQ ID NO: 35 é a sequência de nucleotídeo *2OL-7*.
 SEQ ID NO: 36 é o *2OL-7* codificado pelo SEQ ID NO: 35.
- 15 SEQ ID NO: 37 é a sequência de nucleotídeo *T7-2OL-7*.
 SEQ ID NO: 38 é o *T7-2OL-7* codificado pelo SEQ ID NO: 37.
 SEQ ID NO: 39 é a sequência de nucleotídeo *5*2OL-7*.
 SEQ ID NO: 40 é o *5*2OL-7* codificado pelo SEQ ID NO: 39.
 SEQ ID NO: 41 é a sequência de nucleotídeo *2OL-10*.
- 20 SEQ ID NO: 42 é o *2OL-10* codificado pelo SEQ ID NO: 41.
 SEQ ID NO: 43 é a sequência de nucleotídeo *5*2OL-10*.
 SEQ ID NO: 44 é o *5*2OL-10* codificado pelo SEQ ID NO: 43.
 SEQ ID NO: 45 é a sequência de nucleotídeo *2OL-12A*.
 SEQ ID NO: 46 é o *2OL-12A* codificado pelo SEQ ID NO: 45.
- 25 SEQ ID NO: 47 é a sequência de nucleotídeo *2OL-13*.
 SEQ ID NO: 48 é o *2OL-13* codificado pelo SEQ ID NO: 47.
 SEQ ID NO: 49 é a sequência de nucleotídeo *V5&6*.
 SEQ ID NO: 50 é o *V5&6* codificado pelo SEQ ID NO: 49.
 SEQ ID NO: 51 é a sequência de nucleotídeo *5*V5&6*.
- 30 SEQ ID NO: 52 é a *5*V5&6* codificado pelo SEQ ID NO: 51.
 SEQ ID NO: 53 é a sequência de nucleotídeo *88A-dm3*.
 SEQ ID NO: 54 é o *88A-dm3* codificado pelo SEQ ID NO: 53.

- SEQ ID NO: 55 é a sequência de nucleotídeo FR(1Fa).
SEQ ID NO: 56 é o FR(1Fa) codificado pelo SEQ ID NO: 55.
SEQ ID NO: 57 é a sequência de nucleotídeo FR(1Ac).
SEQ ID NO: 58 é o FR(1Ac) codificado pelo SEQ ID NO: 57.
- 5 SEQ ID NO: 59 é a sequência de nucleotídeo FR(1Ia).
SEQ ID NO: 60 é o FR(1Ia) codificado pelo SEQ ID NO: 59.
SEQ ID NO: 61 é a sequência de nucleotídeo DM23A.
SEQ ID NO: 62 é o DM23A codificado pelo SEQ ID NO: 61.
SEQ ID NO: 63 é a sequência de nucleotídeo 8AF.
- 10 SEQ ID NO: 64 é o 8AF codificado pelo SEQ ID NO: 63.
SEQ ID NO: 65 é a sequência de nucleotídeo 5*cry3A055.
SEQ ID NO: 66 é a 5*Cry3A055 codificado pelo SEQ ID NO: 65.
SEQ ID NO: 67 é uma sequência de nucleotídeo cry3A otimizada de milho.
SEQ ID NO: 68 é o Cry3A codificado pelo SEQ ID NO: 67.
- 15 SEQ ID NO: 69 é a sequência de nucleotídeo cry3A055.
SEQ ID NO: 70 é o Cry3A055 codificado pelo SEQ ID NO: 69.
SEQ ID NO: 71 é uma sequência de nucleotídeo cry1Ab otimizada de milho.
SEQ ID NO: 72 é o Cry1Ab codificado pelo SEQ ID NO: 71.
SEQ ID NO: 73 é uma sequência de nucleotídeo cry1Ba otimizada de milho.
- 20 SEQ ID NO: 74 é o Cry1Ba codificado pelo SEQ ID NO: 73.
SEQ ID NO: 75 é uma sequência de nucleotídeo cry1Fa otimizada de milho.
SEQ ID NO: 76 é o Cry1Fa codificado pelo SEQ ID NO: 75.
SEQ ID NO: 77 é uma sequência de nucleotídeo cry8Aa.
SEQ ID NO: 78 é o Cry8Aa codificado pelo SEQ ID NO: 77.
- 25 SEQ ID NO: 79 é uma sequência de nucleotídeo cry1Ac.
SEQ ID NO: 80 é o Cry1Ac codificado pelo SEQ ID NO: 79.
SEQ ID NO: 81 é uma sequência de nucleotídeo cry1Ia.
SEQ ID NO: 82 é o Cry1Ia codificado pelo SEQ ID NO: 81.
SEQ ID NOs 83-125 são sequências iniciador úteis na presente invenção.
- 30 SEQ ID NOs 126-134 são fragmentos peptídica N-terminal.
SEQ ID NO: 135 é uma proteína Cry3A de extensão completa.
SEQ ID NO: 136-143 são sequências iniciador úteis na presente invenção.

SEQ ID NO: 144 é o código de sequência *T7-8AF*.

SEQ ID NO: 145 é o *T7-8AF* codificado pelo SEQ ID NO: 144.

SEQ ID NO: 146 é o código de sequência *-catG8AF*.

SEQ ID NO: 147 é o *-CatG8AF* codificado pelo SEQ ID NO: 146.

5 SEQ ID NO: 148 é o código de sequência *8AFdm3*.

SEQ ID NO: 149 é o *8AFdm3* codificado pelo SEQ ID NO: 148.

SEQ ID NO: 150 é o código de sequência *8AFlongdm3*.

SEQ ID NO: 151 é o *8AFlongdm3* codificado pelo SEQ ID NO: 150.

SEQ ID NO: 152 é o código de sequência *cap8AFdm3*.

10 SEQ ID NO: 153 é o *cap8AFdm3* codificado pelo SEQ ID NO: 152.

SEQ ID NO: 154 é o código de sequência *8AFdm3T*.

SEQ ID NO: 155 é o *8AFdm3T* codificado pelo SEQ ID NO: 154.

SEQ ID NO: 156 é o código de sequência *8AFlongdm3T*.

SEQ ID NO: 157 é o *8AFlongdm3T* codificado pelo SEQ ID NO: 156.

15 SEQ ID NO: 158 é o código de sequência *cap8AFdm3T*.

SEQ ID NO: 159 é o *cap8AFdm3T* codificado pelo SEQ ID NO: 158.

SEQ ID NO: 160 é o FR8a+34 eHIP.

DEFINIÇÕES

20 Para clareza, alguns termos utilizados na especificação são definidos e apresentados como segue:

"Atividade" dos eHIPs da invenção significa que a função do eHIPs como agentes ativos de controle de insetos oralmente, possuem efeito tóxico, ou são capazes de interromper ou deter a alimentação de insetos, o que pode ou não causar a morte do inseto. Quando um eHIP da invenção
25 é colocado no inseto, o resultado é tipicamente a morte do inseto, ou o inseto não se alimenta da fonte que torna o eHIP disponível ao inseto.

"Associado com/relacionado operacionalmente" refere-se a dois ácidos nucleicos que estão física ou funcionalmente relacionados. Por exemplo, uma sequência de DNA promotora ou reguladora é relatada de
30 estar "associada com" uma sequência de DNA que codifica para RNA ou uma proteína caso as duas sequências sejam relacionadas operacionalmente, ou situada de forma que a sequência reguladora de DNS afetará o nível

de expressão da sequência de codificação ou estrutural de DNA.

No contexto da presente invenção, uma "proteína inseticida química" (CIP) é uma proteína inseticida que compõe um fragmento de peptidila fundido no N-terminal de um eHIP. O fragmento de peptidila pode conferir atividade inseticida ao eHIP, pode aumentar a atividade inseticida do eHIP sobre um eHIP sem fragmento de peptidila, ou tornar o eHIP mais estável do que um eHIP sem o fragmento de peptidila, particularmente contra a larva do milho do oeste. O fragmento de peptidila é uma sequência de aminoácido que é tipicamente heteróloga para (não-derivada de) uma proteína Cry Bt, mas pode ser derivada de uma proteína Cry Bt. Referidos fragmentos peptidila estendem-se dos terminais N da proteína inseticida e não correm de forma natural no N-terminal das proteínas Cry Bt. Um exemplo de um fragmento de peptidila N-terminal possui a sequência de aminoácido MTSNGR-QCAGIRP (Nº ID SEQ: 129) que não é derivada de uma proteína Cry Bt.

Uma "sequência de codificação" é uma sequência de ácido nucleico que é transcrita em um RNA tal como um mRNA, rRNA, tRNA, snRNA, RNA senso ou RNA antissenso. Preferivelmente o RNA é traduzido em um organismo para produzir uma proteína.

No contexto da presente invenção, "conectar" ácidos nucleicos significa juntar dois ou mais ácidos nucleicos utilizando quaisquer meios conhecidos na técnica. Por exemplo, sem-limitação, os ácidos nucleicos podem ser ligados utilizando, por exemplo, ligase de DNA, ou pode ser recozido utilizando PCR. Os ácidos nucleicos podem também ser juntados ao sintetizar quimicamente um ácido nucleico utilizando a sequência de dois ou mais ácidos nucleicos separados.

Para "Controlar" insetos significa inibir, por meio de efeito tóxico, a habilidade das pestes de insetos de sobreviver, crescer, alimentar-se, e/ou se reproduzir ou de limitar os danos nas plantas de safra relacionados a insetos. "Controlar" insetos pode ou não significar matar os insetos, apesar de preferencialmente significar matar os insetos.

No contexto da presente invenção, "correspondente a" significa que quando as sequências de aminoácido de certas proteínas (por exemplo,

proteínas Cry Bt ou proteínas Cry3A modificadas) são alinhadas uma com a outra, os aminoácidos que alinham com certas posições enumeradas em, por exemplo, mas não-limitado a, uma toxina Cry3A (seja SEQ ID NO: 68 ou N° ID SEQ: 134); uma toxina Cry3A055 (SEQ ID NO: 70); ou uma toxina Cry1Ab (N° ID SEQ: 72), mas que não estão necessariamente nestas posições numéricas exatas relativas a sequência de aminoácidos referência, particularmente conforme se relaciona com a identificação dos domínios I, II e III, e/ou os blocos conservados e regiões variáveis, estas posições de aminoácidos "correspondem" uma a outra. Por exemplo, na delineação do domínio I de uma proteína híbrida, aminoácidos 11-244 de uma proteína Cry3A055 (SEQ ID NO: 70) correspondem a aminoácidos 58-290 de uma toxina nativa Cry3A (SEQ ID NO: 135) ou a aminoácidos 11-243 de uma toxina nativa Cry3A (SEQ ID NO: 68) ou a aminoácidos 33-254 de uma toxina nativa Cry1Ab.

No contexto da presente invenção as palavras "proteína Cry" podem ser utilizadas intercambiadamente com as palavras "delta-endotoxina" ou "δ -endotoxina."

No contexto da presente invenção, uma "proteína híbrida inseticida projetada" (eHIP) é uma proteína inseticida criada ao fundir combinações únicas de regiões variáveis e blocos conservados de pelo menos duas proteínas Cry diferentes. Referidas eHIPs recentes podem compor regiões variáveis completas ou parciais, blocos conservados ou domínios de uma proteína Cry3A modificada e uma proteína Cry diferente da proteína Cry3A modificada. Os eHIPs da invenção podem opcionalmente incluir uma região de cauda de proteína Cry Bt ou um fragmento de peptidila N-terminal ou ambos. Por exemplo, sem-limitação, um eHIP é criado ao combinar um N-terminal para direção C-terminal, aminoácidos 1-468 de uma proteína Cry3A055 (SEQ ID NO: 70), que compõe a região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e o N-terminal 24 aminoácidos do bloco conservado 3, e aminoácidos 477-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compõe o C-terminal 24 aminoácidos do bloco conservado 3, região variável 4, bloco conservado 4, região

variável 5, bloco conservado 5 e região variável 6, e uma região de 38 aminoácidos de uma cauda de protoxina Cry1Ab. Um eHIP que compõe o fragmento de peptídila N-terminal pode também ser designado como uma "proteína quimérica inseticida (CIP)."

5 Para "Entregar" um eHIP significa que o eHIP entra em contato com um inseto, resultando em efeito tóxico e controle do referido inseto. O eHIP pode ser entregue de muitas maneiras reconhecidas, por exemplo, por meio de planta transgênica expressando o eHIP, composição(ões) de proteína(s) formulada(s), composição(ões) de proteína(s) borrifável(is), uma matriz isca, ou qualquer outro sistema de entrega de toxina reconhecido.

10 "Quantidade eficaz de controle de inseto" significa que a concentração de um eHIP que inibe por meio de efeito tóxico a habilidade dos insetos de sobreviver, crescer, alimentar-se e/ou se reproduzir ou de limitar os danos ou perdas nas plantas de safra relacionados a insetos. "Quantidade

15 eficaz de controle de inseto" pode ou não significar matar os insetos, apesar de preferencialmente significar matar os insetos.

 "Cassete de expressão" conforme utilizado aqui significa uma sequência de ácido nucleico capaz de direcionar expressão de uma sequência de nucleotídeo em particular em uma célula hospedeira apropriada, compreendendo um promotor operacionalmente relacionado com a sequência de nucleotídeo de interesse que é operacionalmente relacionada a os sinais de terminação. Também tipicamente compõe sequências requeridas para a adequada translação da sequência de nucleotídeo. O cassete de expressão que compõe a sequência de nucleotídeo de interesse pode ter pelo menos

20 um dos seus componentes heterólogos com relação a pelo menos um dos seus outros componentes. O cassete de expressão pode adicionalmente ser de ocorrência natural porém tem sido obtido em forma recombinante útil para expressão heteróloga. Tipicamente, não obstante, o cassete de expressão é heterólogo em relação ao hospedeiro, isto é, a sequência particular de ácido

25 nucleico do cassete de expressão não ocorre naturalmente em uma célula hospedeira e deve ser introduzida na célula hospedeira ou um antecessor da

30 célula hospedeira por meio de evento de transformação. A expressão de

uma sequência de nucleotídeo no cassete de expressão pode estar sob controle de um promotor constitutivo ou de um promotor de indução que inicia a transcrição apenas quando a célula hospedeira é exposta a alguns estímulos externos em particular. No caso de organismo multicelular, tal como uma
5 planta, o promotor pode também ser específico para um tecido em particular, um órgão, ou estágio de desenvolvimento.

Um "gene" é uma região definida que está localizada dentro de um genoma e que, além da acima referida codificação de sequência de ácido nucleico, compõe outros ácidos nucleicos primariamente regulatórios responsáveis pelo controle da expressão, o que significa a transcrição e transla-
10 ção, da porção de codificação. Um gene pode também compor outras sequências 5' e 3' não-traduzidas e sequências de terminação. Elementos adicionais que podem estar presentes podem, por exemplo, ser íntrons. A sequência regulatória de ácido nucleico do gene pode normalmente não ser
15 operacionalmente relacionada a a sequência de ácido nucleico associada conforme encontrado na natureza e portanto seria um gene quimérico.

"Gene de interesse" refere-se a qualquer gene que, quando transferido para uma planta, confere à referida planta a característica desejada tal como uma resistência antibiótica, resistência contra vírus, resistência
20 contra inseto, resistência contra doenças, ou resistência a outras pestes, tolerância herbicida, valor nutricional melhorado, desempenho melhorado em processo industrial ou capacidade reprodutiva alterada. O "gene de interesse" pode ser também um que seja transferido para plantas para a produção de enzimas de valor comercial ou metabólitos na planta.

25 Uma sequência de ácido nucleico "heteróloga" é uma sequência de ácido nucleico não-naturalmente associada com a célula hospedeira na qual é introduzida, incluindo cópias múltiplas de ocorrência não-natural de uma sequência de ácido nucleico de ocorrência natural. Uma sequência heteróloga de aminoácido é aquela que não está naturalmente associada com
30 uma sequência nativa de aminoácido, por exemplo, uma sequência de aminoácido de uma proteína Cry.

Uma sequência de ácido nucleico "homóloga" é uma sequência

de ácido nucleico naturalmente associada com uma célula hospedeira na qual é introduzida.

"Recombinação homóloga" é a troca recíproca de fragmentos de ácido nucleico entre moléculas de ácido nucleico homóloga.

5 "Identidade" ou "identidade de porcentagem" refere-se ao grau de similaridade entre duas sequências de ácido nucleico ou proteína. Para comparação de sequência, tipicamente uma sequência age como sequência de referência com a qual as sequências de teste são comparadas. Ao utilizar um algoritmo de comparação de sequência, as sequências de teste e de re-
10 ferência são introduzidas em um computador, coordenadas de subsequência são designadas caso necessário, e parâmetros de programa de algoritmo de sequência são designados. O algoritmo de comparação de sequência então calcula a identidade de porcentagem de sequência para a(s) sequência(s) de teste(s) relativas à sequência de referência, com base nos parâmetros de
15 programa designados.

 O alinhamento ótimo de sequências para comparação pode ser conduzido, exemplo, pelo algoritmo local de homologia de Smith & Waterman, *Adv. Appl. Math.* 2: 482 (1981), pelo algoritmo de alinhamento de homologia de Needleman & Wunsch, *J. Mol. Biol.* 48: 443 (1970), pelo método
20 de busca de similaridade de Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), por implementações computadorizadas destes algoritmos (GAP, BESTFIT, FASTA, e TFASTA no Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), ou por inspeção visual (ver, Ausubel *et al.*, *infra*).

25 Um exemplo de um algoritmo que é adequado para determinar identidade de porcentagem de sequência e similaridade de sequência é o algoritmo BLAST, que é descrito em Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990). O software para realização de análises BLAST está disponível publicamente por meio do National Center for Biotechnology Information (Centro
30 Nacional de Informação de Biotecnologia) (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo envolve primeiro a identificação de pares de sequência de alta pontuação (HSPs) ao identificar palavras curtas de extensão W na se-

quência de consulta, que combinam ou satisfazem determinada pontuação T de limite com valor positivo quando alinhado com a palavra de mesma extensão em uma sequência de base de dados. T é referido como o limite de pontuação de palavra vizinha (Altschul *et al.*, 1990). Estes indícios iniciais de

5 palavra vizinha agem como sementes para iniciar as buscas para encontrar HSP's mais longos contendo estes. Os indícios de palavra são então estendidos em ambas as direções ao longo de cada sequência até que a pontuação de alinhamento cumulativo possa ser aumentada. As pontuações cumulativas são calculadas utilizando, para as sequências de nucleotídeo, os pa-

10 râmetros M (pontuação recompensa para um par de resíduos que combinam; sempre > 0) e N (pontuação de penalidade para resíduos que não combinam; sempre < 0). Para as sequências de aminoácidos, uma matriz de pontuação pode ser utilizada pra calcular a pontuação cumulativa. A extensão dos indícios de palavra em cada direção é parada quando a pontuação de alinhamento cumulativa cair para a quantidade de X a partir do valor máximo alcançado, a pontuação cumulativa vai a zero ou abaixo em função da

15 acumulação de um ou mais alinhamentos de resíduo de pontuação negativa, ou o final de qualquer uma das sequências seja alcançado. Os parâmetros de algoritmo BLAST W, T, e X determinam a sensibilidade e a velocidade do alinhamento. O programa BLASTN (para sequências de nucleotídeo) utiliza como padrões uma extensão de palavra (W) de 11, uma expectativa (E) de 10, um corte de 100, $M=5$, $N=-4$, e uma comparação de ambos os fitas. Para

20 sequências de aminoácidos, o programa BLASTP utiliza como padrões uma extensão de palavra (W) de 3, uma expectativa (E) de 10, e a matriz de pontuação BLOSUM62 (ver Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89: 10915 (1989)).

25

Adicionalmente para calcular identidade de porcentagem de sequência, o algoritmo BLAST também realiza análise estatística da similaridade entre duas sequências (ver, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)). Uma medida de similaridade fornecida pelo

30 algoritmo BLAST é a probabilidade de menor soma ($P(N)$), que fornece uma indicação da probabilidade pela qual uma combinação entre duas sequên-

cias de aminoácido ou nucleotídeo poderiam ocorrer por acaso. Por exemplo, uma sequência de teste de ácido nucleico é considerada similar a uma sequência de referência caso a probabilidade de menor soma em comparação com a sequência de teste de ácido nucleico com a sequência de ácido nucleico de referência seja menos de aproximadamente 0,1, mais preferencialmente menos de aproximadamente 0,01, e mais preferivelmente menos de aproximadamente 0,001.

Outro programa de computador amplamente utilizado e aceito para a realização de alinhamento de sequências é o CLUSTALW v1.6 (Thompson, *et al.* Nuc. Acids Res., 22: 4673-4680, 1994). O número de bases que combinam ou aminoácidos é dividido pelo número total de bases ou de aminoácidos e multiplicado por 100 para obter uma identidade de porcentagem. Por exemplo, caso duas sequências de par de base de 580 tivessem 145 bases combinadas, seriam 25 por cento idênticas. Caso as duas sequências comparadas sejam de extensões diferentes, o número de combinações é dividido pela mais curta das duas extensões. Por exemplo, caso existissem 100 aminoácidos combinados entre 200 e 400 proteínas de aminoácido, estas são 50 por cento idênticas em relação à sequência mais curta. Caso a sequência mais curta seja menos de 150 bases ou 50 aminoácidos em extensão, o número de combinações é dividido por 150 (para bases de ácido nucleico) ou 50 (para aminoácidos), e multiplicadas por 100 para obter uma identidade de porcentagem.

Outra indicação que os dois ácidos nucleicos são substancialmente idênticos é que as duas moléculas hibridizam uma com a outra sob condições estritas. A frase "hibridizando especificamente a" refere-se a juntar, duplicar, ou hibridizar uma molécula apenas a uma sequência particular de nucleotídeo sob condições estritas quando essa sequência está presente em uma mistura complexa (exemplo, celular total) DNA ou RNA. "Junta substancialmente" refere-se a hibridização complementar entre um ácido nucleico sonda e um ácido nucleico-alvo e engloba combinações mal-sucedidas menores que podem ser acomodadas ao reduzir a severidade do meio de hibridização para obter a detecção desejada da sequência de ácido

nucleico-alvo desejada.

"Condições estritas de hibridização" e "condições estritas de lavagem de hibridização" no contexto dos experimentos de hibridização de ácido nucleico tais como hibridizações do sul e do norte são dependentes de sequências e são diferentes sob diferentes parâmetros ambientais. As sequências mais longas hibridizam especificamente em temperaturas mais elevadas. Um guia extensivo sobre a hibridização de ácidos nucleicos pode ser encontrado em Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes* part I chapter 2 "Visão Geral da Hibridização e da estratégia para estudos de sonda de ácido nucleico" Elsevier, New York. Geralmente, condições mais estritas de hibridização e de lavagem de hibridização são selecionadas para serem aproximadamente 5°C mais baixos que o ponto térmico de derretimento (T_m) para a sequência específica na potência iônica definida e pH. Tipicamente, sob "condições estritas" uma sonda hibridiza na sua subsequência-alvo, mas não a outras sequências.

T_m é a temperatura (sob potência iônica definida e pH) na qual 50% da sequência-alvo hibridiza a uma sonda perfeitamente combinada. Condições muito estritas são selecionadas para serem iguais a T_m para uma sonda em particular. Um exemplo de condições estritas de hibridização de ácidos nucleicos complementares que possuem mais de 100 resíduos complementares em um filtro em um borrão do sul ou do norte é 50% formamida com 1mg de heparina a 42°C, com a hibridização sendo realizada em pernoite. Um exemplo de condições muito estritas de lavagem é 0,1 5M NaCl a 72°C por aproximadamente 15 minutos. Um exemplo de condições estritas de lavagem é lavagem 0,2x SSC a 65°C por 15 minutos (ver, Sambrook, *in-fra*, para obter descrição do SSC tampão). Com frequência, uma lavagem de alta restrição é precedida de uma lavagem de baixa restrição para remover o sinal de sonda de fundo. Um exemplo de lavagem de restrição média para um duplex de, por exemplo, mais de 100 nucleotídeos, é 1x SSC a 45°C por 15 minutos. Um exemplo de lavagem de restrição baixa para um duplex de, por exemplo, mais de 100 nucleotídeos, é 4-6x SSC a 45°C por 15 minutos.

Para sondas curtas (exemplo, aproximadamente 10 a 50 nucleotídeos), as condições estritas geralmente envolvem concentrações de sal de menos que aproximadamente 1,0 M Na⁺ íon, tipicamente aproximadamente concentração de 0,01 a 1,0 M Na⁺ íon (ou outros sais) em pH 7,0 a 8,3, e a temperatura é tipicamente pelo menos aproximadamente 30°C. Condições estritas também podem ser obtidas com a adição de agentes desestabilizadores tais como formaldeído. Em geral, uma taxa de sinal para ruído de 2x (ou maior) que aquela observada para sonda não-relacionada no estudo particular de hibridização indica detecção de uma hibridização específica. Os ácidos nucleicos que não hibridizam uns aos outros sob condições estritas ainda são substancialmente idênticos caso as proteínas que codificam sejam substancialmente idênticas. Isto ocorre, por exemplo, quando uma cópia de ácido nucleico é criada utilizando a degenerância máxima de códon permitida pelo código genético.

A seguir estão exemplos de conjuntos de condições de hibridização/lavagem que podem ser utilizados para clonar as sequências homólogas de nucleotídeo que são substancialmente idênticas às sequências de referência de nucleotídeos da presente invenção: uma sequência referência de nucleotídeo preferivelmente hibridiza para sequência referência de nucleotídeo em 7% sulfato dodecila de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 2X SSC, 0,1% SDS a 50°C, mais desejavelmente em 7% sulfato dodecila de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem a 1X SSC, 0,1% SDS a 50°C, mais desejavelmente ainda em 7% sulfato dodecila de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,5X SSC, 0,1% SDS a 50°C, preferivelmente em 7% sulfato dodecila de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1X SSC, 0,1% SDS a 50°C, mais preferivelmente em 7% sulfato dodecila de sódio (SDS), 0,5 M NaPO₄, 1 mM EDTA a 50°C com lavagem em 0,1X SSC, 0,1% SDS a 65°C.

Uma indicação adicional que dois ácidos nucleicos ou proteínas são substancialmente idênticos é que a proteína codificada pelo primeiro ácido nucleico é imunologicamente de reatividade cruzada com, ou especifi-

camente se junta com, a proteína codificada pelo segundo ácido nucleico. Assim, uma proteína é tipicamente substancialmente idêntica à segunda proteína, por exemplo, onde as duas proteínas diferem apenas por substituições conservadoras.

5 "Inseticida" é definido como uma atividade biológica tóxica capaz de controlar insetos, preferivelmente por meio da morte destes.

Uma sequência de ácido nucleico está "isocodificando com" uma sequência de ácido nucleico referencia quando a sequência de ácido nucleico codifica um polipeptídio tendo a mesma sequência de aminoácido que o polipeptídio codificado pela sequência de ácido nucleico de referência.

10 Uma molécula "isolada" de ácido nucleico ou uma toxina isolada é uma molécula de ácido nucleico ou toxina que pela mão do homem, existe aparte do seu ambiente nativo e, portanto, não é um produto da natureza. Uma molécula isolada de ácido nucleico ou toxina pode existir em forma purificada ou pode existir em um ambiente não-nativo tal como, por exemplo, sem-limitação, uma célula microbial recombinante, célula de planta, tecido de planta ou planta.

Uma "toxina Cry3A modificada" ou "Cry3A055" desta invenção refere-se a uma toxina derivada de Cry3A tendo pelo menos um sítio adicional de reconhecimento de protease que é reconhecido por uma protease de intestino do inseto-alvo, que não ocorre naturalmente em uma toxina Cry3A, como descrito na Patente US 7,030,295, aqui incorporada por referência.

25 Uma "sequência de codificação *cry3A* modificada" de acordo com esta invenção pode ser derivada de uma sequência de codificação *cry3A* nativa ou de uma sequência de codificação *cry3A* sintética e compõe a sequência de codificação de pelo menos um sítio adicional de reconhecimento de protease que não ocorre de forma natural em um gene *cry3A* não modificado.

