



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(11) PI 1007269-1 B1



(22) Data do Depósito: 12/01/2010

(45) Data de Concessão: 29/06/2021

(54) Título: MÉTODO PARA PRODUZIR UM POLIPEPTÍDEO, E, MÉTODO PARA RECONHECER SELETIVAMENTE UMA SEQUÊNCIA DE DNA ALVO PREDETERMINADA

(51) Int.Cl.: C07K 14/415; C12N 15/82; G01N 33/50.

(30) Prioridade Unionista: 12/01/2009 DE 102009004659.6; 13/07/2009 EP 09165328.7; 13/07/2009 US 61/225043.

(73) Titular(es): THOMAS LAHAYE; ULLA BONAS; SEBASTIAN SCHORNACK; JENS BOCH.

(72) Inventor(es): ULLA BONAS; JENS BOCH; SEBASTIAN SCHORNACK; THOMAS LAHAYE.

(86) Pedido PCT: PCT IB2010000154 de 12/01/2010

(87) Publicação PCT: WO 2010/079430 de 15/07/2010

(85) Data do Início da Fase Nacional: 12/07/2011

(57) Resumo: MÉTODO PARA PRODUZIR UM POLIPEPTÍDEO, POLIPEPTÍDEO, MOLÉCULA DE AÇO NUCLEICO, DNA, CASSETTE DE EXPRESSÃO, CÉLULA HOSPEDEIRA OU ORGANISMO NÃO HUMANOS, MÉTODOS PARA RECONHECER SELETIVAMENTE UM PAR DE BASES EM UMA SEQUÊNCIA DE DNA, E PARA MODULAR EXPRESSÃO DE UM GENE ALVO EM UMA CÉLULA, VETOR, E, MÉTODO PARA PRODUZIR UM DNA A presente invenção refere-se aos métodos para seletivamente reconhecer um par de bases em uma sequência de DNA por um polipeptídeo, para polipeptídeos modificados que especificamente reconhecem um ou mais pares de base em uma sequência de DNA e, para o DNA que é modificado de modo que possa ser especificamente reconhecido por um polipeptídeo e para usos do polipeptídeo e DNA no direcionamento de DNA específico assim como aos métodos de modulação da expressão dos genes alvo em uma célula.

“MÉTODO PARA PRODUZIR UM POLIPEPTÍDEO, E, MÉTODO PARA RECONHECER SELETIVAMENTE UMA SEQUÊNCIA DE DNA ALVO PREDETERMINADA ”

CAMPO TÉCNICO DA INVENÇÃO

[0001] A presente invenção se refere a métodos para reconhecer seletivamente um par de bases em uma sequência de DNA alvo por um polipeptídeo, para polipeptídeos modificados que reconhecem especificamente um ou mais pares de base em uma sequência de DNA alvo e, para DNA que é modificado de modo que possa ser especificamente reconhecido por um polipeptídeo e para usos do polipeptídeo e DNA em direcionamento de DNA específico bem como para métodos de modulação de expressão de genes alvo em uma célula.

FUNDAMENTOS DA INVENÇÃO

[0002] Bactérias fitopatogênicas do gênero *Xanthomonas* causam diversas doenças em muitas plantas de safra importante. As bactérias translocam um arsenal de efetores incluindo membros da família efetora semelhante a (TAL)/AvrBs3 semelhante a ativador de transcrição ampla através do sistema de secreção do tipo III em células de plantas (Kay & Bonas (2009) *Curr. Opin. Microbiol.* 12:37-43, White & Yang (2009) *Plant Physiol.* doi:10.1104/pp.1109.139360; Schornack *et al.* (2006) *J. Plant Physiol.* 163:256-272). Efetores TAL, fatores de virulência importantes de *Xanthomonas*, contêm um domínio central de repetições em tandem, sinais de localização nuclear (NLSs), e um domínio de ativação (AD) e agem como fatores de transcrição em células de plantas (Kay *et al.* (2007) *Science* 318:648-651; Römer *et al.* (2007) *Science* 318:645-648; Gu *et al.* (2005) *Nature* 435, 1122-1125; Fig. 1a). O tipo de membro desta família efetora, AvrBs3 de *Xanthomonas campestris* pv. *vesicatoria*, contém 17,5 repetições e induz expressão de genes *UPA* (regulado para cima por AvrBs3) incluindo o gene de resistência *Bs3* em plantas de pimenta (Kay *et al.* (2007) *Science*

318:648-651; Römer *et al.* (2007) *Science* 318:645-648; Marois *et al.* (2002) *Mol. Plant-Microbe Interact.* 15:637-646). O número e ordem de repetições em um efector TAL determina sua atividade específica (Herbers *et al.* (1992) *Nature* 356:172-174). As repetições foram consideradas essenciais para ligação a DNA de AvrBs3 e constituem um domínio de ligação a DNA novo (Kay *et al.* (2007) *Science* 318:648-651). Como este domínio põe em contato o DNA e o que determina a especificidade permanece desconhecido.

[0003] A expressão seletiva de genes é mediada através da interação de fatores de transcrição de proteína com seqüências de nucleotídeos específicas dentro da região regulatória do gene. A maneira na qual os domínios de proteína de ligação a DNA são capazes de discriminar entre seqüências de DNA diferentes é uma questão importante para entender processos cruciais como o controle de expressão de gene em diferenciação e desenvolvimento.

[0004] A capacidade de designer e gerar especificamente domínios de ligação a DNA que reconhecem um alvo de DNA desejado é altamente desejável em biotecnologia. Tal capacidade pode ser útil para o desenvolvimento de fatores de transcrição personalizados com a capacidade de modular expressão de gene mediante ligação a DNA alvo. Exemplos incluem o trabalho extensor feito com o projeto de proteínas de ligação a DNA de dedo de zinco personalizado específicas para uma seqüência de DNA alvo desejada (Choo *et al.* (1994) *Nature* 372:645; Pomerantz *et al.*, (1995) *Science* 267:93-96; Liu *et al.*, *Proc. Natl. Acad. Sci. USA* 94:5525-5530 (1997); Guan *et al.* (2002) *Proc. Natl. Acad. Sci. USA* 99:13296-13301; U.S. Pat. No. 7,273,923; Pat. No. U.S. 7,220,719). Além disso, os polipeptídeos contendo domínios de ligação a DNA *designer* podem ser utilizados para modificar a seqüência de DNA alvo real pela inclusão de domínios modificadores de DNA, como um domínio catalítico de nuclease, dentro do polipeptídeo. Exemplos disto incluem o domínio de ligação a DNA de um

local de reconhecimento de DNA de meganuclease/endonuclease de abrigo em combinação com um domínio de nuclease não-específico (ver Pedido de Patente U.S. 2007/0141038), local de reconhecimento de DNA de meganuclease modificada e/ou domínios de nuclease da mesma ou de meganucleases diferentes (ver Publicação de Pedido de Patente U.S. 20090271881), e domínios de dedo de zinco em combinação com um domínio com atividade de nuclease, geralmente de uma endonuclease de restrição do tipo IIS como *FokI* (Bibikova *et al.* (2003) *Science* 300:764; Urnov *et al.* (2005) *Nature* 435, 646; Skukla, *et al.* (2009) *Nature* 459, 437-441; Townsend *et al.* (2009) *Nature* 459:442445; Kim *et al.* (1996) *Proc. Natl Acad. Sci USA* 93:1156-1160; U.S. Pat. No. 7,163,824). Os métodos atuais utilizados para identificar domínios de ligação a DNA de dedo de zinco personalizado empregam métodos baseados em seleção combinatória utilizando grandes bibliotecas randomizadas (geralmente $>10^8$ em tamanho) para gerar domínios multi-dedos com especificidade de DNA desejada (Greisman & Pabo (1997) *Science* 275:657-661; Hurt *et al.* (2003) *Proc Natl Acad Sci USA* 100:12271-12276; Isalan *et al.* (2001) *Nat Biotechnol* 19:656-660.) Tais métodos exigem muito tempo, demandam técnica e são potencialmente custosos. A identificação de um código de reconhecimento simples para a engenharia de polipeptídeos de ligação a DNA representaria um avanço significativo sobre os métodos atuais para projetar domínios de ligação a DNA que reconheçam um alvo de nucleotídeo desejado.

BREVE SUMÁRIO DA INVENÇÃO

[0005] A presente invenção fornece um método para produzir um polipeptídeo que seletivamente reconhece um par de bases em uma sequência de DNA, o método compreendendo sintetizar um polipeptídeo que compreende um domínio de repetição, em que o domínio de repetição compreende pelo menos uma unidade de repetição derivada de um efector semelhante a ativador de transcrição (TAL), em que a unidade de repetição

compreende uma região hipervariável que determina o reconhecimento de um par de bases na sequência de DNA, em que a unidade de repetição é responsável pelo reconhecimento de um par de bases na sequência de DNA. Estes polipeptídeos da invenção compreendem unidades de repetição da presente invenção e podem ser interpretados por uma abordagem modular pré-montando unidades de repetição em vetores alvo que podem subsequentemente ser montados em um vetor de destino final. A invenção fornece o polipeptídeo produzido por este método bem como sequências de DNA que codificam os polipeptídeos e organismos hospedeiros e células que compreendem tais sequências de DNA.

[0006] A presente invenção fornece um método para reconhecer seletivamente um par de bases em uma sequência de DNA alvo por um polipeptídeo em que tal polipeptídeo compreende pelo menos um domínio de repetição que compreende unidades de repetição em que em tais unidades de repetição cada um compreende uma região hipervariável que determina o reconhecimento de um par de bases em tal sequência de DNA alvo.

[0007] Mais especificamente, os inventores determinaram aqueles aminoácidos em um polipeptídeo de ligação a DNA responsável pelo reconhecimento seletivo de pares de base em uma sequência de DNA alvo. Com esclarecimento do código de reconhecimento, um princípio geral para reconhecer pares de base específicos em uma sequência de DNA alvo por aminoácidos selecionados em um polipeptídeo foi determinado. Os inventores descobriram que tipos distintos de unidades de repetição que são parte de um arranjo de unidades de repetição de comprimento variante têm a capacidade de reconhecer um par de bases definido/específico. Dentro de cada unidade de repetição que forma um domínio de repetição, uma região hipervariável é responsável pelo reconhecimento específico de um par de bases em uma sequência de DNA alvo.

[0008] Desta forma, a presente invenção fornece não apenas um

método para reconhecer seletivamente um par de bases em uma sequência de DNA alvo por um polipeptídeo compreendendo pelo menos um domínio de repetição que compreende unidades de repetição, mas também métodos em que as sequências de DNA alvo podem ser geradas sendo seletivamente reconhecidas por domínios de repetição em um polipeptídeo.

[0009] A invenção também fornece um método para construir polipeptídeos que reconhecem sequências de DNA específicas. Estes polipeptídeos da invenção compreendem unidades de repetição da presente invenção e podem ser construídos por uma abordagem modular pré-montando unidades de repetição em vetores alvo que podem ser subsequentemente montados em um vetor de destino final.

[00010] A invenção também fornece um método para modulação direcionada de expressão de genes através da construção de unidades de repetição modulares específicas para uma sequência de DNA alvo de interesse, modificando um polipeptídeo pela adição de tais unidades de repetição de modo a permitir que tal polipeptídeo agora reconheça o DNA alvo, introduzindo ou expressando tal polipeptídeo modificado em uma célula procariótica ou eucariótica de modo a permitir que tal polipeptídeo modificado reconheça a sequência de DNA alvo, e a modulação da expressão de tal gene alvo na referida célula como um resultado de tal reconhecimento.

[00011] A invenção também fornece um método para modificação direcionada de uma sequência de DNA alvo pela construção de um polipeptídeo incluindo pelo menos um domínio de repetição da presente invenção que reconheça tal sequência de DNA alvo e que tal polipeptídeo também contenha um domínio funcional capaz de modificar o DNA alvo (como através de recombinação, restrição ou integração específica no local de sequências alvo de doador) permitindo assim modificações de DNA direcionadas em genomas complexos.

[00012] A invenção ainda fornece a produção de polipeptídeos

modificados incluindo pelo menos um domínio de repetição compreendendo unidades de repetição em que uma região hipervariável dentro de cada unidade de repetição determina reconhecimento seletivo de um par de bases em uma sequência de DNA alvo.

[00013] Em outra modalidade da invenção, o DNA é fornecido codificando um polipeptídeo que contém um domínio de repetição conforme descrito acima.

[00014] Ainda em outra modalidade da invenção, o DNA é fornecido sendo modificado para incluir um ou mais pares de base localizados em uma sequência de DNA alvo de modo que cada um dos pares de base possa ser especificamente reconhecido por um polipeptídeo que inclui um domínio de repetição tendo unidades de repetição correspondentes, cada unidade de repetição compreendendo uma região hipervariável que determina o reconhecimento do par de bases correspondente em tal DNA.

[00015] Ainda em outra modalidade da invenção, são fornecidos os usos daqueles polipeptídeos e DNAs. Além disso, são fornecidas plantas, partes de plantas, sementes, células de plantas e outras células hospedeiras não-humanas transformadas com as moléculas de ácido nucléico isoladas da presente invenção e as proteínas ou polipeptídeos codificados pelas sequências de codificação da presente invenção. Ainda, os polipeptídeos e DNA descritos no presente documento podem ser introduzidas em células de animal e humanas bem como células de outros organismos como fungos ou plantas.

[00016] Em resumo, a invenção foca em um método para reconhecer seletivamente pares de base em uma sequência de DNA alvo por um polipeptídeo em que tal polipeptídeo compreende pelo menos um domínio de repetição que compreende unidades de repetição em que cada unidade de repetição contém uma região hipervariável que determinar o reconhecimento de um par de bases em tal sequência de DNA alvo em que as unidades de

repetição consecutivas correspondem a pares de base consecutivos em tal sequência de DNA alvo.

BREVE DESCRIÇÃO DAS FIGURAS

Figura 1 | Modelo para especificidade de DNA alvo de efetores TAL.

(A) Efetores TAL contêm unidades de repetição em tandem centrais (vermelho), sinais de localização nucleares (NLS) e um domínio de ativação (AD). Sequência de aminoácidos da primeira repetição de AvrBs3. Aminoácidos hipervariáveis 12 e 13 são sombreados em cinza.

(B) Aminoácidos hipervariáveis na posição 12 e 13 das unidades de repetição 17.5 AvrBs3 são alinhados ao consenso de caixa de UPA (21).

(C) Unidades de repetição de efetores TAL e sequências alvo previstas em promotores de genes induzidos foram alinhadas manualmente. Os nucleotídeos no filamento de DNA superior que correspondem aos aminoácidos hipervariáveis em cada repetição foram contados com base nas seguintes combinações de oito efetores e genes alvo identificados experimentalmente: AvrBs3/Bs3, UPA10, UPA12, UPA14, UPA19, UPA20, UPA21, UPA23, UPA25, AvrBs3Δrep16/Bs3-E, AvrBs3Δrep109/Bs3, AvrHah1/Bs3, AvrXa27/Xa27, PthXo1/Xa13, PthXo6/OsTFX1, PthXo7/OsTFIIAγ1 (ver Fig. 5). As combinações predominantes (n>4) são sombreadas em cinza. Um asterisco indica que o aminoácido 13 está faltando neste tipo de repetição.

(D) Código de especificidade de DNA alvo (R=A/G; N=A/C/G/T) de tipos de repetição com base nos aminoácidos hipervariáveis 12 e 13 (comprovados experimentalmente neste estudo).

Figura 2 | Sequências de DNA alvo de Hax2, Hax3, e Hax4.

(A) Aminoácidos 12 e 13 das unidades de repetição Hax2, Hax3, e Hax4 e especificidades de DNA alvo previsto (caixa de Hax).

(B) Caixas de Hax foram clonados em frente ao promotor mínimo *Bs4* em um vetor reportador de GUS.

(C) Inducibilidade específica das caixas de Hax por efetores Hax. Os construtos reportadores de GUS foram codistribuídos através de *A. tumefaciens* em *N. benthamiana* com *hax2*, *hax3*, *hax4* direcionados por 35S e T-DNA (-) vazio, respectivamente (barras de erro indicam SD; n=3 amostras; 4-MU, 4-metil-umbeliferona). 35S::*uidA* (+) serviu como controle. Os discos de folhas foram marcados com X-Gluc (5-bromo-4-cloro-3-indolil- β -D-glucuronida).

Figura 3 | Especificidades de reconhecimento de par de bases de DNA de tipos de repetição

(A) Derivados de caixa de Hax4 e ArtX foram clonados em frente ao promotor mínimo *Bs4* em um vetor reportador de GUS.

(B) Especificidade de unidades de repetição NG-, HD-, NI- e NS. Inducibilidade de Hax4 de derivados de caixa de Hax4 permutados em bases alvo de tipo de repetição (fundo cinza).

(C) Especificidade de unidades de repetição de NN. Efetor artificial ArtX1 e seqüências de DNA alvo previstas. Inducibilidade de ArtX1 de derivados de caixa de ArtX1 permutados em bases alvo de repetição de NN (fundo cinza).

(D) Efetores artificiais ArtX2 e ArtX3 e seqüências alvo de DNA derivado.

(E) Inducibilidade específica de caixas de ArtX por efetores artificiais.

(A) - (E) Construtos reportadores de GUS foram codistribuídos através de *A. tumefaciens* em *N. benthamiana* com genes *hax4*, *artX1*, *artX2*, ou *artX3* direcionados por 35S, e T-DNA (-) vazio, respectivamente. 35S::*uidA* (+) serviu como controle. Os discos de folhas foram marcados com X-Gluc. Para dados quantitativos ver Fig. 11.

Figura 4 | Um número mínimo de unidades de repetição é exigido para ativação transcricional.

(A) Efetores de ArtHD artificiais com números diferentes (0,5 a 15,5) de unidades de repetição de HD (total de 1,5 a 16,5 de unidades de repetição).

(B) Uma caixa alvo de ArtHD consistindo de TA e 17 C foi clonada em frente ao promotor mínimo *Bs4* em um vetor reportador de GUS.

(C) Ativação do promotor por efetores de ArtHD com número diferente de unidades de repetição. Gene efector direcionado por 35S ou T-DNA (-) vazio foram codistribuídos através de *A. tumefaciens* com o construto reportador de GUS em *N. benthamiana* (barras de erro indicam SD; n=3 amostras; 4-MU). 35S::uidA (+) serviu como controle. Os discos de folha foram marcados com X-Gluc.

Figura 5 | Alinhamento de seqüências de DNA alvo em promotores de genes induzidos com os aminoácidos hipervariáveis 12 e 13 de unidades de repetição de efetores TAL.

(A) As unidades de repetição de AvrBs3, AvrBs3Δrep16, AvrBs3Δrep109 e AvrHah1 foram alinhadas à caixa *UPA* no promotor do gene ECW-30R *Bs3* de pimenta (acesso: EU078684). AvrBs3Δrep16 e AvrBs3Δrep109 são derivados de deleção de AvrBs3 em que as unidades de repetição 11-14 e as unidades de repetição 12-14 foram deletadas, respectivamente. AvrBs3, AvrBs3Δrep109 e AvrHah1, mas não AvrBs3Δrep16 induzem o HR em plantas ECW-30R.

(B) Unidades de repetição de AvrBs3, AvrBs3Δrep16, AvrBs3Δrep109, e AvrHah1 foram alinhadas à caixa de *UPA* não-funcional no promotor do gene ECW *Bs3-E* de pimenta (acesso: EU078683). AvrBs3Δrep16, mas não AvrBs3, AvrBs3Δrep109, ou AvrHah1 induzem HR em plantas ECW de pimenta.

(C) Unidades de repetição de AvrXa27 foram alinhadas a uma

seqüência alvo putativo no promotor do gene *Xa27* de arroz. *Xa27* (acesso: AY986492) é induzido por AvrXa27 em arroz cultivar IRBB27 gerando um HR, mas não *xa27* (acesso: AY986491) em arroz cultivar IR24.

(D) Unidades de repetição de PthXo1 foram alinhadas a uma seqüência alvo putativo no promotor do gene *Xa13/Os8N3* de arroz. *Xa13* (acesso: DQ421396) é induzido por PthXo1 em arroz cultivar IR24 gerando susceptibilidade, mas não *xa13* (acesso: DQ421394) em arroz cultivar IRBB13.

(E) Unidades de repetição de PthXo6 foram alinhadas a uma seqüência alvo putativo no promotor do gene *OsTFX1* de arroz (acesso: AK108319). *OsTFX1* é induzido por PthXo6 em arroz cultivar IR24.

(F) Unidades de repetição de PthXo7 foram alinhadas a uma seqüência alvo putativa no promotor do gene *OsTFIIAγ1* de arroz (CB097192). *OsTFIIAγ1* é induzido por PthXo7 em arroz cultivar IR24.

(A) - (F) Números acima das seqüências de DNA indicam distância de nucleotídeo ao primeiro ATG na região de codificação. Combinações de repetição/base que não correspondem à especificidade alvo prevista (aminoácidos 12/13: NI=A; HD=C; NG=T; NS=A/C/G/T; NN=A/G; IG=T) são coloridas em vermelho. As unidades de repetição com especificidade de DNA alvo desconhecida são coloridas em verde.

Figura 6 | Região de DNA protegida por AvrBs3Δrep16 é 4 bp mais curta do que AvrBs3.

Resumo de análises de pegadas de DNaseI com AvrBs3 e AvrBs3Δrep16 (ver Figs. 7, 8).

(A) Seqüências promotoras de *Bs3* (topo) e *Bs3-E* (meio) protegidas por AvrBs3 e AvrBs3Δrep16, respectivamente. A análise de pegadas de DNaseI revelou que AvrBs3 protegeu 37 nucleotídeos do filamento sentido e 36 nucleotídeos do filamento anti-sentido do promotor de *Bs3*, e AvrBs3Δrep16 protegeu 30 nucleotídeos do filamento sentido e 32

nucleotídeos do filamento anti-sentido do promotor *Bs3-E*. A caixa de *UPA* e a caixa de *AvrBs3Δrep16* prevista são sublinhadas. Sequências promotoras de *UPA20-ubm-r16* (parte inferior) protegidas por *AvrBs3* e *AvrBs3Δrep16*. O promotor de *UPA20-ubmr16* é um derivado de promotor de *UPA20* com uma substituição de 2 bp (GA para CT, **itálico negrito**) que resulta no reconhecimento por ambos *AvrBs3* e *AvrBs3Δrep16*. A análise de pegadas de DNaseI revelou que 35 nucleotídeos do filamento sentido e 34 nucleotídeos do filamento anti-sentido são protegidos por *AvrBs3* (caixa de *UPA* é sublinhada), e 31 nucleotídeos do filamento sentido e 32 nucleotídeos do filamento anti-sentido são protegidos por *AvrBs3Δrep16* (caixa de *AvrBs3Δrep16* é sublinhada). As regiões de DNA sombreadas em verde (*AvrBs3*) ou vermelho (*AvrBs3Δrep16*) se referem às pegadas centrais que foram protegidas por *AvrBs3* e *AvrBs3Δrep16*, respectivamente, em cada experimento, mesmo com quantidades baixas de proteína (molaridade igual de DNA e dímeros de proteína). As regiões de DNA sombreadas em cinza se referem a nucleotídeos que não foram protegidos em todos dos 4 experimentos em todas as concentrações de proteína pelas proteínas fornecidas. Por favor, observe que as extremidades 5' das regiões protegidas por *AvrBs3*- e *AvrBs3Δrep16* são idênticas. As linhas verticais tracejadas indicam as diferenças entre as extremidades 3' das regiões promotoras protegidas por *AvrBs3*- e *AvrBs3Δrep16* que são compatíveis com nosso modelo em que uma repetição cantata um par de bases no DNA.