30 Uma "molécula de ácido nucleico" ou "sequência de ácido nucleico" é um segmento de um DNA ou RNA de fita único ou duplo que pode ser isolado de qualquer fonte. No contexto da presente invenção, a molécula de ácido nucleico é tipicamente um segmento de DNA.

Uma "planta" é qualquer planta em qualquer estágio de desenvolvimento, particularmente uma planta de semente.

Uma "célula de planta" é uma unidade estrutural e fisiológica de uma planta, compreendendo uma protoplasta e uma parede de célula. A célula de planta pode estar na forma de uma célula única isolada ou uma célula de cultura, ou como parte de uma unidade mais organizada como por exemplo, tecido de planta, um órgão de planta, ou uma planta inteira.

"Cultura de células de planta" significa culturas de unidades de planta tais como, por exemplo, protoplastas, células de cultura de células, células em tecido de plantas, pólen, tubos de pólen, óvulos, sacos de embrião, zigotos e embriões em vários estágios de desenvolvimento.

"Material de planta" refere-se a folhas, talos, raízes, flores ou partes de flores, frutas, pólen, ovos de células, zigotos, sementes, cortes, culturas de células ou tecido, ou qualquer outra parte ou produto de uma planta.

Um "órgão de planta" é uma parte distinta e visivelmente estruturada e diferenciada de uma planta tal como a raiz, talo, folha, botão de flor, ou embrião.

"Tecido de planta" conforme aqui utilizando refere-se a um grupo de células de planta organizado em uma unidade estrutural e funcional. Qualquer tecido de planta *in planta* ou em cultura é incluído. Este termo inclui, mas não está limitado a, plantas inteiras, órgãos de planta, sementes de planta, cultura de tecido e quaisquer grupos de células de planta organizados em unidades estruturais e/ou funcionais. O uso deste termo em conjunto com, ou na ausência de, qualquer tipo específico de tecido de planta conforme acima listado ou de outra forma englobado por esta definição não visa ser exclusiva de qualquer outro tipo de tecido de planta.

Um "promotor" é uma sequência não-traduzida de DNA a montante da região de codificação que contém o sítio de junção para polimerase RNA e inicia a transcrição do DNA. A região de promotor pode também incluir outros elementos que agem como reguladores de expressão de gene.

"Elementos regulatórios" referem-se às sequências envolvidas

no controle da expressão de uma sequência de nucleotídeos. Os elementos regulatórios compõem um promotor operacionalmente relacionado com a sequência de nucleotídeos de interesse e os sinais de terminação. Também tipicamente englobam sequências requeridas para a adequada translação da
5 sequência de nucleotídeo.

A "transformação" é um processo para introduzir ácido nucleico heterólogo em uma célula ou organismo hospedeiro. Em particular, "transformação" significa uma integração estável da molécula de DNA em um genoma de um organismo de interesse.

10 "Transformado/transgênico/recombinante" referem-se a um organismo hospedeiro tal como uma bactéria ou uma planta na qual uma molécula heteróloga de ácido nucleico foi introduzida. A molécula de ácido nucleico pode ser integrada de maneira estável no genoma do hospedeiro ou a molécula de ácido nucleico também pode estar presente como molécula de
15 extracromossoma. Tal molécula de extracromossoma pode ser autorreplícante. Células transformadas, tecidos, ou plantas são entendidos de englobar não apenas o produto final do processo de transformação, mas também a prole deste. Um hospedeiro "não-transformado", "não-transgênico", ou "não- recombinante" refere-se a um tipo de organismo selvagem, exemplo,
20 uma bactéria ou planta, que não contém a molécula heteróloga de ácido nucleico.

Os nucleotídeos são indicados por suas bases pelas seguintes abreviações padrão: adenina (A), citosina (C), timina (T), e guanina (G). Os aminoácidos são da mesma forma indicados pelas seguintes abreviações:
25 alanina (Ala; A), arginina (Arg; R), asparagina (Asn; N), ácido aspártico (Asp; D), cisteína (Cys; C), glutamina (Gln; Q), ácido glutâmico (Glu; E), glicina (Gly; G), histidina (His; H), isoleucina (Ile; I), leucina (Leu; L), lisina (Lys; K), metionina (Met; M), fenilalanina (Phe; F), prolina (Pro; P), serina (Ser; S), treonina (Thr; T), triptofan (Trp; W), tirosina (Tyr; Y), e valina (Val; V).

30 DESCRIÇÃO

Esta invenção refere-se a recentes proteínas inseticidas híbridas projetadas (eHIPS), criadas para ter atividade contra pelo menos a larva do

milho do oeste, e pode adicionalmente incluir a larva do milho do norte, larva do milho mexicana e/ou o besouro da batata do Colorado. Alguns eHIPs possuem atividade contra a peste lepidóptera, broca do milho da Europa. Referidas eHIPs recentes são desenvolvidas ao fundir combinações únicas de regiões variáveis e blocos, completos ou parciais, conservados de pelo menos duas proteínas Cry diferentes e opcionalmente incluem uma região de cauda de protoxina de uma proteína Cry Bt no C-terminal ou fragmento terminal peptidil N-terminal ou ambos.[0001] Por exemplo, sem-limitação, ao combinar regiões variáveis completas ou parciais com blocos conservados de uma primeira proteína Cry que possui atividade coleóptera com regiões variáveis, completas ou parciais, com blocos conservados de uma segunda proteína Cry com atividade lepidóptera, diferente da primeira proteína Bt e opcionalmente incluindo uma região de cauda de protoxina de proteína Cry Bt no C-terminal, ou um fragmento de peptidila N-terminal ou ambos, novas proteínas híbridas inseticidas projetadas são criadas possuindo atividade contra um espectro de insetos diferente do que a primeira ou segunda proteína Cry paterna ou de ambas é criado. Referidas eHIPs recentes podem também compor regiões variáveis completas ou parciais, blocos conservados ou domínios de uma proteína Cry3A modificada e uma proteína Cry diferente da proteína Cry3A modificada. O fragmento de peptidila N-terminal ou região de cauda de protoxina pode conferir atividade inseticida ao eHIP, pode aumentar a atividade inseticida do eHIP sobre um eHIP sem fragmento de peptidila ou região de cauda de protoxina, e/ou tornar o eHIP mais estável do que um eHIP sem o fragmento de peptidila ou região de cauda da protoxina, particularmente pelo menos contra a larva do milho do oeste. A sequência de aminoácido do fragmento de peptidila é tipicamente heteróloga a (isto é, não-derivada de) uma proteína Cry Bt. Não obstante, com base nos ensinamentos aqui descritos, uma pessoa com habilidade reconhecerá que o fragmento de peptidila N-terminal pode ser gerado utilizando uma sequência de aminoácido derivada de uma proteína Cry Bt. Os eHIPs da invenção possuem uma surpreendente e inesperada toxicidade para a larva do milho, particularmente para a larva do milho do oeste e a lar-

va do milho mexicana. A presente invenção também refere-se a ácidos nucleicos cuja expressão resulta em eHIPs, e na realização e uso de eHIPs para controlar pestes de insetos. A expressão de ácidos nucleicos resulta em eHIPs que podem ser utilizados para controlar insetos coleópteros tais como a larva do milho do oeste, do norte ou a lavra do milho mexicana, ou usada para controlar insetos lepidópteros tais como a broca do milho da Europa particularmente quando expressa em planta transgênica tal como uma planta transgênica de milho.

Em uma modalidade, a invenção engloba uma proteína inseticida híbrida projetada compreendendo uma sequência de aminoácido de uma primeira proteína (Bt) Cry *Bacillus thuringiensis* compreendendo regiões variáveis completas ou parciais e blocos conservados da primeira proteína Cry fundida a uma sequência de aminoácido a partir de uma segunda proteína Cry Bt diferente da primeira proteína Cry Bt compreendendo regiões variáveis parciais ou completas e blocos conservados da segunda proteína Cry, e opcionalmente compreendendo: (a) uma região de cauda de protoxina de uma proteína Cry Bt localizada no C-terminal ou (b) um fragmento de peptidila N-terminal; ou ambos (a) e (b), em que o eHIP possui atividade contra pelo menos a larva do milho do oeste.

Em outra modalidade, a presente invenção engloba um eHIP compreendendo uma região N-terminal de uma primeira proteína Cry Bt fundida a uma região C-terminal de uma segunda proteína Cry Bt diferente da primeira proteína Cry Bt, em que pelo menos a posição de cruzamento entre a primeira e a segunda proteína Cry Bt está localizada no bloco conservado 2, bloco conservado 3, região variável 4 ou bloco conservado 4, e opcionalmente compreendendo: (a) uma região de cauda de protoxina de uma proteína Cry Bt localizada no C-terminal; ou (b) um fragmento de peptidila N-terminal; ou ambos (a) e (b), em que o eHIP possui atividade inseticida contra pelo menos a larva do milho do oeste.

Em ainda outra modalidade, um eHIP de acordo com a invenção compõe a região variável 1 do N-terminal ao C-terminal ou a porção C-terminal da região variável 1, bloco conservado 1, região variável 2, bloco

conservado 2, região variável 3, e uma porção N-terminal do bloco conservado 3 de uma primeira proteína Cry Bt fundida em uma porção de C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6 de uma segunda proteína Cry Bt.

Em outra modalidade, uma eHIP da invenção compõe pelo menos duas posições de cruzamento entre uma sequência de aminoácidos da primeira proteína Cry Bt e uma sequência de aminoácidos da segunda proteína Cry Bt. Em uma modalidade, uma primeira posição de cruzamento está localizada no bloco conservado 2 uma segunda posição de cruzamento está localizada no bloco conservado 3. Em outra modalidade, a primeira junção de cruzamento está localizada no bloco conservado 3 e a segunda posição de cruzamento está localizada no bloco conservado 4.

Em outra modalidade, um eHIP da invenção compõe C-terminal de região de cauda de protoxina de uma proteína Cry Bt. A região de cauda de protoxina pode conferir atividade inseticida ao eHIP, significando que sem a região de cauda de protoxina o eHIP não seria ativo, pode aumentar a atividade do eHIP sobre um eHIP sem a região de cauda de protoxina, ou pode tornar o eHIP mais estável do que um eHIP sem a região de cauda de protoxina. Em uma modalidade, a região de cauda de protoxina é de uma proteína Cry Bt ativa lepidóptera. Em outra modalidade, a região de cauda de protoxina é de uma proteína Cry1A. Em ainda outra modalidade, a região de cauda de protoxina é de uma proteína Cry1Aa ou de uma proteína Cry1Ab. A região de cauda de protoxina da invenção pode compor uma cauda inteira de protoxina de uma proteína Cry Bt ou qualquer fragmento desta. Em um aspecto desta modalidade, a região de cauda de protoxina de um eHIP compõe pelo menos 38 aminoácidos do N-terminal de uma cauda de protoxina de uma proteína Cry1Ab. Em outro aspecto desta modalidade, a região de cauda de protoxina compõe uma sequência de aminoácido correspondente a aminoácidos 611-648 de SEQ ID NO: 72. Ainda em outro aspecto desta modalidade, a região de cauda de protoxina compõe aminoácidos 611-648 de SEQ ID NO: 72.

Ainda em outra modalidade, um eHIP compõe um fragmento de peptidila N-terminal. O fragmento de peptidila N-terminal pode conferir atividade inseticida ao eHIP, significando que sem o fragmento de peptidila N-terminal a proteína não possui atividade inseticida, ou o fragmento de peptidila N-terminal pode aumentar a atividade inseticida do eHIP sobre um eHIP sem o fragmento de peptidila N-terminal, ou o fragmento de peptidila N-terminal pode tornar o eHIP mais estável do que um eHIP sem o fragmento de peptidila N-terminal. Em um aspecto desta modalidade, o fragmento de peptidila compõe uma sequência de aminoácido que é heteróloga a (isto é, não-derivada de) uma proteína Cry Bt. Em outro aspecto desta modalidade, o fragmento de peptidila N-C-terminal compõe pelo menos 9 aminoácidos. Em ainda outro aspecto desta modalidade, o fragmento de peptidila compõe pelo menos 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 70, 80, 90 ou 100 aminoácidos. Em outro aspecto desta modalidade, o fragmento de peptidila compõe além de 100 aminoácidos. Em ainda outro aspecto desta modalidade, o fragmento de peptidila N-C-terminal compõe a sequência de aminoácido YDGR-QQHRG (SEQ ID NO: 133) ou TSNGRQCAGIRP (SEQ ID NO: 134). Em ainda outro aspecto desta modalidade, o fragmento de peptidila N-C-terminal compõe uma sequência de aminoácido selecionada a partir do grupo consistindo no SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, e SEQ ID NO: 132.

Em ainda outra modalidade, as regiões variáveis e os blocos conservados de uma primeira proteína Cry ativa contra insetos coleópteros são utilizadas para fazer o eHIP da invenção em combinação com regiões variáveis e blocos conservados de uma segunda proteína Cry ativa contra um inseto lepidóptero. As proteínas Cry ativas coleópteras incluem, mas não estão limitadas a Cry3, Cry7, Cry8, e Cry34/Cry35. As proteínas Cry ativas lepidópteras incluem, mas não estão limitadas a Cry1 e Cry9. Em um aspecto desta modalidade, a primeira proteína Cry é uma Cry3A e a segunda proteína Cry é Cry1A. Em outro aspecto, a proteína Cry3A pode ser substituída

por uma proteína Cry3A modificada, por exemplo, sem-limitação, a proteína Cry3A055 descrita na patente US 5,659,123, aqui incorporada por referência. Em ainda outro aspecto desta modalidade, a proteína Cry3A é uma Cry3Aa e a proteína Cry1A é uma Cry1Aa ou uma Cry1Ab. Em outro aspecto desta modalidade, a Cry3Aa é selecionada do seguinte grupo e possui o Número de Acesso no GenBank indicado: Cry3Aa1 (M22472), Cry3Aa2 (J02978), Cry3Aa3 (Y00420), Cry3Aa4 (M30503), Cry3Aa5 (M37207), Cry3Aa6 (U10985), Cry3Aa7 (AJ237900), Cry3Aa8 (AAS79487), Cry3Aa9 (AAW05659), Cry3Aa10 (AAU29411), e Cry3Aa11 (AY882576). Em outro aspecto desta modalidade, a Cry1Aa é selecionada do seguinte grupo e possui o Número de Acesso no GenBank indicado: Cry1Aa1 (M11250), Cry1Aa2 (M10917), Cry1Aa3 (D00348), Cry1Aa4 (X13535), Cry1Aa5 (D17518), Cry1Aa6 (U43605), Cry1Aa7 (AF081790), Cry1Aa8 (I26149), Cry1Aa9 (AB026261), Cry1Aa10 (AF154676), Cry1Aa11 (Y09663), Cry1Aa12 (AF384211), Cry1Aa13 (AF510713), Cry1Aa14 (AY197341), e Cry1Aa15 (DQ062690). Em ainda outro aspecto desta modalidade, a Cry1Ab é selecionada do seguinte grupo e possui o Número de Acesso no GenBank indicado: Cry1Ab1 (M13898), Cry1Ab2 (M12661), Cry1Ab3 (M15271), Cry1Ab4 (D00117), Cry1Ab5 (X04698), Cry1Ab6 (M37263), Cry1Ab7 (X13233), Cry1Ab8 (M16463), Cry1Ab9 (X54939), Cry1Ab10 (A29125), Cry1Ab11 (I12419), Cry1Ab12 (AF059670), Cry1Ab13 (AF254640), Cry1Ab14 (U94191), Cry1Ab15 (AF358861), Cry1Ab16 (AF37560), Cry1Ab17 (AAT46415), Cry1Ab18 (AAQ88259), Cry1Ab19 (AY847289), Cry1Ab20 (DQ241675), Cry1Ab21 (EF683163), e Cry1Ab22 (ABW87320).

Em ainda mais um aspecto desta modalidade, a primeira proteína Cry compõe uma sequência de aminoácido selecionada a partir do grupo consistindo no SEQ ID NO: 68, SEQ ID NO: 70, e SEQ ID NO: 135, e a segunda proteína Cry compõe uma sequência de aminoácido listada no SEQ ID NO: 72.

Em uma modalidade, a presente invenção engloba um eHIP da invenção que compõe pelo menos um cruzamento entre a região N-terminal da primeira proteína Cry e a região C-terminal de uma segunda proteína Cry localizada no bloco conservado 3, região variável 4, ou bloco conservado 4.

Em um aspecto desta modalidade, a posição de cruzamento no bloco conservado 3 está localizada imediatamente em seguida de um aminoácido correspondente a Ser451, Phe454, ou Leu468 de SEQ ID NO: 70. Em outro aspecto desta modalidade, a posição de cruzamento está localizada no bloco conservado 3 logo em seguida de Ser451, Phe454, ou Leu468 de ID SEQ: 70 ou Ser450, Phe453, ou Leu467 do N°de ID de SEQ: 68; ou Ser497, Phe100, Leu114 do SEQ ID NO: 135. As posições de cruzamento em determinadas incorporações de Cry3A/Cry1Ab eHIP ou incorporações modificadas de Cry3A/Cry1Ab eHIP de acordo com a invenção estão indicadas na Figura 2, que indica a identidade de porcentagem.

Em outra modalidade, uma eHIP da invenção compõe pelo menos duas posições de cruzamento entre uma sequência de aminoácidos da primeira proteína Cry Bt e uma sequência de aminoácidos da segunda proteína Cry Bt. Em um aspecto desta modalidade, uma posição de cruzamento entre Cry3A ou Cry3A modificada e uma Cry1Ab ou uma Cry1Aa está localizada no bloco conservado 2 imediatamente depois de um aminoácido correspondente a Asp232 do SEQ ID NO: 70 e uma segunda posição de cruzamento entre Cry1Ab e Cry3A ou Cry3A modificada está localizada no bloco conservado 3 imediatamente após um aminoácido correspondente a Leu476 do N de ID de SEQ: 72. Em outro aspecto desta modalidade, uma posição de cruzamento entre uma Cry3A ou Cry3A modificada e uma Cry1Ab ou uma Cry1Aa está localizada no bloco conservado 2 imediatamente após Asp232 do SEQ ID NO: 70, ou Asp231 do SEQ ID NO: 68, ou Asp278 do N° SEQ ID NO: 135, e uma segunda posição de cruzamento entre Cry1Ab e Cry3A ou Cry3A modificada está localizada no bloco conservado 3 imediatamente após Leu476 do SEQ ID NO: 72.

Em ainda outro aspecto desta modalidade, uma primeira posição de cruzamento entre Cry3A ou Cry3A modificada e uma Cry1Ab está localizada no bloco conservado 3 imediatamente após um aminoácido correspondente a Leu468 do SEQ ID NO: 70 e uma segunda posição de cruzamento entre Cry1Ab e Cry3A ou Cry3A modificada está localizada no bloco conservado 4 imediatamente após um aminoácido correspondente a Ile527 do SEQ

ID NO: 72. Em ainda outro aspecto desta modalidade, a primeira posição de cruzamento entre Cry3A ou Cry3Aa modificada e uma Cry1Ab está localizada no bloco conservado 3 imediatamente após um Leu468 do SEQ ID NO: 70, ou Leu467 do SEQ ID NO: 68, ou Leu114 do SEQ ID NO: 135, e a segunda posição de cruzamento entre Cry1Ab e Cry3A ou Cry3A modificada está localizada no bloco conservado 4 imediatamente após Ile527 do SEQ ID NO: 72. Em ainda outro aspecto desta modalidade, o eHIP compõe a sequência de aminoácido do SEQ ID NO: 28 ou N° ID SEQ: 34.

Em uma modalidade, a presente invenção engloba um eHIP em que a primeira proteína Cry é uma Cry3A ou Cry3A modificada e a segunda proteína Cry é uma Cry1Aa ou Cry1Ab e em que o eHIP compõe uma sequência de aminoácido que possui pelo menos 80% identidade para o SEQ ID NO: 64. Em outra modalidade eHIP compõe uma sequência de aminoácido que possui pelo menos 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 ou 99% identidade para o SEQ ID NO: 64. Um alinhamento de determinadas incorporações de eHIP da invenção com o SEQ ID NO: 64 é ilustrado na Figura 2, que indica a porcentagem de identidade.

Em outra modalidade, a presente invenção engloba um eHIP em que a primeira proteína Cry é uma Cry3A ou Cry3A modificada e a segunda proteína Cry é uma Cry1Aa ou Cry1Ab e em que o eHIP compõe uma sequência de aminoácido que possui pelo menos 75% identidade para o SEQ ID NO: 70. Em outra modalidade o eHIP compõe uma sequência de aminoácido que possui pelo menos 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 ou 99% identidade para o SEQ ID NO: 70. Um alinhamento de determinadas incorporações de eHIP da invenção com o SEQ ID NO: 70 é ilustrado na Figura 1, que indica a porcentagem de identidade.

Em outra modalidade, a presente invenção engloba um eHIP tendo uma primeira posição de cruzamento entre uma Cry3A ou Cry3A modificada e uma Cry1Aa ou Cry1Ab no bloco conservado 2 e uma segunda posição de cruzamento entre uma Cry1Aa ou Cry1Ab e Cry3A ou Cry3A modificada no bloco conservado 3 e em que o eHIP compõe uma sequência

de aminoácido que possui pelo menos 56% de identidade para o SEQ ID NO: 64. Em um aspecto desta modalidade, o eHIP possui pelo menos 60, 70 ou 80% de identidade para o SEQ ID NO: 64. Em outro aspecto desta modalidade o eHIP possui pelo menos 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 ou 99% identidade para o SEQ ID NO: 64.

Em ainda outra modalidade, a presente invenção engloba um eHIP que compõe uma sequência de aminoácido selecionada do grupo que consiste do SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 30, SEQ ID NO: 34, SEQ ID NO: 62; SEQ ID NO: 64, SEQ ID NO: 147, SEQ ID NO: 153, SEQ ID NO: 155, SEQ ID NO: 159 e SEQ ID NO: 160.

Em uma modalidade o eHIPs da invenção possui atividade contra outras pestes de insetos incluindo mas não-limitado a larva do milho do norte, larva do milho mexicana, besouro da batata do Colorado, e/ou a broca de milho da Europa.

Em outra modalidade a presente invenção engloba uma molécula de ácido nucleico compreendendo uma sequência de nucleotídeo que codifica um eHIP da invenção. Em um aspecto desta modalidade a molécula de ácido nucleico compõe uma sequência de nucleotídeo selecionada do grupo consistindo no SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 29, SEQ ID NO: 33, SEQ ID NO: 61; SEQ ID NO: 63, SEQ ID NO: 146, SEQ ID NO: 152, SEQ ID NO: 154 e SEQ ID NO: 158. Ensinaamentos especificamente exemplificados dos métodos para fazer moléculas de ácido nucleico que codificam eHIPs podem ser encontrados nos Exemplos 1-41. Aqueles com habilidade na técnica reconhecerão que modificações podem ser realizadas aos métodos exemplificados para fazer eHIPs englobados pela presente invenção.

A presente invenção ainda engloba cassetes que compõem as moléculas de ácido nucleico e vetores recombinantes e células hospedeiras transgênicas não-humanas tais como células bacterianas ou células de plan-

tas, compreendendo os cassetes de expressão da invenção.

A presente invenção também engloba vetores recombinantes compreendendo os ácidos nucleicos desta invenção. Nos referidos vetores, os ácidos nucleicos são preferencialmente compostos de cassetes de expressão que compõem elementos regulatórios para expressão das sequências de nucleotídeo em uma célula hospedeira capaz de expressar as sequências de nucleotídeo. Referidos elementos regulatórios geralmente compõem sinais de promotor e terminação e preferivelmente também compõem elementos que permitem eficiente translação de polipeptídios codificados pelos ácidos nucleicos da presente invenção. Vetores que compõem os ácidos nucleicos podem ser capazes de ser replicados em determinadas células hospedeiras, preferivelmente como moléculas extracromossomais, e são, portanto, utilizados para aplicar os ácidos nucleicos desta invenção nas células hospedeiras. Em uma modalidade, as células hospedeiras para referidos vetores são micro-organismos, tais como bactérias, em particular *Bacillus thuringiensis* ou *E. coli*. Em outra modalidade, as células hospedeiras de referidos vetores recombinantes são endófitos ou epífitos. Em ainda outra modalidade estes vetores são vetores virais e são utilizados para replicação das sequências de nucleotídeo em determinadas células hospedeiras, por exemplo células de inseto ou células de planta. Os vetores recombinantes também são utilizados para transformação das sequências de nucleotídeo desta invenção em células hospedeiras, onde as sequências de nucleotídeo são integradas estavelmente ao DNA de um hospedeiro transgênico. Em uma modalidade, o hospedeiro transgênico é uma planta tal como a planta do milho.

Os eHIPs da presente invenção possuem atividade de controle de insetos quando testados contra pestes de insetos em bioestudos. Em uma modalidade, os eHIPs da invenção são ativos contra insetos coleópteros ou insetos lepidópteros ou contra ambos os insetos. Em um aspecto desta modalidade, os eHIPs da invenção são ativos contra a larva do milho do oeste, larva do milho do norte, larva do milho mexicana e/ou o besouro da batata do Colorado. Em outro aspecto desta modalidade, os eHIPs da inven-

ção são ativos contra a broca do milho da Europa. As propriedades de controle de inseto dos eHIPs da invenção são adicionalmente ilustrados nos Exemplos 43, 45 e 46.

5 A presente invenção também engloba uma composição compreendendo uma quantidade eficaz de controle de inseto de um eHIP de acordo com a invenção.

Em outra modalidade, a invenção engloba um método de produzir um eHIP que é ativo contra insetos, compreendendo: (a) obter uma célula hospedeira compreendendo um gene, que em si compõe uma sequência de
10 promotor heteróloga que é operacionalmente relacionada a uma molécula de ácido nucleico da invenção; e (b) desenvolver a célula transgênica de maneira a expressar um eHIP que seja ativo contra insetos.

Em ainda outra modalidade, a invenção engloba um método de produzir uma planta transgênica resistente a insetos, compreendendo introdução de uma molécula de ácido nucleico da invenção em uma planta transgênica, em que a molécula de ácido nucleico causa a expressão de eHIP na
15 planta transgênica em quantidade eficaz para controlar insetos. Em um aspecto desta modalidade, os insetos são insetos coleópteros ou insetos lepidópteros. Em um aspecto desta modalidade, os insetos coleópteros são, a larva do milho do oeste, larva do milho do norte, larva do milho mexicana e/ou o besouro da batata do Colorado. Em ainda outro aspecto desta modalidade, os insetos lepidópteros são a broca do milho de Europa.

Em ainda mais uma modalidade, a invenção engloba um método de controlar insetos, que compõe a entrega aos insetos de uma quantidade
25 eficaz de um eHIP da invenção. Em um aspecto desta modalidade, os insetos são insetos coleópteros ou insetos lepidópteros. Em um aspecto desta modalidade, os insetos coleópteros são, a larva do milho do oeste, larva do milho do norte, larva do milho mexicana e/ou o besouro da batata do Colorado. Em ainda outro aspecto desta modalidade, os insetos lepidópteros são a
30 broca do milho de Europa. Tipicamente, o eHIP é entregue aos insetos oralmente. Em um aspecto o eHIP é entregue oralmente por meio de uma planta transgênica que compõe uma sequência de ácido nucleico que expressa um

eHIP da presente invenção.

A presente invenção adicionalmente engloba um método de controle de insetos em que a planta transgênica adicionalmente compõe uma segunda molécula de ácido nucleico ou grupos de moléculas de ácido nucleico que codificam um segundo princípio pesticida. Exemplos destes segundos ácidos nucleicos são aqueles que codificam uma proteína Cry Bt, os que codificam a Proteína Inseticida Vegetativa, descrita nas Patentes US 5,849,870 e 5,877,012, aqui incorporadas por referência, ou que codificam um caminho para a produção do princípio não-proteinácio.

A presente invenção também engloba um método de fazer uma proteína inseticida híbrida projetada (eHIP), compreendendo: (a) obter uma primeira proteína Cry Bt ou proteína modificada Cry Bt; (b) obter uma segunda proteína Cry Bt que é diferente da primeira proteína Cry Bt ou proteína modificada Cry Bt; (c) combinar regiões variáveis completas ou parciais e blocos conservados da primeira proteína Cry Bt ou proteína modificada Cry Bt com regiões variáveis completas ou parciais e blocos conservados da segunda proteína Cry Bt para fazer um eHIP que possua atividade contra pelo menos a larva do milho do oeste; e opcionalmente (d) inserir um fragmento de peptidila no N-terminal ou uma região de cauda de protoxina de uma proteína Cry Bt no C-terminal do eHIP, ou ambos, em que o fragmento de peptidila N-terminal ou a região protoxina C-terminal ou ambos conferem atividade ao eHIP, ou aumentam a atividade inseticida do eHIP ou tornam o eHIP mais estável que um eHIP sem o fragmento de peptidila ou região de cauda de protoxina ou ambos.

Em outra modalidade, a presente invenção também engloba um método de fazer uma proteína inseticida híbrida projetada (eHIP), compreendendo: (a) obter um primeiro ácido nucleico que codifica uma primeira proteína Cry Bt ou proteína modificada Cry Bt e um segundo ácido nucleico que codifica uma segunda proteína Cry diferente da primeira proteína Cry ou proteína modificada Cry Bt; (b) isolar do primeiro e do segundo ácido nucleico, uma sequência de nucleotídeos que codifica regiões variáveis completas ou parciais e blocos conservados da primeira proteína Cry Bt ou proteína modi-

ficada Cry Bt e a segunda proteína Cry Bt; (c) conectar os ácidos nucleicos isolados resultantes do passo (b) de maneira a fazer um novo ácido nucleico híbrido que codifique uma proteína e opcionalmente fundir a uma extensão de 5' referido ácido nucleico híbrido um ácido nucleico que codifica um fragmento de peptidila resultando em extensão de 5' ou fundir a uma extensão de 3' de referido ácido nucleico híbrido um ácido nucleico que codifica uma região de cauda de protoxina de uma proteína Cry Bt resultando em extensão de 3', ou ambos; (d) inserir o ácido nucleico híbrido com ou sem uma ou ambas das extensões de 5' ou 3' em um cassete de expressão; (e) transformar o cassete de expressão em uma célula hospedeira, resultando em referida célula hospedeira produzindo um eHIP; e (f) bio-estudo do eHIP contra pelo menos a larva do milho do oeste que resulta em atividade inseticida contra a larva do milho do oeste.

Em outras incorporações do método da invenção, a primeira proteína Cry Bt ou proteína modificada Cry Bt é uma proteína Cry3A ou modificada Cry3A e a segunda proteína Cry Bt é uma A Cry1Aa ou Cry1Ab.

Em outra modalidade do método da invenção, o fragmento de peptidila N-terminal se compõe de 9 aminoácidos. Em outro aspecto desta modalidade, o fragmento de peptidila N-C-terminal compõe a sequência de aminoácido YDGRQQHRG (SEQ ID NO: 132) ou a sequência de aminoácido TSNGRQCAGIRP (SEQ ID NO: 133). Em outro aspecto desta modalidade, o fragmento de peptidila N-terminal é selecionado a partir do grupo consistindo no SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 130, SEQ ID NO: 131, e SEQ ID NO: 132.

Ainda em outra modalidade dos métodos da invenção, a região de cauda da protoxina é de uma Cry1Aa ou Cry1Ab. Em outro aspecto desta modalidade, a região de cauda da protoxina se compõe de pelo menos 38 aminoácidos. Em outro aspecto desta modalidade, a região de cauda de protoxina compõe uma sequência de aminoácido que corresponde a aminoácidos 611-648 de SEQ ID NO: 72. Ainda em outro aspecto desta modalidade, a região de cauda de protoxina compõe aminoácidos 611-648 de SEQ ID NO: 72.

Ensinos especificamente exemplificados dos métodos da confecção dos ácidos nucleicos híbridos e os eHIPs da invenção podem ser encontrados nos Exemplos 1-41.

Em novas incorporações, as sequências de nucleotídeo da invenção, em particular uma sequência que codifica o fragmento de peptidila, a cauda da protoxina, e/ou blocos conservados 2, 3, e 4, podem ser além disso modificadas pela modalidade de mutações aleatórias em uma técnica conhecida como recombinação *in vitro* ou troca de DNA. Esta técnica é descrita em Stemmer *et al.*, a Nature 370:389-391 (1994) e na Patente US 5,605,793, que são incorporados aqui por referência. Milhões de cópias mutantes de uma sequência de nucleotídeo são produzidas com base em uma sequência do nucleotídeo Original desta invenção e variantes com propriedades melhoradas, como atividade inseticida aumentada, estabilidade realçada, ou a especificidade diferente ou as variedades de insetos objetivo são recuperadas. O método abrange a formação de um polinucleotídeo mutagenizado com fitas duplas de um padrão de polinucleotídeos de fitas duplas que abrangem uma sequência de nucleotídeos desta invenção, em que o padrão de polinucleotídeos de fitas duplas foi dividido em fragmentos aleatórios de fitas duplas de um tamanho desejado, e compreende os passos da soma à população resultante de fragmentos aleatórios de fita dupla um ou vários oligonucleotídeos de fitas únicos ou duplos, em que os oligonucleotídeos compreendem uma área da identidade e uma área de heterologia ao padrão de polinucleotídeos de fita dupla; desnaturar da mistura resultante de fragmentos aleatórios duplamente encaixados e oligonucleotídeos em fragmentos de fita simples; incubação da população resultante de fragmentos de fita simples com polimerase em condições que resultam no recozimento do dito fragmento de fita simples em áreas ditas da identidade para formar pares de fragmentos recozidos, onde áreas da identidade suficientes para um membro de um par para a réplica principal do outro, por meio deste formando um polinucleotídeo mutagenizado de fita dupla; e repetindo os segundos e terceiros passos de pelo menos dois novos ciclos, em que a mistura resultante no segundo passo de um novo ciclo inclui o polinucleotídeo mutageni-

zado de fita dupla do terceiro passo do ciclo anterior, e o novo ciclo forma um novo polinucleotídeo mutagenizado de fita dupla. Em uma modalidade preferida, a concentração de uma espécie única de fragmento aleatório com fita dupla na população de fragmentos aleatórios de fragmentos duplos é menor que 1% por peso no DNA total. Em outra modalidade preferida, o modelo de polinucleotídeo de fita dupla compreende pelo menos cerca de 100 espécies de polinucleotídeos. Em outra modalidade preferida, o tamanho dos fragmentos aleatórios de fita dupla é de cerca de 5 pb a 5 kb. Em outra modalidade preferida, um quarto passo do método compreende repetir o segundo e o terceiro passo por pelo menos 10 ciclos.

Como agentes de controle de insetos biológico, os eHIPs são produzidos pela expressão dos ácidos nucleicos em células hospedeiras heterólogas capazes de expressar os ácidos nucleicos. Em uma modalidade, as células do *B. thuringiensis* que compõe modificações de ácido nucleico desta invenção são feitas. Tais modificações compreendem mutações ou deleções de elementos regulatórios existentes, levando assim à expressão alterada do ácido nucleico, ou à modalidade de novos elementos regulatórios controlando a expressão do ácido nucleico. Em outra modalidade, cópias adicionais de um ou mais dos ácidos nucleicos são adicionadas às células do *Bacillus thuringiensis* seja pela inserção no cromossomo ou pela introdução de moléculas de replicação extracromossômica contendo os ácidos nucleicos.

Em outra modalidade, ao menos um dos ácidos nucleicos da invenção é inserido no cassete da expressão apropriada, compreendendo um sinal promotor e de terminação. A expressão do ácido nucleico é constitutiva, ou é utilizado um promotor indutivo respondendo a vários tipos de estímulos para iniciar a transcrição. Em outra modalidade, a célula na qual o eHIP é expressado é um microorganismo, tal como um vírus, bactéria ou fungo. Ainda em outra modalidade, um vírus, tal como um baculovírus, contém um ácido nucleico da invenção em seu genoma e expressa grandes volumes de proteína inseticida correspondente após a infecção das células eucariontes apropriadas que são adequadas para replicação do vírus e ex-

pressão do ácido nucleico. A proteína inseticida assim produzida é utilizada como agente inseticida. Como alternativa, os baculovírus transformados para incluir o ácido nucleico são utilizados para infectar insetos *in vivo* e os matam ou pela expressão da toxina inseticida ou por uma combinação de infecção viral e expressão da toxina inseticida.

As células bacterianas são também hospedeiras para a expressão dos ácidos nucleicos da invenção. Em uma modalidade, bactérias simbióticas não-patogênicas, que conseguem viver e se reproduzir dentro dos tecidos da planta, as chamadas endófitas, ou bactérias simbióticas não-patogênicas, que conseguem colonizar a filosfera ou a rizosfera, chamadas de epífitas, são utilizadas. Tais bactérias incluem bactérias do gênero *Agrobacterium*, *Alcaligenes*, *Azospirillum*, *Azotobacter*, *Bacillus*, *Clavibacter*, *Enterobacter*, *Erwinia*, *Flavobacter*, *Klebsiella*, *Pseudomonas*, *Rhizobium*, *Serratia*, *Streptomyces* e *Xanthomonas*. Fungos simbióticos, tais como *Trichoderma* e *Gliocladium* são também possíveis hospedeiros para a expressão dos ácidos nucleicos da invenção pelo mesmo propósito.

Técnicas para estas manipulações genéticas são específicas para os diferentes hospedeiros disponíveis e são conhecidas na técnica. Por exemplo, os vetores da expressão pKK223-3 e pKK223-2 podem ser utilizados para expressar genes heterólogos em *E. coli*, ou em fusão transcripcional ou translacional, por trás do promotor tac ou trc. Para a expressão de operons de ORFs de codificação múltipla, o procedimento mais simples é inserir o operon em um vetor tal como pKK223-3 na fusão transcripcional, permitindo ser utilizado o sítio de ligação do ribossomo cognato dos genes heterólogos. Técnicas de sobre-expressão em espécies gram-positivas tais como *Bacillus* são também conhecidas na técnica e podem ser utilizadas no contexto desta invenção (Quax *et al.* In: Industrial Microorganisms: Basic and Applied Molecular Genetics, Eds. Baltz *et al.*, American Society for Microbiology, Washington (1993)). Sistemas alternativos de sobre-expressão se baseiam por exemplo, em vetores de levedura e incluem o uso de *Pichia*, *Saccharomyces* e *Kluyveromyces* (Sreekrishna, In: Industrial microorganisms: basic and applied molecular genetics, Baltz, Hegeman, e Skatrud eds.,

American Society for Microbiology, Washington (1993); Dequin & Barre, *Biotechnology* L2:173- 177 (1994); van den Berg *et al.*, *Biotechnology* 8:135-139 (1990)).

Em uma modalidade, ao menos um dos eHIPs da invenção é expresso em um organismo mais alto tal como uma planta. Neste caso, plantas transgênicas expressando volumes efetivos do eHIP se protegem de pestes de insetos. Quando o inseto começa a alimentar-se em tal planta transgênica, ele também ingere o eHIP expressado. Isto deterá o inseto de continuar mordendo o tecido da planta ou pode mesmo danificar ou matar o inseto. Um ácido nucleico da presente invenção é inserido em um cassete de expressão, que pode então ser estavelmente integrado no genoma da planta. Em outra modalidade, o ácido nucleico é incluído em um vírus autor-replicante não-patogênico. As plantas transformadas conforme a presente invenção podem ser monocotiledôneas ou dicotiledôneas e incluem, mas não são limitadas a, grão, trigo, cevada, centeio, batata doce, feijão, ervilha, chicória, alface, repolho, couve-flor, brócolis, nabo, rabanete, espinafre, aspargo, cebola, alho, pimentão, aipo, abóbora, abóbora moranga, cânhamo, abobrinha, maçã, pêra, marmelo, melão, ameixa, cereja, pêssego, nectarina, damasco, morango, uva, framboesa, amora-preta, abacaxi, abacate, mamão, manga, banana, soja, tomate, sorgo, cana-de-açúcar, beterraba, girassol, colza, trevo, fumo, cenoura, algodão, alfafa, arroz, batata, berinjela, pepino, *Arabidopsis*, e plantas arborizadas como árvores coníferas e decíduas.

Uma vez que um ácido nucleico desejado tenha sido transformado em uma espécie particular de planta, pode se propagar nesta espécie ou se mover em outras variedades da mesma espécie, particularmente incluindo variedades comerciais, utilizando técnicas de cultivo tradicional.

Um ácido nucleico desta invenção é preferivelmente expresso em plantas transgênicas, causando assim a biossíntese do eHIP correspondente nas plantas transgênicas. Desta forma, plantas transgênicas com resistência aumentada a insetos, particularmente contra a larva do milho do oeste, são geradas. Para sua expressão em plantas transgênicas, os ácidos nucleicos da invenção podem requerer outras modificações e otimização.

Ainda que em muitos casos os genes de organismos microbianos possam ser expressados em plantas em altos níveis sem modificação, a baixa expressão em plantas transgênicas pode resultar de ácidos nucleicos microbianos com códons que não são preferidos nas plantas. Sabe-se na técnica

5 que todos os organismos têm preferências específicas para o uso de códon, e os códons dos ácidos nucleicos descritos nesta invenção podem ser alterados conforme as preferências da planta, enquanto mantêm os aminoácidos codificados assim. Além disso, a alta expressão em plantas é melhor conseguida das sequências de codificação que têm ao menos cerca de 35%

10 de conteúdo de GC, de preferência mais de 45%, mais preferivelmente mais de 50%, e mais preferivelmente mais que 60%. Os ácidos nucleicos microbianos que têm baixo conteúdo de GC podem se expressar pobremente em plantas devido à existência de motivos ATTTA que podem desestabilizar mensagens, e motivos AATAAA que podem causar poliadenilação imprópria.

15 Embora as sequências genéticas preferidas possam ser apropriadamente expressas tanto em monocotiledôneas como em espécies de plantas dicotiledôneas, as sequências podem ser modificadas para contar as preferências de códon específicas e as preferências de conteúdo de GC das monocotiledôneas ou dicotiledôneas como essas preferências foram mostradas

20 diferenciando-se (Murray *et al.* Nucl. Acids Res. 17:477-498 (1989)). Além disso, os ácidos nucleicos são protegidos contra a existência de sítios de junção ilegítimos que podem causar o truncamento de mensagem. Todas as modificações que necessitam ser feitas dentro dos ácidos nucleicos como os

25 descritos acima são feitas utilizando as técnicas bem-conhecidas de mutagêneses dirigida ao sítio, PCR, e construção genética sintética que utiliza os métodos descritos nos pedidos de patente publicados EP 0 385 962, EP 0 359 472, e WO 93/07278.

Em uma modalidade da invenção, um *eHIP* codificando a sequência e/ou uma proteína mãe Cry Bt codificando a sequência é(são) feito(s) segundo o procedimento descrito na Patente US 5,625,136, aqui incorporada por referência. Neste procedimento, os códons preferidos do milho, isto é, o códon único que mais frequentemente codifica este aminoácido no

30

milho, é utilizado. O códon preferido do milho para um determinado aminoácido pode ser derivado, por exemplo, de sequências genéticas conhecidas do milho. O uso do códon de milho para 28 genes de plantas de milho é encontrado em Murray *et al.*, Nucleic Acids Research 17:477-498 (1989), cuja
5 divulgação é incorporada aqui por referência.

Desta maneira, as sequências de nucleotídeos podem ser otimizadas para a expressão em qualquer planta. É reconhecido que toda ou qualquer parte da sequência genética pode ser otimizada ou sintética. Ou seja, sequências sintéticas ou parcialmente otimizadas podem também ser
10 utilizadas.

Para início eficaz da translação, as sequências adjacentes da metionina inicial podem requerer modificação. Por exemplo, elas podem ser modificadas pela inclusão de sequências conhecidas como eficientes em plantas. Joshi sugere um consenso apropriado para plantas (NAR 15:6643-
15 6653 (1987)) e Clontech sugere mais outro consenso de iniciador de translação (1993/1994, catálogo, pag. 210). Estes consensos são adequados para uso com ácidos nucleicos desta invenção. As sequências são incorporadas em construções que compreendem os ácidos nucleicos, até e inclusive o ATG (deixando o segundo aminoácido não modificado), ou como alternativa
20 até e inclusive o subsequente GTC ao ATG (com a possibilidade de modificar o segundo aminoácido da transgene).

A expressão dos ácidos nucleicos em plantas transgênicas é direcionada por promotores naquela função em plantas. A escolha do promotor variará dependendo das exigências temporais e espaciais da expressão, e também dependendo das espécies objetivo. Assim, a expressão dos
25 ácidos nucleicos desta invenção em folhas, em talos ou troncos, em orelhas, em inflorescências (por exemplo espinhos, panículas, espigas de milho, etc.), em raízes, e/ou plantas cultivadas de sementes é preferida. Em muitos casos, contudo, a proteção contra mais de um tipo de inseto é buscada, e assim a expressão em múltiplos tecidos é desejável. Embora muitos promo-
30 tores de dicotiledôneas tenham sido mostrados como operacionais em monocotiledôneas e vice-versa, de maneira ideal dicotiledôneas promotoras são

selecionadas para expressão em dicotiledôneas, e monocotiledôneas promotoras da expressão em monocotiledôneas. Contudo, não há nenhuma restrição quanto a proveniência de promotores selecionados; é suficiente que eles sejam operacionais na condução da expressão dos ácidos nucleicos na célula desejada.

Em uma modalidade os promotores são utilizados, o que é expresso constitutivamente incluindo o actina ou ubiquitina ou promotores cmp ou promotores CaMV 35S e 19S. Os ácidos nucleicos desta invenção também podem ser expressos sob a regulação de promotores que são quimicamente regulados. Isto permite aos eHIPs serem sintetizados só quando as plantas colhidas são tratadas com os produtos químicos de indução. A tecnologia preferida para indução química da expressão do gene é detalhada na aplicação publicada EP 0 332 104 (para Ciba-Geigy) e Patente US 5.614.395. Um promotor preferido para indução química é o promotor de tabaco PR-1a.

Em outra modalidade uma categoria de promotores que é indutora de lesões pode ser utilizada. Numerosos promotores foram descritos e que são expressos em sítios de lesão e também nos sítios de infecção de fitopatógenos. Idealmente, referido promotor deve apenas ser ativo localmente nos locais da infecção, e desta forma os eHIPs apenas acumulam em células que precisam sintetizar os eHIPs para matar as pestes de insetos invasores. Os promotores preferidos deste tipo incluem aqueles descritos por Stanford *et al.* Mol. Gen. Genet. 215:200-208 (1989), Xu *et al.* Plant Molec. Biol. 22:573-588 (1993), Logemann *et al.* Plant Cell 1:151-158 (1989), Rohrmeier & Lehle, Plant Molec. Biol. 22:783-792 (1993), Firek *et al.* Plant Molec. Biol. 22:129-142 (1993), e Warner *et al.* Plant J. 3:191-201 (1993).

Os promotores de tecido específico ou tecido preferencial úteis para expressão dos genes que codificam eHIPs em plantas, particularmente milho, são aqueles que direcionam expressão na raiz, folha ou pólen, particularmente raiz. Referidos promotores exemplo, aqueles isolados de PEPC ou trpA, são descritos na Patente US 5,625,136, ou MTL, descrita na Patente US 5,466,785. Ambas as patentes US aqui incorporadas por referência na

íntegra.

As incorporações adicionais são plantas transgênicas expressando os ácidos nucleicos em modo indutor de lesão ou indutor de patógeno de infecção.

5 Adicionalmente aos promotores, uma variedade de terminadores transcrpcionais também estão disponíveis para uso em construção de ácidos nucleicos híbridos utilizando os genes de eHIP da presente invenção. Os terminadores transcrpcionais são responsáveis pela terminação da transcri-
10 ção além do transgene e na sua correta poliadenilação. Terminadores transcrpcionais adequados e aqueles que são conhecidos de funcionar em plantas incluem o terminador CaMV 35S, terminador tml, o terminador sintase nopalina (NOS), o terminador pea rbcS E9 e outros conhecidos na técnica. Estes podem ser utilizados em monocotiledoneas e dicotiledôneas. Qualquer terminador disponível conhecido de funcionar em plantas pode ser utilizado no
15 contexto desta invenção.

 Numerosas outras sequências podem ser incorporadas nos cassetes de expressão descritos nesta invenção. Estes incluem as sequências que mostraram melhorar a expressão tal como sequências de íntron (por exemplo de Adhl e bronzel) e sequências de líder viral (por exemplo de TMV, MCMV e AMV).
20

 Pode ser preferível visar expressão dos ácidos nucleicos da presente invenção para diferentes localizações celulares na planta. Em alguns casos, a localização no citosol pode ser desejável, onde em outros casos, a localização em órgão subcelular pode ser preferida. A localização subcelular
25 de enzimas de transgene codificado é realizada utilizando técnicas bem-conhecidas na técnica. Tipicamente, o DNA codificando o peptídeo alvo de um produto de órgão visado conhecido é manipulado e fundido com o ácido nucleico. Muitas das referidas sequências-alvo são conhecidas pelo cloroplasta e seu funcionamento nas construções tem sido mostrado. A expressão de ácidos nucleicos da presente invenção também é visada ao retículo
30 endoplásmico ou aos vacúolos das células hospedeiras. Técnicas para obter isto são bem-conhecidas na técnica.

Vetores adequados para transformação de planta são descritos em outros pontos desta especificação. Para transformação mediada por agrobactéria, os vetores binários ou vetores que carregam pelo menos uma sequência de borda de T-DNA são adequados, onde para a transferência direta de genes qualquer vetor é adequado e DNA linear contendo apenas a construção de interesse pode ser preferido. No caso de transferência direta de gene, a transformação com uma espécie única de DNA ou cotransformação pode ser utilizada (Schocher *et al.* Biotechnology 4:1093- 1096 (1986)). Tanto para a transferência direta de gene e para a transferência mediada por agrobactéria, a transformação geralmente (mas não necessariamente) é realizada com um marcador selecionável que pode fornecer resistência a um antibiótico (Canamicina, higromicina ou metotrexato) ou um herbicida (basta). Os vetores de transformação de planta compreendendo os genes de eHIP da presente invenção podem também compor genes (por exemplo, fosfomanose *isomerase*; PMI) que providencia seleção positiva das plantas transgênicas conforme descrito nas Patentes US 5.767.378 e 5.994.629, aqui incorporadas por referência. A escolha de marcadores selecionáveis não é, entretanto, crítica para a invenção.

Em outra modalidade, um ácido nucleico da presente invenção é diretamente transformado em um genoma plastídeo. Uma grande vantagem da transformação plastídea é que os plastídeos são geralmente capazes de expressar genes bactericidas sem otimização substancial de códon, e os plastídeos são capazes de expressar quadros múltiplos abertos de leitura sob controle de um único promotor. A tecnologia de transformação plastídea é extensamente descrita nas patentes US 5.451,513. 5.545.817 e 5.545.818, na aplicação PCT Nº WO 95/16783, e em McBride *et al.* (1994) Proc. Nati. Acad. Sci. USA 91, 7301-7305. A técnica básica para transformação cloroplasta envolve introduzir regiões de DNA plastídeo clonadas flanqueando um marcador selecionável junto com o gene de interesse em um tecido-alvo adequado, exemplo, utilizando transformação biolística ou protoplasta (exemplo, cloreto de cálcio ou transformação mediada por PEG). As regiões de flanqueio 1 a 1,5 kb, denominadas sequências-alvo, facilitam a recombinação.

nação homóloga com o genoma plastídeo e assim permite a reposição ou modificação de regiões específicas do plastoma. Inicialmente, mutações de ponto no cloroplasta 16S rRNA e genes *rps12* conferindo resistência a espectinomicina e/ou estreptomicina são utilizadas como marcadores selecionáveis para transformação (Svab, Z., Hajdukiewicz, P., e Maliga, P. (1990) Proc. Natl. Acad. Sci. USA 87, 8526-8530; Staub, J. M., e Maliga, P. (1992) Plant Cell 4, 39-45). Isto resultou em transformantes homoplásmicos estáveis em frequência de aproximadamente um por 100 bombardeios de folhas-alvo. A presença de sítios de clonagem entre estes marcadores permitiu a criação de vetor de alvo plastídeo para introdução de genes estrangeiros (Staub, J.M., e Maliga, P. (1993) EMBO J. 12, 601-606). Aumentos substanciais na frequência de transformação são obtidos por reposição do rRNA recessivo ou genes resistentes a antibióticos proteína-r sem marcador selecionável dominante, o gene bactéria *aadA* codificando a enzima espectinomicina-"cletoxyfying" aminoglicosídeo-3' adeniltransferase (Svab, Z., e Maliga, P. (1993) Proc. Natl. Acad. Sci. USA 90, 913-917). Previamente, este marcador era utilizado para transformação de alta frequência do genoma plastídeo da alga verde *Chlamydomonas reinhardtii* (Goldschmidt-Clermont, M. (1991) Nucl. Acids Res. 19:4083-4089). Outros marcadores selecionáveis que são úteis para a transformação plastídea são conhecidos na técnica e englobados dentro do escopo desta invenção. Tipicamente, aproximadamente 15-20 ciclos de divisão celular em seguida da transformação são requeridos para obter o estado homoplastídico. A expressão plastídea, em genes que são inseridos por recombinação homóloga em todas as milhares de cópias do genoma plastídeo presente em cada célula de planta, tira vantagem do enorme número de cópias sobre genes de expressão nuclear para permitir níveis de expressão que podem prontamente exceder 10% da proteína solúvel total da planta. Em uma modalidade preferida, um ácido nucleico da presente invenção é inserido em um vetor de alvo plastídeo e transformado no genoma plastídeo de uma planta hospedeira desejada. As plantas homoplásticas para genomas plastídeos que contêm um ácido nucleico da presente invenção são obtidas, e são preferencialmente capazes da alta expressão

do ácido nucleico.

Os eHIPs da invenção podem ser utilizados em combinação com outros princípios pesticidas (por exemplo proteína Cry Bt) para aumentar a variedade pestes-alvo. Além disso, o uso de eHIPs da invenção em combinação com toxinas Cry3A modificadas, proteínas de Cry Bt, ou outros princípios CRW-ativos, como RNAi, que têm um modo diferente da ação ou visam um receptor diferente no intestino do inseto, tem uma utilidade em particular na prevenção e/ou gerenciamento na resistência da larva do milho do oeste. Outros princípios inseticidas incluem, mas não se limitam a, lectinas, α -amilase, peroxidase, e oxidase colesterol. Genes vip, como descrito na Patente US N° 5.889.174 e aqui por incorporados por referência, são também úteis em combinação com os eHIPs da presente invenção.

Esta coexpressão de mais de um princípio inseticida na mesma planta transgênica pode ser atingida fazendo uma compreensão de vetor recombinante única que codifica sequências de mais de um princípio inseticida em uma assim chamada pilha molecular e geneticamente projetando uma planta para conter e expressar todos os princípios inseticidas na planta transgênica.

Tais pilhas moleculares podem também ser feitas utilizando micromossomos como descritos, por exemplo na Patente US 7,235,716. Como alternativa, uma planta transgênica compreendendo um ácido nucleico codificando um primeiro princípio inseticida pode ser retransformada com um ácido nucleico diferente codificando um segundo princípio inseticida e assim por diante. Como alternativa, uma planta, Original 1, pode ser geneticamente modificada para a expressão de genes da presente invenção. Uma segunda planta, Original 2, pode ser geneticamente modificada para a expressão de um princípio de controle de insetos suplementar. Cruzando a Original 1 com a Original 2, as plantas progênes são obtidas que expressam todos os genes introduzidos na Original 1 e 2.

A semente transgênica da presente invenção pode também ser tratada com uma semente de cobertura inseticida como descrito nas Patentes US de N°s 5.849.320 e 5.876.739, aqui incorporadas por referência. On-

de tanto o revestimento de semente inseticida como a semente transgênica da invenção estão ativos contra o mesmo inseto-alvo, a combinação é útil (i) em um método para realçar a atividade de um eHIP da invenção contra o inseto-alvo e (ii) em um método para prevenir o desenvolvimento da resistência a um eHIP da invenção fornecendo um segundo mecanismo da ação contra o inseto-alvo. Assim, a invenção fornece um método de realçar a atividade contra ou prevenir o desenvolvimento da resistência em um inseto-alvo, por exemplo, a larva do milho do oeste, compreendendo a aplicação de um revestimento de semente inseticida a uma semente transgênica que compreende um ou vários eHIPs da invenção.