(B) Alinhamento de sequências de DNA alvo de *AvrBs3* e *AvrBs3Δrep16* no promotor de *UPA20-ubm-r16* com regiões de repetição de *AvrBs3* e *AvrBs3Δrep16* (aminoácidos hipervariáveis na posição 12 e 13). As combinações de repetição/base que não correspondem a nossa especificidade alvo prevista (aminoácidos 12/13: NI=A; HD=C; NG=T; NS=A/C/G/T) são coloridas em vermelho.

Figura 7 | Sequências promotoras de *Bs3* e *Bs3-E* protegidas

por AvrBs3 e AvrBs3 Δ rep16, respectivamente.

Um experimento de pegada de DNaseI representativo é mostrado. A pegada de AvrBs3 DNaseI na sequência promotora de *Bs3* (**A**, filamento de DNA sentido/superior; **B**, filamento de DNA anti-sentido/inferior). Pegada de AvrBs3 Δ rep16 DNaseI na sequência promotora de *Bs3-E* (**C**, filamento de DNA sentido, superior; **D**, filamento de DNA anti-sentido).

(**A**) - (**D**) (parte superior) Produto de PCT marcado fluorescentemente foi incubado com um excesso molar de 5x (calculado para dímeros de proteína) de His6::AvrBs3, His6::AvrBs3 Δ rep16, e BSA, respectivamente, tratado com DNaseI e analisado em um seqüenciador capilar. O eixo y do eletroferograma mostra a intensidade de fluorescência relativa ao filamento sentido marcado por 5'-6-FAM (a, c) ou o filamento anti-sentido marcado por 5'-HEX (b, d) do produto de PCR em uma escala arbitrária. Os traços para as reações com His6::AvrBs3 (verde) ou His6::AvrBs3 Δ rep16 (vermelho), respectivamente, e BSA (preto, controle negativo) foram sobrepostos. Uma redução da altura de pico na presença de AvrBs3 ou AvrBs3 Δ rep16, respectivamente, em comparação ao controle negativo corresponde à proteção. A região protegida é indicada por linhas verticais verde (AvrBs3) ou vermelha (AvrBs3 Δ rep16). (meio) Eletroferograma da sequência de DNA. Os picos de cor laranja com números correspondem ao padrão de tamanho do nucleotídeo de DNA. As caixas alvo previstas dos efetores na sequência de DNA são sublinhadas. Os nucleotídeos cobertos são marcados por uma caixa verde (AvrBs3) ou vermelha (AvrBs3 Δ rep16). Os números abaixo se referem a posições de nucleotídeos relativas ao início da transcrição (+1) na presença de AvrBs3 (a, b) ou AvrBs3 Δ rep16 (c, d), respectivamente. (parte inferior) Produto de PCR de DNA usado para pegada de DNaseI, amplificado a partir dos promotores de *Bs3* (a, b) ou *Bs3-E* (c, d), respectivamente. As regiões protegidas nos

filamentos de DNA únicos são indicadas por caixas cinzas. Os números abaixo se referem a posições de nucleotídeos relativas ao início da transcrição (+1) na presença de AvrBs3 (a, b) ou AvrBs3 Δ rep16 (c, d), respectivamente. Os experimentos foram repetidos três vezes com resultados semelhantes.

Figura 8 | Sequência promotora de *UPA20-ubm-r16* protegida por AvrBs3 e AvrBs3 Δ rep16.

[00017] Um experimento de pegada de DNaseI representante. Pegada de DNaseI de AvrBs3 e AvrBs3 Δ rep16 na sequência promotora de *UPA20-ubm-r16* (**A**), filamento de DNA sentido, superior; (**B**) filamento de DNA anti-sentido, inferior). (Produto de PCR marcado fluorescentemente superior foi incubado com um excesso molar de 5x de His6::AvrBs3, His6::AvrBs3 Δ rep16 e BSA (calculado para dímeros de proteína), respectivamente, tratado com DNaseI e analisado em um seqüenciador capilar. O eixo y do eletroferograma mostra a intensidade de fluorescência relativa correspondente ao filamento sentido marcado por 5'-6-FAM (a) ou ao filamento anti-sentido marcado por 5'-HEX (b) do produto de PCR em uma escala arbitrária. Os traços para as reações com His6::AvrBs3 (verde), His6::AvrBs3 Δ rep16 (vermelho) e o controle negativo BSA (preto) foram sobrepostos. Uma redução de altura de pico na presença de AvrBs3 e AvrBs3 Δ rep16 em comparação ao controle negativo corresponde à proteção. As regiões protegidas são indicadas por linhas verticais verde (AvrBs3) e vermelha (AvrBs3 Δ rep16). (meio) Eletroferograma da sequência de DNA. Os picos de cor laranja com números correspondem ao padrão de tamanho do nucleotídeo de DNA. Os nucleotídeos cobertos por AvrBs3 são marcados por linhas verdes e uma caixa verde (com a caixa *UPA* sublinhada), os nucleotídeos cobertos por AvrBs3 Δ rep16 são marcados por linhas vermelhas e uma caixa vermelha (com a caixa de AvrBs3 Δ rep16 sublinhada). A mutação *UPA20-ubm-r16* (GA a CT) é indicada em *itálico*. (parte inferior) Produto de PCR de DNA usado para pegada de DNaseI, ampliado a partir do promotor de

UPA20-ubm-r16. As regiões protegidas nos filamentos de DNA são indicadas por caixas cinzas. Os números abaixo se referem a posições de nucleotídeos relativas ao início da transcrição (+1) do promotor de tipo selvagem de *UPA20* na presença de AvrBs3. O experimento foi repetido três vezes com resultados semelhantes.

Figura 9 | Construtos reportadores de GUS.

[00018] Sequências de DNA alvo (caixa de efetor TAL) foram inseridas 5' da sequência do promotor mínimo de *Bs4* de tomate (41) (*pBs4*; -50 a +25) e transferidas por recombinação GATEWAY no vetor de T-DNA de *A. tumefaciens* T-DNA pGWB330 que constrói uma fusão para um gene *uidA* (β -glucuronidase, GUS) não promotor. Locais de recombinação de attB1, attB2; GATEWAY.

Figura 10 | Especificidade de reconhecimento da repetição putativa 0 em Hax3.

(A) Os aminoácidos 12 e 13 das unidades de repetição de Hax3 e quatro possíveis caixas de Hax3 alvo com permutações na posição correspondente à repetição 0.

(B) As caixas alvo foram clonadas em frente ao promotor mínimo de *Bs4* de tomate em um vetor reportador de GUS.

(C) Atividades de GUS com *hax3* direcionado por 35S ou T-DNA (-) vazio codistribuídas através de *A. tumefaciens* com os construtos reportadores de GUS em células de folhas de *N. benthamiana* (4-MU, 4-metil-umbelliferona; n=3; barras de erro indicam SD). Para análises qualitativas, os discos de folha foram marcados com X-Gluc. O experimento foi realizado duas vezes com resultados semelhantes.

Figura 11 | Especificidades de reconhecimento de par de bases de DNA de tipos de repetição

[00019] Derivados de caixa de Hax4 e ArtX foram clonados em frente ao promotor mínimo de *Bs4* em um vetor reportador de GUS. Dados

quantitativos a Fig. 3.

(A) Especificidade de unidades de repetição de NG-, HD-, NI-, e NS. Inducibilidade de Hax4 de derivados de caixa de Hax4 permutados em bases alvo de tipo de repetição.

(B) Especificidade de unidades de repetição de NN. Inducibilidade de ArtX1 de derivados de caixa de ArtX1 permutados em bases alvo de repetição de NN.

(C) Inducibilidade específica de caixas de ArtX por efetores artificiais ArtX1, ArtX2 e ArtX3, respectivamente.

(A) - (C) Construtos reportadores de GUS foram codistribuídos através de *A. tumefaciens* em células de folha *N. benthamiana* junto com genes *hax4*, *artX1*, *artX2*, *artX3* direcionados por 35S (barras cinzas), e T-DNA vazio (a, b, barras brancas; c, -), respectivamente (n=3; barras de erro indicam SD). 35S::*uidA* (+) serviu como controle. Os experimentos foram realizados três vezes com resultados semelhantes.

Figura 12 | Sequências de DNA alvo previstas para AvrXa10.

(A) Aminoácidos 12 e 13 das unidades de repetição de AvrXa10 e duas possíveis caixas alvo com especificidade do tipo repetição de NN previstas A ou G.

(B) Caixas alvo de AvrXa10 foram clonadas em frente ao promotor mínimo de *Bs4* em um vetor reportador de GUS.

(C) Análise de GUS de *avrXa10*, *hax3* (controle de especificidade) direcionados por 35S, ou T-DNA (-) vazio codistribuído através de *A. tumefaciens* com construtos reportadores de GUS em células de folhas *N. benthamiana*. 35S::*uidA* (+) serviu como controle constitutivo (n=3; barras de erro indicam SD). Para análises qualitativas, discos de folhas foram marcados com X-Gluc. O experimento foi realizado três vezes com resultados semelhantes.

Figura 13 | Especificidade de reconhecimento de IG do tipo

repetição em Hax2.

(A) Aminoácidos 12 e 13 unidades de repetição de Hax2 e quatro possíveis caixas de Hax2 alvo para IG do tipo repetição.

(B) As caixas alvo de Hax2 foram clonadas em frente ao promotor mínimo de *Bs4* em um vetor reportador de GUS.

(C) Análise de GUS do *hax2* direcionado por promotor de 35S ou T-DNA (-) vazio codistribuído através de *A. tumefaciens* com os construtos reportadores de GUS nas células de folhas *N. benthamiana*. 35S::*uidA* (+) serviu como controle constitutivo (n=3; barras de erro indicam SD. Para análises qualitativas, os discos de folhas foram marcados com X-Gluc. O experimento foi realizado três vezes com resultados semelhantes.

Figura14 | Hax2 induz expressão de *PAP1* em *A. thaliana*.

(A) Folhas de *A. thaliana* foram inoculadas com cepas de *A. tumefaciens* que distribuem construtos de T-DNA para expressão de *hax2*, *hax3* e *hax4* direcionados por 35S, respectivamente. A expressão de *hax2*, mas não de *hax3* e *hax4* induziu pigmentação roxa sugestiva de produção de antocianina. A fotografia foi tirada 7 dias após inoculação.

(B) Linha de *A. thaliana* transgênica que transporta *hax2* sob controle de um promotor induzível por etanol. As plantas de uma população de T2 segregante foram borrifadas com 10% de etanol para induzir expressão do transgene. Somente plantas transgênicas *hax2* acumularam antocianina. A fotografia foi tirada 6 dias após o tratamento.

(C) RT-PCR semiquantitativa de *hax2* (29 ciclos), *PAP1* (32 ciclos), e fator de alongamento Tu (*EF-Tu*, 32 ciclos) com cDNA de plantas transgênicas de *hax2*- de três linhas independentes de *A. thaliana* antes de (-) e 24 h após (+) borrifação com 10% de etanol.

(D) Aminoácidos 12 e 13 de unidades de repetição de Hax2 e sequência de DNA alvo de Hax2.

(E) O promotor de *PAP1* de *A. thaliana* Col-0 contém uma

caixa de Hax2 imperfeita. As incompatibilidades à caixa de Hax2 prevista são coloridas em vermelho. Uma caixa de TATA putativa, o local de início de transcrição natural (+1), e o primeiro códon da sequência de codificação de *PAP1* são indicados.

Figura 15 | Tabela I. Sequências de DNA alvo previstas de efetores TAL

[00020] A tabela mostra sequências de repetição de efetores TAL e as sequências de DNA alvo previstas usadas a partir dos aminoácidos 12 d 13 das unidades de repetição.

[00021] As anotações mostram:

(A) *Xcv*, *Xanthomonas campestris* pv. *vesicatoria*; *Xg*, *Xanthomonas gardneri*; *Xca*, *Xanthomonas campestris* pv. *armoraciae*; *Xoo*, *Xanthomonas oryzae* pv. *oryzae*; *Xac*, *Xanthomonas axonopodis* pv. *citri*; *Xau*, *Xanthomonas citri* pv. *aurantifolii*; *Xcm*, *Xanthomonas campestris* pv. *malvacearum*; *Xam*, *Xanthomonas axonopodis* pv. *manihotis*; *Xoc*, *Xanthomonas oryzae* pv. *oryzicola*

(B) Um asterisco (*) indica uma deleção de aminoácido 13

(C) Especificidade de DNA alvo deduzida a partir dos aminoácidos 12 e 13 das unidades de repetição. Um nucleotídeo de timidina é adicionado na extremidade A 5' devido à especificidade da repetição putativa 0. A sequência do filamento (sentido) superior do DNA de filamento duplo é fornecida em código ambíguo (R = A/G; N = A/C/G/T; • = especificidade desconhecida).

Figura 16 | Sequências de proteínas de AvrBs3, Hax2, Hax3, Hax4

[00022] Para cada sequência de proteínas são mostrados o terminal N, terminal C, bem como as sequências de repetição única.

Figura 17 | O efector ARTBs4 induz expressão do promotor mínimo de Bs4

(A) Aminoácidos 12 e 13 das unidades de repetição de Hax4 e especificidade de DNA alvo prevista (caixa de Hax4). A caixa de Hax4(mut) contém quatro trocas de pares de base em comparação à caixa de Hax4.

(B) Aminoácidos 12 e 13 das unidades de repetição de efetor artificial ARTBs4 e especificidade de DNA alvo prevista (caixa de ARTBs4).

(C) A caixa de Hax4 foi clonada em frente ao promotor mínimo de *Bs4* em um vetor reportador de GUS. A caixa de ARTBs4 é naturalmente presente no promotor mínimo de *Bs4*.

(D) Inducibilidade específica das caixas de Hax4 e ARTBs4 por Hax4 e ARTBs4, respectivamente. Os construtos reportadores de GUS foram codistribuídos através de *Agrobacterium tumefaciens* em *N. benthamiana* com *hax4* direcionado por 35S (barras cinzas), *ARTBs4* (barras brancas) e T-DNA vazio (ev, barras pretas), respectivamente (barras de erro indicam SD). 4-MU, 4-metil-umbeliferona. 35S::*uidA* (GUS, barra cinza) serviu como controle. Os discos de folhas foram marcados com X-Gluc (5-bromo-4-cloro-3-indolil- β -D-glucuronida).

Figura 18 | Diagrama para clonagem “Golden gate” de domínios de repetição e efetores

(A) Os blocos de construção que consistem de unidades de repetição individuais (ou outros domínios de proteínas) são subclonados com locais alvo de enzimas de restrição do tipo II flanqueadoras (e.g. *BsaI*) que geram ressaltos específicos. Os ressaltos correspondentes são indicados com letras idênticas (A a O). Tipos de repetição diferentes são clonados como blocos de construção para cada posição (e.g. repetição 1, repetição 2, etc.). As especificidades de repetição são: NI=A, HD=C, NG=T, NN=G ou A.

(B) Os blocos de construção são montados em um vetor alvo pela ligação de ressaltos correspondentes usando clonagem “Golden gate” (ligação de restrição). No geral, o produto de montagem resultante não contém qualquer dos locais alvo para clonagem.

Figura 19 | Método alternativa para geração de efetores designer através de clonagem Golden Gate

[00023] As figuras 19 A a D descrevem vários vetores descritos nos métodos divulgados no Exemplo 3 abaixo bem como fornecem um esquema do método.

LISTAGEM DE SEQUÊNCIAS

[00024] As sequências de nucleotídeos e aminoácidos listadas nas figuras em anexo e a listagem de sequências são mostradas usando abreviações de letras padrão para bases de nucleotídeos, e código de uma letra para aminoácidos. As sequências de nucleotídeos seguem a convenção padrão de início na 5' extremidade da sequência seguindo para a extremidade 3' (i.e., da esquerda para direita em cada linha). Apenas um filamento de cada sequência de ácidos nucleicos é mostrada, mas o filamento complementar é entendido como devendo ser induzido por qualquer referência ao filamento exibido. As sequências de aminoácidos seguem a convenção padrão de início no terminal amino da sequência seguindo para o terminal carboxi (i.e., da esquerda para direita em cada linha).

DESCRIÇÃO DETALHADA DA INVENÇÃO

[00025] A presente invenção agora será descrita mais detalhadamente com referência aos desenhos em anexo, em que algumas, mas não todas as modalidades da invenção são mostradas. De fato, estas invenções podem ser incorporadas de formas muito diferentes e não devem ser interpretadas como limitadas às modalidades estipuladas no presente documento; em vez disto, estas modalidades são fornecidas de modo que esta divulgação cumprirá exigências legais aplicáveis. Números semelhantes se referem a elementos semelhantes do começo ao fim.

[00026] Muitas modificações e outras modalidades da invenção estipuladas no presente documento virão à mente para aqueles versados na técnica a qual pertencem estas invenções tendo o benefício das instruções

apresentadas nas descrições anteriores e nos desenhos em anexo. Portanto, deve ser entendido que as invenções não devem ser limitadas às modalidades específicas divulgadas e que as modificações e outras modalidades se destinam a ser incluídas no escopo das reivindicações em anexo. Embora termos específicos sejam empregados no presente documento, eles são usados em sentido genérico e descritivo apenas e não para propósitos de limitação.

[00027] Uma série de termos que são usados nesta divulgação é definida abaixo no presente documento.

[00028] O termo “domínio de repetição” é usado para descrever o domínio de reconhecimento de DNA a partir de um efetor TAL, ou versão artificial do mesmo que é feito usando os métodos divulgados, consistindo de unidades de repetição modulares que quando presentes em um polipeptídeo conferem especificidade de DNA alvo. Um domínio de repetição composto de unidades de repetição pode ser adicionado a qualquer polipeptídeo em que direcionamento de sequência de DNA é desejado e não são limitadas a uso em efetores TAL.

[00029] O termo “unidade de repetição” é usado para descrever a porção modular de um domínio de repetição a partir de um efetor TAL, ou uma versão artificial do mesmo, que contém um aminoácido ou dois aminoácidos adjacentes que determinam o reconhecimento de um par de bases em uma sequência de DNA alvo. As unidades de repetição retiradas juntas reconhecem uma sequência de DNA alvo definida e constituem um domínio de repetição. As unidades de repetição podem ser adicionadas a qualquer polipeptídeo em que é desejado direcionamento de sequência de DNA e não são limitadas a uso em efetores TAL.

[00030] O termo “código de reconhecimento” é usado para descrever a relação entre os aminoácidos nas posições 12 e 13 de uma unidade de repetição e o par de bases de DNA correspondente em uma sequência alvo de DNA de modo que os aminoácidos confirmem reconhecimento de, como segue:

HD para reconhecimento de C/G; NI para reconhecimento de A/T; NG para reconhecimento de T/A; NS para reconhecimento de C/G ou A/T ou T/A ou G/C; NN para reconhecimento de G/C ou A/T; IG para reconhecimento de T/A; N para reconhecimento de C/G; HG para reconhecimento de C/G ou T/A; H para reconhecimento de T/A; e NK para reconhecimento de G/C.

[00031] Conforme usado no presente documento, “efetor” (ou “proteína efetora” ou “polipeptídeo efetor”) se refere a construtos ou seus produtos de polipeptídeo codificados em que tal polipeptídeo é capaz de reconhecer uma sequência de DNA alvo. A proteína efetora inclui um domínio de repetição composto de 1,5 ou mais unidades de repetição e também pode incluir um ou mais domínios funcionais como um domínio regulatório. Em modalidades preferenciais da invenção, o “efetor” é adicionalmente capaz de exercer um efeito, como regulação da expressão de genes. Embora a presente invenção não dependa de um mecanismo biológico específico, acredita-se que proteínas ou polipeptídeos da invenção que reconhecem uma sequência de DNA alvo se ligam à sequência de DNA alvo.

[00032] O termo “de ocorrência natural” é usado para descrever um objeto que pode ser encontrado na natureza diferente de como é produzido pelo homem. Por exemplo, uma sequência de polipeptídeos ou polinucleotídeos que está presente em um organismo (incluindo vírus) que pode ser isolado de uma fonte na natureza e que não foi intencionalmente modificado no laboratório é de ocorrência natural. Geralmente, o termo “de ocorrência natural” se refere a um objeto como presente em um indivíduo do tipo selvagem, como seria típico para a espécie.

[00033] Os termos “expressão moduladora”, “expressão inibidora” e “expressão ativadora” de um gene se referem a uma capacidade de um polipeptídeo da presente invenção ativar ou inibir a transcrição de um gene. A ativação inclui prevenção de inibição transcricional subsequente (i.e., prevenção de repressão de expressão de genes) e a inibição inclui prevenção

de ativação transcricional subsequente (i.e., prevenção de uma ativação de gene). A modulação pode ser analisada determinando qualquer parâmetro que seja indiretamente ou diretamente afetado pela expressão do gene alvo. Tais parâmetros incluem, ex., mudanças em níveis de RNA ou proteína, mudanças em atividade de proteínas, mudanças em níveis de produtos, mudanças em expressão de genes a jusante, mudanças em transcrição de genes reportadores (luciferase, CAT, beta-galactosidase, GFP (ver, ex., Mistili & Spector (1997) *Nature Biotechnology* 15:961-964); mudanças em transdução de sinal, fosforilação e desfosforilação, interações receptor-ligante, concentrações de segundo mensageiro (ex., cGMP, cAMP, IP3 e Ca²⁺), crescimento celular, neovascularização, in vitro, in vivo e ex vivo. Tais efeitos funcionais podem ser medidos por quaisquer meios conhecidos por aqueles versados na técnica, ex., medição de níveis de RNA ou proteínas, medição de estabilidade de RNA, identificação de expressão de gene reportador ou a jusante, ex., através de quimioluminescência, fluorescência, reações calorimétricas, ligação a anticorpos, marcadores induzíveis, análises de ligação a ligantes; mudanças em segundos mensageiros intracelulares como cGMP e inositol trifosfato (IP3); mudanças em níveis de cálcio intracelular; liberação de citosina e semelhantes.