Mesmo onde o revestimento de semente inseticida é ativo contra um inseto diferente, o revestimento da semente inseticida é útil para estender a variedade do controle de insetos, por exemplo, acrescentando uma cobertura à semente inseticida que tem a atividade contra insetos lepidópteros à semente transgênica da invenção, que tem a atividade contra insetos coleópteros, a cobertura das sementes transgênicas produzem controles tanto contra lepidópteros como contra insetos coleópteros.

EXEMPLOS

A invenção será mais descrita por referência aos seguintes exemplos detalhados. Estes exemplos são fornecidos com o objetivo de ilustração apenas, e não destinados a limitar a menos que de outra maneira especificado. O padrão recombinante de DNA e técnicas de clonagem moleculares aqui utilizados são bem-conhecidos na técnica e descritos por J. Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 3d Ed., Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2001); por T.J. Silhavy, M.L. Berman, e L.W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) e por Ausubel, F.M. *et al.*, *Current Protocols in Molecular Biology*, New York, John Wiley and Sons Inc., (1988), Reiter, *et al.*, *Methods in Arabidopsis Research*, World Scientific Press (1992), e Schultz *et al.*, *Plant Molecular Biology Manual*, Kluwer Academic Publishers (1998).

Exemplo 1: Sequências de Codificação Parental

Sequências codificadas *cry3Aa*, *cry1Ab*, *cry1Ba* e *cry1Fa* otimizadas de milho; aqui designadas *mocry3Aa*, *mocry1Ab*, *mocry1Ba* e *mocry1Fa*, respectivamente, foram feitas de acordo com o procedimento descrito na Patente US 5,625,136, cujo conteúdo é aqui incorporado por citação.

5 A sequência de codificação *cry3A055* (SEQ ID NO: 67), que codifica uma proteína Cry3A055 (SEQ ID NO: 68) foi feita modificando-se a sequência de codificação *mocry3A* inserindo-se uma sequência de nucleotídeo que codifica um sítio de reconhecimento de protease Catepsina G no domínio I de acordo com a Patente US 7,030,295, cujo conteúdo é aqui incorporado por citação.

A sequência de codificação *mocry3Aa* (SEQ ID NO: 67), que codifica a proteína fixada no SEQ ID NO: 68, *cry3A055* (SEQ ID NO: 69), que codifica a proteína fixada no SEQ ID NO: 70, *mocry1Ab* (SEQ ID NO: 71), que codifica a proteína fixada no SEQ ID NO: 72, *mocry1Ba* (SEQ ID NO: 73), que codifica a proteína fixada no SEQ ID NO: 74, *mocry1Fa* (SEQ ID NO: 75), que codifica a proteína fixada no SEQ ID NO: 76, *cry8Aa* (SEQ ID NO: 77), que codifica a proteína fixada no SEQ ID NO: 78, *cry1Ac* (SEQ ID NO: 79), que codifica a proteína fixada no SEQ ID NO: 80, and *cry1Ia* (SEQ ID NO: 81), que codifica a proteína fixada no SEQ ID NO: 82, foram utilizadas na construção dos ácidos nucleicos híbridos e das proteínas que eles codificaram e descritos nos seguintes Exemplos.

Exemplo 2: Uso de iniciadores PCR para construir ácidos nucleicos híbridos.

A Reação de Cadeia de Polimerase (PCR) é uma síntese repetitiva, enzimática e causada de uma sequência de ácido nucleico. Este procedimento é bem-conhecido e comumente utilizado por aqueles versados nesta técnica (*Vide* Mullis, Patentes US N^{os} 4.683.195, 4.683.202, e 4,800,159; Saiki, Randall K., Stephen Scharf, Fred Faloona, Kary B. Mullis, Glenn T. Horn, Henry A. Erlich, Norman Arnheim [1985] "Enzymatic Amplification of .beta.-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia," Science 230:1350-1354.). A PCR é baseada na amplificação enzimática de um fragmento de DNA de interesse que é flanqueado por dois iniciadores de oligonucleotídeo que hibridizam as margens

opostas da sequência objetivada. Os iniciadores são orientados com as extremidades 3' apontando umas para as outras. Os ciclos repetitivos de desnaturação por calor do modelo, recozimento dos iniciadores às suas sequências complementares, e extensão dos iniciadores recozidos com uma polimerase DNA resultam na amplificação do segmento definido pelas extremidades 5' dos iniciadores PCR. Uma vez que o produto de extensão de cada iniciador pode servir como um modelo para outro iniciador, cada ciclo essencialmente dobra o montante do fragmento de DNA produzido no ciclo anterior. Isto resulta no acúmulo exponencial do fragmento objetivado específico, em até vários milhões de duplicações em poucas horas. Utilizando uma polimerase DNA termoestável tal como uma polimerase Taq, que é isolada da bactéria *Thermus aquaticus* termofílica, o processo de amplificação pode ser completamente automatizado.

As sequências de codificação quiméricas descritas nos exemplos a seguir foram construídas utilizando várias combinações dos iniciadores exemplificados exibidos na Tabela 1. As misturas de reação PCR e protocolos termocíclicos PCR utilizados nos experimentos são listadas nas Tabelas 2 e 3, respectivamente. Em cada um dos exemplos a seguir, os iniciadores PCR são referidos pelo nome e "SEQ ID NO:" e as misturas de reação PCR e os protocolos termocíclicos são referidos por seus respectivos números. Será reconhecido por pessoas versados que outros iniciadores PCR e condições de reação PCR podem ser utilizados para construir as sequências de codificação quiméricas da invenção e listando os iniciadores e condições PCR exemplificados que foram utilizados na presente invenção não pretendem ser limitativos de forma alguma.

Tabela 1: Iniciadores utilizados para construir sequências de codificação codificando eHIPs.

Nome do Iniciador	Sequência	SEQ ID NO:
5'3A-1-bam	5'-CCGGATCCATGACGGCCGACAACAACACCGAGGC-3'	SEQ ID NO:83
C3-3A-6	5'-CAGGGGCAGCTGGGTGATCT-3'	SEQ ID NO:84

Nome do Iniciador	Sequência	SEQ ID NO:
C3-1Ab-3	5'-AGATCACCCAGATCCCCCTG-3'	SEQ ID NO:85
1Ab-6-sac	5'- CCGAGCTCAGCTCCTACAC- CTGATCGATGTGGTAGTCGG-3'	SEQ ID NO:86
8A-atg-delRI	5'- CCGGATCCACCATGACTAGTAACGG- CCGCCAGTGTGCTGGTATTCGCCCTTA- TGAC-3'	SEQ ID NO:87
C2-3A-4	5'-GTCCAGCACGGTCAGGGTCA-3'	SEQ ID NO:88
reverse	5'-GCGTGCAAGTCAGATC-3'	SEQ ID NO:89
FR8a-OL-1	5'- GGTGTGTTGTCGGCCGTCATAGGG- CGAATACCAGCAC-3'	SEQ ID NO:90
FR8a-OL-2	5'- GCCGACAACAACCGAGGCCCTGGA- CAGCAGCACCACC-3'	SEQ ID NO:91
C1-3A-2	5'- CAGGTGGGTGTTGGCGGCCTGGGCGTA- 3'	SEQ ID NO:92
5'FR8a	5'-GGATCCACCATGACTAGTAAC-3'	SEQ ID NO:93
5'FR8a-12aa	5'- CCGGATCCACCATGTATGACGGCCGA- CAACAACACC-3'	SEQ ID NO:94
C2-3A-3	5'-TGACCCTGACCGTGCTGGAC-3'	SEQ ID NO:95
3'1Ab-dm3	5'- GAGCTCCTAGGTCACCTCGGCGGGCAC- 3'	SEQ ID NO:96
5'FR-del6	5'- GGATCCACCATGTGTGCTGGTATTCG- CCCTAT-3'	SEQ ID NO:97
5'1Ab-bam	5'- CCGGATCCATGGACAACAACCCCAACAT- CAAC-3'	SEQ ID NO:98
C3-3A-8	5'-GATGTCGCCGCCGGTGAAGC-3'	SEQ ID NO:99
C3-3A-7	5'-GCTTCACCGGCGGCGACATC-3'	SEQ ID NO:100
1B-5	5'- CCGCCGCGACCTGACCCTGGGCGTG- CTGGAC-3'	SEQ ID NO:101

Nome do Iniciador	Sequência	SEQ ID NO:	
1B-10	5'-CCGAGCTCCTAGAACAGGGCGTTCAC-3'	SEQ NO:102	ID
3A-22	5'-GGCCTTCACCAGGGGCAGCTGGGTGAT-3'	SEQ NO:103	ID
1B-7	5'-ATCACCCAGATCCCCATGGTGAAGGCC-3'	SEQ NO:104	ID
C3-1Ab-2	5'-CAGGGGGATCTGGGTGATCT-3'	SEQ NO:105	ID
C3-3A-5	5'-AGATCACCCAGCTGCCCCCTG-3'	SEQ NO:106	ID
3A-12-sac	5'-CCGAGCTCAGCTCAGATCTAGTT-CACGGGGATGAACTCGATCTT-3'	SEQ NO:107	ID
C4-3A-10	5'-TGGTGCTGGCGTAGTGGATGCGG-3'	SEQ NO:108	ID
C4-3A-9	5'-CCGCATCCACTACGCCAGCACCA-3'	SEQ NO:109	ID
C1-1Ab-1	5'-TACGTGCAGGCCGCCAACCTGCACCTG-3'	SEQ NO:110	ID
5'8Aa-dm3	5'-AGATCACCCAGCTGCCCCCTGGTAAAGG-GAGACATGTTATATC-3'	SEQ NO:111	ID
3'8Aa-dm3	5'-GAGCTCCTATGTCTCATCTACTGGGA-TGAA-3'	SEQ NO:112	ID
tant-OL-2	5'-GAGGGTGTGGGCCTTCACCAGGGGCAG-CTGGGT-3'	SEQ NO:113	ID
tant-OL-1	5'-ACCCAGCTGCCCCCTGGTGAAGGCCCA-CACCCTC-3'	SEQ NO:114	ID
tant-3'sac	5'-GAGCTCTAGCTTAAGCAG-TCCACGAGGTT-3'	SEQ NO:115	ID
1Ac-OL-2	5'-TAAAAAGAAAGTTTCCCTTCACCAGGGG-CAGCTGGGT-3'	SEQ NO:116	ID

Nome do Iniciador	Sequência	SEQ ID NO:	
1Ac-OL-1	5'- ACCCAGCTGCCCCTGGTGAAGGGAAAC- TTTCTTTTAA-3'	SEQ NO:117	ID
1Ac-3'sac	5'- GAGCTCCTATGTTGCAGTAACTGGAA- TAAA-3'	SEQ NO:118	ID
1la-OL-2	5'- AAGACAGATTGAAAGCTTTTACTCA- GGGGCAGCTGGGT-3'	SEQ NO:119	ID
1la-OL-1	5'- ACCCAGCTGCCCCTGAGTAAAAGCTTT- CAATCTGTCTT-3'	SEQ NO:120	ID
1la-3'sac	5'- GAGCTCCTACATGTTACGCTCAATATG- GAGT-3'	SEQ NO:121	ID
FR-1Ab-2	5'- GATGTTGTTGAACTCGGCGCTCTTG- TGGGTCCA-3'	SEQ NO:122	ID
FR-1Ab-1	5'- TGGACCCACAAGAGCGCCGAGTTCAA- CAACATC-3'	SEQ NO:123	ID
FR-1Ab-4	5'- GGCTCGTGGGGATGATGTTGTTGAAG- TCGACGCTCTTGTGG-3'	SEQ NO:124	ID
FR-1Ab-3	5'- CCACAAGAGCGTCGACTTCAACACAT- CATCCCCAGCAGCC-3'	SEQ NO:125	ID
CMS94	5'- GGCGCGCCACCATGGCTAGCATGAC- TGGTGG-3'	SEQ NO:136	ID
CMS95	5'-GCAGGAACAGGTGGGTGTTG-3'	SEQ NO:137	ID
CMS96	5'-CCTGAACACCATCTGGCCCA-3'	SEQ NO:138	ID
CMS97	5'- CTGGCTGCTGGGGATGATGTTGTTGAAG- TCGACGCTCTT-3'	SEQ NO:139	ID
CMS98	5'-GAGCTCTTAGGTCACCTCGGC-3'	SEQ NO:140	ID

Nome do Iniciador	Sequência	SEQ ID NO:
CMS99	5'- AAGAGCGTCGACTTCAACAACAT- CATCCCCAGCAGCCAG-3'	SEQ ID NO:141
CMS100	5'- GAAGTACCGCGCCCGCATCCGCTACG- CCAGCACCACCAAC-3'	SEQ ID NO:142
CMS101	5'- GTTGGTGGTGCTGGCGTAGCGGATG- CGGGCGCGGTACTTC-3'	SEQ ID NO:143

Tabela 2: Misturas de reação PCR.

Mistura 1	Mistura 2	Mistura 3
50 a 100ng modelo de DNA 0,8 µM iniciador 1 0,8 µM iniciador 2 1X Pfu tampão 0,4 mM dNTPs 2% formamida 1,25 unidades Pfu Polimerase (Estratagene) 2,5 unidades Taq Polimerase (Qiagen) água a um volume total de 50 µl	50 a 100ng modelo de DNA 0,8 µM iniciador 1 0,8 µM iniciador 2 1X Taq tampão 0,4 mM dNTPs 2% formamida 2,5 unidades Taq Polimerase (Qiagen) água a um volume total de 50 µl	50 a 100ng modelo de DNA 0,8 µM iniciador 1 0,8 µM iniciador 2 1X cDNA tampão de Vantagem 0,4 mM dNTPs x unidades cDNA Vantagem Polimerase (Clontech) água a um volume total de 50 µl
Mistura 4	Mistura 5	
50 a 100ng modelo de DNA 0,4 µM iniciador 1 0,4 µM iniciador 2 1X PCR tampão (Invitrogen) 0,4 mM dNTPs 2,5 unidades "HotStart Taq Polimerase" água a um volume total de 50 µl	50 a 100ng modelo de DNA 0,4 µM iniciador 1 0,4 µM iniciador 2 1X Pfu tampão (Estratagene) 0,2 mM dNTPs 1,25 unidades Pfu Polimerase Turbo água a um volume total de 50 µl	

Tabela 3: Perfis termocíclicos PCR.

Perfil Termocíclico 1	Perfil Termocíclico 2	Perfil Termocíclico 3
94°C - 5 minutos 20 ciclos: 94°C - 30 segundos 65°C - 30 segundos 72°C - 30 segundos 72°C - 7 minutos mantido a 4°C	94°C - 5 minutos 20 ciclos: 94°C - 30 segundos 55°C - 30 segundos 72°C - 30 segundos 72°C - 7 minutos mantido a 4°C	94°C - 5 minutos 20 ciclos: 94°C - 30 segundos 55°C - 30 segundos 68°C - 30 segundos 68°C - 7 minutos mantido a 4°C

Perfil Termocíclico 4	Perfil Termocíclico 5	Perfil Termocíclico 6
94°C - 15 minutos 20 ciclos: 94°C - 30 segundos 50-70°C - 30 segundos 72°C - 30 segundos 72°C - 7 minutos mantido a 4°C	94°C - 5 minutos 20 ciclos: 94°C - 30 segundos 55-75°C - 30 segundos 72°C - 1 minuto 72°C - 15 minutos mantido a 4°C	94°C - 5 minutos 20 ciclos: 94°C - 30 segundos 55-75°C - 30 segundos 72°C - 2 minutos 72°C - 15 minutos mantido a 4°C

A Tabela 4 apresenta a relação entre os três domínios de Cry3A055, Cry1Ab e Cry3A com suas respectivas regiões variáveis e blocos conservados. Os aminoácidos compreendidos nos domínios, blocos conservados e regiões variáveis são exibidos para cada proteína.

Tabela 4:

Domínio	Região	Cry3A055 (SEQ ID NO: 70)	Cry1Ab (SEQ ID NO: 72)	Cry3A (SEQ ID NO: 68)	Cry3A (SEQ ID NO: 131)
	V1	1-10	1-32	1-10	1-57
I	V1	11-142	33-152	11-141	58-188
	CB1	143-172	153-182	142-171	189-218
	V2	173-192	183-202	172-191	219-238
	CB2	193-244	203-254	192-243	239-290
II		245-259	255-269	244-258	291-305
	V3	260-444	270-452	259-443	306-490
	CB3	445-454	453-462	444-453	491-500
III		455-492	463-500	454-491	501-538
	V4	493-513	501-520	492-512	539-559
	CB4	514-523	521-531	513-522	560-569

Domínio	Região	Cry3A055 (SEQ ID NO: 70)	Cry1Ab (SEQ ID NO: 72)	Cry3A (SEQ ID NO: 68)	Cry3A (SEQ ID NO: 131)
	V5	524-586	532-596	523-585	570-632
	CB5	587-598	597-606	586-597	633-644
	V6		607-610		
	Protó- xina		611-648		

Exemplo 3: Construção de 2OL-8a.

Um primeiro fragmento de ácido nucleico codificando uma porção N-terminal de uma proteína Cry3A055 (SEQ ID NO: 70) foi amplificada PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e C3-3A-6 (SEQ ID NO: 84) e a Mistura de reação PCR 1 e o Perfil termocíclico 1. Esta reação PCR induziu uma mutação de ponto pela exclusão do nucleotídeo 28 do SEQ ID NO: 69 (*cry3A055*), que causou uma troca estrutural na estrutura de leitura *cry3A055*, e excluiu o sítio BamHI e sequência Kozak (Kozak, M., 1986. Cell 44:283-92) na extremidade 5' do amplicon resultante.

Um segundo fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry1Ab (SEQ ID NO: 72) foi amplificada PCR de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores C3-1Ab-3 (SEQ ID NO: 85) e 1Ab-6-*Sac* (SEQ ID NO: 86) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Os primeiro e segundo ácidos nucleicos descritos acima foram conectados utilizando-os como modelos em uma reação PCR sobreposta (Horton *et al.*, 1989. Gene 77: 61-68) utilizando os iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e 1Ab-6-*Sac* (SEQ ID NO: 86) utilizando a Mistura de reação PCR 2 e Perfil termocíclico 1, exceto um gradiente de 45-65°C foi utilizado para a temperatura de recozimento.

O amplicon resultante foi ligado a um fragmento de extremidade inflexível a um vetor pCR2.1-TOPO (Invitrogen, Carlsbad, CA) cortado com *Sma*I para formar plasmídeo p2OL8a/CR2.1. Um fragmento BamHI-*Sac*I de p2OL8a/CR2.1 foi ligado ao pET21a (EMD Biosciences, Inc., San Diego, CA), que foi cortado com BamHI-*Sac*I, e transformado em *E. coli*. O frag-

mento BamHI-SacI de p2OL8a/CR2.1 composto de 40 nucleotídeos derivados do vetor pCR2.1-TOPO adjacente ao amplicon estrutural externo da primeira reação PCR. Ligando este fragmento BamHI-SacI ao pET21a criado em uma estrutura de leitura aberta iniciada com o códon inicial (ATG) de uma etiqueta T7 e terminando com o sítio SacI do DNA inserido. Esta estrutura de leitura aberta foi designada 2OL-8a (SEQ ID NO: 1) e codifica a proteína inseticida quimérica 2OL-8a (SEQ ID NO: 2). Então, a proteína inseticida quimérica 2OL-8a compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MASM-TGGQQMGRGSTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 126), aminoácidos 10-468 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina Cry1Ab.

Os nucleotídeos que codificam os aminoácidos 1-14 do fragmento de peptidila são derivados da etiqueta T7 e sítio de clivagem BamHI do vetor pET21a. Os nucleotídeos que codificam os aminoácidos 15-26 do fragmento de peptidila são derivados do vetor pCR2.1-TOPO. E os nucleotídeos que codificam os aminoácidos 27- 35 do fragmento de peptidila são derivados de *cry3A055* que estão fora da estrutura com o remanescente da sequência de codificação *cry3A055*.

25 Exemplo 4: Construção de FR8a.

A sequência codificada *FR8a* foi construída pelo posicionamento de uma sequência Kozak (ACC) e um códon inicial (ATG) logo abaixo de um sítio BamHI N-terminal em 2OL-8a (*Vide Exemplo 3*). Adicionalmente, um sítio EcoRI em 2OL-8a foi interrompido para auxiliar na vetorização futura de *FR8a*. Todas estas mudanças foram realizadas utilizando-se uma reação PCR com 2OL-8a como o modelo e os iniciadores: *8a-atg-delRI* (SEQ ID NO: 87) e *C2-3A-4* (SEQ ID NO: 88) utilizando a Mistura de reação PCR 2 e

Perfil termocíclico 2.

O amplicon resultante foi ligado a um vetor pCR2.1-TOPO (Invitrogen). Um fragmento BamHI-PpuMI do produto PCR clonado foi então ligado a um fragmento PpuMI-NcoI de 2OL8a/pCR2.1 (*Vide Exemplo 3*) e um
 5 fragmento NcoI-BamHI de 2OL8a/pCR2.1 para criar *FR8a* (SEQ ID NO: 3) que codifica a proteína inseticida quimérica *FR8a* (SEQ ID NO: 4). Então, a proteína inseticida quimérica *FR8a* compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-468
 10 de uma proteína *Cry3A055* (SEQ ID NO: 70), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína *Cry1Ab* (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, blo-
 15 co conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina *Cry1Ab*.

O *FR8a eHIP* foi bastante ativo contra verme de raiz de milho ocidental conforme mostrado na Tabela 5. Então, a eliminação da sequência de aminoácido T7 do fragmento de peptidila N-terminal de 2OL-8a *eHIP* não
 20 teve um impacto negativo sobre a atividade inseticida.

Adicionando aminoácidos 34 adicionais ao N-termino de *FR8a* criou um *eHIP*, designado *FR8a+34* (SEQ ID NO: 160), com um fragmento de peptidila N-terminal de aminoácidos 56 (SEQ ID NO: 131). O fragmento de peptidila N-terminal de aminoácido 56 não teve efeito negativo sobre a
 25 atividade de *FR8a* contra verme de raiz de milho ocidental (*Vide Tabela 5*).

Exemplo 5: Construção de FRCG.

Para determinar se um sítio de reconhecimento de protease G catepsina era necessário para atividade inseticida de uma proteína híbrida compreendendo um fragmento N-terminal de *Cry3A055*, um constructo foi
 30 realizado que eliminou o sítio de G catepsina da proteína híbrida *FR8a* (*Exemplo 4*). Um primeiro fragmento de ácido nucleico MluI-PpuMI de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) e um segundo fragmento

de ácido nucleico Mlul-PpuMI de um plasmídeo compreendendo *mocry3Aa* (SEQ ID NO: 67) foram ligados utilizando técnicas de biologia molecular padrão para criar *FRCG* (também designado *FR8a-catg*) (SEQ ID NO: 5) que codifica a proteína híbrida *FRCG* (SEQ ID NO: 6). Então, a proteína inseticida quimérica *FRCG* compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCA-GIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-467 de uma proteína *Cry3A* (SEQ ID NO: 68), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína *Cry1Ab* (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina *Cry1Ab*.

A proteína *FRCG* foi tão ativa contra o verme de raiz de milho ocidental quanto a proteína *FR8a* (Vide Tabela 5) sugerindo que o sítio de protease de catespina G não é requerido para atividade inseticida de um eHIP.

Exemplo 6: Construção de *FR8a-9F*.

Um primeiro fragmento de ácido nucleico de aproximadamente 323 pb foi amplificado PCR de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) utilizando *reversão de* iniciadores (SEQ ID NO: 89) e *FR8a-OL-1* (SEQ ID NO: 90) e Mistura de reação PCR 2 e Perfil termocíclico 2. Um segundo fragmento de ácido nucleico de aproximadamente 470 pb foi amplificado PCR de um plasmídeo compreendendo *FR8a* utilizando iniciadores *FR8a-OL-2* (SEQ ID NO: 91) e *C1-3A-2* (SEQ ID NO: 92) e Mistura de reação PCR 2 e Perfil termocíclico 2. Os dois amplicons resultantes foram conectados utilizando-os como modelos em uma reação PCR sobreposta com iniciadores *5'FR8a* (SEQ ID NO: 93) e *C1-3A-2* (SEQ ID NO: 92) utilizando Mistura de reação PCR 2 e Perfil termocíclico 2 para amplificar a extremidade 5' de *FR8a-9F*. O produto PCR sobreposto foi clonado em um vetor pCR2.1-TOPO (Invitrogen) designado *5'FR-9F/pCR2.1*. Um fragmento

BamHI/PpuMI de 5'FR-9F/pCR2.1 foi então ligado a um fragmento PpuMI/BamHI de *FR8a* para criar *FR8a-9F* (SEQ ID NO: 7) que codifica a proteína quimérica *FR8a-9F* (SEQ ID NO: 8). Então, a proteína inseticida quimérica *FR8a-9F* compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRP (SEQ ID NO: 129), aminoácidos 1-468 de uma proteína *Cry3A055* (SEQ ID NO: 70), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína *Cry1Ab* (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina *Cry1Ab*.

A *FR8a-9F* eHIP foi levemente menos ativa contra o verme de raiz de milho ocidental do que *FR8a* eHIP (Vide tabela 5) sugerindo que os aminoácidos 9 C-terminal do fragmento de peptidila do SEQ ID NO: 127 desempenham uma função na concessão de atividade inseticida total a *FR8a*.

Exemplo 7: Construção de *FR-9F-catg*.

A sequência codificada *FR-9F-catg* foi criada para posicionar os nucleotídeos derivados *cry3A055* fora de estrutura de *FR8a* novamente em estrutura e eliminar o sítio de reconhecimento de protease G cathepsina. Um fragmento BamHI/PpuMI de 5'FR-9F/pCR2.1 (Vide Exemplo 6) foi ligado com um fragmento PpuMI/BamHI de *FRCG* (Vide Exemplo 5) para criar a sequência codificada *FR-9F-catg* (SEQ ID NO: 9) que codifica a proteína quimérica *FR-9F-catg* (SEQ ID NO: 10). Então, a proteína quimérica *FR-9F-catg* compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRP (SEQ ID NO: 129), aminoácidos 1-467 de uma proteína *Cry3Aa* (SEQ ID NO: 68), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína *Cry1Ab* (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3,

região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina Cry1Ab.

A FR8a-9F-catg eHIP forneceu o mesmo nível de atividade de FR8a contra verme de raiz de milho ocidental (Vide Tabela 5) confirmando
5 que um eHIP pode ser feito a partir de Cry3A modificado ou uma sequência Cry3 nativa.

Exemplo 8: Construção de FR8a-12aa.

Os aminoácidos 2-13 codificando nucleotídeos do fragmento de peptidila compreendido em FR8a (SEQ ID NO: 4) foram removidos utilizando
10 PCR. Um fragmento foi amplificado PCR de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) utilizando iniciadores 5'*FR8a-12aa* (SEQ ID NO: 94) e C1-3A-2 (SEQ ID NO: 90) e Mistura de Reação PCR 1 e Perfil termocíclico 1. O amplicon resultante foi clonado em pCR2.1-TOPO (Invitrogen). Um fragmento BamHI-PpuMI do clone pCR2.1-TOPO foi então ligado com um fragmento PpuMI-BamHI de um plasmídeo compreendendo *FR8a* para criar
15 *FR8a-12aa* (SEQ ID NO: 11) que codifica a proteína inseticida quimérica FR8a-12aa (SEQ ID NO: 12). Então, a proteína inseticida quimérica FR8a-12aa compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MYDGRQQHRG (SEQ ID NO:
20 128), aminoácidos 10-468 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3,
25 região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina Cry1Ab.

A FR8a-12aa eHIP forneceu o mesmo nível de atividade contra verme de raiz de milho ocidental do que FR8a (Vide tabela 5) sugerindo que os aminoácidos 12 N-terminal do fragmento de peptidila do SEQ ID NO: 127
30 não são necessários para atividade inseticida total de FR8a.

Exemplo 9: Construção de Wr-9mut.

Um fragmento de ácido nucleico foi amplificado PCR de

FR8a/pCR2.1 (Exemplo 2) utilizando iniciadores 5'*FR8a-12aa* (SEQ ID NO: 94) e *C1-3A-2* (SEQ ID NO: 92) e Mistura de Reação PCR 1 e Perfil termocíclico 2. O amplicon resultante foi clonado em pCR2.1-TOPO (Invitrogen). Um fragmento BamHI/PpuMI foi então ligado a um fragmento PpuMI/BamHI de *FR8a* (SEQ ID NO: 3) para criar *Wr-9mut* (SEQ ID NO: 13) que codifica a proteína WR-9mut (SEQ ID NO: 14), que compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MYDGRQQHRG (SEQ ID NO: 128), e aminoácidos 10-598 de uma proteína Cry3A055 (SEQ ID NO: 70). Então a proteína WR-9mut é Cry3A055 com um fragmento de peptidila N-terminal da invenção.

A proteína WR-9mut não foi ativa contra o verme de raiz de milho ocidental. Então, a adição de um fragmento de peptidila N-terminal a uma proteína Cry3a modificada não híbrida destruiu a atividade inseticida. Isto sugere que pode haver alguma interação entre a porção C-terminalry1Ab de *FR8a* e o fragmento de peptidila N-terminal que confere atividade inseticida total a *FR8a*.

Exemplo 10: Construção de FRD3.

A extremidade 3' desta sequência de codificação foi feita por amplificação PCR de um fragmento de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) utilizando iniciadores *C2-3A-3* (SEQ ID NO: 95) e *3'1Ab-dm3* (SEQ ID NO: 96) e Mistura de Reação PCR 2 e Perfil termocíclico 2. O amplicon resultante foi clonado em pCR2.1-TOPO (Invitrogen). Um fragmento 364bp Apal/SacI do amplicon clonado, designado 3'*FRD3/pCR2.1*, foi ligado com um fragmento SacI/Apal de *FR8a* para criar *FRD3* (SEQ ID NO: 15) que codifica a proteína quimérica *FRD3* (SEQ ID NO: 16). A proteína quimérica *FRD3* compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-468 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-610 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do

bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6. Então, a proteína inseticida quimérica proteína inseticida quimérica FRD3 é uma variante de uma proteína inseticida quimérica FR8a com a região de aminoácido 38 da proteína dorsal
5 excluída.

A FRD3 eHIP forneceu o mesmo nível de atividade que FR8a contra verme de raiz de milho ocidental (Vide Tabela 5) sugerindo que a região dorsal de protoxina de aminoácido 38 de FR8a não é necessária para atividade inseticida total.

10 Exemplo 11: Construção de FR-12-cg-dm3

Um fragmento 3082bp SacI/PpuMI de uma plasmídeo compreendendo *FR8a-12* (Vide Exemplo 8), um fragmento 721bp PpuMI/MluI de *FRCG* (Vide Exemplo 5) e um fragmento 923bp MluI/SacI de *FRD3* (Vide Exemplo 10) foram combinados para criar a sequência de codificação *FR-12-cg-dm3* (SEQ ID NO: 17) que codifica a proteína quimérica FR-12-cg-dm3 (SEQ ID NO: 18). A proteína quimérica FR-12-cg-dm3 compreende, do N-terminal ao C-terminal, um fragmento de peptidila compreendendo a sequência de aminoácido MYDGRQQHRG (SEQ ID NO: 129), aminoácidos 10-467 de uma proteína Cry3Aa (SEQ ID NO: 70), que compreende região
15 variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-610 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6.
20 Então, a proteína quimérica FR-12-cg-dm3 é uma variante de FR8a com aminoácidos 12 N-terminal do fragmento de peptidila, o sítio de reconhecimento de protease G catepsina, e com a região de aminoácido 38 da proteína dorsal excluída.

A FR-12-cg-dm3 eHIP não teve atividade contra os verme de raiz de milho ocidental como FR8a (Vide Tabela 5) sugerindo que alguma interação entre a porção C-terminal de FR8a e o fragmento de peptidila N-terminal é requerido para atividade inseticida total.
30

Exemplo 12: Construção de 9F-cg-del6

A extremidade 5' desta sequência de codificação era feita pela amplificação PCR de um fragmento de um plasmídeo compreendendo *FR-9F-catg* (Vide Exemplo 7) utilizando iniciadores 5'*FR-del6* (SEQ ID NO: 97) e C1-3A-2 (SEQ ID NO: 92) e Mistura de Reação PCR 3 e Perfil termocíclico 3. O amplicon resultante foi clonado em pCR2.1-TOPO. Um fragmento 215bp BamHI/PpuMI foi então ligado com um fragmento 4668bp PpuMI/BamHI de *FR-9F-catg* para criar *FR-9F-cg-del6* (SEQ ID NO: 19) que codifica a proteína quimérica FR-9F-cg-del6 (SEQ ID NO: 20). A proteína quimérica FR-9F-cg-del6 compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MCAGIRP (SEQ ID NO: 130), aminoácidos 1-467 de uma proteína Cry3A (SEQ ID NO: 68), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina Cry1Ab. Então, a proteína quimérica FR-9F-cg-del6 é uma variante de FR8a-9F-catg com aminoácidos 2 a 7 do fragmento de peptidila excluído.

A FR-9F-cg-del6 não foi ativa contra o verme de raiz de milho ocidental sugerindo que o fragmento de peptidila N-terminal precisa de pelo menos 7 aminoácidos dos aminoácidos C-terminal 9 do SEQ ID NO: 127 para ser ativo contra o verme de raiz de milho ocidental.

25 Exemplo 13: Construção de FR-cg-dm3

Um fragmento 3839bp MluI/SacI de *FRCG* (Exemplo 5) e um fragmento 923bp MluI/SacI de *FRD3* (Exemplo 10) foram ligados para criar FR-cg-dm3 (SEQ ID NO: 21) que codifica a proteína FR-cg-dm (SEQ ID NO: 22). A proteína inseticida quimérica FR-cg-dm3 compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-467 de uma proteína Cry3A (SEQ ID NO: 68), que compreende re-

giões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-610 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6.

A FRD3 eHIP teve o mesmo nível de atividade contra o verme de raiz de milho ocidental que FR8a (Vide Tabela 5) confirmando que o sítio de catepsina G e a região dorsal de protoxina de FR8a não eram requeridos para atividade inseticida total contra o verme de raiz de milho ocidental.

Exemplo 14: Construção de 9F-cg-dm3.

Um fragmento MluI/SacI de um plasmídeo compreendendo *FR-9F-cg* (Vide Exemplo 7) foi ligado com um fragmento 923bp MluI/SacI de um plasmídeo compreendendo *FRD3* (Vide Exemplo 10) para criar *9F-cg-dm3* (SEQ ID NO: 23) que codifica a proteína quimérica 9F-cg-dm3 (SEQ ID NO: 24). A proteína inseticida quimérica 9F-cg-dm3 compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRP (SEQ ID NO: 129), aminoácidos 1-467 de uma proteína Cry3A (SEQ ID NO: 68), que compreende regiões variáveis 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-610 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6.

A 9F-cg-dm3 eHIP forneceu o mesmo nível de atividade contra o verme de raiz de milho (Vide Tabela 5) confirmando que os aminoácidos C-terminal 9 do fragmento de peptidila podem conferir atividade quando o domínio I do eHIP era composto de regiões variáveis e blocos conservados de Cry3A (Cry3A055) ou regiões variáveis e blocos conservados de Cry3A.

Exemplo 15: Construção de B8a.

Um fragmento de ácido nucleico codificando uma porção N-terminal de uma proteína Cry3A055 (SEQ ID NO: 70), foi amplificado PCR

de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores *5'3A-1-bam* (SEQ ID NO: 83) e *C3-3A-8* (SEQ ID NO: 99) e Mistura de reação PCR 1 e Perfil termocíclico 1. Um fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry1Ab (SEQ ID NO: 72), compreendendo regiões variáveis 4-6, foi amplificado de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores *C3-3A-7* (SEQ ID NO: 100) e *1Ab-6-sac* (SEQ ID NO: 86) e Mistura de Reação PCR 1 e Perfil termocíclico 1. O amplicon resultante foi designado *2OL-8b*.

Um fragmento de ácido nucleico codificando uma porção N-terminal da proteína Cry3A055 (SEQ ID NO: 70), foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores *5'3A-1-bam* (SEQ ID NO: 83) e *C2-3A-4* (SEQ ID NO: 88) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Um segundo fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry1Ba (SEQ ID NO: 74) foi amplificada PCR de um plasmídeo compreendendo *mocry1Ba* (SEQ ID NO: 73) utilizando iniciadores *1B-5* (SEQ ID NO: 101) e *1B-10* (SEQ ID NO: 102) e Mistura de reação PCR 1 e Perfil termocíclico 1, exceto uma temperatura de recozimento de 60°C foi utilizada.

Os dois produtos PCR descritos acima foram então utilizados como os modelos em uma reação PCR sobreposta com iniciadores *5'3A-1-bam* (SEQ ID NO: 83) e *1B-10* (SEQ ID NO: 102) utilizando Mistura de Reação PCR 1 e Perfil termocíclico 2. O amplicon resultante foi designado *B10*.

A seguir, um fragmento de ácido nucleico de *cry3A055* (SEQ ID NO: 69) foi amplificado PCR utilizando *2OL-8b* (vide acima) como o modelo e iniciadores *5'3A-1-bam* (SEQ ID NO: 83) e *3A-22* (SEQ ID NO: 103) com as seguintes condições PCR: Mist. 1, perfil termocíclico: 94°C - 45 segundos, 50°C-70°C gradiente - 45 segundos, 72°C-90 segundos por 30 ciclos. Outro fragmento de ácido nucleico foi amplificado PCR utilizando *B10* (vide acima) como o modelo e iniciadores *1B-7* (SEQ ID NO: 104) e *1B-10* (SEQ ID NO: 102) utilizando Mistura de reação PCR 1 e Perfil termocíclico 2, exceto uma temperatura de recozimento de 60°C foi utilizada. Os dois produtos

PCR resultantes foram então utilizados como os modelos em uma reação PCR sobreposta com iniciadores *5'3A-1-bam* (SEQ ID NO: 83) e *1B-10* (SEQ ID NO: 102) utilizando as seguintes condições PCR: Mist. 2, perfil termocíclico: 94°C - 30 segundos, 40°C-60°C gradiente - 30 segundos, 72°C - 60 segundos por 30 ciclos.

O produto PCR resultante foi ligado a um vetor pCR2.1-TOPO (Invitrogen) e designado B10/pCR2.1. Um fragmento BamHi-SacI de B8a/pCR2.1 foi então ligado a pET21a (Novagen), que foi cortado com BamHi/SacI, para criar a sequência de codificação *B8a* (SEQ ID NO: 25) que codifica a toxina híbrida B8a (SEQ ID NO: 26). A proteína híbrida B8a compreende, do N-termino ao C-termino, 1-468 aminoácidos de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 505-656 de uma proteína Cry1Ba (SEQ ID NO: 74).

Exemplo 16: Construção de 5*B8a.

Um fragmento BamHI-XbaI de um plasmídeo compreendendo *2OL-8a* (Vide Exemplo 3) e um fragmento XbaI-SacI de um plasmídeo compreendendo *B8a* (Vide Exemplo 15) foram ligados para criar *5*B8a* (SEQ ID NO: 27) que codifica a proteína quimérica 5*B8a (SEQ ID NO: 28). A proteína 5*B8a compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGR-QQHRG (SEQ ID NO: 127), aminoácidos 10-467 de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 505-656 de uma proteína Cry1Ba (SEQ ID NO: 74). Então, a proteína quimérica 5*B8a é a proteína híbrida B8a a qual um fragmento de peptidila N-terminal foi adicionado.

Exemplo 17: Construção de V3A.

Este gene foi amplificado PCR utilizando 3 fragmentos juntos como modelos: o primeiro fragmento foi amplificado de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores *5'3A-1-bam* (SEQ ID NO: 83) e *C2-3A-4* (SEQ ID NO: 88) e Mistura de reação PCR 1 e Perfil termocíclico 1; o segundo fragmento foi amplificado de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores *C2-3A-3* (SEQ ID NO: 95) e *C3-1Ab-2* (SEQ ID NO: 105) e Mistura de reação PCR 1

e Perfil termocíclico 1; e o terceiro fragmento foi amplificado de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores C3-3A-5 (SEQ ID NO: 106) e 3A-12-sac (SEQ ID NO: 107) e Mistura de reação PCR 1 e Perfil termocíclico 1. Estes 3 produtos PCR foram então utilizados
 5 como modelos em uma reação PCR sobreposta com iniciadores 5'3A-1-bam (SEQ ID NO: 83) e 3A-12-sac (SEQ ID NO: 107) utilizando Mistura de reação PCR 1 e Perfil termocíclico 1, para produzir a sequência de codificação v3A (SEQ ID NO: 29) que codifica a proteína híbrida V3A (SEQ ID NO: 30). A proteína híbrida V3A compreende, do N-termino ao C-termino, aminoácidos 1-226 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende
 10 região variável 1, bloco conservado 1, região variável 2, e os aminoácidos 34 N-terminal do bloco conservado 2, e aminoácidos 237-474 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 33 C-terminal do bloco conservado 2, região variável 3, e os aminoácidos 20 N-terminal do
 15 bloco conservado 3, e aminoácidos 467-598 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende os aminoácidos 28 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5 e bloco conservado 5.

A V3A eHIP compreende duas posições de cruzamento. O primeiro cruzamento entre Cry3A055 e Cry1Ab é localizado no bloco conservado 2 e o segundo cruzamento entre Cry1Ab e Cry3A055 está localizado no bloco conservado 3. Então, V3A é uma variante de Cry3A055 onde toda a região variável 3 foi substituída pela região variável 3 de uma proteína Cry1Ab. A V3A eHIP não foi tão ativa contra o verme de raiz de milho ocidental quanto FR8a, sugerindo que ter a sequência Cry1Ab no bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5 e/ou região variável 6 é importante para a atividade inseticida total de FR8a.
 25

A sequência de codificação v3A foi alinhada a um vetor pCR2.1-TOPO e então subclonada em pET21a utilizando um fragmento BamHI/SacI. A proteína V3A expressa pelo vetor pET21a possui uma etiqueta T7 no N-termino. Esta proteína era designada T7-V3A.
 30

Exemplo 18: Construção de V4F.

Um primeiro fragmento de ácido nucleico codificando regiões variáveis 1-3 de uma Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e C3-3A-6 (SEQ ID NO: 84) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Um segundo fragmento de ácido nucleico codificando região variável 4 de uma Cry1Ab foi amplificado PCR de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores C3-1Ab-3 (SEQ ID NO: 85) e C4-3A-10 (SEQ ID NO: 108) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Um terceiro fragmento de ácido nucleico codificando regiões variáveis 5-6 de uma Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores C4-3A-9 (SEQ ID NO: 109) e 3A-12-*sac* (SEQ ID NO: 107) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Todos os três amplicons PCR foram combinados e utilizados como os modelos em uma reação PCR sobreposta com iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e 3A-12-*sac* (SEQ ID NO: 107) utilizando as seguintes condições PC: Mist. 1 e perfil termocíclico: 94°C - 30 segundos, 50°C-70°C gradiente - 30 segundos, 72°C - 30 segundos por 20 ciclos. O amplicon resultante, designado a sequência de codificação *v4F* (SEQ ID NO: 31) que codifica a toxina híbrida V4F (SEQ ID NO: 32), foi clonado em um vetor pCR2.1-TOPO e designado *v4F/pCR2.1*. A proteína híbrida V4F compreende, do N-termino ao C-termino, 1-468 aminoácidos de uma proteína Cry3A055 (SEQ ID NO: 70), aminoácidos 477-520, compreendendo região variável 4, de uma proteína Cry1Ab (SEQ ID NO: 72), e aminoácidos 512-598 de uma proteína Cry3A055 (SEQ ID NO: 70).

A proteína V4F teve duas posições de cruzamento. O primeiro cruzamento entre Cry3A055 e Cry1Ab está localizado no bloco conservado 3 e o segundo cruzamento entre Cry1Ab e Cry3A055 está localizado no bloco conservado 4. Então, V4F é uma variante de Cry3A055 onde toda a região

variável 4 foi substituída pela região variável 4 de uma proteína Cry1Ab. A proteína híbrida V4F não foi ativa contra o verme de raiz de milho ocidental sugerindo que a sequência Cry1Ab na porção C-terminal de FR8a contribui para a atividade inseticida de FR8a.

- 5 Um fragmento BamHI-SacI de v4F/pCR2.1 foi sub-clonado em pET21. A proteína expressa pelo plasmídeo resultante foi designada T7-V4F.

Exemplo 19: Construção de 5*V4F.

- Um fragmento BamHI-XbaI de um plasmídeo compreendendo
 10 *FR8a* (*Vide* Exemplo 4) e um fragmento XbaI-SacI de V4F/pCR2.1 (*Vide* Exemplo 18) foram ligados a pET21 cortado com BamHI-SacI para formar 5*V4F/pET21. A sequência de codificação 5*V4F (SEQ ID NO: 33) codifica a proteína quimérica 5*V4F (SEQ ID NO: 34). A proteína inseticida quimérica 5*V4F compreende, do N-terminal ao C-terminal, um fragmento de peptidila
 15 compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGR-QQHRG (SEQ ID NO: 127), aminoácidos 10-491 de uma proteína Cry3A055 (SEQ ID NO: 70), aminoácidos 501-520, compreendendo a região variável 4, de uma proteína Cry1Ab (SEQ ID NO: 72), e aminoácidos 512-598 de uma proteína Cry3A055 (SEQ ID NO: 70).

- 20 A 5*V4F eHIP é a proteína híbrida V4F com um fragmento de peptidila N-terminal (SEQ ID NO: 127) adicionado. A 5*V4F eHIP forneceu atividade inseticida contra o verme de raiz de milho ocidental embora não no mesmo nível de FR8a. Então, o N-C-terminal conferiu atividade inseticida a V4F confirmando que pode haver interação contribuinte entre a porção C-
 25 terminal e o fragmento de peptidila N-terminal de FR8a.

A proteína expressa pelo plasmídeo 5*V4F/pET21 foi designada T7-5*V4F e possui uma etiqueta T7 N-terminal ao fragmento de peptidila 5*V4F.

Exemplo 20: Construção de 2OL-7.

- 30 Um fragmento de ácido nucleico codificando região variável 1 de uma Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores 5'3A-1-bam (SEQ ID NO:

83) e *C1-3A-2* (SEQ ID NO: 92) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Um fragmento de ácido nucleico codificando regiões variáveis 2-6 de uma *Cry1Ab* foi amplificado PCR de um plasmídeo compreendendo
 5 *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores *C1-1Ab-1* (SEQ ID NO: 110) e *1Ab-6-sac* (SEQ ID NO: 86) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Os dois amplicons resultantes foram utilizados como os modelos em uma reação PCR sobreposta com iniciadores *5'3A-1-bam* (SEQ ID NO: 83) e *1Ab-6-sac* (SEQ ID NO: 86) utilizando Mistura de reação PCR 2 e Perfil termocíclico 1, para criar a sequência de codificação *2OL-7* (SEQ ID NO: 35) que codifica a proteína híbrida *2OL-7* (SEQ ID NO: 36). A proteína híbrida *2OL-7* compreende, do N-termino ao C-termino, 1-156 aminoácidos de uma proteína *Cry3A055* (SEQ ID NO: 70), que compreende região variável 1
 15 e os aminoácidos 14 N-terminal do bloco conservado 1, e aminoácidos 167-648 de uma proteína *Cry1Ab* (SEQ ID NO: 72), que compreende os aminoácidos 15 C-terminal do bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5 e região variável 6, e 38
 20 aminoácidos das regiões dorsais de protoxina *Cry1Ab*. Então, *2OL-7* é uma variante de uma proteína *Cry1Ab* com região variável 1 substituída pela região variável 1 de uma proteína *Cry3A055*.

A sequência de codificação *2OL-7* foi clonada em pCR2.1-TOPO (Invitrogen) e então movida para pET21a utilizando *Bam*HI/*Sac*I que foi designado *2OL-7/pET21a*. A sequência de codificação em *2OL-7/pET21a* foi
 25 designada *T7-2OL-7* (SEQ ID NO: 37). A proteína expressa pelo vetor *2OL-7/pET21a* foi designada *T7-2OL-7* (SEQ ID NO: 38).

Exemplo 21: Construção de 5*2OL-7.

Um fragmento *Bam*HI/*Xba*I de *FR8a* (Vide Exemplo 4), um fragmento *Ppu*MI/*Sac*I *2OL-7* (Vide Exemplo 20) e um fragmento *Bam*HI/*Sac*I de pET21a foram ligados para produzir *5*2OL-7/pET21a*. A sequência de codificação *5*2OL-7* (SEQ ID NO: 39) codifica a proteína quimérica *5*2OL-7*
 30

(SEQ ID NO: 40). A proteína 5*2OL-7 compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-156 de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 167-643 de uma proteína Cry1Ab (SEQ ID NO: 72). Então, a proteína híbrida 5*2OL-7 é a proteína híbrida 2OL-7 com um fragmento de peptidila N-terminal adicionado.

Exemplo 22: Construção de 2OL-10.

Um fragmento de ácido nucleico codificando uma porção N-terminal de uma proteína Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e C2-3A-4 (SEQ ID NO: 88) e Mistura de reação PCR 1 e Perfil termocíclico 1. Um fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry1Ab foi amplificado PCR de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores C2-3A-3 (SEQ ID NO: 95) e 1Ab-6-*sac* (SEQ ID NO: 86) e Mistura de reação PCR 1 e Perfil termocíclico 1. Estes 2 produtos PCR foram então utilizados como modelos em uma reação PCR sobreposta com iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e 1Ab-6-*sac* (SEQ ID NO: 86) utilizando as seguintes condições PCR: Mist. 2, perfil termocíclico: 94°C - 30 segundos, 45°C-65°C gradiente - 30 segundos, 72°C - 30 segundos por 20 ciclos, resultando na sequência de codificação 2OL-10 (SEQ ID NO: 41) que codifica a toxina híbrida 2OL-10 (SEQ ID NO: 42). A proteína 2OL-10 compreende, do N-termino ao C-termino, 1-232 aminoácidos de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 243-648 de uma proteína Cry1Ab (SEQ ID NO: 72). Então, a proteína híbrida 2OL-10 é substancialmente Domínio I de uma proteína Cry3A055 e Domínios II e III de uma proteína Cry1Ab.

A sequência de codificação 2OL-10 foi clonada em pCR2.1-TOPO (Invitrogen) e então movida a pET21a utilizando BamHI/SacI. A proteína expressa por 2OL-10/pET21a foi designada T7-2OL-10.

Exemplo 23: Construção de 5*2OL-10.

Um fragmento BamHI-XbaI de um plasmídeo compreendendo

FR8a (Vide Exemplo 4) e um fragmento *XbaI-SacI* de 2OL-10/pCR2.1 (Vide Exemplo 22) foram ligados a pET21 cortado com *BamHI-SacI* para formar 5*2OL-10/pET21. A sequência de codificação 5*2OL-10 (SEQ ID NO: 43) codifica a proteína quimérica 5*2OL-10 (SEQ ID NO: 44). A proteína 5*2OL-10 compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-232 de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 243-648 de uma proteína Cry1Ab (SEQ ID NO: 72). Então, a proteína quimérica 5*2OL-10 é a proteína híbrida 2OL-10 com um fragmento de peptidila N-terminal adicionado.

Exemplo 24: Construção de 2OL-12A.

Um primeiro fragmento de ácido nucleico codificando uma porção N-terminal de Cry1Ab foi amplificado PCR de um plasmídeo compreendendo *mcry1Ab* (SEQ ID NO: 71) utilizando iniciadores 5'1Ab-bam (SEQ ID NO: 98) e C3-1Ab-2 (SEQ ID NO: 105) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Um segundo fragmento de ácido nucleico codificando uma porção C-terminal de Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores C3-3A-5 (SEQ ID NO: 106) e 3A-12-sac (SEQ ID NO: 107) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Os primeiro e segundo fragmentos de ácido nucleico descritos acima foram conectados utilizando-os como modelos em uma reação PCR sobreposta com iniciadores 5'1Ab-bam (SEQ ID NO: 98) e 3A-12-sac (SEQ ID NO: 107) utilizando Mistura 1 e Perfil termocíclico 1, para criar a sequência de codificação 2OL-12A (SEQ ID NO: 45) que codifica a proteína híbrida 2OL-12A eHIP (SEQ ID NO: 46). A proteína 2OL-12A compreende, do N-termino ao C-termino, 1-476 aminoácidos de uma proteína Cry1Ab (SEQ ID NO: 72), e aminoácidos 469-598 de uma proteína Cry3A055 (SEQ ID NO: 70).

A 2OL-12A eHIP não foi ativa contra o verme de raiz de milho ocidental, porém foi ativa contra a broca de milho européia (Vide Tabela 6).

Isto demonstra que eHIP pode ser construída utilizando proteínas Cry lepidopteran ativa e coleopteran ativa sem perda de atividade contra as espécies de insetos lepidopteran.

5 A sequência de codificação 2OL-12A foi clonada em pCR2.1-TOPO (Invitrogen) e então movida a pET21a com BamHI/SacI. A proteína expressa pelo vetor 2OL-12A/pET21a foi designada T7-2OL-10A.

Exemplo 25: Construção de 2OL-13.

Quatro fragmentos de ácido nucleico foram gerados conforme a seguir: o fragmento 1 foi amplificado PCR de um plasmídeo compreendendo
 10 *cry3A055* (SEQ ID NO: 69) utilizando iniciadores 5'3A-1-bam (SEQ ID NO: 83) e C1-3A-2 (SEQ ID NO: 92) e Mistura de reação PCR 1 e Perfil termocíclico 1; o fragmento 2 foi amplificado de um plasmídeo compreendendo *mo-cry1Ab* (SEQ ID NO: 71) utilizando iniciadores C2-3A-3 (SEQ ID NO: 95) e C3-1Ab-2 (SEQ ID NO: 105) e Mistura de reação PCR 1 e Perfil termocíclico
 15 1; o fragmento 3 foi PCR amplificado de um plasmídeo compreendendo *mo-cry1Ab* (SEQ ID NO: 71) utilizando iniciadores C3-1Ab-3 (SEQ ID NO: 85) e C4-3A-10 (SEQ ID NO: 108) e Mistura de reação PCR 1 e Perfil termocíclico 1; e o fragmento 4 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores C4-3A-9 (SEQ ID NO: 109)
 20 e 3A-12-sac (SEQ ID NO: 107) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Todos os quatro fragmentos foram então utilizados como modelos em uma reação PCR sobreposta utilizando iniciadores 5'3A--bam (SEQ ID NO: 83) e 3A-12-sac (SEQ ID NO: 107) utilizando Mistura de reação PCR
 25 1 e Perfil termocíclico 1, para criar a sequência de codificação 2OL-13 (SEQ ID NO: 47) que codifica a toxina híbrida 2OL-13 (SEQ ID NO: 48). A proteína 2OL-13 compreende, do N-termino ao C-termino, 1-159 aminoácidos de uma proteína Cry3A055 (SEQ ID NO: 70), aminoácidos 170-522 de uma proteína Cry1Ab (SEQ ID NO: 72), e aminoácidos 515-598 de uma proteína
 30 Cry3A055 (SEQ ID NO: 70). Então, a toxina híbrida 2OL-13 é composta de V1 e a porção N-terminal de CB1 de uma proteína Cry3A055; a porção C-terminal de CB1, V2, CB2, V3, CB3 e V4 de uma proteína Cry1Ab; e CB4,

V5 e CB5 de uma proteína Cry3A055.

A sequência de codificação 2OL-13 foi clonada em pCR2.1-TOPO (Invitrogen) e então movida a pET21a utilizando BamHI/SacI. A proteína expressa pelo vetor 2OL-13/pET21a foi designada T7-2OL-13.