[00034] Um “domínio regulatório” se refere a uma proteína ou uma subsequência de proteínas que tem atividade de modulação transcricional. Geralmente, um domínio regulatório é covalentemente ou não-covalentemente ligado a um polipeptídeo da presente invenção para modular transcrição. Como alternativa, um polipeptídeo da presente invenção pode agir sozinho, sem um domínio regulatório, ou com múltiplos domínios regulatórios para modular transcrição. Os polipeptídeos de fator de transcrição a partir dos quais é possível obter um domínio regulatório incluem aqueles que são envolvidos em transcrição regulada e basal. Tais polipeptídeos incluem fatores de transcrição, seus domínios efetores, coativadores,

silenciadores, receptores de hormônios nucleares (ver, ex., Goodrich *et al.* (1996) *Cell* 84:825 30 para uma revisão de elementos de proteínas e ácidos nucléicos envolvidos em transcrição; fatores de transcrição em geral são revisados em Barnes & Adcock (1995) *Clin. Exp. Allergy* 25 Suppl. 2:46 9 and Roeder (1996) *Methods Enzymol.* 273:165 71). Bancos de dados dedicados a fatores de transcrição são conhecidos (ver, ex., *Science* (1995) 269:630). Os fatores de transcrição de receptores de hormônios nucleares são descritos em, por exemplo, Rosen *et al.* (1995) *J. Med. Chem.* 38:4855 74. A família C/EBP de fatores de transcrição é revisada em Wedel *et al.* (1995) *Immunobiology* 193:171 85. Os coativadores e corepressores que mediam a regulação de transcrição por receptores de hormônios nucleares são revisados em, por exemplo, Meier (1996) *Eur. J. Endocrinol.* 134(2):158 9; Kaiser *et al.* (1996) *Trends Biochem. Sci.* 21:342 5; and Utley *et al.* (1998) *Nature* 394:498 502). Os fatores de transcrição de GATA, que são envolvidos na regulação de hematopoiese, são descritos em, por exemplo, Simon (1995) *Nat. Genet.* 11:9 11; Weiss *et al.* (1995) *Exp. Hematol.* 23:99-107. A proteína de ligação de caixa de TATA (TBP) e seus polipeptídeos TAF relacionados (que incluem TAF30, TAF55, TAF80, TAF110, TAF150 e TAF250) são descritos em Goodrich & Tjian (1994) *Curr. Opin. Cell Biol.* 6:403 9 and Hurley (1996) *Curr. Opin. Struct. Biol.* 6:69 75. A família de fatores de transcrição é revisada em, por exemplo, Barahmand-Pour *et al.* (1996) *Curr. Top. Microbiol. Immunol.* 211:121 8. Os fatores de transcrição envolvidos na doença são revisados em Aso *et al.* (1996) *J. Clin. Invest.* 97:1561 9. Quinases, fosfatases e outras proteínas que modificam polipeptídeos envolvidos em regulação de genes também são úteis como domínios regulatórios para polipeptídeos da presente invenção. Tais modificadores são com frequência envolvidos na ativação e desativação de transcrição mediada por, por exemplo, hormônios. As quinases envolvidas na regulação de transcrição são revisadas em Davis (1995) *Mol. Reprod. Dev.* 42:459 67,

Jackson *et al.* (1993) *Adv. Second Messenger Phosphoprotein Res.* 28:279 86, and Boulikas (1995) *Crit. Rev. Eukaryot. Gene Expr.* 5:1 77, enquanto as fosfatases são revisadas em, por exemplo, Schonthal & Semin (1995) *Cancer Biol.* 6:239 48. Tirosina quinases nucleares são descritas em Wang (1994) *Trends Biochem. Sci.* 19:373 6. Domínios úteis também podem ser obtidos a partir dos produtos de genes de oncogenes (ex., membros das famílias myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos) e seus fatores e modificadores relacionados. Os oncogenes são descritos em, por exemplo, Cooper, *Oncogenes*, 2nd ed., The Jones and Bartlett Series in Biology, Boston, Mass., Jones and Bartlett Publishers, 1995. Os fatores de transcrição de ets são revisados em Waslylk *et al.* (1993) *Eur. J. Biochem.* 211:7 18 and Crepieux *et al.* (1994) *Crit. Rev. Oncog.* 5:615 38. Os oncogenes myc são revisados, por exemplo, em Ryan *et al.* (1996) *Biochem. J.* 314:713 21. Os fatores de transcrição jun e fos são descritos em, por exemplo, The Fos and Jun Families of Transcription Factors, Angel & Herrlich, eds. (1994). O oncogene max é revisado em Hurlin *et al.* *Cold Spring Harb. Symp. Quant. Biol.* 59:109 16. A família do gene myb é revisada em Kanei-Ishii *et al.* (1996) *Curr. Top. Microbiol. Immunol.* 211:89 98. A família mos é revisada em Yew *et al.* (1993) *Curr. Opin. Genet. Dev.* 3:19 25. Os polipeptídeos da presente invenção podem incluir domínios regulatórios obtidos a partir de enzimas de recepção de DNA e seus fatores e modificadores relacionados. Os sistemas de reparação de DNA são revisados em, por exemplo, Vos (1992) *Curr. Opin. Cell Biol.* 4:385 95; Sancar (1995) *Ann. Rev. Genet.* 29:69 105; Lehmann (1995) *Genet. Eng.* 17:1 19; and Wood (1996) *Ann. Rev. Biochem.* 65:135 67. As enzimas de rearranjo de DNA e seus fatores e modificadores associados podem também ser usados como domínios regulatórios (ver, ex., Gangloff *et al.* (1994) *Experientia* 50:261 9; Sadowski (1993) *FASEB J.* 7:760 7).

[00035] Semelhantemente, os domínios regulatórios podem ser

derivados de enzimas modificadoras de DNA (ex., DNA metiltransferases, topoisomerases, helicases, ligases, quinases, fosfatases, polimerases) e seus fatores e modificadores relacionados. Helicases são revisadas em Matson *et al.* (1994) *Bioessays* 16:13 22, e metiltransferases são descritas em Cheng (1995) *Curr. Opin. Struct. Biol.* 5:4 10. Proteínas associadas a cromatina e seus modificadores (ex., quinases, acetilases e deacetilases), como uma histona deacetilase (Wolffe *Science* 272:371 2 (1996)) também são úteis como domínios para adição ao efector de escolha. Em uma modalidade preferencial, o domínio regulatório é um DNA metil transferase que age como um repressor transcricional (ver, ex., Van den Wyngaert *et al.* *FEBS Lett.* 426:283 289 (1998); Flynn *et al.* *J. Mol. Biol.* 279:101 116 (1998); Okano *et al.* *Nucleic Acids Res.* 26:2536 2540 (1998); e Zardo & Caiafa, *J. Biol. Chem.* 273:16517 16520 (1998)). Em outra modalidade preferencial, endonucleases como Fok1 são usadas como repressores transcricionais, que agem através de clivagem de genes (ver, ex., WO95/09233; e PCT/US94/01201). Fatores que controlam cromatina e estrutura, movimento e localização de DNA e seus fatores e modificadores relacionados; fatores derivados de micróbios (ex., procariotos, eucariotos e vírus) e fatores que se associam ou os modificam também podem ser usados para obter proteínas quiméricas. Em uma modalidade, as recombinases e integrases são usadas como domínios regulatórios. Em uma modalidade, histona acetiltransferase é usada como um ativador transcricional (ver, ex., Jin & Scotto (1998) *Mol. Cell. Biol.* 18:4377 4384; Wolffe (1996) *Science* 272:371 372; Taunton *et al.* *Science* 272:408 411 (1996); and Hassig *et al.* *PNAS* 95:3519 3524 (1998)). Em outra modalidade, histona deacetilase é usada como um repressor transcricional (ver, ex., Jin & Scotto (1998) *Mol. Cell. Biol.* 18:4377 4384; Syntichaki & Thireos (1998) *J. Biol. Chem.* 273:24414 24419; Sakaguchi *et al.* (1998) *Genes Dev.* 12:2831 2841; and Martinez *et al.* (1998) *J. Biol. Chem.* 273:23781 23785).

[00036] Conforme usado no presente documento, “gene” se refere a uma molécula de ácido nucléico ou porção da mesma que compreende uma seqüência de codificação, opcionalmente contendo íntrons, e regiões de controle que regulam a expressão da seqüência de codificação e a transcrição de porções não-traduzidas da transcrição. Assim, o termo “gene” inclui, além da seqüência de codificação, seqüência regulatória como o promotor, realçador, regiões 5’ não traduzidas, região 3’ não traduzida, sinais de terminação, região de poli adenilação e semelhantes. A seqüência regulatória de um gene pode ser localizada próxima a, dentro ou distante da região de codificação.

[00037] Conforme empregado aqui, “gene alvo” se refere a um gene cuja expressão deve ser modulada por um polipeptídeo da presente invenção.

[00038] Conforme empregado no presente documento, “planta” se refere a quaisquer de vários organismos fotosintéticos, multi-celulares eucarióticos do reino das Plantas, caracteristicamente produzindo embriões contendo cloroplastos, tendo paredes de células de celulose e desprovidos de locomoção. Conforme usado no presente documento, “planta” inclui qualquer planta ou parte de uma planta em qualquer estágio de desenvolvimento, incluindo sementes, culturas de suspensão, embriões, regiões meristemáticas, tecido callus, folhas, raízes, brotos, gametófitos, esporófitos, pólen, microesporos e sua progenitura. Também são incluídos cortes e culturas de células ou tecidos. Conforme usado junto com a presente invenção, o termo “tecido de planta” inclui, mas não é limitado a, plantas inteiras, células de plantas, órgãos de plantas, ex., folhas, caules, raízes, meristemas, sementes de plantas, callus, culturas de células e quaisquer grupos de células de plantas organizados em unidades estruturais e/ou funcionais.

[00039] Conforme usado no presente documento, “modular a expressão de um gene alvo em células de plantas” se refere a aumentar

(ativação) ou diminuir (repressão) a expressão do gene alvo em células de plantas com um polipeptídeo da presente invenção, sozinho ou em combinação com outros fatores regulatórios de transcrição e/ou translacionais, ou ácidos nucleicos codificando tal polipeptídeo, em células de plantas.

[00040] Conforme usado no presente documento, uma “sequência de DNA alvo” se refere a uma porção de DNA de filamento duplo a qual é desejado reconhecimento por uma proteína. Em uma modalidade, uma “sequência de DNA alvo” é todo ou parte de um elemento de controle transcricional para um gene para qual um resultado fenotípico desejado pode ser obtidos alterando ou grau de sua expressão. Um elemento de controle transcricional inclui elementos de controle positivo e negativo, como um promotor, um realçador, outros elementos de resposta, ex., elemento de resposta esteróide, elemento de resposta de choque térmico, elemento de resposta de metal, um local de ligação repressor, operador e/ou um silenciador. O elemento de controle transcricional pode ser viral, eucariótico ou procariótico. Uma “sequência de DNA alvo” também inclui uma sequência a jusante ou a montante que pode ser ligar a uma proteína e assim modular, geralmente prevenir, uma transcrição.

[00041] O uso do termo “DNA” ou “sequência de DNA” no presente documento não deve ser destinado a limitar a presente invenção a moléculas de polinucleotídeos que compreendem DNA. Aqueles versados na técnica reconhecerão que os métodos e composições da invenção compreendem moléculas de polinucleotídeos compostas de deoxiribonucleotídeos (i.e., DNA), ribonucleotídeos (i.e., RNA) ou combinações de ribonucleotídeos e deoxiribonucleotídeos. Tais deoxiribonucleotídeos e ribonucleotídeos incluem moléculas de ocorrência natural e análogos sintéticos incluindo, mas não sendo limitados a, análogos de nucleotídeos, resíduos de estrutura principal modificados ou ligações, que são sintéticos, de ocorrência natural e de ocorrência não-natural, que têm propriedades de ligação semelhantes ao ácido

nucléico de referência, e que são metabolizados de forma semelhante aos nucleotídeos de referência. Exemplos de tais análogos incluem, sem limitação, fosforotioatos, fosforamidatos, metil fosfonatos, quiral-metil fosfonatos, 2-O-metil ribonucleotídeos, ácidos nucléicos-peptídeos (PNAs). As moléculas de polinucleotídeos da invenção também compreendem todas as formas de moléculas de polinucleotídeo incluindo, mas não se limitando a, formas de filamento único, formas de filamento duplo, hairpinas, estruturas de haste e laço e semelhantes. Além disso, é entendido por aqueles versados na técnica que as seqüências de DNA divulgadas no presente documento também compreendem o complemento daquela seqüência de nucleotídeos exemplificada.

[00042] Conforme usado no presente documento, “se liga especificamente a uma seqüência de DNA alvo” significa que a afinidade de ligação de um polipeptídeo da presente invenção a uma seqüência de DNA alvo especificada é estatisticamente maior do que a afinidade de ligação do mesmo polipeptídeo a uma seqüência de DNA geralmente comparável, mas não-alvo. Também se refere à ligação de um domínio de repetição da presente invenção a uma seqüência de DNA alvo especificada em um grau detectavelmente maior, ex., pelo menos 1,5 vezes sobre o fundo, do que sua ligação a seqüências de DNA não-alvo e à exclusão substancial de seqüências de DNA não-alvo. Um polipeptídeo do Kd da presente invenção para cada seqüência de DNA pode ser comparado para avaliar a especificidade de ligação do polipeptídeo a uma seqüência de DNA alvo específica.

Conforme usado no presente documento, uma “seqüência de DNA alvo dentro de um gene alvo” se refere a uma relação funcional entre a seqüência de DNA alvo e o gene alvo em que o reconhecimento de um polipeptídeo da presente invenção à seqüência de DNA alvo modular a expressão do gene alvo. A seqüência de DNA alvo pode ser fisicamente localizada em qualquer local dentro dos limites do gene alvo, ex., 5'

extremidades, região de codificação, 3' extremidades, regiões a montante e a jusante foram da região codificada por cDNA, ou dentro do realçador ou outra região regulatória, e pode ser proximal ou distal do gene alvo.

[00043] Conforme usado no presente documento, “endógeno” se refere a sequência de ácido nucléico ou proteína naturalmente associada ao gene alvo ou a uma célula hospedeira na qual é introduzida.

[00044] Conforme usado no presente documento, “exógeno” se refere a sequência de ácido nucléico ou proteína não naturalmente associada a um gene alvo ou a uma célula hospedeira na qual é introduzida, incluindo cópias múltiplas de ocorrência não-natural de um ácido nucléico de ocorrência natural, ex., sequência de DNA, ou sequência de ácidos nucléicos de ocorrência natural localizada em uma localização de genoma de ocorrência não-natural.

[00045] Conforme usado no presente documento, “planta geneticamente modificada (ou planta transgênica)” se refere a uma planta que compreende dentro de seu genoma um polinucleotídeo exógeno. Geralmente e preferencialmente, o polinucleotídeo exógeno é estavelmente integrado dentro do genoma de modo que o polinucleotídeo é passado para gerações sucessivas. O polinucleotídeo exógeno pode ser integrado no genoma sozinho ou como parte de um cassette de expressão recombinante. “Transgênico” é usado no presente documento para incluir qualquer célula, linhagem celular, callus, tecido, parte de planta ou planta, cujo genótipo foi alterado pela presença de ácido nucléico exógeno incluindo aqueles transgênicos inicialmente alterados desta forma bem como aqueles criados por cruzamentos sexuais ou propagação assexual a partir do transgênico inicial. O termo “transgênico” conforme usado no presente documento não compreende a alteração do genoma (cromossomal ou extra-cromossomal) por métodos de semeadura de plantas convencionais ou por eventos de ocorrência natural como fertilização cruzada aleatória, infecção viral não-recombinante,

transformação bacteriana não-recombinante, transposição não-recombinante ou mutação espontânea.

[00046] Conforme usado no presente documento, “promotor mínimo” ou termo substancialmente semelhante se refere a um elemento promotor, especificamente um elemento de TATA, que é inativo ou que tem atividade promotora significativamente reduzida na ausência de ativação a montante. Na presença de um fator de transcrição adequado, o promotor mínimo funciona para permitir transcrição.

[00047] Conforme usado no presente documento, “proteína repressora” ou “repressor” se refere a uma proteína que se liga ao operador de DNA ou a RNA para impedir transcrição ou translação, respectivamente.

[00048] Conforme usado no presente documento, “repressão” se refere à inibição de transcrição ou translação através da ligação da proteína repressora a local específico no DNA ou mRNA. Preferencialmente, a repressão inclui uma mudança significativa no nível de transcrição ou translação de pelo menos 1,5 vezes, mais preferencialmente pelo menos duas vezes, e ainda mais preferencialmente pelo menos cinco vezes.

[00049] Conforme usado no presente documento, “proteína ativadora” ou “ativador” se refere a uma proteína que se liga ao operador de DNA ou a RNA para aumentar a transcrição ou translação, respectivamente.

[00050] Conforme usado no presente documento, “ativação” se refere ao aumento de transcrição ou translação através de ligação da proteína ativadora a local específico em DNA ou mRNA. Preferencialmente, a ativação inclui uma mudança significativa em transcrição de pelo menos 1,5 vezes, mais preferencialmente pelo menos duas vezes, e ainda mais preferencialmente pelo menos cinco vezes.

[00051] Conforme usado no presente documento, “derivado” ou “análogo” de uma molécula se refere a uma porção derivada de ou uma versão modificada da molécula.

[00052] Conforme usado no presente documento, uma “unidade de repetição derivada de um efetor semelhante a ativador de transcrição (TAL)” se refere a uma unidade de repetição de um efetor TAL ou uma versão modificada ou artificial de um ou mais efetores TAL que é produzido por qualquer um dos métodos divulgados no presente documento.

[00053] A seguir, a invenção é especificamente descrita com relação à família efetora semelhante ao ativador de transcrição (TAL) que é translocada através do sistema de secreção do tipo III para células de plantas. O membro do tipo desta família efetora é AvrBs3. Portanto, a família efetora TAL é também denominada família de proteínas semelhante a AvrBs3. Ambas as expressão são usadas sinonimamente e podem ser alternadas. Exemplos não limitantes da família semelhante a AvrBs3 são os seguintes: AvrBs4 e os membros da sub-família Hax , Hax2, Hax3, e Hax4 bem como Brg11. AvrBs3 e os outros membros de sua família são caracterizados por sua capacidade de ligação a seqüências de DNA específicas em regiões promotoras de genes alvo e indução de expressão destes genes. Eles conservaram características estruturais que os permitem agir como ativadores transcricionais de genes de plantas. A família semelhante a AvrBs3 e efetores homólogos geralmente têm em sua região de terminal C seqüências de localização nuclear (NLS) e um domínio de ativação transcricional (AD). A região central contém unidades de repetição de geralmente 34 ou 35 aminoácidos. As unidades de repetição são quase idênticas, mas variáveis em certas posições e agora foi descoberto como estas posições determinam a especificidade de ligação à seqüência de nucleotídeos das proteínas.

[00054] Foi mostrado para AvrBs3 que as unidades de repetição são responsáveis pela ligação a DNA. A especificidade de ligação a DNA de AvrBs3 e provavelmente outros membros da família AvrBs3 parece ser mediada pelo domínio de repetição central das proteínas. Este domínio de repetição consiste em AvrBs3 de 17,5 unidades de repetição e em proteínas

homólogas é composto de 1,5 a 33,5 unidades de repetição que têm geralmente 34 aminoácidos cada. Outros comprimentos de unidade de repetição também são conhecidos (e.g. 30, 33, 35, 39, 40, 42 aminoácidos). A última repetição no domínio de repetição é geralmente apenas uma metade de repetição de 19 ou 20 aminoácidos de comprimento. As unidades de repetição individuais geralmente não são idênticas. Elas variam em certas posições de aminoácidos variáveis, entre estas posições 12 e 13 são hipervariáveis enquanto as posições 4, 11, 24, e 32 variam com alta frequência, mas em uma frequência inferior a 12 e 13 (variações em outras posições também podem ocorrer, mas com menos frequência). A comparação de proteínas semelhante a AvrBs3 diferentes a partir de *Xanthomonas* revela 80 a 97 % de identidade de sequência geral com a maioria das diferenças confinada ao domínio de repetição. Por exemplo, AvrBs3 e o membro da família semelhante a AvrBs3, AvrBs4, diferem exclusivamente em sua região de domínio de repetição, com a exceção de uma deleção de quatro aminoácidos no terminal C de AvrBs4 com relação a AvrBs3.

[00055] Na Figura 16, as seqüências de aminoácidos de AvrBs3 bem como as seqüências de aminoácidos dos membros da sub-família Hax são mostrado. De importância específica para a presente invenção são as unidades de repetição, que são idênticas exceto pelos aminoácidos hipervariáveis nas posições 12 e 13 e os aminoácidos variáveis nas posições 4 e 24. Portanto, cada unidade de repetição destas proteínas é fornecida separadamente.

[00056] Conforme determinado acima, já foi descrito que as unidades de repetição dentro dos domínios de repetição determinam reconhecimento ou capacidade de ligação e especificidade das proteínas efetoras do tipo III da família AvrBs3. No entanto, o princípio subjacente não era conhecido até a presente invenção.

[00057] Os inventores descobriram que uma unidade de repetição dentro de um domínio de repetição é responsável pelo reconhecimento de um

par de bases de DNA específico em uma sequência de DNA alvo. Esta descoberta é, no entanto, apenas um elemento da invenção. Os inventores, além disso, descobriram que uma região hipervariável dentro da cada unidade de um domínio de repetição é responsável pelo reconhecimento de um par de bases DNA específico em uma sequência de DNA alvo. Dentro de uma unidade de repetição, a região hipervariável (corresponde às posições de aminoácidos 12 e 13) são geralmente responsáveis por esta especificidade de reconhecimento. Portanto, cada variação nestes aminoácidos reflete uma variação correspondente em reconhecimento de DNA alvo e preferencialmente também capacidade de reconhecimento.

[00058] Conforme usado no presente documento, “região hipervariável” significa as posições 12 e 13 ou posição equivalente em uma unidade de repetição da presente invenção. É reconhecido que as posições 12 e 13 da invenção correspondem às posições 12 e 13 em unidades de repetição de comprimento completo de AvrBs3 e outros efetores TAL conforme divulgado no presente documento. Ainda é reconhecido que por “posições equivalentes” significa posições que correspondem às posições 12 e 13, respectivamente, em uma unidade de repetição da presente invenção. É possível determinar imediatamente tais posições equivalentes alinhando qualquer unidade de repetição com uma unidade de repetição de comprimento completo de AvrBs3.

[00059] Foi, portanto, mostrado pela primeira vez que uma unidade de repetição em um domínio de repetição de uma proteína de ligação a DNA reconhece um par de bases no DNA alvo, e que um aminoácido ou dois resíduos de aminoácidos adjacentes em uma unidade de repetição, geralmente dentro das regiões hipervariáveis de uma unidade de repetição, determinam qual par de bases no DNA alvo é reconhecido. Com base nesta descoberta, aqueles versados na técnica seriam capazes de direcionar especificamente pares de base em uma sequência de DNA alvo de interesse modificando um

polipeptídeo dentro de suas unidades de repetição do domínio de repetição para direcionar especificamente pares de base na sequência de DNA alvo desejada. Com base nesta descoberta, os inventores identificaram um código de reconhecimento para especificidades de DNA alvo de diferentes tipos de repetição e foram capazes de prever sequências de DNA alvo de diversos efetores TAL que poderiam ser confirmadas experimentalmente. Isto, além disso, facilitará a identificação de genes hospedeiros que são regulados por efetores TAL. O arranjo linear de unidades de repetição que reconhece uma sequência linear de bases no DNA alvo é uma interação de DNA-proteína nova. A arquitetura modular do domínio de repetição e o código de reconhecimento identificado pelos inventores para direcionar DNA com alta especificidade permite o projeto eficaz de domínios de ligação a DNA específico para uso em uma variedade de campos tecnológicos.

[00060] Em uma modalidade da presente invenção, os domínios de repetição são incluídos em um fator de transcrição, por exemplo, em fatores de transcrição ativos em plantas, especificamente preferenciais em proteínas efetoras do tipo III, e.g. em efetores da família semelhante a AvrBs3. No entanto, após ter descoberto da correlação entre as unidades de repetição em um domínio de repetição por um lado e a sequência de base no DNA alvo por outro lado, a arquitetura modular do domínio de repetição pode ser usada em qualquer proteína que será usada para direcionar sequências de DNA específicas. Introduzindo domínios de repetição compreendendo unidades de repetição em um polipeptídeo em que as unidades de repetição são modificadas para compreender uma região hipervariável por unidade de repetição e em que a região hipervariável determina o reconhecimento de um par de bases em uma sequência de DNA alvo, o reconhecimento de uma grande variedade de proteínas para sequências de DNA alvo pré-determinadas será disponível.

[00061] Como uma unidade de repetição dentro de um domínio de

repetição foi descoberto como sendo responsável pelo reconhecimento específico de um par de bases em um DNA, diversas unidades de repetição podem ser combinadas entre si sendo que cada unidade de repetição inclui uma região hipervariável que é responsável pelo reconhecimento de cada unidade de repetição a um par de bases específico em uma sequência de DNA alvo.

[00062] Técnicas para modificar especificamente seqüências de DNA para obter um códon específico para um aminoácido específico são conhecidas na técnica.

[00063] Métodos para mutagenese e alterações de polinucleotídeos foram amplamente descritos. Ver, por exemplo, Kunkel (1985) Proc. Natl. Acad. Sci. USA 82:488-492; Kunkel *et al.* (1987) Methods in Enzymol. 154:367-382; U.S. Pat. No. 4,873,192; Walker and Gaastra, eds. (1983) Techniques in Molecular Biology (MacMillan Publishing Company, New York) e as referências citadas no presente documento. Todas estas publicações são incorporadas no presente documento por referência.

[00064] Os seguintes exemplos fornecem métodos para construir novas unidades de repetição e testar as atividades de ligação específicas de unidades de repetição construídas artificialmente especificamente reconhecendo pares de base em uma sequência de DNA alvo.

[00065] O número de unidades de repetição a ser usado em um domínio de repetição pode ser determinado por aqueles versados na técnica por experimentação de rotina. Geralmente, pelo menos 1,5 unidades de repetição são consideradas como um mínimo, embora geralmente pelo menos 8 unidades de repetição serão usadas. As unidades de repetição não têm de ser unidades de repetição completas, visto que as unidades de repetição com metade do tamanho podem ser usadas. Além disso, os métodos e polipeptídeos divulgados no presente documento não dependem de domínios de repetição com um número específico de unidades de repetição. Assim, um

polipeptídeo da invenção pode compreender, por exemplo, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, 10, 10,5, 11, 11,5, 12, 12,5, 13, 13,5, 14, 14,5, 15, 15,5, 16, 16,5, 17, 17,5, 18, 18,5, 19, 19,5, 20, 20,5, 21, 21,5, 22, 22,5, 23, 23,5, 24, 24,5, 25, 25,5, 26, 26,5, 27, 27,5, 28, 28,5, 29, 29,5, 30, 30,5, 31, 31,5, 32, 32,5, 33, 33,5, 34, 34,5, 35, 35,5, 36, 36,5, 37, 37,5, 38, 38,5, 39, 39,5, 40, 40,5, 41, 41,5, 42, 42,5, 43, 43,5, 44, 44,5, 46, 46,5, 47, 47,5, 48, 48,5, 49, 49,5, 50, 50,5 ou mais unidades de repetição. Geralmente, AvrBs3 contém 17,5 unidades de repetição e induz expressão de genes *UPA* (regulados para cima por AvrBs3). O número e ordem de unidades de repetição determinarão a atividade correspondente e especificidade de reconhecimento de DNA. Conforme outros exemplos, os membros da família AvrBs3 Hax2 incluem 21,5 unidades de repetição, Hax3 11,5 unidades de repetição e Hax4 14,5 unidades de repetição. Preferencialmente, um polipeptídeo da invenção compreende cerca de 8 à cerca de 39 unidades de repetição. Mais preferencialmente, um polipeptídeo da invenção compreende cerca de 11,5 à cerca de 33,5 unidades de repetição.

[00066] Uma sequência consenso típica de uma repetição com 34 aminoácidos (em código de uma letra) é mostrada abaixo:

LTPEQVVAIASNGGGKQALETQRLLPVLCQAHG

[00067] Outra sequência consenso para uma unidade de repetição com 35 aminoácidos (em código de uma letra) é a seguinte:

LTPEQVVAIASNGGGKQALETQRLLPVLCQAPHD

[00068] As unidades de repetição que podem ser usadas em uma modalidade da invenção têm uma identidade com as sequências consenso descritas acima de pelo menos 35%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90% ou 95%. Em modalidades preferenciais, as sequências de repetição de AvrBs3, Hax2, Hax3 e Hax4 e outros membros da família AvrBs3 são usados. As sequências de unidade de repetição destes membros são indicadas na Fig. 16. Estas sequências de unidade de repetição podem ser modificadas

trocando um ou mais dos aminoácidos. As seqüências de unidade de repetição modificadas têm uma identidade com a seqüência de repetição original do membro original da seqüência da família AvrBs3 de pelo menos 35%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90% ou 95%. Em modalidades preferenciais, os aminoácidos nas posições 12 e 13 são alterados. Ainda em outras modalidades, os aminoácidos nas posições 4, 11, 24 e 32 são alterados. Preferencialmente, o número de aminoácidos por repetição está em uma faixa de 20 a 45 aminoácidos, ou 32 a 40 aminoácidos, ainda 32 a 39 aminoácidos, e ainda opcionalmente 32, 34, 35 ou 39 aminoácidos por unidade de repetição.

[00069] Especificamente, a região hipervariável em uma unidade de repetição determina o reconhecimento específico de um par de bases em uma seqüência de DNA alvo. Mais especificamente, os inventores descobriram a seguinte correlação de especificidade de reconhecimento entre aminoácidos encontrados nas posições 12 e 13 em uma unidade de repetição e pares de base na seqüência de DNA alvo:

- HD para reconhecimento de C/G
- NI para reconhecimento de A/T
- NG para reconhecimento de T/A
- NS para reconhecimento de C/G ou A/T ou T/A ou G/C
- NN para reconhecimento de G/C ou A/T
- IG para reconhecimento de T/A
- N para reconhecimento de C/G ou T/A
- HG para reconhecimento de T/A
- H para reconhecimento de T/A
- NK para reconhecimento de G/C.

[00070] Deve ser observado que os aminoácidos são representados no código de letra única. Os nucleotídeos são fornecidos como pares de base, em que a primeira base é localizada no filamento superior e a segunda base é

localizada no filamento inferior; por exemplo, C/G significa que C está localizado no filamento superior, G no filamento inferior.

[00071] Com relação aos aminoácidos únicos N e H, respectivamente, o aminoácido 13 de AvrBs3 parece estar faltando na unidade de repetição quando comparado por múltiplos alinhamentos de sequência de aminoácidos com as outras unidades de repetição.

[00072] Em uma modalidade da invenção, o domínio de terminal N de proteínas semelhantes a AvrBs3 confere especificidade de reconhecimento para um T, 5' da especificidade de reconhecimento de tal repetição.

[00073] Em uma modalidade especificamente preferencial da invenção, as unidades de repetição da família de proteínas AvrBs3 são usadas. Exemplos para os membros desta família de proteínas foram especificados acima. Especificamente, os membros da família de proteínas têm uma homologia de aminoácidos de pelo menos 95%, pelo menos 90%, pelo menos 80%, pelo menos 85%, pelo menos 70%, pelo menos 75%, pelo menos 60%, pelo menos 50%, pelo menos 40% ou pelo menos 35% à sequência de aminoácidos de AvrBs3, especificamente à sequência de aminoácidos da unidade de repetição de AvrBs3. Tendo isto em mente, a região hipervariável em uma unidade de repetição pode ser deduzida por uma comparação de aminoácido entre os membros da família AvrBs3. Em modalidades especificamente preferenciais, os aminoácidos estão nas posições 12 e 13 de uma unidade de repetição de AvrBs3. No entanto, as regiões variáveis também podem ser localizadas em diferentes posições de aminoácidos. Exemplos para posições variáveis são números de aminoácidos 4, 11, 24 e 32. Ainda em outra modalidade da invenção, os aminoácidos responsáveis pelo reconhecimento específico de um par de bases em uma sequência de DNA são localizados em posições que geralmente não variam entre os membros da família AvrBs3 ou em posições que são variáveis mas não hipervariáveis.

[00074] Para resumir, os inventores descobriram que as unidades de

repetição determinam o reconhecimento de um par de bases em uma sequência de DNA e que a região hipervariável dentro de uma unidade de repetição determina a especificidade de reconhecimento da unidade de repetição correspondente. Portanto, a sequência de unidades de repetição correlaciona com uma ordem linear específica de pares de base em uma sequência de DNA alvo. Os inventores descobriram esta correlação com relação a AvrBs3 e a verificaram com relação a um número de membros representante da família semelhante a AvrBs3 de proteínas. Com relação aos membros da família semelhante a AvrBs3, os resíduos de aminoácidos nas posições 12 e 13 em uma unidade de repetição de 34 ou outro comprimento de aminoácidos correlacionam com especificidades de ligação definidas de proteínas semelhantes a AvrBs3. A descoberta deste princípio central fornece uma poderosa ferramenta para customizar um polipeptídeo com seu modelo de DNA alvo cognato para uma variedade de aplicações incluindo, entre outras, modulação de expressão de gene e engenharia de genoma direcionada.

[00075] Na presente invenção, os polipeptídeos podem ser designados compreendendo um domínio de repetição com unidades de repetição em que nas unidades de repetição são incluídas regiões hipervariáveis que determinam o reconhecimento de um par de bases em uma sequência de DNA alvo. Em uma modalidade da invenção, cada unidade de repetição inclui uma região hipervariável que determina o reconhecimento de um par de bases em uma sequência de DNA alvo. Em outra modalidade, 1 ou 2 unidades de repetição em um domínio de repetição são incluídas não reconhecendo especificamente um par de bases em uma sequência de DNA alvo. Considerando o código de reconhecimento encontrado pelos inventores, um arranjo modular de unidades repetidas é viável em que cada unidade de repetição é responsável pelo reconhecimento específico de um par de bases em uma sequência de DNA alvo. Conseqüentemente, uma sequência de unidades de repetição corresponde a uma sequência de pares de base em uma

seqüência de DNA alvo de modo que 1 unidade de repetição corresponde a um par de bases.

[00076] Desde que uma seqüência de DNA alvo seja conhecida e a qual reconhecimento por uma proteína seja desejado, aqueles versados na técnica serão capazes de construir especificamente uma série modular de unidades de repetição, incluindo seqüências de aminoácidos de reconhecimento específico, e montar estas unidades de repetição em um polipeptídeo na ordem apropriada para permitir o reconhecimento de e ligação à seqüência de DNA alvo desejada. Qualquer polipeptídeo pode ser modificado sendo combinado com um domínio de ligação a DNA de unidade de repetição modular da presente invenção. Tais exemplos incluem polipeptídeos que são proteínas repressoras e ativadoras de transcrição, proteínas mediadoras de resistência, nucleases, topoisomerasas, ligases, integrases, recombinases, resolvases, metilases, acetilases, demetilases, deacetilases e qualquer outro polipeptídeo capaz de modificar DNA, RNA ou proteínas.

[00077] O domínio de ligação a DNA da unidade de repetição modular da presente invenção pode ser combinado com sinais de localização de compartimento celular como sinais de localização nuclear, para funcionar em quaisquer outras regiões regulatórias, incluindo, entre outras, regiões regulatórias transcricionais e regiões de terminação translacional.

[00078] Em outra modalidade da invenção, estas unidades de repetição designadas modularmente são combinadas com um domínio de endonuclease capaz de clivar DNA quando colocado próximo a DNA em consequência de ligação pelo domínio de repetição. Tais quebras endonucleolíticas são conhecidas para estimular a taxa de recombinação homóloga em eucariotos, incluindo fungos, plantas e animais. A capacidade de estimular recombinação homóloga em um local específico em consequência de uma quebra endonucleolítica de local específico permite a recuperação de células

transformadas que integraram uma sequência de DNA de interesse no local específico, em uma frequência muito maior do que é possível sem ter feito a quebra de local específico. Além disso, quebras endonucleolíticas como aquelas causadas por polipeptídeos formados a partir de um domínio de repetição e um domínio de endonuclease são as vezes reparadas pelo maquinário metabólico de DNA celular de uma forma que altera a sequência no local da quebra, por exemplo, causando uma inserção curta no local da quebra em comparação à sequência inalterada. Estas alterações de sequência podem causar inativação da função de um gene ou proteína, por exemplo, alterando uma sequência de codificação de proteína para fazer uma proteína não funcional, modificando um local de splica de modo que a transcrição do gene não seja apropriadamente clivada, fazendo uma transcrição não-funcional, mudando a sequência promotora de um gene de forma que esta não possa mais ser transcrita apropriadamente etc.

[00079] Quebrar DNA usando endonucleases específicas pode aumentar a taxa de recombinação homóloga na região da quebra. Em algumas modalidades, a endonuclease FokI (*Flavobacterium okeanoikoites*) pode ser utilizada em um efector para induzir quebras de DNA. O domínio de endonuclease FokI funciona independentemente do domínio de ligação DNA e corta um DNA de filamento duplo geralmente como um dímero (Li *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A* 89 (10):4275-4279, and Kim *et al.* (1996) *Proc. Natl. Acad. Sci. U.S.A* 93 (3):1156-1160; cuja divulgações são incorporadas no presente documento por referência integralmente). Um dímero FokI de cadeia única também foi desenvolvido e poderia também ser utilizado (Mino *et al.* (2009) *J. Biotechnol.* 140:156-161). Um efector poderia ser construído contendo um domínio de repetição para reconhecimento de uma sequência de DNA alvo desejada bem como um domínio de endonuclease FokI para induzir quebra de DNA em ou próximo à sequência de DNA alvo semelhante ao trabalho anterior feito empregando nucleases de

dedo de zinco (Townsend *et al.* (2009) *Nature* 459:442-445; Shukla *et al.* (2009) *Nature* 459, 437-441, todos sendo incorporados no presente documento por referência integralmente). A utilização de tais efetores poderia permitir a geração de mudanças direcionadas em genomas que incluem adições, deleções e outras modificações, análogos àqueles usos reportados para nucleases de dedo de zinco conforme Bibikova *et al.* (2003) *Science* 300, 764; Urnov *et al.* (2005) *Nature* 435, 646; Wright *et al.* (2005) *The Plant Journal* 44:693-705; e Pat. N°s U.S. 7.163.824 e 7.001.768, que são incorporadas no presente documento por referência integralmente.

[00080] O domínio de endonuclease *FokI* pode ser clonado por PCR a partir do DNA genômico das bactérias marinhas *Flavobacterium okeanokoites* (ATCC) preparadas por métodos padrão. A sequência da endonuclease *FokI* é disponível em Pubmed (No. de acesso M28828 e No. de acesso J04623, cujas divulgações são incorporadas no presente documento por referência integralmente).

[00081] Endonuclease I-Sce I a partir da levedura *Saccharomyces cerevisiae* foi usada para produzir quebras de DNA que aumentam a taxa de recombinação homóloga. I-Sce I é uma endonuclease codificada por um íntron mitocondrial que tem uma sequência de reconhecimento de 18 bp, e portanto uma frequência muito baixa de locais de reconhecimento dentro de um determinado DNA, mesmo dentro de genomas grandes (Thierry *et al.* (1991) *Nucleic Acids Res.* 19 (1):189-190; cuja divulgação é incorporada no presente documento por referência integralmente). A infrequência de locais de clivagem reconhecidos por I-SceI a torna adequada para uso para potencializar recombinação homóloga. A descrição adicional a respeito do uso de I-Sce I para induzir tais quebras de DNA pode ser encontrada no Pedido de Pat. U.S. 20090305402, que é incorporado no presente documento por referência integralmente.

[00082] O local de reconhecimento para I-Sce I foi introduzido em

uma variedade de sistemas diferentes. O corte subsequente deste local com I-Sce I aumenta a recombinação homóloga na posição onde o local foi introduzido. O aumento de frequência de recombinação homóloga foi obtido com locais de I-Sce I introduzidos no DNA extra-cromossomal em *Xenopus* oocytes, genoma de camundongo, e DNA genômico da planta de tabaco *Nicotiana plumbaginifolia*. Ver, por exemplo, Segal *et al.* (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92 (3):806-810; Chouluka *et al.* (1995) *Mol. Cell Biol.* 15 (4):1968-1973; and Puchta *et al.* (1993) *Nucleic Acids Res.* 21 (22):5034-5040; cujas divulgações são incorporadas neste documento por referência integralmente. Será apreciado que qualquer outro domínio de endonuclease que funciona com domínios de ligação a DNA heterólogo pode ser utilizado em um efector diferente e que a endonuclease I-Sce I é um exemplo não-limitante. A limitação do uso de endonucleases que têm um reconhecimento de DNA e domínio de ligação como I-Sce I é que o local de reconhecimento tem de ser introduzido por métodos padrão de recombinação homóloga no local desejado antes do uso de tal endonuclease para aumentar recombinação homóloga naquele local, se tal local não já estiver presente na localização desejada. Foram reportados métodos que permitem o projeto e síntese de endonucleases novas, como, modificando endonucleases conhecidas ou fazendo versões quiméricas de uma ou mais tais endonucleases, que reconhecem seqüências de DNA alvo novas, abrindo caminho para a geração de tais domínios de endonuclease engenheirada para clivar seqüências de DNA alvo endógenas de interesse (Chevalier *et al.* (2002) *Molecular Cell* 10:895-905; WO2007/060495; WO2009/095793; Fajardo-Sanchez *et al.* (2008) *Nucleic Acids Res.* 36:2163-2173, ambas sendo incorporadas por referência integralmente). Como tal, poderia ser previsto que tais domínios de endonuclease poderiam ser semelhantemente engenheirados de modo a gerar uma atividade de ligação a DNA não funcional mas deixando a função de clivagem de DNA ativa e utilizar tal domínio de clivagem de endonuclease

semelhantemente engenheirado em um efetor para induzir quebras de DNA semelhantes ao uso de FokI acima. Em tais aplicações, o reconhecimento de sequência de DNA alvo seria preferencialmente fornecido pelo domínio de repetição do efetor, mas a clivagem de DNA seria realizada pelo domínio de endonuclease engenheirado.

[00083] Como mencionado acima, um efetor inclui um domínio de repetição com reconhecimento específico para uma sequência alvo específica desejada. Em modalidades preferenciais, o efetor especificamente se liga a uma sequência de DNA cromossomal endógena. A sequência de ácido nucléico específica ou mais preferencialmente a sequência cromossomal endógena específica pode ser qualquer sequência em uma região de ácido nucléico onde é desejado aumentar recombinação homóloga. Por exemplo, a região de ácido nucléico pode ser uma região que contém um gene em que é desejado introduzir uma mutação, como uma mutação ou deleção de ponto, ou uma região em que é desejado introduzir um gene conferindo um fenótipo desejado.

[00084] Outras modalidades se referem a métodos para gerar uma planta modificada em que uma adição desejada foi introduzida. Os métodos podem incluir obter uma célula de planta que incluir uma sequência de DNA alvo endógena em que é desejado introduzir uma modificação; gerando um corte de filamento duplo dentro da sequência de DNA alvo endógena com um efetor que inclui um domínio de repetição que se liga a uma sequência de DNA alvo endógena e um domínio de endonuclease; introduzindo um ácido nucléico exógeno que inclui uma sequência homóloga a pelo menos uma porção do DNA alvo endógeno na célula de planta sob condições que permitem que recombinação homóloga ocorra entre o ácido nucléico exógeno e a sequência de DNA alvo endógena; e gerando uma planta a partir da célula de planta em que a recombinação homóloga ocorreu. Outras modalidades se referem a células geneticamente modificadas e plantas feitas de acordo com o

método descrito acima e no presente documento. Deve ser observado que a sequência de DNA alvo poderia ser de ocorrência natural ou artificial. Será apreciado que tais métodos poderiam ser usados em qualquer organismo (tais organismos não-limitantes para incluir animais, humanos, fungos, bactérias omócitas e vírus) usando técnicas e métodos conhecidos na técnica e utilizados para tais fins em tais organismos.

[00085] Em outra modalidade da invenção, estes domínios de repetição designados modularmente são combinados com um ou mais domínios responsáveis pela modulação ou controle da expressão de um gene, por exemplo, de genes de plantas, genes de animais, genes de fungos, genes omócitos, genes virais ou genes de humanos. Os métodos para modular expressão de gene através da geração de polipeptídeos de ligação a DNA contendo domínios de dedo de zinco são conhecidos na técnica (Pat. Nos. U.S. 7.285.416, 7.521.241, 7.361.635, 7.273.923, 7.262.054, 7.220.719, 7.070.934, 7.013.219, 6.979.539, 6.933.113, 6.824.978, cada uma sendo incorporada no presente documento por referência integralmente). Por exemplo, estes efetores da família semelhante a AvrBs3 são modificados para se ligar a seqüências de DNA alvo específicas. Tais polipeptídeos podem, por exemplo, ser ativadores de transcrição ou proteínas repressoras de transcrição que são modificadas pelo método da presente invenção para se ligar especificamente a regiões de controle genético em um promotor de ou outra região regulatória para um gene de interesse para ativar, reprimir ou de forma diversa modular transcrição de tal gene.

[00086] Ainda em outra modalidade da invenção, as seqüências de DNA alvo são modificadas para ser especificamente reconhecidas por um domínio de repetição de ocorrência natural ou por um domínio de repetição modificado. Como um exemplo, as seqüências de DNA alvo para membros da família semelhante a AvrBs3 podem ser inseridas nos promotores para gerar promotores controláveis novos que podem ser induzidos pelo efector

AvrBs3 correspondente. Sistemas induzíveis secundários podem ser construídos usando um trans-ativador e um gene alvo, em que o trans-ativador é um polipeptídeo em que tal polipeptídeo compreende pelo menos um domínio de repetição que compreende unidades de repetição da presente invenção que se ligam ao referido gene alvo e induzem expressão. O transativador e o gene alvo podem ser introduzidos em uma linhagem celular, mas podem também estar presentes em diferentes linhagens celulares e ser introgressados posteriormente. Em outra modalidade, plantas resistentes à doença podem ser construídas inserindo a seqüência de DNA alvo de um domínio de repetição contendo polipeptídeo da presente invenção em frente a um gene que após expressão leva a uma reação de defesa da planta ativando um gene mediador de resistência.