5 Exemplo 26: Construção de 2OL-20.

Um fragmento BamHI/Nspl de um plasmídeo compreendendo *mocry3A* (SEQ ID NO: 67), um fragmento Nspl/HindIII de um plasmídeo compreendendo 2OL-8A (SEQ ID NO: 1), e um fragmento HindIII/BamHI de pET21a foram ligados para fazer 2OL-20/pET21a.

10 Exemplo 27: Construção de V5&6.

Um fragmento de ácido nucleico codificando uma porção N-terminal de Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores 5'3A-1-bam (SEQ ID NO: 83) e C4-3A-10 (SEQ ID NO: 108) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Um fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry1Ab foi amplificado PCR de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores C4-3A-9 (SEQ ID NO: 109) e 1Ab-6-sac (SEQ ID NO: 86) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Estes dois produtos PCR foram então utilizados como os modelos em uma reação PCR sobreposta com iniciadores 5'3A-1-bam (SEQ ID NO: 83) e 1Ab-6-sac (SEQ ID NO: 86) utilizando Mistura de reação PCR 1 e Perfil termocíclico 2, para criar a sequência de codificação V5&6 (SEQ ID NO: 49) que codifica a toxina híbrida V5&6 (SEQ ID NO: 50). A proteína V5&6 compreende, do N-termino ao C-termino, 1-524 aminoácidos de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende V1, CB1, V2, CB2, V3, CB3, V4 e CB4, e aminoácidos 533-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende V5, CB5 e V6, e 38 aminoácidos de uma região dorsal de protoxina Cry1Ab.

A sequência de codificação V5&6 foi clonada em pCR2.1-TOPO e então movida para pET21 com BamHI/SacI. A proteína expressa por

V5&6/pET21a foi designada T7-V5&6.

Exemplo 28: Construção de 5*V5&6.

Um fragmento BamHI/XbaI de *FR8a* (Vide Exemplo 4), um fragmento XbaI/SacI de V5&6 (Vide Exemplo 27) e um fragmento BamHI/SacI de pET21a foram ligados para formar 5*V5&6/pET21. A sequência de codificação 5*V5&6 (SEQ ID NO: 51) codifica a proteína quimérica 5*V5&6 (SEQ ID NO: 52). A proteína inseticida quimérica 5*V5&6 compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-524 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende V1, CB1, V2, CB2, V3, CB3, V4 e CB4, e aminoácidos 533-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende V5, CB5 e V6, e 38 aminoácidos de uma região dorsal de protoxina Cry1Ab. Então, a proteína inseticida quimérica 5*V5&6 é a proteína híbrida V5&6 com um fragmento de peptidila N-terminal adicionado.

Exemplo 29: Construção de 88A-dm3.

Um fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry8Aa (SEQ ID NO: 78) foi amplificada PCR de um plasmídeo compreendendo *cry8Aa* (SEQ ID NO: 77) utilizando iniciadores 5'8Aa-dm3 (SEQ ID NO: 111) e 3'8Aa-dm3 (SEQ ID NO: 112) e Mistura de Reação PCR 2 e Perfil termocíclico 2. O amplicon resultante foi clonado em pCR2.1-TOPO (Invitrogen) e designado 88A-dm3/pCR2.1.

Um fragmento MluI/SacI de 88A-dm3/pCR2.1 e um fragmento SacI/MluI de um plasmídeo compreendendo *FR8a* (Vide Exemplo 4) foram ligados para criar a sequência de codificação 88A-dm3 (SEQ ID NO: 53) que codifica a proteína híbrida 88A-dm3 (SEQ ID NO: 54). A proteína 88A-dm3 compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-468 de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 532-664 de uma proteína Cry8Aa (SEQ ID NO: 78).

A sequência de codificação 88A-dm3 foi também transformada em pET21a utilizando uma digestão restrita e ligação BamHI/SacI. A proteí-

na expressa por 88A-dm3/pET21a foi designada T7-88A-dm3.

Exemplo 30: Construção de FR(1Fa)

Um fragmento de ácido nucleico codificando uma porção N-terminal de FR8a (*Vide* Exemplo 3) foi amplificado PCR de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 1) utilizando iniciadores C2-3A-3 (SEQ ID NO: 95) e *tant-OL-2* (SEQ ID NO: 113) e Mistura de reação PCR 3 e Perfil termocíclico 3.

Um fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry1Fa (SEQ ID NO: 76) foi amplificada PCR de um plasmídeo compreendendo *mocry1Fa* (SEQ ID NO: 75) utilizando iniciadores *tant-OL-1* (SEQ ID NO: 114) e *tant-3'sac* (SEQ ID NO: 115) e Mistura de reação PCR 3 e Perfil termocíclico 3.

Estes dois produtos PCR resultantes foram então utilizados como modelos em uma reação PCR sobreposta com iniciadores C2-3A-3 (SEQ ID NO: 95) e *tant-3'sac* (SEQ ID NO: 115) utilizando Mistura de Reação PCR 3 e Perfil termocíclico 3. O produto PCR resultante foi clonado em pCR2.1-TOPO (Invitrogen). Um fragmento de BamHI/MluI de um plasmídeo compreendendo *FR8a*, um fragmento MluI/SacI do produto PCR sobreposto em pCR2.1 e um fragmento BamHI/SacI de pET21a foram então ligados para criar FR(1Fa)/pET21a. A sequência de codificação *FR(1Fa)* (SEQ ID NO: 55) codifica a proteína quimérica FR(1Fa) (SEQ ID NO: 56). A proteína FR(1Fa) compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-468 de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 470-649 de uma proteína Cry1Fa (SEQ ID NO: 76).

Exemplo 31: Construção de FR(1Ac)

Os Domínios I & II de FR8a foram amplificados PCR de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 1) utilizando iniciadores C2-3A-3 (SEQ ID NO: 95) e *1Ac-OL-2* (SEQ ID NO: 116) e Mistura de reação PCR 3 e Perfil termocíclico 3. O Domínio III de Cry1Ac (SEQ ID NO: 80) foi amplificado PCR de um plasmídeo compreendendo *cry1Ac* (SEQ ID NO: 79) utilizando iniciadores *1Ac-OL-1* (SEQ ID NO: 117) e *1Ac-3'sac* (SEQ ID NO:

118) e Mistura de reação PCR 3 e Perfil termocíclico 3.

Estes 2 produtos PCR resultantes foram utilizados como modelos em uma reação PCR sobreposta com iniciadores C2-3A-3 (SEQ ID NO: 95) e 1Ac-3'sac (SEQ ID NO: 118) e as seguintes condições: Mist. 3 e perfil termocíclico: 94°C - 30 segundos, 68°C - 30 segundos, 68°C - 30 segundos por 20 ciclos. O produto PCR sobreposto foi clonado em pCR2.1-TOPO (Invitrogen). Um fragmento de BamHI/MluI de um plasmídeo compreendendo *FR8a*, o fragmento MluI/SacI do produto PCR sobreposto em pCR2.1 e um fragmento BamHI/SacI de pET21a foram ligados para criar FR(1Ac)/pET21a.

A proteína FR(1Ac) compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MTSNGRQCA-GIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-468 de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 477-608 de uma proteína Cry1Ac (SEQ ID NO: 80).

15 Exemplo 32: Construção de FR(1la)

Um fragmento de nucleotídeo codificando os Domínios I e II de *FR8a* foi amplificado PCR de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) utilizando iniciadores C2-3A-3 (SEQ ID NO: 95) e 1la-OL-2 (SEQ ID NO: 119) e Mistura de reação PCR 3 e Perfil termocíclico 3. Um segundo

fragmento nucleotídeo codificando o Domínio III de uma proteína Cry1la (SEQ ID NO: 82) foi amplificada PCR de um plasmídeo compreendendo *cry1la* (SEQ ID NO: 81) utilizando iniciadores 1la-OL-1 (SEQ ID NO: 120) e 1la-3'sac (SEQ ID NO: 121) e Mistura de reação PCR 3 e Perfil termocíclico 3. Estes dois produtos PCR foram utilizados como modelos em uma reação

PCR sobreposta com iniciadores C2-3A-3 (SEQ ID NO: 95) e 1la-3'sac (SEQ ID NO: 121) e Mistura de reação PCR 3 e perfil termocíclico: 94°C - 30 segundos, 68°C - 45 segundos por 20 ciclos. O produto PCR sobreposto foi clonado em pCR2.1-TOPO (Invitrogen). O fragmento de BamHI/MluI de um plasmídeo compreendendo *FR8a*, o fragmento MluI/SacI do produto PCR

sobreposto em pCR2.1 e um fragmento BamHI/SacI de pET21a foram ligados para criar FR(1la)/pET21a. A proteína FR(1la) compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequên-

cia de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO: 127), aminoácidos 10-468 de uma proteína Cry3A055 (SEQ ID NO: 70), e aminoácidos 513-719 de uma proteína Cry1Ia (SEQ ID NO: 82).

Exemplo 33: Construção de Dm2-3A

5 Parte da extremidade 5' desta sequência de codificação foi amplificada PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores C2-3A-3 (SEQ ID NO: 95) e FR-1Ab-2 (SEQ ID NO: 122) e Mistura de reação PCR 3 e Perfil termocíclico 2. Um fragmento de nucleotídeo codificando o Domínio III de Cry1Ab foi PCR amplificado de um

10 plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores FR1Ab-1 (SEQ ID NO: 123) e 1Ab-6-sac (SEQ ID NO: 86) e Mistura de reação PCR 3 e Perfil termocíclico 2. Estes dois produtos PCR foram utilizados como modelos em uma reação PCR sobreposta com iniciadores C2-3A-3 (SEQ ID NO: 95) e 1Ab-6-sac (SEQ ID NO: 86) e Mistura de Reação PCR 3

15 e Perfil termocíclico 2. O amplicon resultante foi clonado em pCR2.1-TOPO (Invitrogen). FR8a BamHI/MluI, e o produto PCR acima em pCR2.1-TOPO AflIII, FR8a AflIII/SacI foram ligados a pET21a BamHI/SacI. Toda a sequência de codificação (BamHI/SacI) foi então movida para 1454. A proteína inseticida quimérica DM2-3A compreende, do N-termino ao C-termino, um frag-

20 mento de peptidila compreendendo a sequência de aminoácido MTSNGRQCAGIRPYDGRQQHRG (SEQ ID NO : 127), aminoácidos 10-451 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 7 N-terminal do bloco conservado 3, e aminoácidos 460-648

25 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 41 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6. Então, a DM2-3A eHIP foi uma junção cruzada entre Cry3A055 e Cry1Ab localizada no bloco conservado 3 seguindo seguindo imediatamente Ser451 que está a jusante

30 da junção do domínio II e domínio III. A DM2-3A teve atividade inseticida contra o verme de raiz de milho ocidental, porém a atividade foi menor do que de a 8AF e FR8a eHIPs conforme apresentado na Tabela 5.

Exemplo 34: Construção de T7-8AF.

Um fragmento de ácido nucleico codificando uma porção N-terminal de uma proteína Cry3A055 (SEQ ID NO: 70) foi amplificada PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e C3-3A-6 (SEQ ID NO: 84) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Um fragmento de ácido nucleico codificando uma porção C-terminal de uma proteína Cry1Ab (SEQ ID NO: 72) foi amplificada PCR de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores C3-1Ab-3 (SEQ ID NO: 85) e 1Ab-6-*Sac* (SEQ ID NO: 86) e Mistura de reação PCR 1 e Perfil termocíclico 1.

Os dois produtos PCR descritos acima foram depois utilizados como modelos em uma reação PCR sobreposta com iniciadores 5'3A-1-*bam* (SEQ ID NO: 83) e 1Ab-6-*Sac* (SEQ ID NO: 86) utilizando Mistura de reação PCR 2 e Perfil termocíclico 1.

O amplicon resultante foi ligado como um fragmento de extremidade inflexível a um vetor pCR2.1-TOPO (Invitrogen, Carlsbad, CA) cortado com *Sma*I para formar plasmídeo p8AF/CR2.1. Um fragmento *Bam*HI-*Sac*I de p8AF/CR2.1 foi então ligado a pET21a (EMD Biosciences, Inc., San Diego, CA), que foi cortado com *Bam*HI-*Sac*I, e transformado em *E. coli*. A estrutura de leitura aberta foi designada T7-8AF (SEQ ID NO: 144) e codifica a proteína híbrida T7-8AF (SEQ ID NO: 145). A proteína híbrida T7-8AF compreende, do N-termino ao C-termino, um fragmento de peptidila compreendendo a sequência de aminoácido MASMTGGGQMGRGS (aminoácidos 1-14 de SEQ ID NO: 126), aminoácidos 1-468 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal da protoxina Cry1Ab. A proteína híbrida T7-8AF teve pouca ou nenhuma

atividade inseticida contra o verme de raiz de milho ocidental.

Exemplo 35: Construção de 8AF.

Um fragmento BamHI-SacI do plasmídeo p8AF/CR2.1 (Vide Exemplo 34) foi ligado a um plasmídeo contendo um promotor Cry1Ac constitutivo que foi modificado do descrito por Schnepf *et al.* (1985. J. Biol. Chem. 260:6264-6272) para corrigir um códon inicial ATG interno que existe no promotor de Schnepf *et al.* a um códon ATC, que foi cortado com BamHI-SacI, e transformado em *E. coli*. A estrutura de leitura aberta foi designada 8AF (SEQ ID NO: 63) e codifica a 8AF eHIP (SEQ ID NO: 64). A 8AF eHIP é similar a FR8a eHIP, porém não contém o fragmento de peptidila N-terminal opcional. A 8AF eHIP compreende, do N-termino ao C-termino, 1-468 aminoácidos de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 24 N-terminal do bloco conservado 3, e aminoácidos 477-648 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende os aminoácidos 24 C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal de uma protoxina Cry1Ab. Então, a 8AF eHIP teve uma junção cruzada entre Cry3A055 e Cry1Ab localizada no bloco conservado 3 imediatamente seguindo Leu468 do SEQ ID NO: 70 que é a jusante da junção do domínio II e domínio III. A 8AF eHIP teve alta atividade contra o verme de raiz de milho ocidental.

Exemplo 36: Construção de -catG8AF.

Um constructo foi feito sem o sítio Catepsina G (Cat G) para determinar se o sítio Cat G no domínio I da 8AF eHIP foi necessário para atividade contra o verme de raiz. Um fragmento 1359 pb BamHI/Sall de um plasmídeo compreendendo *moCry3A* (SEQ ID NO: 67) e um fragmento 3483 pb BamHI/Sall de um plasmídeo compreendendo *20L-8a* (SEQ ID NO: 1) foram ligados para criar -catG8AF (SEQ ID NO: 146) que codifica -catG8AF eHIP (SEQ ID NO: 147).

A -catG8AF eHIP foi muito ativa contra o verme de raiz de milho ocidental demonstrando que o sítio de reconhecimento de protease Catepsi-

na G na 8AF eHIP não é requerido para atividade inseticida.

Exemplo 37: Construção de 8AFdm3

A 8AF eHIP descrita no Exemplo 35 teve um ponto cruzado entre Cry3A055 e Cry1Ab localizado em CB3 a jusante da junção dos domínios II/III, resultando no domínio III da 8AF eHIP tendo uma pequena região N-terminal do domínio III de Cry3A055 e o remanescente do domínio III sendo sequência de domínio III Cry1Ab. Para determinar se a pequena região N-terminal do domínio III de Cry3A055 era requerida para atividade inseticida em 8AF, outro constructo foi feito tendo o cruzamento entre Cry3A055 e Cry1Ab localizado em CB3 exatamente na junção domínio II-domínio III.

Um fragmento de ácido nucleico codificando porção do domínio I e domínio II de Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) utilizando iniciadores *CMS96* (SEQ ID NO: 138) e *CMS97* (SEQ ID NO: 139) e Mistura de reação PCR 5 e Perfil termocíclico 5.

Um fragmento de ácido nucleico codificando o domínio III de *mocry1Ab* foi amplificado PCR de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores *CMS98* (SEQ ID NO: 140) e *CMS99* (SEQ ID NO: 141) e Mistura de reação PCR 5 e Perfil termocíclico 5.

Os dois amplicons resultantes foram utilizados como modelos em uma reação PCR sobreposta com iniciadores *CMS96* (SEQ ID NO: 138) e *CMS98* (SEQ ID NO: 140) utilizando a Mistura de Reação PCR 5 e Perfil termocíclico 6. O amplicon resultante foi clonado em pCR4 Blunt (Invitrogen, Carlsbad, CA). Um fragmento 1633bp *StuI/SacI* do amplicon clonado, designado pCR4Blunt-OLWrdm3, e um fragmento *StuI/SacI* de 3089 pb de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) foram combinados para criar *8AFdm3* (SEQ ID NO: 148) que codifica a proteína híbrida 8AFdm3 (SEQ ID NO: 149).

A proteína híbrida 8AFdm3 compreende, do N-termino ao C-termino, aminoácidos 1-454 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende os domínios I e II, que compreende a região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3,

e os aminoácidos 10 N-terminal do bloco conservado 3, e aminoácidos 463-610 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende todo o domínio III, compreendendo os 38 aminoácidos C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6.

Então, a proteína 8AFdm3 teve uma junção cruzada entre Cry3A055 e Cry1Ab imediatamente após Phe454 do SEQ ID NO: 70 que está na junção do domínio II e domínio III. A proteína 8AFdm3 não teve atividade contra o verme de raiz de milho ocidental. Isto sugere que a região N-terminal de aminoácido 24 de CB3 de Cry3A055 ou Cry3A, uma vez que possuem a mesma sequência nesta região, são necessárias para a atividade de um 8AF eHIP.

Exemplo 38: Construção de 8AFlongdm3

Para determinar se a localização da junção cruzada em CB3 entre Cry3A ou Cry3A005 e Cry1Ab era crítica para atividade de verme de raiz, um constructo foi feita onde a junção cruzada foi posicionada em CB4 imediatamente após do aminoácido 519 de uma proteína Cry3A055.

Um primeiro fragmento de ácido nucleico codificando parte do domínio I e todo o domínio II e parte do domínio III de Cry3A055 foi amplificado PCR de um plasmídeo compreendendo *cry3A055* (SEQ ID NO: 69) utilizando iniciadores *CMS96* (SEQ ID NO: 138) e *CMS101* (SEQ ID NO: 143) e Mistura de reação PCR 5 e Perfil termocíclico 5.

Um fragmento de ácido nucleico codificando parte do domínio III de Cry1Ab foi amplificado PCR de um plasmídeo compreendendo *mocry1Ab* (SEQ ID NO: 71) utilizando iniciadores *CMS98* (SEQ ID NO: 140) e *CMS100* (SEQ ID NO: 142) e Mistura de reação PCR 5 e Perfil termocíclico 5.

Os dois amplicons resultantes foram utilizados como modelos em uma reação PCR sobreposta com iniciadores *CMS96* (SEQ ID NO: 138) e *CMS98* (SEQ ID NO: 140) utilizando a Mistura de Reação PCR 5 e Perfil termocíclico 6. O amplicon resultante foi clonado em pCR4 Blunt (Invitrogen, Carlsbad, CA). Um fragmento *Sall/SacI* de aproximadamente 460bp do amplicon clonado, designado pCR4Blunt-OL8AFlongdm3, e um fragmento Sa-

II/SacI de aproximadamente 4265bp de um plasmídeo compreendendo *8AFdm3* (SEQ ID NO: 147) foram combinados para criar *8AFlongdm3* (SEQ ID NO: 150) que codifica a proteína híbrida *8AFlongdm3* (SEQ ID NO: 151).

A proteína híbrida *8AFlongdm3* compreende, do N-termino ao C-termino, aminoácidos 1-519 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende os domínios I e II, que compreende a região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, bloco conservado 3, região variável 4 e os aminoácidos 6 N-terminal do bloco conservado 4, e aminoácidos 528-610 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende uma região C-terminal do domínio III, compreendendo os aminoácidos 4 C-terminal do bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6.

Então, a proteína *8AFlongdm3* teve uma junção cruzada entre Cry3A055 e Cry1Ab no bloco conservado 4 imediatamente após Ile519 do SEQ ID NO: 70. A proteína Cry híbrida *8AFlongdm3* não teve atividade contra o verme de raiz de milho ocidental. Isto sugere que uma região crítica para atividade de verme de raiz de milho de uma Cry3A-Cry1A eHIP repousa em uma região entre aminoácidos correspondentes a aminoácido 6 de CB3 a aminoácido 7 de CB4.

20 Exemplo 39: Construção de cap8AFdm3

Um fragmento BamHI/Sall de aproximadamente 1363 pb de um plasmídeo compreendendo *8AFdm3* (SEQ ID NO: 148) e um fragmento BamHI/Sall de aproximadamente 3362bp de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) foram ligados para criar *cap8AFdm3* (SEQ ID NO: 152) que codifica a *cap8AFdm3* eHIP (SEQ ID NO: 153).

A proteína *cap8AFdm3* teve alguma atividade contra o verme de raiz de milho ocidental conforme indicado na Tabela 5. A única diferença entre a proteína híbrida *8AFdm3*, que não foi inseticida, e a *cap8AFdm3* eHIP é a presença de um fragmento de peptidila N-terminal (SEQ ID NO: 127). Então, adicionando um fragmento de peptidila a uma proteína Cry híbrida não-ativa criou uma proteína inseticida híbrida projetada ativa contra o verme de raiz.

Exemplo 40: Construção de 8AFdm3T

Um fragmento PmlI/SacI de aproximadamente 4654bp de um plasmídeo compreendendo *8AFdm3* (SEQ ID NO: 148) e um fragmento PmlI/SacI de aproximadamente 190bp de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) foram ligados para criar *8AFdm3T* (SEQ ID NO: 154) que codifica 8AFdm3T eHIP (SEQ ID NO: 155). A 8AFdm3T eHIP compreende, do N-termino ao C-termino, aminoácidos 1-454 de uma proteína Cry3A055 (SEQ ID NO: 70), que compreende os domínios I e II, que compreende a região variável 1, bloco conservado 1, região variável 2, bloco conservado 2, região variável 3, e os aminoácidos 10 N-terminal do bloco conservado 3, e aminoácidos 463-610 de uma proteína Cry1Ab (SEQ ID NO: 72), que compreende todo o domínio III, compreendendo os 38 aminoácidos C-terminal do bloco conservado 3, região variável 4, bloco conservado 4, região variável 5, bloco conservado 5, e região variável 6; e 38 aminoácidos da região dorsal de uma protoxina Cry1Ab.

A única diferença entre a proteína híbrida 8AFdm3 e a 8AFdm3T eHIP é a adição de região dorsal de protoxina Cry1Ab de aminoácido 38 indicando que a adição de região dorsal de protoxina pode alterar uma proteína Cry híbrida não-ativa em uma eHIP ativa.

20 Exemplo 41: Construção de 8AFlongdm3T

Um fragmento PmlI/SacI de aproximadamente 4693bp de um plasmídeo compreendendo *8AFlongdm3* (SEQ ID NO: 150) e um fragmento PmlI/SacI de aproximadamente 190bp de um plasmídeo compreendendo *FR8a* (SEQ ID NO: 3) foram ligados para criar *8AFlongdm3T* (SEQ ID NO: 156) que codifica a proteína Cry híbrida 8AFlongdmT (SEQ ID NO: 157).

A única diferença entre a proteína Cry híbrida 8AFlongdm3 e a proteína Cry híbrida 8AFlongdm3T, que não foi ativada contra o verme de raiz de milho ocidental, é a adição de uma região dorsal de protoxina Cry1Ab de aminoácido 38 indicando que a região de protoxina não foi por si só suficiente para conferir atividade inseticida à proteína Cry híbrida 8AFlongdm3. Isto indica que a combinação das regiões variáveis e blocos conservados em adição a uma região dorsal de protoxina e/ou um fragmento de peptidila N-

terminal pode ser necessária para criar algumas eHIPs.

Exemplo 42: Construção de cap8AFdm3 T

Um fragmento PmlI/SacI de aproximadamente 4693bp de um plasmídeo compreendendo *cap8AFdm3* (SEQ ID NO: 152) e um fragmento
 5 PmlI/SacI de aproximadamente 190bp de um plasmídeo compreendendo *FR8A* (SEQ ID NO: 3) foram ligados para criar *cap8AFdm3T* (SEQ ID NO: 158) que codifica a cap8AFdm3T eHIP (SEQ ID NO: 159).

A proteína cap8AFdm3T teve atividade aprimorada contra o verme de raiz de milho ocidental sobre a cap8AFdm3 eHIP conforme indicado na Tabela 5. A única diferença entre cap8AFdm3 eHIP, que teve alguma
 10 atividade inseticida contra o verme de raiz de milho, e a cap8AFdm3T eHIP é a presença da região dorsal de protoxina de aminoácido 38 de Cry1Ab. Então, algumas proteínas Cry híbridas podem ser ativadas pela adição de um fragmento de peptidila N-terminal a uma região dorsal de protoxina.

15 Exemplo 43: Teste de proteínas híbridas por atividade inseticida

Verme de Raiz de Milho Ocidental

As proteínas híbridas geradas nos Exemplos descritos acima foram testadas por atividade inseticida contra o verme de raiz de milho ocidental em bioanálises laboratoriais. As bioanálises foram realizadas utilizando um método de modalidade de dieta. Clones *E. coli* que expressam uma
 20 das proteínas que cresceram durante a noite. 500 µl de uma cultura noturna foram sonificados e o montante de proteína a ser testado foi determinado. A solução de proteína foi então misturada com 500 µl de dieta artificial fundida similar a aquela descrita em Marrone *et al.* (1985, J. of Economic Entomology 78:290-293). Depois que a dieta solidificou, ela foi depositada em um
 25 prato petri e 20 vermes de raiz de milho recém-nascidos foram posicionados na dieta. Os pratos petri foram mantidos em aproximadamente 30°C. A mortalidade foi registrada após 6 dias.

Os resultados das bioanálises são apresentados na Tabela 5. A
 30 coluna 1 indica os nomes das proteínas Cry híbridas, proteínas inseticidas híbridas projetadas e inseticidas projetadas quiméricas. A coluna 2 indica os níveis relativos de atividade de verme de raiz de milho ocidental ("-" = < 40%

- de mortalidade; "+" = 40-49% de mortalidade; "++" = 50- 59% de mortalidade; "+++" = 60-80% de mortalidade; e "++++" = >80% de mortalidade). A coluna 3 indica níveis relativos das proteínas apropriadas detectadas por Western blot. A coluna 4 indica a presença de fragmento de peptidila ("-" =
- 5 Sem fragmento de peptidila; #1 = SEQ ID NO: 126; #2 = SEQ ID NO: 127; #3 = SEQ ID NO: 128; #4 = SEQ ID NO: 129; #5 = SEQ ID NO: 130; #6 = SEQ ID NO: 131; #7 = SEQ ID NO: 132). As colunas 5-7 apresentam as combinações e arranjos das regiões variáveis (V1-V6), blocos conservados (C1-C5) e domínios associados (Domínio I-III) de uma primeira proteína Cry
- 10 Bt ou proteína Cry modificada e uma segunda proteína Cry Bt diferente da primeira proteína Cry ou proteína Cry modificada que forma uma proteína híbrida central, que não é ativada contra verme de raiz de milho ocidental e eHIPs, que possui atividade contra verme de raiz de milho ocidental. A coluna 8 indica o número de aminoácidos na região dorsal de uma protoxina e
- 15 proteína Cry da qual a região dorsal é derivada ("1Ab-38" = 38 aminoácidos de região dorsal de uma protoxina; "1Ba-18" = 18 aminoácidos de região dorsal de uma protoxina Cry1Ba).

Tabela 5: Resultados de bioanálises de verme de raiz de milho ocidental.