[00087] Em outra modalidade, polipeptídeos de ligação a DNA personalizados podem ser construídos rearranjando tipos de unidade de repetição permitindo assim a geração de domínios de repetição da especificidade de ligação a DNA alvo nova. Unidades de repetição individuais são quase idênticas no nível de DNA que impede estratégias clássicas de clonagem. A presente invenção fornece uma estratégia rápida e barata para montar polipeptídeos personalizados com domínios de repetição da presente invenção. Para melhorar a versatilidade de clonagem de tais polipeptídeos, um método de montagem de duas etapas foi projetado. Este método foi usado para montar polipeptídeos com tipos de repetição novos para estudar seu reconhecimento de DNA alvo e especificidade de ligação.

[00088] Resumidamente, qualquer seqüência de DNA pode ser modificada para permitir ligação por um domínio de repetição contendo polipeptídeo da presente invenção introduzindo pares de base em qualquer região de DNA ou regiões específicas de um gene ou um elemento de controle genético para especificamente direcionar um polipeptídeo tendo um domínio de repetição composto de unidades de repetição que se ligará a tal

seqüência de DNA modificada para facilitar reconhecimento específico e ligação uma à outra.

[00089] Os inventores demonstraram que um polipeptídeo verdadeiramente reconhecedor de DNA e preferencialmente de ligação pode ser produzido de modo eficaz, em que o motivo de ligação de tal polipeptídeo é um domínio de repetição composto de unidades de repetição que são selecionadas com base em sua capacidade de reconhecimento de uma combinação de pares de base específicos. Consequentemente deve estar dentro da capacidade daqueles versados na técnica projetar um polipeptídeo capaz de se ligar a qualquer seqüência de DNA alvo desejada simplesmente considerando a seqüência de pares de base presente no DNA alvo e combinando nas unidades de repetição apropriadas como motivos de ligação tendo as características necessárias para ligação. Quanto maior o comprimento da seqüência conhecida do DNA alvo, maior o número de unidades de repetição modulares que podem ser incluídas no polipeptídeo. Por exemplo, se a seqüência conhecida só tiver 9 bases, então as nove unidades de repetição conforme definido acima poderão ser incluídas no polipeptídeo. Se a seqüência conhecida tiver 27 bases, até 27 unidades de repetição poderão ser incluídas no polipeptídeo. Quanto maior a seqüência de DNA, menor a probabilidade de sua ocorrência em qualquer outra determinada porção de DNA em outro local no genoma.

[00090] Além disso, aquelas unidades de repetição selecionadas para inclusão no polipeptídeo poderiam ser modificadas artificialmente para modificar suas características de ligação. Como alternativa (ou além disso) o comprimento e seqüência de aminoácidos da unidade de repetição poderia ser variado desde que suas características de ligação não fossem afetadas.

[00091] Geralmente, será preferível selecionar aquelas unidades de repetição tendo alta afinidade e alta especificidade para a seqüência de DNA alvo.

[00092] Conforme descrito no presente documento, os efetores podem ser designados para reconhecer qualquer local alvo adequado, para regulação de expressão de qualquer gene endógeno de escolha. Exemplos de genes endógenos adequados para regulação incluem VEGF, CCR5, ER.alfa., Her2/Neu, Tat, Rev, HBV C, S, X e P, LDL-R, PEPCK, CYP7, Fibrinogen, ApoB, Apo E, Apo(a), renin, NF-.kappa.B, I-.kappa.B, TNF-.alfa., ligante FAS, proteína precursora amilóide, fator naturético atrial, ob-leptin, ucp-1, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-12, G-CSF, GM-CSF, Epo, PDGF, PAF, p53, Rb, hemoglobina fetal, distrofina, eutrofina, GDNF, NGF, IGF-1, VEGF receptores flt e flk, topoisomerase, telomerase, bcl-2, ciclinas, angiostatina, IGF, ICAM-1, STATS, c-myc, c-myb, TH, PTI-1, poligalacturonase, EPSP sintase, FAD2-1, delta-12 desaturase, delta-9 desaturase, delta-15 desaturase, acetil-CoA carboxilase, acil-ACP-tioesterase, ADP-glicose pirofosforilase, amido sintase, celulose sintase, sucrose sintase, genes associados a senescência, quelatores de metal pesado, liase de hidroperóxido de ácido graxo, genes virais, genes protozoais, genes fúngicos e genes bacterianos. No geral, genes adequados para serem regulados incluem citosinas, linfocinas, fatores de crescimento, fatores mitogênicos, fatores quimiotáticos, fatores onco ativos, receptores, canais de potássio, proteínas G, moléculas de transdução de sinal, genes de resistência à doença e outros genes relacionados a doenças.

[00093] Em outro aspecto, um método para modular a expressão de um gene alvo em uma célula é fornecido. A célula pode ser preferencialmente uma célula de planta, uma célula humana, célula de animal, célula de fungo ou qualquer outra célula viva. As células contêm um polipeptídeo em que tal polipeptídeo compreende pelo menos um domínio de repetição compreendendo unidades de repetição, e estas unidades de repetição contêm uma região hipervariável e cada unidade de repetição é responsável pelo reconhecimento de 1 par de bases em tal seqüência de DNA alvo. Tal

polipeptídeo é introduzido como DNA que codifica o polipeptídeo ou o polipeptídeo é introduzido per se na célula por métodos conhecidos na técnica. Independente de como é introduzido, o polipeptídeo deve incluir pelo menos um domínio de repetição que reconhece especificamente e preferencialmente se liga a uma seqüência de DNA alvo de pares de base e modula a expressão de um gene alvo. Em uma modalidade preferencial, todas as unidades de repetição contêm uma região hipervariável que determina o reconhecimento de pares de base em uma seqüência de DNA alvo.

[00094] Exemplos seqüências de peptídeos que podem ser ligados a um efector da presente invenção, para facilitar absorção de efetores em células, incluem, mas não são limitados a: um peptídeo de aminoácido 11 da proteína tat de HIV; uma seqüência de peptídeo do resíduo 20 que corresponde aos aminoácidos 84 a 103 da proteína p16 (ver Fahraeus *et al.* (1996) *Current Biology* 6:84); a terceira hélice do homeodomínio de 60 aminoácidos de comprimento de Antennapedia (Derossi *et al.* (1994) *J. Biol. Chem.* 269:10444); a região h de um peptídeo de sinal como região h fator de crescimento de fibroblasto Kaposi (K-FGF); ou o domínio de translocação VP22 de HSV (Elliot & O'Hare (1997) *Cell* 88:223 233). Outras porções químicas adequadas que fornecem absorção celular realçada também podem ser quimicamente ligadas a efetores.

[00095] Moléculas de toxina também têm a capacidade de transportar polipeptídeos pelas membranas celulares. Com freqüência, tais moléculas são compostas de pelo menos duas partes (chamadas “toxinas binárias”): um domínio ou polipeptídeo de translocação ou ligação e um domínio ou polipeptídeo de toxina separada. Geralmente, o domínio ou polipeptídeo de translocação se liga a um receptor celular, e em seguida a toxina é transportada para a célula. Diversas toxinas bacterianas, incluindo toxina *Clostridium perfringens* iota, toxina de difteria (DT), exotoxina *Pseudomonas* A (PE), toxina de pertussis (PT), toxina *Bacillus anthracis* e ciclase adenilato

pertussis (CYA), foram usadas em tentativas de distribuir peptídeos para o citosol celular como fusões internas ou de amino-terminal (Arora *et al.* (1993) *J. Biol. Chem.* 268:3334 3341; Perelle *et al.* (1993) *Infect. Immun.* 61:5147 5156 (1993); Stenmark *et al.* (1991) *J. Cell Biol.* 113:1025 1032 (1991); Donnelly *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:3530 3534; Carbonetti *et al.* (1995) *Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295; Sebo *et al.* (1995) *Infect. Immun.* 63:3851 3857; Klimpel *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:10277 10281; and Novak *et al.* (1992) *J. Biol. Chem.* 267:17186 17193).

[00096] Os efetores também podem ser introduzidos em uma célula de animal, preferencialmente uma célula de mamífero, através de lipossomos e derivados de lipossomos como imunolipossomos. O termo “lipossomo” se refere a vesículas compostas de uma ou mais bicamadas de lipídeos ordenadas concentricamente, que encapsulam uma fase aquosa. A fase aquosa geralmente contém o composto a ser distribuído para a célula, neste caso um efector. O lipossomo é fundido com a membrana plasmática, liberando assim o efector no citosol. Como alternativa, o lipossomo é fagocitado ou absorvido pela célula em uma vesícula de transporte. Uma vez no endossomo ou fagossomo, o lipossomo degrada ou se funde com a membrana da vesícula de transporte e libera seus conteúdos.

[00097] A invenção especificamente se refere ao campo de tecnologia de plantas ou agricultura. Em um aspecto, a presente invenção é direcionada a um método para modular a expressão de um gene alvo em células de plantas, método o qual compreende fornecer células de plantas com um polipeptídeo modificado de acordo com a invenção, tal polipeptídeo sendo capaz de reconhecer especificamente uma sequência de nucleotídeo alvo, ou um filamento complementar da mesma, dentro de um gene alvo, e permitir que tal polipeptídeo reconheça e especificamente se ligue a tal sequência de nucleotídeos alvo, através da qual a expressão de tal gene alvo em células de

plantas é modulada.

[00098] O polipeptídeo pode ser fornecido a células de plantas através de quaisquer métodos adequados conhecidos na técnica. Por exemplo, a proteína pode ser adicionada de forma exógena a células de plantas e as células de plantas são mantidas sob condições de modo que o polipeptídeo seja introduzido na célula de planta, se ligue à seqüência de nucleotídeos alvo e regule a expressão do gene alvo nas células de plantas. Como alternativa, uma seqüência de nucleotídeos, ex., DNA ou RNA, que codifica o polipeptídeo pode ser expressa nas células de plantas e as células de plantas são mantidas sob condições de modo que o polipeptídeo expresso se ligue à seqüência de nucleotídeos alvo e regule a expressão do gene alvo nas células de plantas.

[00099] Um método preferencial para modular a expressão de um gene alvo em células de plantas compreende as seguintes etapas: a) fornecer células de plantas com um sistema de expressão para um polipeptídeo modificado de acordo com a invenção, tal polipeptídeo sendo capaz de reconhecer especificamente, e preferencialmente se ligar a uma seqüência de nucleotídeos alvo, ou um filamento complementar da mesma, dentro de um elemento de controle de expressão de um gene alvo, preferencialmente um promotor; e b) cultivar tais células de plantas sob condições em que tal polipeptídeo é produzido e se liga à tal seqüência de nucleotídeos alvo, através da qual a expressão de tal gene alvo em tais células de plantas é introduzida.

[000100] Qualquer seqüência de nucleotídeos alvo pode ser modulada pelo presente método. Por exemplo, a seqüência de nucleotídeos alvo pode ser endógena ou exógena ao gene alvo. Em uma modalidade da invenção, a seqüência de nucleotídeos alvo pode estar presente em uma célula viva ou presente *in vitro*. Em uma modalidade específica, a seqüência de nucleotídeos alvo é endógena à planta. A seqüência de nucleotídeos alvo pode ser localizada em qualquer local adequado em relação ao gene alvo. Por exemplo,

a seqüência de nucleotídeos alvo pode ser a montante e a jusante da região de codificação do gene alvo. Como alternativa, a seqüência de nucleotídeos alvo fica dentro da região de codificação do gene alvo. Preferencialmente, a seqüência de nucleotídeos alvo é uma promotora de um gene.

[000101] Qualquer gene alvo pode ser modulado pelo presente método. Por exemplo, o gene alvo pode codificar um produto que afeta biosíntese, tráfego celular, metabolismo e degradação de um peptídeo, uma proteína, um oligonucleotídeo, um ácido nucléico, uma vitamina, um oligossacarídeo, um carboidrato, um lipídeo, ou uma molécula pequena. Além disso, os efetores podem ser usados para engenheirar plantas para traços como aumento de resistência a doenças, modificação de polissacarídeos estruturais e de armazenagem, saborizantes, proteínas, ácidos graxos, amadurecimento de frutas, rendimento, cor, características nutricionais, melhora de capacidade de armazenagem e semelhantes.

[000102] Portanto, a invenção fornece um método para alterar a expressão de um gene de interesse em uma célula alvo, compreendendo: determinar (se necessário) pelo menos parte da seqüência de DNA da região estrutural e/ou uma região regulatória do gene de interesse; designer um polipeptídeo incluindo as unidades de repetição modificadas em conformidade com a invenção para reconhecer pares de base específicos no DNA de seqüência conhecida, e fazer com que tal polipeptídeo modificado esteja presente na célula alvo (de preferência em seu núcleo). (Ficará evidente que a seqüência de DNA não precisa ser determinada se já for conhecida).

[000103] A região regulatória poder ser bem remota a partir da região estrutural do gene de interesse (e.g. uma seqüência realçadora distante ou semelhante).

[000104] Além disso, o polipeptídeo pode vantajosamente compreender domínios funcionais a partir de outras proteínas (e.g. domínios catalíticos de endonucleases de restrição, recombinases, replicases, integrases e

semelhantes) ou até domínios efetores “sintéticos”. O polipeptídeo também pode compreender sinais de ativação ou processamento, como sinais de localização nuclear. Estes são de inutilidade específica em direcionar o polipeptídeo ao núcleo da célula para aumentar a ligação do polipeptídeo a um alvo intranuclear (como DNA genômico).

[000105] O polipeptídeo modificado pode ser sintetizado *in situ* na célula como resultado de distribuição à célula de DNA que direciona expressão do polipeptídeo. Os métodos para facilitar a distribuição de DNA são bem conhecidos para aqueles versados na técnica e incluem, por exemplo, vetores virais recombinantes (e.g. retrovírus, adenovírus), lipossomos e semelhantes. Como alternativa, o polipeptídeo modificado poderia ser feito fora da célula e em seguida distribuído à mesma. A distribuição poderia ser facilitada incorporando o polipeptídeo em lipossomos etc. ou ligando o polipeptídeo a uma porção de direcionamento (como a porção de ligação de uma molécula de anticorpo ou hormônio, ou um domínio de transcrição de membrana ou o domínio de translocação de um efector fúngico ou omócito, ou o domínio B de ligação celular da família A-B clássica de toxinas bacterianas). De fato, uma vantagem significativa das proteínas modificadas da invenção em controlar expressão de genes seria a distribuição sem vetor de proteína para células alvo.

[000106] Para o melhor conhecimento dos inventores, o projeto de um polipeptídeo contendo unidades de repetição modificadas capazes de reconhecer especificamente pares de base em uma sequência de DNA alvo e seu uso bem sucedido em modulação de expressão de genes (conforme descrito no presente documento) nunca foi demonstrado anteriormente. Assim, a importante descoberta da presente invenção conforme divulgado no presente documento apresenta diversas possibilidades que se estendem além de usos em plantas. Em uma modalidade da invenção, polipeptídeos efetores são designados para uso terapêutico e/ou profilático na regulação da

expressão de genes relacionados a doenças. Por exemplo, tais polipeptídeos poderiam ser usados para inibir a expressão de genes exógenos (ex., os genes de patógenos bacterianos ou virais) em humanos, outros animais, ou plantas, ou para modificar a expressão de genes hospedeiros mutados (como oncogenes).

[000107] A invenção, portanto, também fornece um polipeptídeo efetor capaz de inibir a expressão de um gene relacionado à doença. Geralmente o polipeptídeo não será um polipeptídeo de ocorrência natural, mas será designado especificamente para inibir a expressão do gene relacionado à doença. Convenientemente, o polipeptídeo efetor será designado por qualquer dos métodos da invenção.

[000108] A invenção também se refere ao campo de engenharia de genomas. Um polipeptídeo efetor pode ser gerado de acordo com a invenção para direcionar uma sequência de DNA específica em um genoma. Tal polipeptídeo pode ser modificado para conter uma atividade que direciona modificação da sequência de DNA alvo (e.g. recombinação de local específico ou integração de sequências alvo). Este método permite modificações de DNA direcionadas em genomas complexos.

[000109] Ainda em outra modalidade da invenção, é fornecido um polipeptídeo sendo modificado para incluir pelo menos um domínio de repetição que compreende unidades de repetição, as unidades de repetição tendo região hipervariável para determinar reconhecimento seletivo de um par de bases em uma sequência de DNA.

[000110] Em uma modalidade preferencial, o polipeptídeo inclui dentro de tal unidade de repetição uma região hiper-variável que é selecionada a partir do seguinte grupo a fim de determinar o reconhecimento de um ou mais dos seguintes pares de base:

- HD para o reconhecimento de C/G
- NI para o reconhecimento de A/T

- NG para o reconhecimento de T/A
- NS para o reconhecimento de C/G ou A/T ou T/A ou G/C
- NN para o reconhecimento de G/C ou A/T
- IG para o reconhecimento de T/A
- N para o reconhecimento de C/G ou T/A
- HG para o reconhecimento de T/A
- H para o reconhecimento de T/A
- NK para o reconhecimento de G/C.

[000111] A invenção também compreende DNA que codifica para qualquer um dos polipeptídeos descritos acima.

[000112] Ainda em outra modalidade, é fornecido DNA que é modificado para incluir um par de bases localizado em uma sequência de DNA alvo de forma que tal par de bases pode ser especificamente reconhecido por um polipeptídeo que inclui ao menos um domínio de repetição que compreende unidades de repetição, as unidades de repetição tendo uma região hiper-variável que determina o reconhecimento de tal par de bases em tal DNA. Em uma modalidade opcional, tal par de bases está localizada em uma sequência de controle de expressão de gene. Devido a montagem modular do domínio de repetição, uma sequência de pares de base pode ser especificamente um alvo por tal domínio de repetição.

[000113] Em uma modalidade alternativa da invenção, tal DNA é modificado por um par de bases selecionado dentre o seguinte grupo a fim de receber um reconhecimento seletivo e determinado por uma das seguintes regiões hiper variáveis:

- C/G para reconhecimento por HD
- A/T para reconhecimento por NI
- T/A para reconhecimento por NG
- CT ou A/T ou T/A ou G/C para reconhecimento por NS
- G/C ou A/T para reconhecimento por NN

- T/A para reconhecimento por IG.
- C/G ou T/A para reconhecimento por N
- T/A para reconhecimento por HG
- T/A para reconhecimento por H
- G/C para reconhecimento por NK.

[000114] Ainda em outro aspecto a invenção fornece um método para modificar uma sequência de ácido nucleico de interesse presente em uma mistura de amostra pela ligação com um polipeptídeo de acordo com a invenção, que compreende o contato da mistura de amostra com tal polipeptídeo com afinidade para ao menos uma porção da sequência de interesse, de forma a permitir que o polipeptídeo reconheça e preferencialmente se ligue especificamente na sequência de interesse.

[000115] O termo “modificando” como usado no presente documento é designado para significar que a sequência é considerada modificada simplesmente pela ligação do polipeptídeo. Não se pretende sugerir que a sequência de nucleotídeos é alterada, apesar de que tais mudanças (e outras) poderiam surgir após a ligação do polipeptídeo com o ácido nucleico de interesse. Convenientemente a sequência de ácido nucleico é DNA.

[000116] A modificação do ácido nucleico de interesse (no sentido de ligação deste com um polipeptídeo modificado para conter unidades de repetição modulares) poderia ser detectada em qualquer um de uma série de métodos (por exemplo, ensaios de mudança de mobilidade gel, uso de polipeptídeos marcados – marcadores poderiam incluir marcadores radioativos, fluorescentes, enzima ou biotina/estreptavidina).

[000117] A modificação de uma sequência de ácido nucleico de interesse (e detecção da mesma) pode ser tudo que é necessário (por exemplo, no diagnóstico de uma doença). Desejavelmente, no entanto, adicional processamento da amostra é realizado. Convenientemente o polipeptídeo (e sequência de ácido nucleicos especificamente ligada no mesmo) é separado

do resto da amostra. Vantajosamente o complexo polipeptídeo-DNA é ligado em um apoio de fase sólida, para facilitar tal separação. Por exemplo, o polipeptídeo pode estar presente em uma acrilamida ou matriz de gel agarose ou, mais preferencialmente, é imobilizado na superfície de uma membrana ou nos poços de uma placa de microtitulação.

[000118] Em uma modalidade da invenção, tal domínio de repetição que compreende unidades de repetição é inserido em um polipeptídeo de bactérias, vírus, fungos, oomicetos, humanos, plantas ou animais para obter um reconhecimento de alvo e preferencialmente a ligação de um ou mais pares de bases especificados em uma sequência de DNA, e opcionalmente em que tais unidades de repetição são retiradas dos domínios de repetição de Famílias de proteínas tipo AvrBs3 que são ainda opcionalmente modificadas a fim de obter uma atividade de ligação específica pré-selecionada para um ou mais pares de base em uma sequência de DNA.

[000119] A invenção engloba isolado ou composições substancialmente purificadas de polinucleotídeo ou proteína. Um polinucleotídeo “isolado” ou “purificado” ou proteína, ou porção biologicamente ativa do mesmo, é substancialmente ou essencialmente livre de componentes que normalmente acompanham ou interagem com o polinucleotídeo ou proteína com encontrado no seu ambiente de ocorrência natural. Assim, um polinucleotídeo isolado ou purificado ou proteína é substancialmente livre de outros materiais celulares ou meios de cultivo quando produzidos por meio de técnicas recombinantes, ou substancialmente livres de precursores químicos ou outras substâncias químicas quando são quimicamente sintetizados. Idealmente, um polinucleotídeo “isolado” é livre de sequências (idealmente sequências de codificação de proteínas) que flaqueia naturalmente o polinucleotídeo (isto é, sequências localizadas nas extremidades 5' e 3' do polinucleotídeo) no DNA genômico do organismo a partir do qual o polinucleotídeo é derivado. Por exemplo, em várias modalidades, o

polinucleotídeo isolado pode conter menos do que cerca de 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb, ou 0,1 kb da sequência de nucleotídeo que naturalmente flanqueia o polinucleotídeo em DNA genômico da célula a partir da qual o polinucleotídeo é derivado. A proteína que é substancialmente livre de materiais celulares inclui preparações de proteína tendo menos do que cerca de 30%, 20%, 10%, 5%, ou 1% (por peso seco) da proteína de contaminação. Quando a proteína da invenção ou porção biologicamente ativa da mesma é produzida de forma recombinante, idealmente o meio de cultivo representa menos do que cerca de 30%, 20%, 10%, 5%, ou 1% (por peso seco) de precursores químicos ou substâncias químicas não-proteínas-de-interesse.