Proteínas testadas	Acti-vida-de CRW	Pro-teí-nas ex-pres-sas	Frag-mento de Pep-tidila	Domínio I				Domínio II		Domínio III						Região de Protoxina
				V1	C1	V2	C2	V3	C3	V4	C4	V5	C5	V6		
8AF	++++	++	-	3A055				3A055		3A055		1Ab				1Ab-38
T7-8AF	-	+	#7	3A055				3A055		3A055		1Ab				1Ab-38
-CatG8AF	++++	++	-	3A				3A		3A		1Ab				1Ab-38
8AFdm3	-	+	-	3A055				3A055		1Ab						-
8AFdm3T	+++	++	-	3A055				3A055		1Ab						1Ab-38
8AFlongdm3	-	+	-	3A055				3A055		3A055		1Ab				-
8AFlongdm3 T	-	+	-	3A055				3A055		3A055		1Ab				1Ab-38
Cap8AFdm3	+	+	#2	3A055				3A055		1Ab						-
Cap8AFdm3 T	++	++	#2	3A055				3A055		1Ab						1Ab-38
2OL-8a	++++	++	#1	3A055				3A055		3A055		1Ab				1Ab-38
FR8a +34	++++	++	#6	3A055				3A055		3A055		1Ab				1Ab-38
FR8a	++++	++	#2	3A055				3A055		3A055		1Ab				1Ab-38
FRCG	++++	++	#2	3A				3A		3A		1Ab				1Ab-38

Tabela 5: (continuação)

Proteínas testadas	Ati-vida-de CRW	Pro-teí-nas ex-pres-sas	Frag-mento de Pep-tidila	Domínio I				Domínio II		Domínio III						Região de Protoxina
				V1	C1	V2	C2	V3	C3	V4	C4	V5	C5	V6		
FR8a-9F	+++	++	#5	3A055				3A055		3A055	1Ab				1Ab-38	
FR8a-9F-catg	++++	++	#5	3A				3A		3A	1Ab				1Ab-38	
FR8a-12aa	++++	++	#3	3A055				3A055		3A055	1Ab				1Ab-38	
Cry3A055	++++	++	-	3A055				3A055		3A055				-		
5*Cry3A055	-	++	#2	3A055				3A055		3A055				-		
Wr-9mut	-	++	#3	3A055				3A055		3A055				-		
FRD3	++++	++	#2	3A055				3A055		3A055	1Ab				-	
FR-12-cg-dm3	++	++	#3	3A055				3A055		3A055	1Ab				-	
9F-cg-del6	-	++	#5	3A				3A		3A	1Ab				1Ab-38	
FR-cg-dm3	++++	++	#2	3A				3A		3A	1Ab				-	
9F-cg-dm3	++++	++	#5	3A				3A		3A	1Ab				-	
B8a	-	+	-	3A055				3A055		3A055	1Ba				1Ba-18	
5*B8a	-	+	#2	3A055				3A055		3A055	1Ba				1Ba-18	

Tabela 5: (continuação)

Proteínas testadas	Acti-vida-de CRW	Pro-teí-nas ex-pres-sas	Frag-mento de Pep-tidila	Domínio I				Domínio II		Domínio III						Região de Protoxina
				V1	C1	V2	C2	V3	C3	V4	C4	V5	C5	V6		
V3A	++	+	-	3A055			1Ab		3A055						-	
V4F	-	++	-	3A055			3A055		1Ab	3A055					-	
5*V4F	++	+	#2	3A055			3A055		1Ab	3A055					-	
2OL-7	-	++	-	3A055	1Ab		1Ab		1Ab						1Ab-38	
5*2OL-7	-	+	#2	3A055	1Ab		1Ab		1Ab						1Ab-38	
2OL-10	-	+	-	3A055			1Ab		1Ab						1Ab-38	
5*2OL-10	+/-	+/-	#2	3A055			1Ab		1Ab						1Ab-38	
2OL-12A	-	++	-	1Ab			1Ab		3A						-	
2OL-13	-	-	-	3A055			1Ab		1Ab	3A055					-	
2OL-20	-	+	-	3A			3A		3A	1Ab					1Ab-38	
V5&6	-	++	-	3A055			3A055		3A055	1Ab					1Ab-38	
5*V5&6	-	++	#2	3A055			3A055		3A055	1Ab					1Ab-38	
88A-dm3	-	++	#2	3A055			3A055		3A055	8Aa					-	
FR(1Fa)	-	++	#2	3A055			3A055		3A055	1Fa					-	

Tabela 5: (continuação)

Proteínas testadas	Acti-vida-de CRW	Pro-teí-nas ex-pres-sas	Frag-mento de Pep-tidila	Domínio I				Domínio II		Domínio III						Região de Protoxina
				V1	C1	V2	C2	V3	C3	V4	C4	V5	C5	V6		
FR(1Ac)	-	+	#2	3A055				3A055		3A055	1Ac				-	
FR(1Ia)	-	-	#2	3A055				3A055		3A055	1Ia				-	
DM23A	+	+	#2	3A055				3A055		3A055	1Ab				1Ab-38	

As proteínas inseticidas quiméricas, 2OL-8a e FR8a, e a 2OL-12A eHIP, foram testadas contra diversas espécies de insetos para determinar o espectro de atividade. Os insetos testados incluíam verme de raiz de milho ocidental (WCR), verme de raiz de milho do norte (NCR), verme de raiz de milho do sul (SCR), besouro de batata do Colorado (CPB), e broca de milho européia (ECB). Os resultados das análises são apresentados na Tabela 6. Um "+" indica atividade inseticida. Um "-" indica sem atividade. As CIPs 2OL-8a e FR8a foram ativadas contra WCR, NCR e CPB. A 2OL-12A eHIP foi surpreendentemente ativada contra ECB.

10 **Tabela 6:** Espectro de atividades de CIPs.

	Espectro de Atividade				
Proteína	WCR	NCR	SCR	CPB	ECB
2OL-8a	+	+	-	+	-
FR8a	+	+	-	+	-
2OL-12A	-	nt	nt	nt	+
Cry3A055	+	+	-	+	-
Cry3A	-	-	-	+	-
Cry1Ab	-	-	-	-	+

Exemplo 44: Inserção de genes codificando eHIPs em plantas

Três genes codificando as proteínas inseticidas quiméricas FR8a, FRCG e FRD3 foram selecionados para transformação em plantas de milho. Umum cassete de expressão compreendendo a sequência de codificação FR8a ou FRCG ou FRD3 foi transferida a um vetor adequado para transformação de milho mediada por *Agrobacterium*. Para este exemplo, os seguintes vetores foram utilizados nos experimentos de transformação: 12207 (figura 3), 12161 (figura 4), 12208 (figura 5), 12274 (figura 6), 12473 (figura 7) e 12474 (figura 8).

20 A transformação de embriões de milho imaturos foi realizada essencialmente conforme descrito em Negrotto *et al.*, 2000, Plant Cell Reports 19: 798-803. Para este exemplo, todos os constituintes de meio foram essencialmente conforme descrito em Negrotto *et al.*, *supra*. Contudo, vários constituintes de meio conhecidos na técnica podem ser utilizados.

Os genes utilizados para transformação foram clonados em um vetor adequado para transformação de milho. Os vetores utilizados neste exemplo contêm o gene fosfomanose *isomerase* (PMI) para seleção de linhas transgênicas (Negrotto *et al.*, *supra*).

5 Resumidamente, a cepa *Agrobacterium* LBA4404 (pSB1) contendo um plasmídeo de transformação de planta foi cultivada em YEP (extra-
to de levedura (5 g/L), peptona (10g/L), NaCl (5g/L), 15g/l agar, pH 6,8) meio
sólido por 2 - 4 dias a 28°C. Aproximadamente $0,8 \times 10^9$ *Agrobacterium* fo-
ram suspensas em meio LS-inf suplementado com 100 µM As (Negrotto *et*
10 *al.*, *supra*). As bactérias foram pré-induzidas neste meio por 30-60 minutos.

Embriões imaturos de A188 ou outros genótipos adequados fo-
ram cortados de 8 - 12 dias de idade em LS-inf + 100 µM As líquido. Os em-
briões foram lavados uma vez com meio de infecção fresco. Solução de
Agrobacterium foi então adicionada e os embriões foram vertidos por 30 se-
15 gundos e deixados para assentar com a bactéria por 5 minutos. Os embriões
foram então transferidos scutellum lateralmente para meio LSAs e cultivados
no escuro por dois a três dias. Subsequentemente, entre 20 e 25 embriões
por prato petri foram transferidos a meio LSDc suplementado com cefotaxi-
ma (250 mg/l) e nitrato de prata (1,6 mg/l) e cultivados no escuro a 28°C por
20 10 dias.

Os embriões imaturos, produzindo callus embriogênico foram
transferidos para meio LSD1M0.5S. As culturas foram selecionadas neste
meio por cerca de 6 semanas com um passo de subcultura em cerca de 3
semanas. Os calli sobreviventes foram transferidos para meio Reg1 suple-
25 mentado com mannose. Seguindo cultura na luz (regime de 16 horas de luz/
8 horas no escuro), tecidos verdes foram transferidos a meio Reg2 sem re-
guladores de cultura e incubados por cerca de 1-2 semanas. As plantas fo-
ram transferidas para caixas Magenta GA-7 (Magenta Corp, Chicago Ill.)
contendo meio Reg3 e cultivadas à luz. Após cerca de 2-3 semanas, as
30 plantas foram testadas pela presença de gene *pmi* e os genes *FR8a* ou
FRCG por PCR. Plantas positivas da análise PCR foram transferidas para a
estufa e testadas para resistência ao verme de raiz de milho.

Exemplo 45: Análise de Plantas de Milho Transgênicas para Eficácia contra Vermes de Raiz de Milho: Bioanálise de Excisão de Raiz

Normalmente, as plantas de milho são amostradas conforme são transferidas das caixas Magenta GA-7 para o solo. Isto permite que as raízes sejam amostradas a partir de um ambiente razoavelmente estéril para condições de solo. A amostragem consiste do corte de uma pequena parte da raiz (ca. 2-4 cm comp.) e posicioná-la em phytagar enriquecido (phytagar, 12 g., sacarose, 9 g., sais MS, 3 ml., vitaminas MS, 3 ml., Nistatina(25mg/ml), 3 ml., Cefotaxima (50mg/ml), 7 ml., Aureomicina (50 mg/ml), 7 ml., Estreptomicina (50mg/ml), 7 ml., dH₂O, 600 ml) em um pequeno prato petri. Controles negativos são plantas transgênicas que são negativas PCR para o gene *FR8a* ou *FRCG* do mesmo experimento de transformação, ou de plantas não-transgênicas (de um tamanho similar das plantas de teste) que estão sendo cultivadas em phytotron.

As raízes são também amostradas depois que as plantas tenham crescido no solo. Se as amostras de raízes são do solo, os pedaços de raiz são lavados com água para remover resíduo de solo, mergulhadas em solução de Nistatina (5mg/ml), removidas do mergulho, secas com papel toalha e posicionadas em um prato phytagar conforme acima.

As amostras de raiz são inoculadas com vermes de raiz de milho ocidental posicionando-se cerca das 10 primeiras larvas na superfície interna de um controle de cada prato phytagar e os controles são então suavemente soltos sobre o pedaço de raiz exposto. As larvas foram manejadas utilizando uma camada fina de tinta. Depois que todos os pratos foram inoculados, a bandeja de pratos foi colocada no escuro a temperatura ambiente até a data da coleta.

Em cerca de 2-4 dias após a inoculação de raiz, os dados foram coletados. A mortalidade percentual das larvas foi calculada com classificação de dano visual das raízes. Danos de alimentação foram classificados através da observação do número de orifícios de alimentação (FH) no pedaço de raiz causados pelas larvas de vermes de raiz e foram classificados como alto, moderado, baixo ou ausente e dado um valor numérico de cate-

goria 3, 2 ou 1, respectivamente (com Categoria 1 incluindo classificações de dano de ausente e/ou baixo). As plantas de Categoria 1 tiveram normalmente 0-FH a 2-FH; plantas de Categoria 2 tiveram 3 a 4-FH; e plantas de Categoria 3 tiveram >5-FH. As amostras de raiz tendo uma classificação de dano na Categoria 1 foram consideradas excelentes, categoria 2: médias e categoria 3: pobres. As plantas de Categoria 1 foram selecionadas para maiores testes na estufa e em campo.

Os resultados na Tabela 7 mostram as plantas expressando uma das raízes protegidas FR8a e FRCG eHIPs de danos de alimentação causados pelo verme de raiz de milho ocidental. A maioria dos eventos expressando a proteína inseticida quimérica foram considerados plantas de categoria 1, considerando que plantas de controle não expressaram uma proteína inseticida quimérica quando em categoria 3. As plantas expressando FRD3 eHIP forneceram níveis de controle comparáveis de verme de raiz de milho ocidental.

Tabela 7: Eficácia de plantas transgênicas expressando FR8a e FRCG contra WCR.

Vetor	Evento	Classif. De Dano (Nº FH)	Categoria	Vetor	Evento	Classif. de Dano (Nº FH)	Categoria
12161	1	1	1	12207	1	0	1
	2	1	1		2	2	1
	3	0	1		3	1	1
	4	3	2		4	2	1
	5	2	1		5	1	1
	6	0	1		6	2	1
	7	0	1		7	4	2
	8	0	1		8	3	2
	9	1	1		9	4	2
	10	1	1				
	11	4	2	Controle	1	7	3
	12	0	1		2	9	3
					3	8	3
Controle	1	6	3		4	11	3
	2	6	3		5	7	3

Vetor	Evento	Classif. De Dano (Nº FH)	Categoria	Vetor	Evento	Classif. de Dano (Nº FH)	Cate-goria
	3	6	3		6	12	3
	4	18	3				
12208	1	0	1	12274	1	3	2
	2	0	1		2	0	1
	3	3	2		3	3	2
	4	4	2		4	3	2
	5	1	1		5	0	1
	6	4	2		6	3	2
	7	0	1		7	3	2
	8	4	2		8	0	1
	9	4	2		9	3	2
	10	1	1		10	3	2
	11	1	1		11	0	1
	12	1	1				
	13	0	1	Controle	1	10	3
					2	10	3
Controle	1	10	3		3	10	3
	2	5	3		4	7	3
	3	7	3		5	8	3
	4	7	3		6	6	3
	5	8	3				
	6	8	3				

Exemplo 46: Análise de Plantas de Milho Transgênicas para Eficácia no Campo contra Vermes de Raiz de Milho

Algumas plantas positivas identificadas utilizando a bioanálise de excisão de raiz descrita acima foram avaliadas no campo. Dezoito plantas de cada evento foram removidas dos planos de campo e avaliadas por danos nas raízes. Os danos na raiz foram classificados utilizando a escala de dano de raiz linear de 0 a 3 do Estado de Iowa (Oleson, J.D. *et al.*, 2005. J. Econ Entomol. 98(1): 1-8), onde 0,00 = sem dano alimentar (menor classificação que pode ser dada); 1,00 = nenhum nódulo (círculo de raízes), ou o equivalente de um nódulo inteiro, mordidas dentro de aproximadamente 1½ pole-

gadas do caule (linha de solo no 7º nódulo); 2,00 = dois nódulos completos mordidos; 3,00 = três ou mais nódulos mordidos (maior classificação que pode ser dada); e dano entre nódulos completos mordidos foi notado conforme o percentual do nódulo faltante, isto é, 1,50 = 1 1/2 nódulos mordidos,

5 0,25 = 1/4 de um nódulo mordido, etc.

Os resultados de exames de campo contra o verme de raiz de milho ocidental e do norte são apresentados na Tabela 8 e contra o verme de raiz de milho mexicano na Tabela 9. Todo o milho transgênico expressando a proteína inseticida quimérica FR8a teve desempenho melhor do que

10 um inseticida comercial padrão contra verme de raiz de milho ocidental, do norte e mexicano.

Tabela 8: Resultados de testes de campo de verme de raiz de milho ocidental e do norte.

Evento	Plasmídeo	Classificação de Raiz
1	12161 (ubi:FR8a)	0,08
2	12161	0,05
3	12161	0,09
4	12161	0,04
5	12274 (cmp:FR8a)	0,04
6	12274	0,08
7	12274	0,05
Químico		0,15
Verif. Negativa		0,87

Tabela 9: Resultados de testes de verme de raiz de milho mexicano.

Evento	Plasmídeo	Classificação de Raiz
1	12161 (ubi:FR8a)	0,04
5	12274 (cmp:FR8a)	0,22
6	12274	0,05
Químico		0,15
Verif. Negativa		1,04

REIVINDICAÇÕES

1. Proteína inseticida híbrida projetada, **caracterizada** pelo fato de que compreende, de N-terminus a C-terminus, uma região N-terminal de uma proteína Cry3A ou Cry3A modificada fundida a uma região C-terminal de uma proteína Cry1Aa ou Cry1Ab, em que pelo menos uma posição cruzada entre a proteína Cry3A e a proteína Cry1Aa ou Cry1Ab está localizada no bloco conservado 2, bloco conservado 3, região variável 4 ou bloco conservado 4, e opcionalmente compreendendo:

(a) uma região dorsal de protoxina de uma proteína Cry Bt no C-terminus; ou

(b) um fragmento de peptidila no N-terminus; ou

(c) ambos (a) e (b),

em que a proteína inseticida híbrida projetada compreende uma sequência de aminoácidos selecionada do grupo consistindo na SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 30, SEQ ID NO: 34, SEQ ID NO: 44; SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 147, SEQ ID NO: 153; SEQ ID NO: 155; SEQ ID NO: 159 e SEQ ID NO: 160 e possui atividade inseticida contra o verme de raiz de milho ocidental.

2. Proteína inseticida híbrida projetada, de acordo com a reivindicação 1, **caracterizada** pelo fato de que a proteína inseticida é ativa contra o verme de raiz de milho do norte ou o verme de raiz de milho mexicano.

3. Molécula de ácido nucleico codificadora de uma proteína inseticida híbrida projetada como definida na reivindicação 1 ou 2, **caracterizada** pelo fato de que compreende uma sequência de nucleotídeos selecionada do grupo consistindo na SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 29, SEQ ID NO: 33, SEQ ID NO: 43; SEQ ID NO: 61, SEQ ID NO: 63, SEQ ID NO: 146, SEQ ID NO: 152, SEQ ID NO: 154 e SEQ ID NO: 158.

4. Cassete de expressão, **caracterizado** pelo fato de que

compreende uma sequência promotora heteróloga ligada operacionalmente a uma molécula de ácido nucleico como definida na reivindicação 3.

5. Vetor recombinante, **caracterizado** pelo fato de que compreende o cassete de expressão como definido na reivindicação 4.

5 6. Célula hospedeira transgênica, **caracterizada** pelo fato de que compreende o cassete de expressão como definido na reivindicação 4, a qual é uma célula bacteriana.

7. Composição inseticida, **caracterizada** pelo fato de que compreende a proteína inseticida como definida na reivindicação 1 ou 2.

10 8. Método de produção de uma proteína inseticida híbrida projetada que é ativa contra insetos, **caracterizado** pelo fato de que compreende:

(a) obtenção da célula hospedeira transgênica compreendendo o cassete de expressão como definido na reivindicação 4; e

15 (b) cultivo da célula hospedeira transgênica sob condições que permitam que o hospedeiro expresse uma proteína inseticida híbrida projetada que seja ativa contra insetos.

9. Método de produção de uma planta transgênica resistente a insetos, **caracterizado** pelo fato de que compreende a introdução de uma molécula de ácido nucleico como definida na reivindicação 3 em uma planta, assim produzindo uma planta transgênica, em que uma proteína inseticida híbrida projetada é expressa na planta transgênica em uma quantidade eficaz para controlar insetos.

20

10. Método de controle de insetos, **caracterizado** pelo fato de que compreende fornecer aos insetos uma quantidade eficaz de uma toxina inseticida como definida na reivindicação 1 ou 2.

25

Fig. 1A

Alinhamento: alinhamento de Proteínas global contra a molécula de referência
Parâmetros: matriz de pontuação: BLOSUM 62

Molécula de Referência: Cry3A055, Região 1-598
Nº de seqüências a alinhar: 19

Pos	Nome	Seqüência	Início	Fim	Comprimento	Pares	% Identidade
Ref 1	Cry3A055	(SEQ ID NO: 70)	1	598	598 aa		
2	moCry3A	(SEQ ID NO: 68)	1	597	597 aa	594	99
3	Native Cry3A	(SEQ ID NO: 131)	1	644	644 aa	594	92
4	8AF	(SEQ ID NO: 64)	1	640	640 aa	524	81
5	T7-8AF	(SEQ ID NO: 133)	1	654	654 aa	524	80
6	2OL-8A	(SEQ ID NO: 2)	1	668	668 aa	517	77
7	FR8a	(SEQ ID NO: 4)	1	653	653 aa	518	79
8	FRCG	(SEQ ID NO: 6)	1	652	652 aa	514	78
9	FRD3	(SEQ ID NO: 16)	1	615	615 aa	518	84
10	FR-cg-dm3	(SEQ ID NO: 18)	1	614	614 aa	514	83
11	FR8a-12aa	(SEQ ID NO: 12)	1	641	641 aa	515	80
12	WR-9mut	(SEQ ID NO: 14)	1	599	599 aa	589	98
13	DM23A	(SEQ ID NO: 62)	1	653	653 aa	510	78
14	FR8a-9F	(SEQ ID NO: 8)	1	653	653 aa	524	80
15	FR-9F-cg-del6	(SEQ ID NO: 20)	1	646	646 aa	520	80
16	FR-9F-catg	(SEQ ID NO: 10)	1	652	652 aa	520	79
17	V4F	(SEQ ID NO: 32)	1	598	598 aa	571	95
18	5*V4F	(SEQ ID NO: 34)	1	611	611 aa	565	92
19	Cry1Ab	(SEQ ID NO: 72)	1	648	648 aa	213	31

Fig. 1B

Cry3A055	1		MTADNNTALDSSTTKDV
moCry3A	1		MTADNNTALDSSTTKDV
Cry3A	1	MNPNNRSEHDIKTENNEVPTNHVQYPLAETPNPTLEDLNYKEFLRMTADNNTALDSSTTKDV	
8AF	1		MTADNNTALDSSTTKDV
T7-8AF	1		MASMTGGQQMGRGSM
2OL-8A	1		MTADNNTALDSSTTKDV
FR8a	1	MASMTGGQQMGRGSGSTSNGRQCAGIRPYDGRQQHRLDSSTTKDV	
FRCG	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDV
FRD3	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDV
FR-cg-dm3	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDV
FR8a-12aa	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDV
WR-9mut	1		MYDGRQQHRLDSSTTKDV
DM23A	1		MYDGRQQHRLDSSTTKDV
FR8a-9F	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDV
FR-9F-cg-del6	1		MTSNGRQCAGIRPMTADNNTALDSSTTKDV
FR-9F-catg	1		MCAGIRPMTADNNTALDSSTTKDV
V4F	1		MTSNGRQCAGIRPMTADNNTALDSSTTKDV
5*V4F	1		MTADNNTALDSSTTKDV
Cry1Ab	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDV
			MDNPNINIECIPYNCLSNPEVEV
Cry3A055	19	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
moCry3A	19	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
Cry3A	66	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
8AF	19	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
T7-8AF	33	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
2OL-8A	47	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FR8a	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FRCG	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FRD3	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FR-cg-dm3 pr	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FR8a-12aa	20	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
WR-9mut	20	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
DM23A	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FR8a-9F	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FR-9F-cg-del6	26	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
FR-9F-catg	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
V4F	19	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
5*V4F	32	-----IQKGISVVGDLGTVGFPFGGAL--VSFYTNFLNTIWPSDEP--WKA	FMEQVEALMDQKIADYAKNKALAE
Cry1Ab	24	LGGERIFTGYTPIDISLSLTQFLLSFVPGAGFVLGLVDI IWGIFGSPQWDAFLVQIEQLINQR	IEEFARNQAI SRLEGLSNLYQIYAESFREWEADPT----NPALREE

Fig. 1C

			CB1		CB2
Cry3A055	120	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
moCry3A	119	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
Cry3A	166	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
8AF	120	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
T7-8AF pro	134	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
20L-8A	148	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FR8a	133	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FRCG	132	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FRD3	133	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FR-cg-dm3	132	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FR8a-12aa	121	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
WR-9mut	121	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
DM23A	133	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FR8a-9F	133	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FR-9F-cg-del6	126	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
FR-9F-catg	132	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
V4F	120	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
5*V4F	133	IRELFSQAESHFRNSMPFAISG	YEVFLFTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTLT
Cry1Ab	130	MRIQENDMNSALTTAIPFAVQN	YQVPLLSYVYQAANTHLFLLKDAQIYGEEW	GFDAAATINSRYNDLTRLIGN	YTDHAVRWYNTGLERVWGPDSRDWIRYNQFRRELTLT

Domain I

Domain II

Cry3A055	230	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
moCry3A	229	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
Cry3A	276	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
8AF	230	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
T7-8AF	244	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
20L-8A	258	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FR8a	243	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FRCG	242	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FRD3	243	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FR-cg-dm3	242	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FR8a-12aa	231	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
WR-9mut	231	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
DM23A	243	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FR8a-9F	243	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FR-9F-cg-del6	236	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
FR-9F-catg	242	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
V4F	230	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
5*V4F	243	VLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYGGNDSFNWWSGNVSTRPSIGSNDIITSFFYGNK
Cry1Ab	240	VLDIVSLFPNYDSRTYPIRTVSQLTREIYT	NPV--LENFDGSRGSAQIEGSIKPSHMLDILNSITTYTDAHRGEY-----YWSGHQIMASPVGFGPEFTPLYGTM

Fig. 1D

Cry3A055 339 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
moCry3A 338 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
Cry3A 385 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
8AF 339 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
T7-8AF 353 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
20L-8A 367 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FR8a 352 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FRCG 351 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FRD3 352 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FR-cg-dm3 351 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FR8a-12aa 340 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
WR-9mut 340 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
DM23A 352 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FR8a-9F 352 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FR-9F-cg-del 345 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
FR-9F-catg 351 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
V4F 339 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
5*V4F 352 -SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQTDASTQTYDSKRNVGAVSW-----DSIDQLPPETTTDEPLEKGYSHQLNYVMCFLMQGSRG
Cry1Ab 342 GNAAPQQRIVAQLGQGVYRTLST---LYRRPFNIGINNQQLSVL-DGTEFA----YGTSSNLPASVYRKSGTVDSLOEIPQNNVPPRQGFSHRLSHVSMFRSGFSNS

		Domain II	Domain III	CB3	CB4	
Cry3A055	441	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLVKAYKLQSGASVVAGPRFTGGDII	QCTENGSAATIYVTPDVSYSC	KYRARIHYAS	TSQITFTLSLDGAPFNQYYFDK
moCry3A	440	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLVKAYKLQSGASVVAGPRFTGGDII	QCTENGSAATIYVTPDVSYSC	KYRARIHYAS	TSQITFTLSLDGAPFNQYYFDK
Cry3A	487	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLVKAYKLQSGASVVAGPRFTGGDII	QCTENGSAATIYVTPDVSYSC	KYRARIHYAS	TSQITFTLSLDGAPFNQYYFDK
8AF	441	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
T7-8AF	455	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
20L-8A	469	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FR8a	454	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FRCG	453	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FRD3	454	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FR-cg-dm3	453	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FR8a-12aa	442	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
WR-9mut	442	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLVKAYKLQSGASVVAGPRFTGGDII	QCTENGSAATIYVTPDVSYSC	KYRARIHYAS	TSQITFTLSLDGAPFNQYYFDK
DM23A	454	TI-----PV	LTWTHKSAEFNNIIPSSQITQIPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FR8a-9F	454	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FR-9F-cg-del6	447	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
FR-9F-catg	453	TI-----PV	LTWTHKSVDFENMIDSKKITQLPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA
V4F	441	TI-----PV	LTWTHKSVDFENMIDSKKITQLALTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TSQITFTLSLDGAPFNQYYFDK
5*V4F	454	TI-----PV	LTWTHKSVDFENMIDSKKITQLALTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TSQITFTLSLDGAPFNQYYFDK
Cry1Ab	444	SVSIIRAFM	FSWIHRSAEFNNIIPSSQITQIPLTKSTNLGSGTSVVKPGPGFTGGDIL	RRTSPGQISTLRVNITAPLSC	RYRVIRIYAS	TTNLQFHTSIDGRPINQGNFSA

Fig. 1E

		CBS	
Cry3A055	546	TINKGDTLTYN SFN LASFSTPFELSGNNLQIGVTG--LSAGDK	VYIDKIEFIPVN
moCry3A	545	TINKGDTLTYN SFN LASFSTPFELSGNNLQIGVTG--LSAGDK	VYIDKIEFIPVN
Cry3A	592	TINKGDTLTYN SFN LASFSTPFELSGNNLQIGVTG--LSAGDK	VYIDKIEFIPVN
8AF	546	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
T7-8AF	560	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
2OL-8A	574	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
FR8a	559	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
FRCG	558	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
FRD3	559	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VT
FR-cg-dm3	558	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VT
FR8a-12aa	547	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
WR-9mut	547	TINKGDTLTYN SFN LASFSTPFELSGNNLQIGVTG--LSAGDK	VYIDKIEFIPVN
DM23A	559	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
FR8a-9F	559	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
FR-9F-cg-del6	552	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
FR-9F-catg	558	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV
V4F	546	TINKGDTLTYN SFN LASFSTPFELSGNNLQIGVTG--LSAGDK	VYIDKIEFIPVN
5*V4F	559	TINKGDTLTYN SFN LASFSTPFELSGNNLQIGVTG--LSAGDK	VYIDKIEFIPVN
Cry1Ab	554	TMSSGSNLQSGSFRTVGFTT PFNFSGSSVFTLSAHVENS GNE	VYIDRIEFVPAE VTFEAEYDLERAQKAVNEL FTSSNQIGLKT DVT DYHIDQV

Fig. 2A

Alinhamento: alinhamento de Proteínas global contra a molécula de referência
Parâmetros: matriz de pontuação: BLOSUM 62

Molécula de Referência: Proteína 8AF, Aminoácidos 1-640
Número de sequências a alinhar: 18

Pos	Nome	Sequência ID	Início	Fim	Comprimento	Pares	% Identidade
Ref 1	8AF	SEQ ID NO: 64	1	640	640 aa		
2	8AFdm3T	SEQ ID NO: 155	1	640	640 aa	634	99
3	-CatG8AF	SEQ ID NO: 147	1	639	639 aa	636	99
4	FR8a-9F	SEQ ID NO: 8	1	653	653 aa	640	98
5	FR8a-12 AA	SEQ ID NO: 12	1	641	641 aa	631	98
6	FR8a	SEQ ID NO: 4	1	653	653 aa	634	97
7	FR-9F-catg	SEQ ID NO: 10	1	653	653 aa	636	97
8	cap8AFdm3T	SEQ ID NO: 159	1	653	653 aa	628	96
9	FRCG	SEQ ID NO: 16	1	652	652 aa	630	96
10	DM23A	SEQ ID NO: 62	1	653	653 aa	626	95
11	ZOL-8A	SEQ ID NO: 2	1	668	668 aa	633	94
12	8AFdm3	SEQ ID NO: 149	1	602	602 aa	596	93
13	FR8a +34	SEQ ID NO: 160	1	687	687 aa	634	92
14	FR-12-cg-dm3	SEQ ID NO: 18	1	603	603 aa	589	91
15	9F-cg-dm3	SEQ ID NO: 24	1	614	614 aa	598	91
16	Cap8AFdm3	SEQ ID NO: 153	1	615	615 aa	590	90
17	5+V4F	SEQ ID NO: 34	1	611	611 aa	545	83
18	V3A	SEQ ID NO: 30	1	596	596 aa	368	56

Fig. 2B

8AF	1		MTADNNTALDSSTTKDVIQKGISVVG
8AFdm3T	1		MTADNNTALDSSTTKDVIQKGISVVG
-CatG8AF	1		MTADNNTALDSSTTKDVIQKGISVVG
FR8a-9F	1		MTSNGRQCAGIRPMTADNNTALDSSTTKDVIQKGISVVG
FR8a-12 AA	1		MYDGRQQHRLDSSTTKDVIQKGISVVG
FR8a	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
FR-9F-catg	1		MTSNGRQCAGIRPMTADNNTALDSSTTKDVIQKGISVVG
cap8AFdm3T	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
FRCG	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
DM23A	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
2OL-8A	1		MASMTGGQQMGRSGSTSNRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
8AFdm3	1		MTADNNTALDSSTTKDVIQKGISVVG
FR8a +34	1		MKETAATAAKFERQHMDSPDLGTLVPRGSMADIGSTMTSNGRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
FR-12-cg-dm3	1		MYDGRQQHRLDSSTTKDVIQKGISVVG
9F-cg-dm3	1		MTSNGRQCAGIRPMTADNNTALDSSTTKDVIQKGISVVG
Cap8AFdm3	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
5*V4F	1		MTSNGRQCAGIRPYDGRQQHRLDSSTTKDVIQKGISVVG
V3A	1		MTADNNTALDSSTTKDVIQKGISVVG

8AF	29	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
8AFdm3T	29	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
-CatG8AF	29	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPVSS-RNPHSQGRIRELFSQAE
FR8a-9F	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
FR8a-12 AA	30	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
FR8a	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
FR-9F-catg	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPVSS-RNPHSQGRIRELFSQAE
cap8AFdm3T	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
FRCG	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPVSS-RNPHSQGRIRELFSQAE
DM23A	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
2OL-8A	57	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
8AFdm3	29	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
FR8a +34	76	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
FR-12-cg-dm3	30	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPVSS-RNPHSQGRIRELFSQAE
9F-cg-dm3	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPVSS-RNPHSQGRIRELFSQAE
Cap8AFdm3	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
5*V4F	42	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE
V3A	29	LLGVVGFPPGGALVSFYTNFLNTIWPSEDPWKAEMEVEALMDQKIADYAKNKALAEQLQNNVEDYVSALSSWQKNPAAPFRNPHSQGRIRELFSQAE

Fig. 2C

		CB1	CB2		
8AF	129	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
8AFdm3T	129	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
-CatG8AF	128	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
FR8a-9F	142	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
FR8a-12 AA	130	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
FR8a	142	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
FR-9F-catg	141	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
cap8AFdm3T	142	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
FRCG	141	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
DM23A	142	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
2OL-8A	157	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
8AFdm3	129	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
FR8a +34	176	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
FR-12-cg-dm3	129	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
9F-cg-dm3	141	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
Cap8AFdm3	142	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
5*V4F	142	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL
V3A	129	SHFRNSMPSFAISG	YEVLFLTTYAQAANTHLFLLKDAQIYGEEW	GYEKEDIAEFYKRQLKLTQE	YTDHCVKWYNVGLDKLRGSSYESWVNFNRYRREMTL

Domain I

Domain II

8AF	229	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
8AFdm3T	229	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
-CatG8AF	228	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
FR8a-9F	242	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
FR8a-12 AA	230	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
FR8a	242	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
FR-9F-catg	241	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
cap8AFdm3T	242	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
FRCG	241	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
DM23A	242	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
2OL-8A	257	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
8AFdm3	229	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
FR8a +34	276	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
FR-12-cg-dm3	229	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
9F-cg-dm3	241	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
Cap8AFdm3	242	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
5*V4F	242	TVLDLIALFPLYDVRLYPKEVKTELTRDVL	DPIVGNNLRG-YGTTFSNIENYIRKPHLFDYLHRIQFHTRFQPGYYGNDSFNYWSGNYVSTRPSIGSN
V3A	229	TVLDIVSLFPNYSRTYPIRTVSQLTREIYT	NPVL--ENFDGSRGSAQIEGSIKPSHLMIDILNSITITDAHRGEY-----YWSGHQIMASPVGFSG

Fig. 2D

8AF	328	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
8AFdm3T	328	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
-CatG8AF	327	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
FR8a-9F	341	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
FR8a-12 AA	329	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
FR8a	341	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
FR-9F-catg	340	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
cap8AFdm3T	341	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
FRCG	340	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
DM23A	341	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
2OL-8A	356	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
8AFdm3	328	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
FR8a +34	375	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
FR-12-cg-dm3	328	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
9F-cg-dm3	340	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
Cap8AFdm3	341	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
5*V4F	341	DIITSPFYGNK-SSEPQNLEFN-GEKVYRAVANTNLAVWPSAVYSGVTKVEFSQYNDQDEASTQTYDSKRNVGAVSW-----DSIDQLPPETTDEPL
V3A	321	PEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLST---LYRPFENIGINNQLSVL-DGTEFA---YGTSSNLPSAVYRKSGTVDSLDEIPQNNNVPP

			Domain II	Domain III		
			←	↓	→	CB3
8AF	420	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
8AFdm3T	420	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFIIPSSQITQIPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
-CatG8AF	419	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
FR8a-9F	433	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
FR8a-12 AA	421	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
FR8a	433	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
FR-9F-catg	432	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
cap8AFdm3T	433	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFIIPSSQITQIPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
FRCG	432	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
DM23A	433	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSAEFNFIIPSSQITQIPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
2OL-8A	448	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
8AFdm3	420	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFIIPSSQITQIPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
FR8a +34	467	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
FR-12-cg-dm3	420	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
9F-cg-dm3	432	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
Cap8AFdm3	433	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFIIPSSQITQIPLTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
5*V4F	433	EKGYSHQNLNYVMCFLMQGSRGTI-----PV	LTWTHKSVDFNFNIDSKKITQLALTKSTNLGSGTSVVKGPFTGGDII	RRTSPGQISTLRVNITAPLSQ		
V3A	413	RQGFSHRLSHVSMFRSGFSNSVSIIRAPM	FSWIHRSAEFNFIIPSSQITQIPLVKAYKLGSGASVVGPRFTGGDII	OCTENGSAATIYVTPDVSYSQ		

Fig. 2E

CB4		CB5	
8AF	514	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
8AFdm3T	514	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
-CatG8AF	513	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
FR8a-9F	527	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
FR8a-12 AA	515	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
FR8a	527	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
FR-9F-catg	526	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
cap8AFdm3T	527	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
FRCG	526	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
DM23A	527	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
20L-8A	542	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
8AFdm3	514	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
FR8a +34	561	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
FR-12-cg-dm3	514	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
9F-cg-dm3	526	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
Cap8AFdm3	527	RYRVRIRYAS	TTNLQFHTSIDGRPINQGNFSATMSSGSNLQSGSFRTVGFTT
5*V4F	527	RYRVRIRYAS	TSQITFTLSLDGAPFNQYYFDKTIKGDTLTYNSFNLSFSTPFELSGNN--LQIGVTGLSAGDK
V3A	512	KYRARIHYAS	TSQITFTLSLDGAPFNQYYFDKTIKGDTLTYNSFNLSFSTPFELSGNN--LQIGVTGLSAGDK
8AF	615	AVNELETTSSNOIGLKTVDVTDYHIDOV	
8AFdm3T	615	AVNELETTSSNOIGLKTVDVTDYHIDOV	
-CatG8AF	614	AVNELETTSSNOIGLKTVDVTDYHIDOV	
FR8a-9F	628	AVNELETTSSNOIGLKTVDVTDYHIDOV	
FR8a-12 AA	616	AVNELETTSSNOIGLKTVDVTDYHIDOV	
FR8a	628	AVNELETTSSNOIGLKTVDVTDYHIDOV	
FR-9F-catg	627	AVNELETTSSNOIGLKTVDVTDYHIDOV	
cap8AFdm3T	628	AVNELETTSSNOIGLKTVDVTDYHIDOV	
FRCG	627	AVNELETTSSNOIGLKTVDVTDYHIDOV	
DM23A	628	AVNELETTSSNOIGLKTVDVTDYHIDOV	
20L-8A	643	AVNELETTSSNOIGLKTVDVTDYHIDOV	
8AFdm3	603	-----	
FR8a +34	662	AVNELETTSSNOIGLKTVDVTDYHIDOV	
FR-12-cg-dm3	603	-----	
9F-cg-dm3	615	-----	
Cap8AFdm3	628	-----	
5*V4F	610	-----	
V3A	595	-----	

Fig. 3

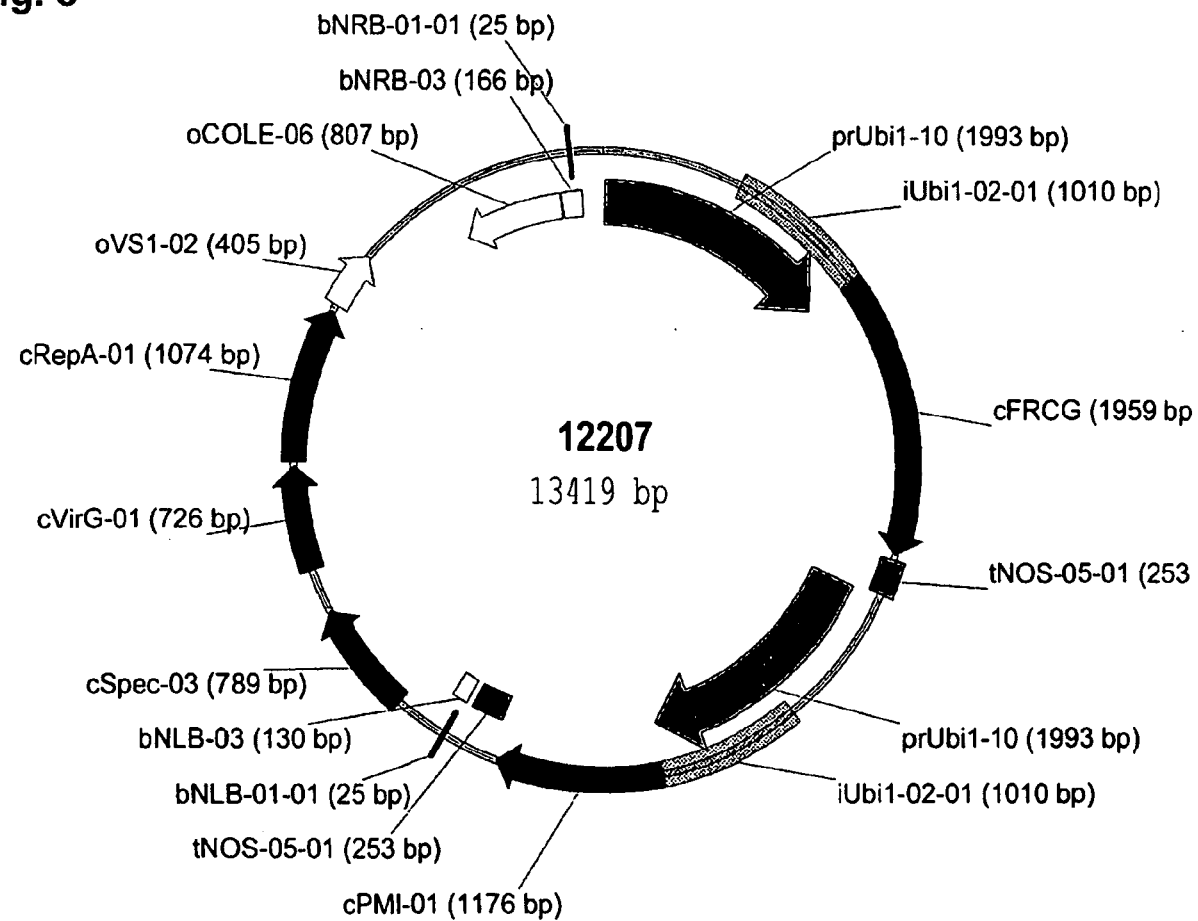


Fig. 4

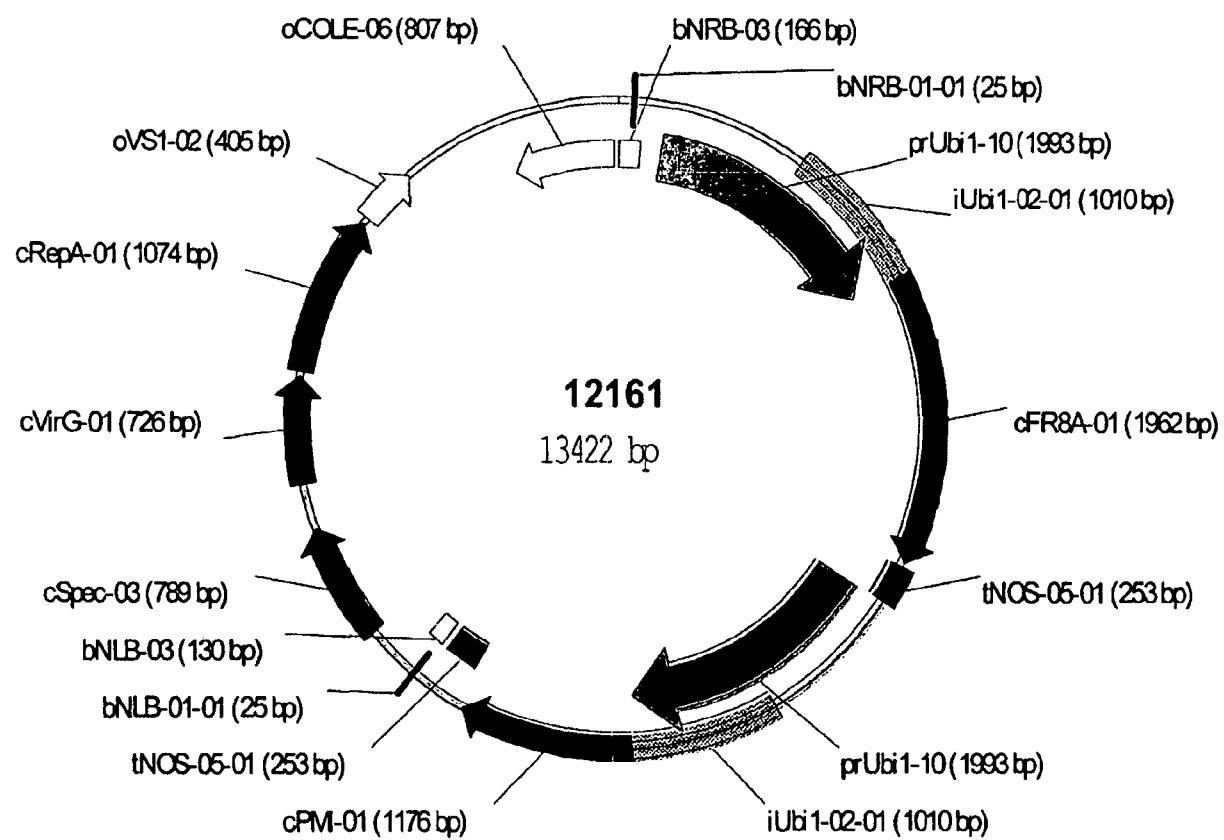


Fig. 5

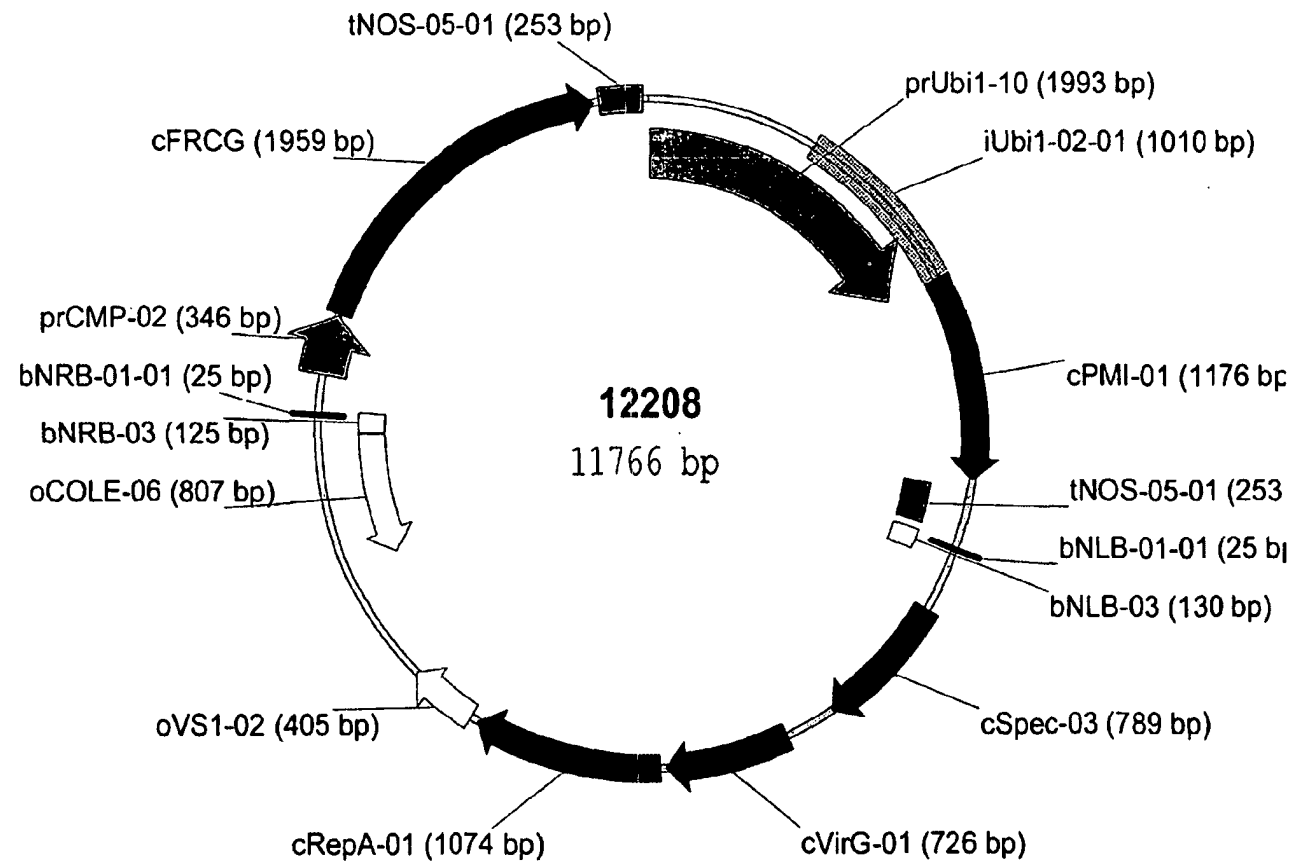


Fig. 6

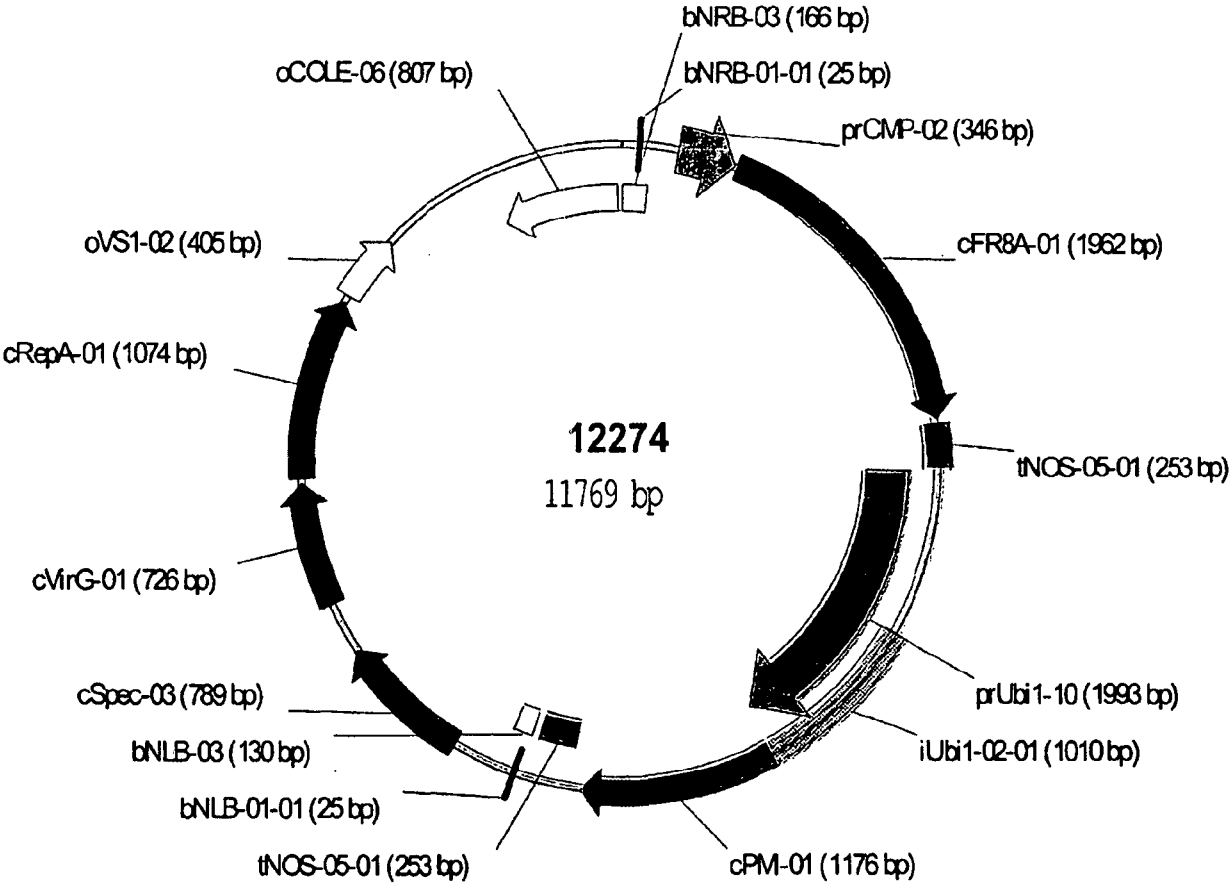


Fig. 7

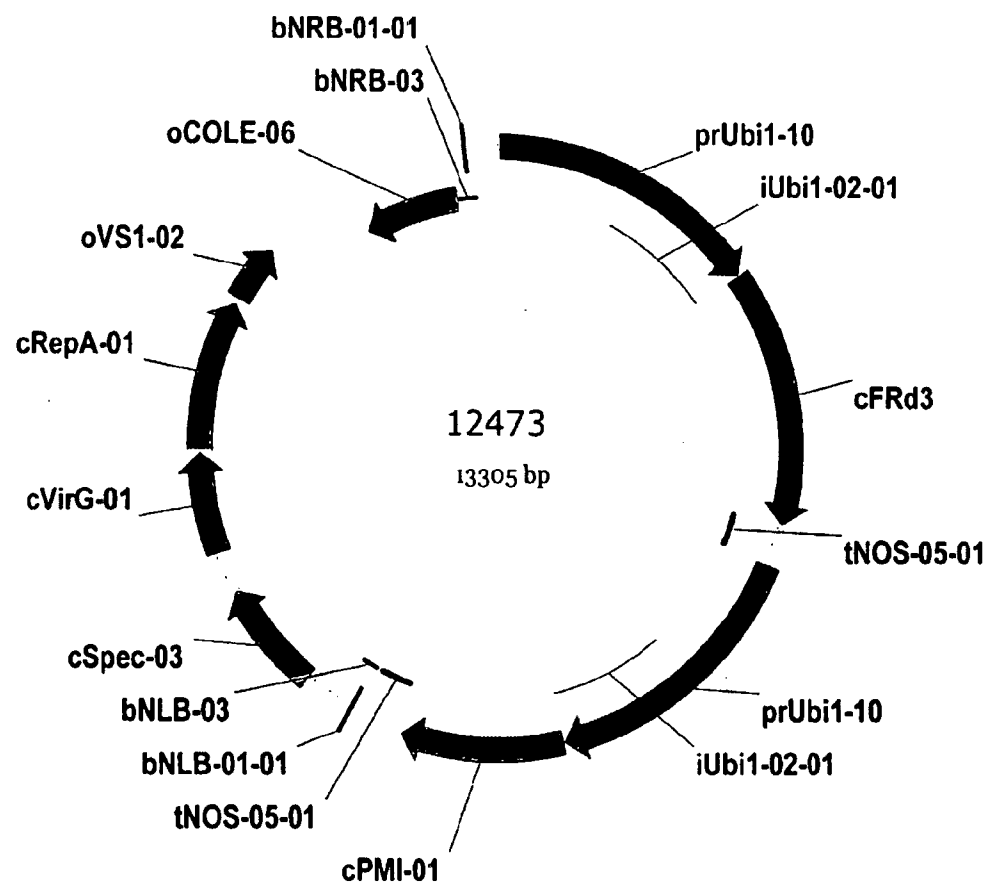


Fig. 8