[000120] Fragmentos e variantes das Sequências de DNA e proteínas codificadas pelas mesmas apresentadas são também englobados pela presente invenção. Pelo termo “fragmento” compreende-se uma porção da sequência de DNA ou uma porção da sequência de aminoácido e assim, proteínas codificadas pelas mesmas. Fragmentos de uma sequência de DNA que compreende sequências de codificação podem codificar fragmentos de proteína que retêm atividade biológica da proteína nativa e assim o reconhecimento de DNA ou atividade de ligação para uma sequência de DNA alvo como descrito no presente documento. Alternativamente, os fragmentos de uma sequência de DNA que são úteis como testes de hibridização geralmente não codificam proteínas que retêm atividade biológica ou não retêm atividade promotora. Portanto, os fragmentos de uma sequência de DNA podem variar desde ao menos cerca de 20 nucleotídeos, cerca de 50 nucleotídeos, cerca de 100 nucleotídeos, e até o polinucleotídeo completo da invenção.

[000121] O termo “variantes” é designado para significar sequências substancialmente similares. Para as sequências de DNA, um variante compreende uma sequência de DNA com eliminações (isto é, truncamentos) na extremidade 5' e/ou 3'; eliminação e/ou adição de um ou mais nucleotídeos

em um ou mais locais internos no polinucleotídeo nativo; e/ou substituição de um ou mais nucleotídeos em um ou mais locais no polinucleotídeo nativo. Como usado no presente documento, uma sequência de DNA ou polipeptídeo “nativo” compreende sequência de DNA de ocorrência natural ou sequência de aminoácido, respectivamente. Para sequências de DNA, variantes conservadores incluem aquelas sequências que, por causa da degeneração do código genético, codifica a sequência de aminoácido de um dos polipeptídeos da invenção. Sequências de DNA variantes também incluem sequências de DNA sinteticamente derivadas, tal como aquelas geradas, por exemplo, por meio do uso de mutagênese de local direcionado, mas que ainda codifica a proteína da invenção. Geralmente, variantes de uma particular sequência de DNA da invenção terão ao menos cerca de 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% ou maior identidade com a sequência para o particular polinucleotídeo como determinado por parâmetros e programas de alinhamento de sequência como descritos em outro local no presente documento.

[000122] Variantes de uma particular sequência de DNA da invenção (isto é, a sequência de DNA de referência) também podem ser avaliada pela comparação de identidade percentual da sequência entre o polipeptídeo codificado por uma sequência de DNA variante e o polipeptídeo codificado por uma sequência de DNA de referência. A Identidade percentual da sequência entre qualquer um dos dois polipeptídeos pode ser calculada usando os parâmetros e programas de alinhamento de sequência descritos em outro local no presente documento. Quando qualquer par de polinucleotídeos da invenção é avaliado pela comparação de identidade percentual da sequência compartilhada pelos dois polipeptídeos que eles codificam, a identidade percentual da sequência entre os dois polipeptídeos codificados é ao menos cerca de 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% ou têm maior identidade com a sequência.

[000123] “Variante” de proteína é designado para significar uma proteína derivada da proteína nativa por meio de eliminação (chamada truncação) de um ou mais aminoácidos na extremidade da terminação N e/ou terminação C da proteína nativa; eliminação e/ou adição de um ou mais aminoácidos em um ou mais locais internos na proteína nativa; ou substituição de um ou mais aminoácidos em um ou mais locais na proteína nativa. As proteínas variantes englobadas pela presente invenção são biologicamente ativas, isto é, elas continuam compreendendo a desejada atividade biológica da proteína nativa como foi descrito no presente documento. Tais variantes podem resultar de, por exemplo, polimorfismo genético ou de manipulação humana. Variantes biologicamente ativos da proteína da invenção terão ao menos cerca de 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% ou maior identidade com a sequência de aminoácido para a proteína nativa como determinado por parâmetros e programas de alinhamento de sequência descritos em outro local no presente documento. Um variante biologicamente ativo da proteína da invenção pode ser diferente daquela proteína por tão poucos quanto 1-15 resíduos de aminoácido, tão pouco quanto 1-10, tanto quanto 6-10, tão pouco quanto 5, tão pouco quanto 4, 3, 2, ou até mesmo 1 resíduo de aminoácido.

[000124] As proteínas da invenção podem ser alteradas de várias maneiras incluindo substituições de aminoácidos, eliminações, truncamentos e inserções. Métodos para tais manipulações são geralmente conhecidas na técnica. Por exemplo, sequência de aminoácido variantes e Fragmentos das proteínas podem ser preparados por mutações no DNA. Os métodos para mutagênese e alterações do polinucleotídeo são bem conhecidos na técnica. Ver, por exemplo, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Métodos in Enzymol.* 154:367-382; Patente No. U.S. 4,873,192; Walker e Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) e as referências citadas nele.

As orientações para adequadas substituições de aminoácidos que não afetam a atividade biológica da proteína de interesse podem ser encontradas no modelo de Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), incorporado por referência no presente documento. Substituições conservadoras, como por exemplo, a troca de um aminoácido com outras propriedades similares, podem ser ideais.

[000125] Não se espera que as eliminações, inserções, e substituições das sequências de proteínas englobadas no presente documento possam produzir radicais alterações nas características da proteína. No entanto, quando é difícil prever o efeito exato da substituição, eliminação, ou inserção antes de fazê-lo, a pessoa versada na técnica apreciará que o efeito será avaliado por ensaios de varredura de rotina como descritos em outro local no presente documento ou conhecidos na técnica.

[000126] Sequências de DNA variantes e proteínas também englobam sequências e proteínas derivadas a partir de um procedimento mutagênico e recombinogênico como o embaralhamento de DNA. As estratégias para tal embaralhamento de DNA são conhecidas na técnica. Ver, por exemplo, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Cramer *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Cramer *et al.* (1998) *Nature* 391:288-291; e Patentes Nos. U.S. 5,605,793 e 5,837,458.

[000127] Nas abordagens de PCR, iniciadores de oligonucleotídeo podem ser designados para uso em reações de PCR para amplificar as correspondentes sequências de DNA de cDNA ou DNA genômico extraído de qualquer organismo de interesse. Métodos para projetar os iniciadores de PCR e clonagem de PCR são geralmente conhecidos na técnica e são divulgados em Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New

York). Ver também Innis *et al.*, eds. (1990) *PCR Protocols: A Guide to Métodos e Applications* (Academic Press, New York); Innis e Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); e Innis e Gelfand, eds. (1999) *PCR Métodos Manual* (Academic Press, New York). Os métodos conhecidos para PCR incluem, mas não estão limitados a, métodos usando iniciadores pareados, iniciadores aninhados, únicos iniciadores específicos, iniciadores de degeneração, iniciadores gene-específicos, iniciadores específicos de vetores, iniciadores parcialmente incompatíveis, e semelhantes.

[000128] Em técnicas de hibridização, todo ou parte de um polinucleotídeo conhecido é usado como um teste que se hibridiza seletivamente com outros polinucleotídeos correspondentes presentes em uma população de fragmentos de DNA genômico clonados ou fragmentos de cDNA (isto é, bibliotecas genômicas ou de cDNA) a partir de um organismo escolhido. Os testes de hibridização podem ser fragmentos de DNA genômico, fragmentos de cDNA, fragmentos de RNA, ou outros oligonucleotídeos, e podem ser marcados com um grupo detectável como o ^{32}P , ou qualquer outro marcador detectável. Assim, por exemplo, testes para hibridização podem ser feitos por meio da marcação de oligonucleotídeos sintéticos com base nas sequências de DNA da invenção. Os métodos para preparação de testes para hibridização e para construção de cDNA e bibliotecas genômicas são geralmente conhecidos na técnica e estão descritos em Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

[000129] A hibridização de tais sequências pode ser realizada sob condições severas. Com “condições severas” ou “condições severas de hibridização” designa condições sob as quais um teste vai se hibridizar com a sua sequência alvo para um grau de maior detecção do que outras sequências (por exemplo, ao menos 2-vezes sobre fundamento). As condições severas dependem da sequência e serão diferentes em circunstâncias diferentes. Pelo

controle da severidade das condições de hibridização e/ou lavagem, as sequências alvo que são 100% complementares com o teste podem ser identificadas (sondagem homóloga). Alternativamente, condições de severidade podem ser ajustadas para permitir alguma incompatibilidade em sequências de forma que menores graus de similaridade são detectados (sondagem heteróloga). Geralmente, a sondagem tem menos do que cerca de 1000 nucleotídeos em comprimento, idealmente menos do que 500 nucleotídeos em comprimento.

[000130] Tipicamente, condições severas serão aquelas em que a concentração de sal é menos do que cerca de 1,5 M Na íon, tipicamente cerca de 0,01 até 1,0 M Na íon de concentração (ou outros sais) em pH 7,0 até 8,3 e a temperatura é ao menos cerca de 30°C para sondagens curtas (por exemplo, 10 até 50 nucleotídeos) e ao menos cerca de 60°C para sondagens longas (por exemplo, maiores do que 50 nucleotídeos). Condições severas podem também ser obtidas com a adição de agentes de desestabilização como a formamida. Exemplos de baixas condições de severidade incluem hibridização com uma solução tampão de 30 até 35% formamida, 1 M NaCl, 1% SDS (dodecilsulfato de sódio) em 37°C, e uma lavagem em 1X até 2X SSC (20X SSC = 3,0 M NaCl/0,3 M citrato trissódico) em 50 até 55°C. Exemplos de condições de severidade moderadas incluem hibridização em 40 até 45% formamida, 1,0 M NaCl, 1% SDS em 37°C, e uma lavagem em 0,5X to 1X SSC em 55 até 60°C. Exemplos de altas condições de severidade incluem hibridização em 50% formamida, 1 M NaCl, 1% SDS a 37°C, e uma lavagem em 0,1X SSC em 60 até 65°C. Opcionalmente, tampões de lavagem podem compreender cerca de 0,1% até cerca de 1% SDS. A duração da hibridização é geralmente menor do que cerca de 24 horas, normalmente cerca de 4 até cerca de 12 horas. A duração do tempo de lavagem será ao menos uma duração de tempo suficiente para alcançar o equilíbrio.

[000131] A especificidade é tipicamente a função das lavagens pós-

hibridização, os fatores críticos sendo a força iônica e a temperatura da solução de lavagem final. Para híbridos de DNA-DNA, o T_m pode ser aproximado da equação de Meinkoth e Wahl (1984) *Anal. Biochem.* 138:267-284: $T_m = 81,5^{\circ}\text{C} + 16,6 (\log M) + 0,41 (\%GC) - 0,61 (\% \text{ forma}) - 500/L$; onde M é a molaridade de cátions monovalentes, $\%GC$ é o percentual de nucleotídeos de guanina e citosina no DNA, $\% \text{ forma}$ é o percentual de formamida na solução de hibridização, e L é o comprimento do híbrido em pares de bases. T_m é a temperatura (em determinada força iônica e pH) na qual 50% de uma sequência alvo complementar se hibridiza para uma sonda perfeitamente combinada. T_m é reduzido em torno de 1°C para cada 1% de incompatibilidade; portanto, T_m , hibridização, e/ou as condições de lavagem podem ajustadas para hibridizar as sequências da identidade desejada. Por exemplo, se as sequências com $\geq 90\%$ de identidade são buscadas, o T_m pode ser diminuído em 10°C . Geralmente, as condições severas são selecionadas estando em cerca de 5°C abaixo do ponto de fusão térmica (T_m) para a específica sequência e seu complemento em uma definida força iônica e pH. No entanto, as condições altamente severas podem utilizar a hibridização e/ou lavagem em 1, 2, 3, ou 4°C abaixo do ponto de fusão térmica (T_m); condições severas moderadas podem usar a hibridização e/ou lavagem em 6, 7, 8, 9, ou 10°C abaixo do ponto de fusão térmica (T_m); baixas condições de severidade podem utilizar uma hibridização e/ou lavagem em 11, 12, 13, 14, 15, ou 20°C abaixo do ponto de fusão térmica (T_m). Usando a equação, hibridização e composições de lavagem, e o desejado T_m , as pessoas versadas na técnica compreenderão que as variações da severidade da hibridização e/ou soluções de lavagem são escritas de forma inerente. Se o grau desejado de incompatibilidade resultar em um T_m com menos do que 45°C (solução aquosa) ou 32°C (solução de formamida), seria ideal aumentar a concentração de SSC de forma que uma temperatura mais alta possa ser usada. Um aprofundado guia para a hibridização de ácidos nucleicos é encontrado em

Tijssen (1993) *Laboratory Techniques in Biochemistry e Molecular Biology-Hybridization with nucleic acid Probes*, Part I, Chapter 2 (Elsevier, New York); e Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Chapter 2 (Greene Publishing e Wiley-Interscience, New York). Ver Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

[000132] É reconhecido que as sequências de DNA e proteínas da invenção englobam moléculas de polinucleotídeo e proteínas que compreendem um nucleotídeo ou uma sequência de aminoácido que é suficientemente idêntica às sequências de DNA ou à sequência de aminoácido apresentada no presente documento. O termo “suficientemente idêntica” é usado no presente documento para se referir a um primeiro aminoácido ou sequência de nucleotídeo que contém um número suficiente ou mínimo de resíduos aminoácido ou nucleotídeos idênticos ou equivalentes (por exemplo, com uma cadeia lateral semelhante) para um segundo aminoácido ou sequência de nucleotídeo de forma que o primeiro ou segundo aminoácido ou sequência de nucleotídeos possuem um domínio estrutural em comum e/ou atividade funcional em comum. Por exemplo, aminoácido ou sequência de nucleotídeos que contém um domínio estrutural em comum tendo ao menos cerca de 70% de identidade, preferencialmente 75% de identidade, mais preferencialmente 85%, 90%, 95%, 96%, 97%, 98% ou 99% de identidade são definidos no presente documento como suficientemente idênticos.

[000133] Para determinar a identidade percentual de duas sequências de aminoácidos ou de dois ácidos nucleicos, as sequências são alinhadas para fins de comparação ideal. A identidade percentual entre as duas sequências é uma função do número de posições idênticas compartilhadas pelas sequências (*isto é*, identidade percentual = número de posições idênticas/número total de posições (*por exemplo*, posições sobrepostas) x 100). Em uma modalidade, as duas sequências têm o mesmo comprimento. A identidade percentual entre

duas sequências pode ser determinada usando técnicas similares aquelas descritas abaixo, permitindo ou não lacunas. No cálculo da identidade percentual, tipicamente compatibilidades exatas são contadas.

[000134] A determinação da identidade percentual entre duas sequências pode ser obtida usando um algoritmo matemático. Um preferido exemplo não limitador de um algoritmo matemático utilizado para a comparação de duas sequências é o algoritmo de Karlin e Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264, modificado como em Karlin e Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Tal algoritmo é incorporado nos programas de NBLAST e XBLAST de Altschul *et al.* (1990) *J. Mol. Biol.* 215:403. As buscas por nucleotídeo BLAST podem ser realizadas com o programa NBLAST, pontuação = 100, tamanho da palavra = 12, para obter as sequências de nucleotídeo homólogas as moléculas de polinucleotídeo da invenção. As buscas por proteínas BLAST podem ser realizadas com o programa XBLAST, pontuação = 50, tamanho da palavra = 3, para obter as sequências de aminoácidos homólogas as moléculas de proteína da invenção. Para obter alinhamentos com lacunas para fins de comparação, Gapped BLAST pode ser utilizado como foi descrito em Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389. Alternativamente, PSI-Blast pode ser usado para realizar uma busca iterativa que detecta relações distantes entre as moléculas. Ver Altschul *et al.* (1997) *supra*. Quando se utiliza programas BLAST, Gapped BLAST, e PSI-Blast, os parâmetros padrão dos respectivos programas (por exemplo, XBLAST e NBLAST) podem ser usados. Ver, <http://www.ncbi.nlm.nih.gov>. Outro exemplo preferencial não limitador de um algoritmo matemático utilizado para a comparação de sequências é o algoritmo de Myers e Miller (1988) *CABIOS* 4:11-17. Tal algoritmo é incorporado no programa ALIGN (versão 2.0), que é parte do pacote de software de alinhamento de sequência GCG. Quando se utiliza o programa ALIGN programa para comparação das sequências de aminoácidos, uma

tabela PAM120 de resíduo de peso, uma penalidade de comprimento de lacuna de 12, e penalidade de lacuna de 4 pode ser usada. O alinhamento também é realizado manualmente por meio de inspeção.

[000135] A menos que declarado de outra forma, os valores de identidade/similaridade da seqüência fornecidos no presente documento se referem ao valor obtido usando as seqüências completas da invenção e usando múltiplos alinhamentos por meio do algoritmo Clustal W (Acid Nucleic Research, 22(22):4673-4680, 1994) usando o programa AlignX incluindo no pacote de software Vector NTI Suite Versão 7 (InforMax, Inc., Bethesda, MD, USA) usando os parâmetros padrão; ou qualquer programa equivalente do mesmo. Por “programa equivalente” designam-se qualquer programa de comparação de seqüência que, para qualquer uma das duas seqüências em questão, gera um alinhamento tendo compatibilidades idênticas de nucleotídeo ou resíduo de aminoácido e uma idêntica identidade percentual da seqüência quando comparado com o alinhamento correspondente gerado por CLUSTALW (Version 1.83) usando os parâmetros padrão (disponível no site de European Bioinformatics Institute: <http://www.ebi.ac.uk/Tools/clustalw/index.html>).

[000136] As seqüências de DNA da invenção podem ser fornecidas em cassetes de expressão para expressão em qualquer célula procariótica ou eucariótica e/ou organismo de interesse incluindo, mas não limitado a, bactéria, fungo, alga, plantas e animais. O cassete incluirá 5' e 3' seqüências reguladoras operacionalmente ligadas para uma seqüência de DNA da invenção. “operacionalmente ligada” é designado para significar uma ligação funcional entre dois ou mais elementos. Por exemplo, uma ligação operável entre um polinucleotídeo ou gene de interesse e uma seqüência reguladora (isto é, um promotor) é uma ligação funcional que permite a expressão do polinucleotídeo de interesse. Elementos operacionalmente ligados podem ser contínuos ou não contínuos. Quando usados para se referir à união de duas

regiões de codificação de proteína, operativamente ligado significa que as regiões de codificação estão na mesma estrutura de leitura. O cassete pode conter adicionalmente ao menos um gene adicional para ser cotransformado no organismo. Alternativamente, o gene adicional(is) pode ser fornecido em cassetes de múltipla expressão. Tal cassete de expressão é fornecido com uma pluralidade de locais de restrição e/ou locais de recombinação para inserção da sequência de DNA estando sob a regulação transcricional das regiões reguladoras. O cassete de expressão pode conter adicionalmente genes de marcação selecionáveis.

[000137] O cassete de expressão incluirá na direção 5'-3' de transcrição, uma região de iniciação transcricional e translacional (isto é, um promotor), uma sequência de DNA da invenção, e uma região de terminação transcricional e translacional (isto é, região de terminação) funcional em plantas ou outros organismos ou célula hospedeira não humana. As regiões reguladoras (isto é, promotoras, regiões reguladoras transcricionais, e regiões de terminação translacional) e/ou a Sequência de DNA da invenção podem ser nativos /análogos a célula hospedeira ou um do outro. Alternativamente, as regiões reguladoras e/ou Sequência de DNA da invenção podem ser heterólogas a célula hospedeira ou uma da outra. Como usado no presente documento, “heterólogo” em referência a uma sequência é uma sequência que se origina a partir de uma espécie estranha, ou, se for da mesma espécie, é substancialmente modificado a partir da sua forma nativa em composição e/ou locus genômico por intervenção humana deliberada. Por exemplo, um promotor operacionalmente ligado a um polinucleotídeo heterólogo é de uma espécie diferente da espécie a partir da qual o polinucleotídeo foi derivado, ou, se for da mesma/análoga espécie, um ou ambos serão substancialmente modificados a partir da sua forma original e/ou locus genômico, ou o promotor não é o promotor nativo para o polinucleotídeo operacionalmente ligado. Como usado no presente

documento, um gene quimérico compreende uma sequência de codificação operacionalmente ligada em uma região de iniciação de transcrição que é heteróloga a sequência de codificação.

[000138] A região de terminação pode ser nativa com a região de iniciação transcricional, pode ser nativa com a sequência de DNA de interesse operacionalmente ligada, pode ser nativa com o hospedeiro, ou pode ser derivada a partir de outra fonte (isto é, estranha ou heteróloga) para o promotor, a sequência de DNA de interesse, o hospedeiro de planta, ou qualquer combinação dos mesmos. Convenientes regiões de terminação para uso em plantas são disponíveis a partir de Ti-plasmídeo de *A. tumefaciens*, como as regiões de terminação de octopina sintase e nopalina sintase. Ver também Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; e Joshi *et al.* (1987) *Nucleic Acids Res.* 15:9627-9639.

[000139] Onde for apropriado, os polinucleotídeos podem ser otimizados para aumentar a expressão em um organismo transformado. Isto é, os polinucleotídeos podem ser sintetizados usando códons preferidos pelo hospedeiro para uma expressão melhorada. Ver, por exemplo, Campbell e Gowri (1990) *Plant Physiol.* 92:1-11 para uma discussão de uso de códons preferidos pelo hospedeiro. Os métodos estão disponíveis na técnica para a síntese de genes preferidos pelo hospedeiro, particularmente genes preferidos pelas plantas. Ver, por exemplo, Patentes Nos. U.S. 5,380,831, e 5,436,391, e Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, incorporadas por referência no presente documento.

[000140] Adicionais modificações da sequência são conhecidas por melhorar expressão de gene em um hospedeiro celular. Estes incluem a eliminação de sequências codificando espúrios de sinais de poliadenilação,

locais de sinais de Splice de exon-intron, repetições do tipo transposon, e outros tais como as sequências bem caracterizadas que podem ser deletérias para a expressão de gene. O conteúdo G-C de uma sequência pode ser ajustado para nivelar a média para um determinado hospedeiro celular, como calculado por referência para conhecidos genes expressos na célula hospedeira. Quando for possível, a sequência é modificada para evitar as previstas estruturas de mRNA secundárias de hairpin.

[000141] Os cassetes de expressão podem conter adicionalmente 5' sequências líder. Tais sequências líder podem agir para aumentar a translação. Os líderes de translação são reconhecidos na técnica e incluem: picornavirus líderes, por exemplo, EMCV líder (Encefalomiocardite 5' região de não codificação) (Elroy-Stein *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6126-6130); potyvirus líderes, por exemplo, TEV líder (Vírus “ETCH” do tabaco) (Gallie *et al.* (1995) *Gene* 165(2):233-238), MDMV líder (Vírus do mosaico anão do milho) (*Virology* 154:9-20), e proteína de ligação de cadeia pesada de imunoglobulina humana (BiP) (Macejak *et al.* (1991) *Nature* 353:90-94); líder não traduzido de proteína de revestimento de mRNA do vírus mosaico da alfalfa (AMV RNA 4) (Jobling *et al.* (1987) *Nature* 325:622-625); vírus mosáico do tabaco líder (TMV) (Gallie *et al.* (1989) em *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256); e vírus do mosqueado clorótico do milho líder (MCMV) (Lommel *et al.* (1991) *Virology* 81:382-385). Ver também, Della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968.

[000142] Na preparação do cassete de expressão, os vários fragmentos de DNA podem ser manipulados de forma a fornecer as sequências de DNA na orientação correta e, conforma for adequado, na estrutura de leitura adequada. Para esta finalidade, os adaptadores ou ligantes podem ser usados para unir os fragmentos de DNA ou outras manipulações podem ser envolvidas para os convenientes locais de restrição, remoção de DNA

supérfluo, remoção de locais de restrição, ou semelhantes. Para essa finalidade, mutagenese *in vitro*, reparo de iniciador, restrição, anelamento, resubstituições, por exemplo, transições e transversões, podem ser envolvidas.

[000143] Um número de promotores pode ser usado na pratica da invenção. Os promotores podem ser selecionados com base no hospedeiro de interesse e no resultado desejado. Os ácidos nucleicos podem ser combinados com os promotores constitutivos, os preferidos para o tecido ou outros promotores para expressão em plantas. Tais promotores constitutivos incluem, por exemplo, o promotor de núcleo CaMV 35S (Odell *et al.* (1985) *Nature* 313:810-812); actina de arroz (McElroy *et al.* (1990) *Plant Cell* 2:163-171); ubiquitina (Christensen *et al.* (1989) *Plant Mol. Biol.* 12:619-632 e Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689); pEMU (Last *et al.* (1991) *Theor. Appl. Genet.* 81:581-588); MAS (Velten *et al.* (1984) *EMBO J.* 3:2723-2730); promotor ALS (Patente No. U.S. 5,659,026), e semelhantes. Outros promotores constitutivos incluem, por exemplo, Patentes No. U.S. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; e 6,177,611.

[000144] Os promotores preferidos para o tecido podem ser usados tendo como alvo a expressão ampliada dentro de um tecido hospedeiro particular. Tais promotores preferidos para o tecido para uso nas plantas incluem, mas não estão limitados a, promotores preferidos para folha, promotores preferidos para a raiz, promotores preferidos para a semente, e promotores preferidos para o haste. Os promotores preferidos para o tecido incluem Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265; Kawamata *et al.* (1997) *Plant Cell Physiol.* 38(7):792-803; Hansen *et al.* (1997) *Mol. Gen Genet.* 254(3):337-343; Russell *et al.* (1997) *Transgenic Res.* 6(2):157-168; Rinehart *et al.* (1996) *Plant Physiol.* 112(3):1331-1341; Van Camp *et al.* (1996) *Plant Physiol.* 112(2):525-535; Canevascini *et al.* (1996) *Plant Physiol.* 112(2):513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol.*

35(5):773-778; Lam (1994) *Results Probl. Cell Differ.* 20:181-196; Orozco *et al.* (1993) *Plant Mol Biol.* 23(6):1129-1138; Matsuoka *et al.* (1993) *Proc Natl. Acad. Sci. USA* 90(20):9586-9590; e Guevara-Garcia *et al.* (1993) *Plant J.* 4(3):495-505. Tais promotores podem ser modificados, se necessário, para uma expressão fraca.

[000145] Geralmente, será benéfico expressar o gene de um promotor induzível, particularmente a partir de um promotor induzível de patógeno. Tais promotores incluem aqueles de proteínas relacionadas com patogênese (proteínas PR), que são induzidos em seguida a infecção por um patógeno; por exemplo, proteínas PR, proteínas SAR, beta-1,3-glucanase, chitinase, etc. Ver, por exemplo, Redolfi *et al.* (1983) *Neth. J. Plant Pathol.* 89:245-254; Uknes *et al.* (1992) *Plant Cell* 4:645-656; e Van Loon (1985) *Plant Mol. Virol.* 4:111-116. Ver também WO 99/43819, incorporado por referência no presente documento.

[000146] Os promotores de interesse são aqueles expressos localmente ou perto do local da infecção por patógeno. Ver, por exemplo, Marineau *et al.* (1987) *Plant Mol. Biol.* 9:335-342; Matton *et al.* (1989) *Molecular Plant-Microbe Interactions* 2:325-331; Somsisch *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:2427-2430; Somsisch *et al.* (1988) *Mol. Gen. Genet.* 2:93-98; e Yang (1996) *Proc. Natl. Acad. Sci. USA* 93:14972-14977. Ver também, Chen *et al.* (1996) *Plant J.* 10:955-966; Zhang *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:2507-2511; Warner *et al.* (1993) *Plant J.* 3:191-201; Siebertz *et al.* (1989) *Plant Cell* 1:961-968; Patente nº U.S. 5,750,386 (nematode-induzível); e as referências citadas nele. O de interesse particular é o promotor induzível para gene PRms de milho, cuja expressão é induzida pelo patógeno *Fusarium moniliforme* (ver, por exemplo, Cordero *et al.* (1992) *Physiol. Mol. Plant Path.* 41:189-200).

[000147] Os promotores quimicamente regulados podem ser usados para modular a expressão de um gene em uma planta através da aplicação de um

regulador químico exógeno. Dependendo do objetivo, o promotor pode ser um promotor químico induzível, onde a aplicação da substância química induz a expressão de gene, ou um promotor quimicamente repressor, onde a aplicação da substância química reprime a expressão de gene. Os promotores químicos induzíveis são conhecidos na técnica e incluem, mas não estão limitados a, o promotor de milho In2-2 que é ativado por meio de protetores herbicidas de benzenosulfonamida, o promotor de milho GST que é ativado por meio de compostos eletrofílicos hidrofóbicos que são utilizados como herbicidas pré-emergentes, e o promotor de tabaco PR-1um que é ativado por meio de ácido salicílico. Outros promotores quimicamente regulados de interesse incluem promotores de resposta para esteróides (ver, por exemplo, o promotor induzível por glucocorticóide em Schena *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 e McNellis *et al.* (1998) *Plant J.* 14(2):247-257) e os promotores induzíveis por tetraciclina e reprimidos por tetraciclina (ver, por exemplo, Gatz *et al.* (1991) *Mol. Gen. Genet.* 227:229-237, e Patentes Nos. U.S. 5,814,618 e 5,789,156), incorporado por referência no presente documento.

[000148] O cassete de expressão também pode compreender um gene marcador selecionável para seleção de células transformadas. Os genes de marcação selecionáveis são usados para seleção de células transformadas ou tecidos. Genes de marcação incluem genes que codificam a resistência a antibióticos, tal como aqueles codificando neomicina fosfotransferase II (NEO) e higromicina fosfotransferase (HPT), bem como genes conferindo resistência para compostos herbicidas, tais como glufosinato de amônio, bromoxinil, imidazolinonas, e 2,4-diclorofenoxyacetato (2,4-D). Adicionais marcadores selecionáveis incluem marcadores fenotípicos tais como β -galactosidase e proteínas fluorescentes tais como proteína fluorescente verde (GFP) (Su *et al.* (2004) *Biotechnol Bioeng* 85:610-9 e Fetter *et al.* (2004) *Plant Cell* 16:215-28), proteína ciano florescente (CYP) (Bolte *et al.* (2004) *J.*

Cell Science 117:943-54 e Kato *et al.* (2002) *Plant Physiol* 129:913-42), e proteína amarelo fluorescente (PhiYFP™ da Evrogen, ver, Bolte *et al.* (2004) *J. Cell Science* 117:943-54). Para mais marcadores selecionáveis, ver geralmente, Yarranton (1992) *Curr. Opin. Biotech.* 3:506-511; Christopherson *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-6318; Yao *et al.* (1992) *Cell* 71:63-72; Reznikoff (1992) *Mol. Microbiol.* 6:2419-2422; Barkley *et al.* (1980) in *The Operon*, pp. 177-220; Hu *et al.* (1987) *Cell* 48:555-566; Brown *et al.* (1987) *Cell* 49:603-612; Figge *et al.* (1988) *Cell* 52:713-722; Deuschle *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:5400-5404; Fuerst *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:2549-2553; Deuschle *et al.* (1990) *Science* 248:480-483; Gossen (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-1921; Labow *et al.* (1990) *Mol. Cell. Biol.* 10:3343-3356; Zambretti *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-3956; Baim *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-5076; Wyborski *et al.* (1991) *Nucleic Acids Res.* 19:4647-4653; Hillenand-Wissman (1989) *Topics Mol. Struc. Biol.* 10:143-162; Degenkolb *et al.* (1991) *Antimicrob. Agents Chemother.* 35:1591-1595; Kleinschmidt *et al.* (1988) *Biochemistry* 27:1094-1104; Bonin (1993) Ph.D. Thesis, University of Heidelberg; Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Oliva *et al.* (1992) *Antimicrob. Agents Chemother.* 36:913-919; Hlavka *et al.* (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.* (1988) *Nature* 334:721-724. Tais divulgações estão incorporadas por referência no presente documento.

[000149] A lista acima de genes de marcação selecionáveis não se destina à limitação. Qualquer marcador selecionável de gene pode ser usado na presente invenção.

[000150] Vários vetores de transformação de plantas e métodos para transformação de plantas estão disponíveis. Ver, por exemplo, An, G. *et al.* (1986) *Plant Physiol.*, 81:301-305; Fry, J., *et al.* (1987) *Plant Cell Rep.* 6:321-

325; Block, M. (1988) *Theor. Appl Genet.* 76:767-774; Hinchey, *et al.* (1990) *Stadler. Genet. Symp.* 203/212.203-212; Cousins, *et al.* (1991) *Aust. J. Plant Physiol.* 18:481-494; Chee, P. P. e Slightom, J. L. (1992) *Gene.* 118:255-260; Christou, *et al.* (1992) *Trends. Biotechnol.* 10:239-246; D'Halluin, *et al.* (1992) *Bio/Technol.* 10:309-314; Dhir, *et al.* (1992) *Plant Physiol.* 99:81-88; Casas *et al.* (1993) *Proc. Nat. Acad Sci. USA* 90:11212-11216; Christou, P. (1993) *In Vitro Cell. Dev. Biol.-Plant*; 29P:119-124; Davies, *et al.* (1993) *Plant Cell Rep.* 12:180-183; Dong, J. A. e Mchughen, A. (1993) *Plant Sci.* 91:139-148; Franklin, C. I. e Trieu, T. N. (1993) *Plant. Physiol.* 102:167; Golovkin, *et al.* (1993) *Plant Sci.* 90:41-52; *Guo Chin Sci. Bull.* 38:2072-2078; Asano, *et al.* (1994) *Plant Cell Rep.* 13; Ayeres N. M. e Park, W. D. (1994) *Crit. Rev. Plant. Sci.* 13:219-239; Barcelo, *et al.* (1994) *Plant. J.* 5:583-592; Becker, *et al.* (1994) *Plant. J.* 5:299-307; Borkowska *et al.* (1994) *Acta. Physiol Plant.* 16:225-230; Christou, P. (1994) *Agro. Food. Ind. Hi Tech.* 5: 17-27; Eapen *et al.* (1994) *Plant Cell Rep.* 13:582-586; Hartman, *et al.* (1994) *Bio-Technology* 12: 919923; Ritala, *et al.* (1994) *Plant. Mol. Biol.* 24:317-325; e Wan, Y. C. e Lemaux, P. G. (1994) *Plant Physiol.* 104:3748.

[000151] Os métodos da invenção envolvem a introdução de um construto de polinucleotídeo que compreende a Sequência de DNA dentro da célula hospedeira. Pela “introdução” compreende-se a apresentação para a planta do construto de polinucleotídeo de tal forma que o construto ganha acesso ao interior da célula hospedeira. Os métodos da invenção não dependem de um método partículas para introdução de um construto de polinucleotídeo dentro da célula hospedeira, apenas aquele construto de polinucleotídeo ganha acesso ao interior de uma das células do hospedeiro. Métodos para introduzir o construto de polinucleotídeos em bactérias, plantas, fungos e animais são conhecidos na técnica incluindo, mas não limitado a, métodos de transformação estável, métodos de transformação transientes, e métodos mediados por vírus.

[000152] Com “transformação estável” compreende-se que o construto de polinucleotídeo introduzido em plantas integra o genoma do hospedeiro e é capaz de ser herdado pelos seus descendentes. Com “transformação transiente” compreende-se que um construto de polinucleotídeo introduzido no hospedeiro não integra o genoma do hospedeiro.

[000153] Para a transformação de plantas e células de plantas, as sequências de DNA da invenção são inseridas usando técnicas padrão em qualquer vetor conhecido na técnica que é adequado para a expressão das sequências de DNA em uma célula hospedeira ou organismo de interesse. A seleção do vetor depende da preferida técnica de transformação e da espécie do hospedeiro alvo a ser transformado.

[000154] As metodologias para a construção de cassetes de expressão de planta e a introdução de ácidos nucleicos exógenos em plantas são geralmente conhecidas na técnica e forma previamente descritas. Por exemplo, DNA exógeno pode ser introduzido em plantas, usando vetores de plasmídeos de indução de tumores (Ti). Outros métodos usados para a entrega de DNA exógeno envolve o uso da transformação protoplasto mediada por PEG, electroporation, capilares (*whiskers*) de microinjeção, e bombardeio Biolístico ou de microprojéteis para a absorção direta de DNA. Tais métodos são conhecidos na técnica. (U.S. Pat. No. 5,405,765 to Vasil *et al.*; Bilang *et al.* (1991) *Gene* 100: 247-250; Scheid *et al.*, (1991) *Mol. Gen. Genet.*, 228: 104-112; Guerche *et al.*, (1987) *Plant Science* 52: 111-116; Neuhaus *et al.*, (1987) *Theor. Appl Genet.* 75: 30-36; Klein *et al.*, (1987) *Nature* 327: 70-73; Howell *et al.*, (1980) *Science* 208:1265; Horsch *et al.*, (1985) *Science* 227: 1229-1231; DeBlock *et al.*, (1989) *Plant Physiology* 91: 694-701; *Methods for Plant Molecular Biology* (Weissbach e Weissbach, eds.) Academic Press, Inc. (1988) e *Métodos in Plant Molecular Biology* (Schuler e Zielinski, eds.) Academic Press, Inc. (1989). O método de transformação depende da célula da planta que será transformada, a estabilidade dos vetores usados, o nível de

expressão dos produtos de genes e outros parâmetros.

[000155] As sequências de DNA da invenção podem ser introduzidas em plantas pelo contato das plantas com um vírus ou ácidos nucleicos virais. Geralmente, tais métodos envolvem a incorporação de uma construção de polinucleotídeo da invenção dentro de uma molécula de DNA ou RNA viral. É reconhecido que a proteína da invenção pode ser inicialmente sintetizada como parte de uma poliproteína viral que mais tarde pode ser processada por proteólise *in vivo* ou *in vitro* para produzir a desejada proteína recombinante. Além disso, é reconhecido que promotores da invenção também englobam promotores utilizados para transcrição por polimerases de RNA viral. Métodos para introdução de construtos de polinucleotídeos em plantas e expressando a proteína codificada nele, envolvendo moléculas de DNA ou RNA viral, são conhecidos na técnica. Ver, por exemplo, Patentes No. U.S. 5,889,191, 5,889,190, 5,866,785, 5,589,367 e 5,316,931; incorporado por referência no presente documento.

[000156] Em específicas modalidades, as sequências de DNA da invenção podem ser fornecidas para uma planta usando uma variedade de métodos de transformação transiente. Tais métodos de transformação transiente incluem, mas não estão limitados a, introdução da proteína ou variantes e fragmentos dos mesmos diretamente na planta a introdução de um transcrito codificando a proteína dentro da planta. Tais métodos incluem, por exemplo, microinjeção ou bombardeamento de partículas. Ver, por exemplo, Crossway *et al.* (1986) *Mol Gen. Genet.* 202:179-185; Nomura *et al.* (1986) *Plant Sci.* 44:53-58; Hepler *et al.* (1994) *Proc. Natl. Acad. Sci.* 91: 2176-2180 e Hush *et al.* (1994) *The Journal of Cell Science* 107:775-784, em que todos estes são incorporados por referência no presente documento. Alternativamente, o polinucleotídeo pode ser transformado transientemente dentro da planta usando técnicas conhecidas na técnica. Tais técnicas incluem expressão transiente mediada por *Agrobacterium tumefaciens* como foi

descrito abaixo.

[000157] As células que foram transformadas podem ser cultivadas em plantas de acordo com os meios convencionais. Ver, por exemplo, McCormick *et al.* (1986) *Plant Cell Reports* 5:81-84. Essas plantas podem ser cultivadas, e tanto polinizadas com a mesma cepa transformada ou cepas diferentes, e o híbrido resultante tendo a expressão constitutiva da característica fenotípica desejada é identificado. Duas ou mais gerações podem ser cultivadas para garantir que a expressão da desejada característica fenotípica seja mantida estável e herdada e então as sementes colhidas para garantir a expressão das desejadas características fenotípicas serão obtidas. Dessa maneira, a presente invenção fornece uma semente transformada (também referida como “semente transgênica”) tendo uma construção de polinucleotídeo da invenção, por exemplo, um cassete de expressão da invenção, estavelmente incorporados no seu genoma.

[000158] A presente invenção pode ser usada para transformação de qualquer espécie de planta, incluindo, mas não limitado a, monocotiledôneas e dicotiledôneas. Plantas de interesse particular incluem, mas não estão limitadas a, e plantas de grãos que fornecem sementes de interesse, plantas de sementes oleaginosas, plantas leguminosas, e *Arabidopsis thaliana*. As sementes interesse incluem sementes de grãos, tais como milho, trigo, cevada, arroz, sorgo, centeio, etc. Plantas de sementes oleaginosas incluem algodão, soja, soja, cártamo, girassol, Brassica, milho, alfafa, palma, coco, etc. Leguminosas incluem feijões e ervilhas. Feijões incluem guar, alfarroba, soja, feno-grego, feijão vagem, feijão de corda, feijão mungo, feijão de lima, feijão de fava, lentilha, grão de bico, etc.

[000159] Como usado no presente documento, o termo planta inclui células de plantas, protoplastos de plantas, culturas de tecidos celulares de planta, a partir do qual plantas podem ser geradas, calos de plantas, pedaços de plantas, e células de plantas que estão intactas em plantas ou partes de

plantas, tais como embriões, pólen, óvulos, sementes, folhas, flores, ramos, frutas, raízes, pontas das raízes, anteras, e semelhantes. Progenie, variantes, mutantes e das plantas regeneradas são também incluídos dentro do escopo da invenção, desde que essas partes compreendam os polinucleotídeos introduzidos.

[000160] A presente invenção ainda engloba a introdução das sequências de DNA da invenção dentro de célula hospedeiras que não são de plantas, incluindo, mas não limitado a, células de bactérias, células de leveduras, outras células de fungos, células humanas, e outras células animais. Além disso, a invenção engloba a introdução das sequências de DNA em animais e outros organismos tanto pelos métodos de métodos de transformação estável e transiente.

[000161] Como discutido no presente documento, a sequência de DNA da presente invenção pode ser expressa nesses sistemas eucarióticos. A síntese dos polinucleotídeos heterólogos em leveduras é bem conhecida (Sherman *et al.* (1982) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory). Duas leveduras amplamente utilizadas para a produção de proteínas eucarióticas são *Saccharomyces cerevisiae* e *Pichia pastoris*. Vetores, cepas e protocolos para a expressão em *Saccharomyces* e *Pichia* são conhecidos na técnica e disponíveis por fornecedores comerciais (por exemplo, Invitrogen). Adequados vetores normalmente possuem sequências de controle de expressão, tais como promotores, incluindo 3-fosfoglicerato quinase ou oxidase de álcool, e uma origem de replicação, sequências de terminação e semelhantes conforme desejado.

[000162] A sequências da presente invenção também pode ser ligada em vários vetores de expressão para uso na transfecção de culturas de células de origem de mamíferos ou insetos. Culturas de células ilustrativas úteis para a produção de peptídeos são células de mamíferos. Um adequado número de linhagens de célula hospedeira capazes de expressar proteínas intactas foi

desenvolvido na técnica e incluem as linhagens celulares HEK293, BHK21, e CHO. Vetores de expressão para essas células podem incluir sequências de controle de expressão, como uma origem de replicação, um promotor (por exemplo, o promotor CMV, um promotor HSV tk ou promotor pgk (fosfoglicerato quinase), um realçador (Queen *et al.* (1986) Immunol. Rev. 89:49), e locais de informações de processamento necessários, tais como locais de ligação de ribossomo, locais de Splice de RNA, locais de poliadenilação (por exemplo, um local de adição de SV40 largo T Ag poli A), e sequências terminadoras transcriptionais. Outras células de animais úteis para a produção de proteínas da presente invenção são disponíveis, por exemplo, a partir de American Type Culture Collection.

[000163] Adequados vetores para expressarem proteínas da presente invenção em células de insetos são normalmente derivados a partir do baculovírus SF9. Adequadas linhagens celulares de insetos incluem larvas de mosquito, bicho da seda, lagarta militar, traça e linhagem celulares de *Drosophila* tais como uma linhagem celular de Schneider (ver, Schneider (1987) J. Embryol. Exp. Morphol. 27:353-365).

[000164] Assim como com a levedura, quando células hospedeiras superiores de animais ou plantas são usadas, as sequências terminadoras de poliadenilação ou transcrição são tipicamente incorporadas dentro do vetor. Um exemplo de uma sequência terminadora é a sequência de poliadenilação do gene de hormônio de crescimento bovino. As sequências para um encaixe com exatidão do transcrito podem também ser incluídas. Um exemplo de uma sequência de encaixe é o íntron VP 1 de SV40 (Sprague *et al.* (1983) *J. Virol.* 45:773-781). Adicionalmente, as sequências de genes para controlar a replicação na célula hospedeira podem ser incorporadas dentro do vetor assim como aquelas encontradas nos vetores do tipo de vírus de papiloma (Saveria-Campo (1985) DNA Cloning Vol. II a Practical Approach, D. M. Glover, Ed., IRL Press, Arlington, Va., pp. 213-238).

[000165] Células hospedeiras de animal e eucariótica inferior (por exemplo, leveduras) são competentes ou produzidas de forma competente para transfecção de várias maneiras. Existem vários métodos bem conhecidos para introdução de DNA dentro de células animais. Estas incluem: precipitação de fosfato de cálcio, fusão das células receptoras com protoplastos bacterianos contendo o DNA, tratamento das células recipientes com lipossomas contendo o DNA, dextrins DEAE, eletroporação, biolística, e micro-injeção de DNA diretamente dentro das células. As células transfectadas são cultivadas por meios bem conhecidos na técnica (Kuchler (1997) *Biochemical Methods in Cell Culture and Virology*, Dowden, Hutchinson e Ross, Inc.).

[000166] Procariontes são mais freqüentemente representados por várias cepas de *E. coli*; no entanto, outras cepas microbianas podem também ser usadas no método da invenção. As sequências de controle procariótico comumente usadas que são definidas no presente documento para incluir promotores para iniciação da transcrição, opcionalmente com um operador, junto com as sequências de ligação de ribossomo, incluem tais promotores comumente usados como os sistemas promotores de beta lactamase (penicillinase) e lactose (*lac*) (Chang *et al.* (1977) *Nature* 198:1056), o sistema promotor de triptofano (*trp*) (Goeddel *et al.* (1980) *Nucleic Acids Res.* 8:4057) e o promotor P_L derivado de lambda e local de ligação de ribossomo N-gene (Shimatake *et al.* (1981) *Nature* 292:128). A inclusão de marcadores de seleção em vetores de DNA transfectados em *E. coli*. também é útil. Exemplos de tais maracadores incluem genes que especificam resistência à ampicilina, tetraciclina, ou cloramfenicol.

[000167] Este vetor é selecionado para permitir a introdução dentro da adequada célula hospedeira. Os vetores bacterianos são tipicamente de origem de plasmídeo ou fago. As células bacterianas adequadas são infectadas com partículas de vetores de fagos ou transfectadas com vetor de fago de DNA

desnudo. Se um vetor de plasmídeo for usado as células bacterianas são transfectadas com o DNA de vetor de plasmídeo. Os sistemas de expressão para expressar a proteína da presente invenção estão disponíveis usando *Bacillus sp.* e *Salmonella* (Palva *et al.* (1983) *Gene* 22:229-235); Mosbach *et al.* (1983) *Nature* 302:543-545).

[000168] Com respeito às proteínas de fusão, “operacionalmente ligado” é designado para significar uma ligação funcional entre dois ou mais elementos ou domínios. Se ele reconheceu que um ligante de um ou mais aminoácidos pode ser inserido entre cada um dos dois elementos ou mais para manter a função desejada de dois ou mais elementos.

[000169] Em uma modalidade da invenção, as proteínas de fusão compreendem um domínio de repetição da invenção operacionalmente ligado em ao menos uma proteína ou parte ou domínio da mesma. Em certas modalidades da invenção, a proteína ou parte ou domínio da mesma compreende a proteína ou parte funcional ou domínio da mesma, que é capaz de modificar DNA ou RNA. Em outras modalidades, a proteína ou parte funcional ou domínio da mesma é capaz de funcionar como um ativador transcricional ou um repressor transcricional. As proteínas preferidas incluem, mas não estão limitadas a, ativadores de transcrição, repressores de transcrição, um proteína mediadora de resistência, nucleases, topoisomerasas, ligases, integrases, recombinases, resolvases, metilases, acetilases, demetilases, e deacetilases.

[000170] Os seguintes exemplos são oferecidos apenas para fins de ilustração e não para fins de limitação.

EXEMPLOS

Exemplo 1: Identificação da Base para especificidade de DNA de efetores TAL

[000171] O fato de que AvrBs3 liga diretamente UPA-caixa, um elemento promotor em genes alvo induzidos (Kay *et al.* (2007) *Science* 318,

648-651; Römer *et al.* (2007) *Science* 318:645-648), nos levou a investigar a base para Especificidade de sequência de DNA. Cada região de repetição geralmente consiste de 34 aminoácidos, e as unidades de repetição são quase idênticas; no entanto, aminoácidos 12 e 13 são hipervariáveis (Schornack *et al.* (2006) *J. Plant Physiol.* 163:256-272; Fig. 1A). A principal repetição de C-terminal de AvrBs3 mostra a similaridade de sequência com outras unidades de repetição apenas nos seus primeiros 20 aminoácidos e é portanto referida como meia repetição. As unidades de repetição podem ser classificadas em diferentes tipos de repetição com base nos seus 12º e 13º aminoácidos hipervariáveis (Fig. 1B). Como o tamanho da UPA-caixa (18 (20)/19 (21) bp) quase corresponde ao número de unidades de repetição (17.5) em AvrBs3, nós consideramos a possibilidade de que uma unidade de repetição de AvrBs3 contate um específico par de bases DNAm. Quando os tipos de repetição de AvrBs3 (aminoácido 12 e 13 de cada repetição) são projetados sobre a caixa UPA, torna-se evidente que certos tipos de repetição se correlacionam com específicos pares de base no DNAlvo. Por exemplo, unidades de repetição de HD e NI têm uma forte preferência por C e A, respectivamente (Fig. 1B). Para maior simplicidade, nós designamos apenas as bases na cepa de DNA superior (sentido). Nosso modelo de especificidade de reconhecimento é apoiado pelo fato de que o derivado de eliminação de repetição de AvrBs3 AvrBs3 Δ rep16 que não tem quatro unidades de repetição (Δ 11-14; Fig. 5A, B) reconhece uma sequência de DNA alvo curta e diferente (Figs. 5 até 8). Com base nas comparações da sequência de UPA-caixas de genes de pimenta induzidos por AvrBs3 e análise mutacional, o DNA caixa alvo de AvrBs3 aparenta ter 1 bp a mais do que o número de unidades de repetição em AvrBs3. Além disso, um T é conservado na extremidade 5' da caixa UPA imediatamente antes da especificidade de reconhecimento previsto da primeira repetição (Fig. 1). Curiosamente, previsões estruturas secundárias da região da proteína antes da primeira repetição e a região de repetição

apresentou similaridades, apesar da ausência da conservação da sequência de aminoácido. Isto sugere uma repetição, denominada repetição 0 (Fig. 1B).

[000172] Para melhor fundamentar e ampliar o nosso modelo (Fig. 1B), nós previmos a então desconhecida sequência de DNA alvo de efetores TAL de *Xantomonas* baseados na sequência das suas unidades de repetição, e inspecionamos os promotores de conhecidos genes de alvo TAL e seus alelos para a presença de locais de ligação putativos. Nós identificamos sequências compatíveis com a especificidade prevista em promotores de alelos que são induzidos em resposta ao correspondente efector TAL, mas não em alelos não induzidos (Fig. 5C-F). A presença dessas caixas sugere que os genes induzidos são alvos diretos dos correspondentes efetores TAL. Com base na frequência de bases do DNA de diferentes tipos de repetição em uma sequência de DNA alvo usando oito efetores TAL nós deduzimos um código para a especificidade do DNA alvo de certos tipos de repetição (Fig. 1C, D; Fig. 5).

[000173] Para validar experimentalmente nosso modelo nós previmos sequências de DNA alvos para os efetores TAL Hax2 (21,5 unidades de repetição), Hax3 (11,5 unidades de repetição), e Hax4 (14,5 unidades de repetição) de patógeno de *Brassicaceae* *X. campestris* pv. *armoraciae* (22). Primeiro, nós derivamos caixas de DNA alvo para Hax3 e Hax4, porque eles contêm exclusivamente os tipos de repetição presentes em AvrBs3 (aminoácido 12/13: NI, HD, NG, NS; Fig. 1A, Fig. 2A) para o qual a ligação de DNA e ativação de gene foram mostrados experimentalmente. As caixas alvo Hax3 e Hax4 foram colocadas na frente de um mínimo promotor (-55 até +25) de tomate *Bs4*, que tem atividade basal muito fraca (Schornack *et al.* (2005) *Mol. Plant-Microbe Interact.* 18:1215-1225; Fig. 2B; Fig. 9), levando um gene repórter sem promotor *uidA* (β -glucuronidase, GUS). Para estudos de expressão transiente, nós transfectamos o construto repórter em conjunto com genes efetores de *hax3* e *hax4* acionado por promotor 35S de vírus

mosaico de couve-flor em folhas de *Nicotiana benthamiana* usando distribuição de T-DNA mediada por *Agrobacterium*. Ensaios GUS Qualitativos e quantitativos demonstraram que os promotores contendo a caixa Hax3- ou Hax4- foi fortemente e especificamente induzida na presença do efetor correspondente (Fig. 2C). Da mesma forma, nós direcionamos a importância do primeiro nucleotídeo (T) na sequência prevista sequência de DNA alvo de Hax3 e geramos quatro diferentes Hax3-caixas tanto com A, C, G ou T na extremidade 5' (Fig. 10A, B). A coexpressão de *hax3* e do construto de repórter em *N. benthamiana* demonstrou que apenas um promotor contendo uma Hax3-caixa com 5' T foi fortemente induzido na presença de Hax3 enquanto os outros levaram ao enfraquecimento da ativação (Fig. 10C). Isso indica que a posição 0 contribui para a especificidade do promotor de ativação Hax3 e da mesma forma outros efetores TAL. Para abordar a possibilidade de que alguns tipos de repetição conferem maior especificidade, isto é, reconhecer mais de uma base, nós permutamos Hax4-caixa (Fig. 3A, B). Ensaios GUS transientes mostraram que unidades de repetição NI-, HD-, e NG- em Hax4 favorecem fortemente o reconhecimento de bases A, C, e T, respectivamente, enquanto as unidades de repetição NS- reconhecem todas as quatro bases (Fig. 3B; Fig. 11). Como vários efetores TAL contém unidades de repetição NN- (Fig. 5 e Fig. 15, Tabela 1), nós geramos ArtX1, um efetor TAL artificial com unidades de repetição NN- e deduzimos um correspondente reconhecimento de sequência de DNA usando nosso código (Fig. 3C). A análise de derivados de Caixa de ArtX1 demonstrou que as unidades de repetição NN- reconhecem tanto A e G, com preferência para G (Fig. 3C). esse resultado confirma a nossa previsão de natural AvrXa27-caixa em arroz que contém tanto um A ou um G nas posições correspondendo as unidades de repetição NN- (Fig. 5C). Além disso, nós derivamos dois possíveis caixas de AvrXa10 com A ou G em posições correspondentes as unidades de repetição NN- em AvrXa10. Ambos os

construtos repórter foram induzidos de forma eficiente por AvrXa10 (Fig. 12). Juntos, estes dados sugerem fortemente que alguns tipos de repetição reconhecem específicos pares de base enquanto outros são mais flexíveis.

[000174] Um excepcional efector TAL é Hax2 porque ele contém 35 aminoácidos por repetição ao invés das típicas unidades de repetição de 34 aminoácidos (Kay *et al.* (2005) *Mol. Plant-Microbe Interact.* 18:838-848). Além disso, Hax2 contém uma rara combinação de aminoácidos na sua segunda repetição (aminoácidos 12/13: IG; Fig. 2A). Nós permutamos a correspondente terceira base de Hax2-caixa e analisamos a ativação do gene repórter com o efector Hax2 usando o ensaio transiente. Isso mostrou que uma repetição IG confere especificidade para T (Fig. 13). Hax2-caixa só leva à ativação do promotor Hax2, mas não por Hax3 ou Hax4 (Fig. 2C). Isso demonstrou que as unidades de repetição de 35 aminoácidos funcionam como unidades de repetição de 34 aminoácidos. Isso é sustentado pelo fato de que o efector TAL AvrHah1 que contém unidades de repetição de 35 aminoácido, induz a resistência mediada por *Bs3* (Schornack *et al.* (2008) *New Phytol.* 179:546-556). Os tipos de repetição de AvrHah1 são compatíveis com UPA-caixa no promotor *Bs3* (Fig. 5A, B).

[000175] Curiosamente, a expressão de *hax2* em *Arabidopsis thaliana* gera folhas de coloração roxa, indicando a acumulação de antocianina (Fig. 14A, B). Para identificar os genes alvo de Hax2 nós analisamos as regiões de promotor de genoma *A. thaliana* usando busca padrão (Patmatch, TAIR; www.arabidopsis.org) com sequências degeneradas de Hax2-caixa. Um dos supostos genes alvo de Hax2 codifica o fator de transcrição MYB de PAP1 (At1G56650) que controla a biosíntese de antocianina (Borevitz *et al.* (2000) *Plant Cell* 12:2383-2394). Análise semiquantitativa do nível de transcrição de *PAP1* demonstrou que expressão de *PAP1* é fortemente induzida por Hax2 (Fig. 14C). A inspeção visual da região de promotor de *PAP1* revelou a presença abaixo do ideal de Hax2-caixa (Fig. 14D, E). Com base no código

para os tipos de repetição de Efetor TAL (Fig. 1D) e os dados descritos acima nós previmos as possíveis sequências de DNA alvo para adicionais Efetores TAL, alguns dos quais são importantes fatores de virulência (Fig. 15, Tabela 1).

[000176] Como o número de repetição em efetores TAL varia de 1,5 a 28,5, uma questão chave é se efetores com poucas unidades de repetição podem ativar a expressão do gene. Por isso, testamos como o número de unidades de repetição influencia a expressão do gene alvo. Para isso, construímos efetores artificiais contendo as regiões N- e C-terminais de Hax3 e um domínio de repetição com 0,5 a 15,5 HD-unidades de repetição (especificidade para C). Por razões técnicas, a primeira repetição em todos os casos foi NI (especificidade para A). A caixa de DNA alvo correspondente consiste em 17 C-resíduos precedidos por TA (Fig. 4A, B). A ativação do promotor pelos efetores artificiais foi medida usando o ensaio GUS do promotor *Bs4* transiente em *N. benthamiana*. Enquanto pelo menos 6,5 unidades de repetição eram necessárias para a indução do gene, 10,5 ou mais unidades de repetição levaram a forte ativação do gene repórter (Fig. 4C). Estes dados demonstram que um número mínimo de unidades de repetição é necessário para reconhecer a caixa de DNA alvo artificial e ativar a expressão do gene. Os resultados também sugerem que efetores com menor número de repetição são em grande parte inativos. Nós mostramos que a região de repetição dos efetores TAL tem uma natureza sequencial que corresponde a uma sequência de DNA alvo consecutiva. Assim, deverá ser viável gerar efetores com as novas especificidades de ligação de DNA. Três efetores de efeitos artificiais foram gerados (ArtX1, ArtX2, ArtX3), cada um aleatoriamente montados com 12,5 unidades de repetição (Fig. 3C, D), e testados para a indução de fusões do repórter do promotor de *Bs4* contendo as sequências de DNA alvo previstas. Todos os três efetores artificiais fortemente e, especificamente, induziram o repórter GUS apenas na presença

da caixa de DNA alvo correspondente (Fig. 3E; Fig. 11). O nosso modelo de reconhecimento de especificidade dos efetores TAL no qual uma unidade de repetição entra em contato com um par de bases no DNA através dos aminoácidos 12 e 13 de cada repetição permite prever a especificidade de ligação dos efetores TAL e identificação dos genes alvo da planta. Como muitos efetores TAL são fatores de virulência importantes o conhecimento dos genes alvo da planta irá aumentar consideravelmente nossa compreensão do desenvolvimento da doença da planta causada por xantomonas. Além disso, conseguimos projetar com sucesso efetores artificiais que atuam como fatores de transcrição com domínios de ligação de DNA específicos. Anteriormente, os fatores de transcrição de dedo de zinco contendo um arranjo tandem de unidades de dedo de zinco foram projetados para ligar as seqüências de DNA alvo específicas.

[000177] Da mesma forma, os efetores TAL têm uma especificidade de ligação de DNA linear que pode ser facilmente reorganizada. Não escapou à nossa atenção que a estrutura superhelicoidal direita postulada das regiões de repetição nos efetores TAL sugere imediatamente um possível mecanismo para a interação com a hélice direita do material genético. Será importante determinar a estrutura do novo domínio de ligação de DNA dos efetores TAL complexados com o DNA alvo.

[000178] Os parágrafos seguintes descrevem modalidades adicionais da invenção:

1) Previsão das especificidades de ligação de DNA de proteínas homólogas AvrBs3- de ocorrência natural e geração de plantas resistentes.

[000179] As unidades de repetição do domínio de repetição dos efetores de ocorrência natural da família AvrBs3 codificam uma especificidade de ligação de DNA correspondente. Essas seqüências de reconhecimento podem ser previstas com o código de reconhecimento.

[000180] A inserção artificial das seqüências de reconhecimento previstas na frente de um gene em plantas transgênicas leva à expressão do gene se o efector semelhante a AvrBs3 correspondente é translocado para dentro da célula da planta (por exemplo, durante uma infecção bacteriana).

[000181] Se a seqüência de reconhecimento é inserida na frente de um gene cuja expressão leva a uma reação de defesa (gene de mediação de resistência) da planta, tais plantas transgênicas construídas são resistentes contra uma infecção das bactérias patogênicas da planta o que transloca o efector correspondente.

2) A identificação dos genes da planta cuja expressão é induzida por um efector específico da família AvrBs3

[000182] A previsão das seqüências de DNA alvo de um efector correspondente da família de AvrBs3 na região promotora dos genes de plantas é uma indicação para a expressão induzível destes genes pelo efector. Utilizando o método de acordo com a invenção, é possível prever os genes de plantas induzíveis. Previsões são particularmente simples nos genomas seqüenciados.

3) Utilização de outros efetores como ativadores de transcrição em sistemas de expressão

[000183] Análogo ao uso de Hax3 e Hax4, as seqüências de ligação de DNA previstas de outros membros da família AvrBs3 podem ser inseridas em promotores para gerar novos promotores controláveis que podem ser induzidos pelo efector correspondente.

4) Construção de um sistema induzível secundário

[000184] Dois constructos são introduzidos em plantas. Primeiro, um gene *hax3* cuja expressão está sob controle de um promotor induzível. Em segundo lugar, um gene alvo que contém a caixa Hax3 no promotor. Indução da expressão de *hax3* leva a produção da proteína Hax3 que, então, induz a expressão do gene alvo. A construção de dois componentes descrita leva a

uma chave de expressão dupla que permite uma expressão variável do gene alvo. O transativador e o gene alvo também podem estar presentes primeiro em diferentes linhagens de plantas e podem ser introgressados à vontade. Análoga a esta, Hax4 e a caixa xHax4 correspondente podem ser usados. Este sistema também pode ser usado com outros membros da família AvrBs3 ou derivados artificiais e seqüências de DNA alvo previstas. A funcionalidade do sistema já poderia ser verificada. As plantas *Arabidopsis thaliana* transgênicas foram construídas, as quais contêm um gene *avrBs3* induzível, bem como um gene *Bs3* sob o controle do seu promotor nativo, cuja expressão pode ser induzida por AvrBs3. A indução da expressão de *avrBs3* leva à expressão de *Bs3* e, portanto, à morte celular. Ver, WO 2009/042753, aqui incorporada por referência.

5) Construção das plantas resistentes à doenças

[000185] Se a seqüência de DNA alvo de um efector similar a AvrBs3 é inserida na frente de um gene cuja expressão leva a uma reação de defesa (gene de mediação de resistência) da planta, correspondentemente as plantas transgênicas construídas serão resistentes contra a infecção de organismos patogênicos de plantas, que tornam este efector disponível. Como um gene de mediação de resistência pode, por exemplo, levar a uma morte celular local que impede a propagação de organismos/patógenos, ou induzir a resistência basal ou sistêmica da célula da planta.

6) Geração de domínios de repetição para a detecção de uma seqüência de DNA específica e a indução de transcrição dos seguintes genes

[000186] A arquitetura modular do domínio de repetição central permite a construção específica das especificidades de ligação de DNA definitivas e com isso a indução da transcrição dos genes de plantas selecionados. As especificidades de ligação do DNA podem ser inseridas artificialmente na frente dos genes alvo para que as novas variantes da caixa de DNA efectoras

são geradas para a expressão induzível de genes alvo. Além disso, os domínios de repetição podem ser construídos que reconhecem uma sequência de DNA de ocorrência natural nos organismos. A vantagem dessa abordagem é que a expressão de qualquer gene em organismos não-transgênicos pode ser induzida se um efector correspondente da invenção está presente nas células do organismo.

[000187] A introdução do efector pode ser feita de diferentes maneiras:

(1) transferência através das bactérias com um sistema de transporte de proteína (por exemplo, sistema de secreção do tipo III);

(2) bombardeio de célula com uma proteína AvrBs3 artificial;

(3) transferência de um segmento de DNA que leva à produção do efector, através de introgressão, *Agrobacterium*, vetores virais ou bombardeio de célula; ou

(4) outros métodos que resultam na absorção da proteína efetora pela célula alvo

[000188] O domínio de repetição central dos efetores da família AvrBs3 é um novo tipo de domínio de ligação de DNA (Kay *et al.*, 2007). A decodificação da especificidade das unidades de repetição únicas agora permite a adaptação direcionada da especificidade de ligação de DNA desta região. A região de ligação de DNA pode ser translacionalmente fundida a outros domínios funcionais para gerar efetores de sequência específica. A seguir, quatro exemplos de tais fusões de proteínas são dados.

7) Construção de ativadores de transcrição para a expressão induzível de genes em células de organismos vivos

[000189] Os efetores da família semelhante a AvrBs3 induzem a expressão dos genes em células de planta. Para isso, o C-terminal da proteína é essencial, que contém um domínio de ativação de transcrição e sequências de localização nucleares que mediam a importação de proteínas para o núcleo da planta. O C-terminal da proteína homóloga AvrBs3 pode ser modificado de

tal maneira que este media a expressão de genes em sistemas de fungos, animais ou humanos. Assim, os efetores podem construir aquela função como ativadores transcricionais em humanos, outros animais ou fungos. Assim, os métodos de acordo com a invenção podem ser aplicados não só às plantas, mas também para outros organismos vivos.

8) Uso de efetores como repressores de transcrição

[000190] A especificidade de ligação do DNA do domínio de repetição pode ser usada junto com outros domínios em fusões de proteínas para a construção de efetores que atuam como repressores específicos. Estes efetores exibem uma especificidade de ligação de DNA que foi gerada de tal maneira que eles se ligam aos promotores dos genes alvo. Em contraste com os efetores TAL que são ativadores de transcrição, estes efetores são construídos para bloquear a expressão dos genes alvo. Como com os repressores clássicos, espera-se que estes efetores cubram as seqüências promotoras pelo seu reconhecimento de, ou ligação a, uma seqüência de DNA alvo e as torna inacessíveis para os fatores que de outra forma controlam a expressão dos genes alvo. Alternativamente, ou adicionalmente, os domínios de repetição podem ser fundidos a um domínio de repressão de transcrição, como um motivo EAR (Ohta *et al. Plant Cell* 13:1959-1968 (2001)).

9) Uso de domínios de repetição para rotulagem e isolamento de seqüências específicas

[000191] A capacidade de um domínio de repetição reconhecer uma seqüência de DNA alvo específica pode ser utilizada em conjunto com outros domínios para rotular as seqüências de DNA específicas. C-terminalmente um GFP (“proteína fluorescente verde”) pode, por exemplo, ser fundida a um domínio de repetição artificial que detecta uma seqüência de DNA desejada. Esta proteína de fusão se liga *in vivo* e *in vitro* a uma seqüência de DNA correspondente. A posição desta seqüência no cromossomo pode ser localizada usando a proteína GFP fundida. De forma análoga, outros domínios

de proteínas que permitem uma localização celular da proteína (por exemplo, por FISH) podem ser fundidos a um domínio de repetição artificial específico que visa a proteína para uma sequência de DNA correspondentes no genoma da célula. Além disso, a especificidade de reconhecimento do DNA dos domínios de repetição da invenção pode ser usada para isolar sequências de DNA específicas. Para isso, a proteína semelhante a AvrBs3 pode ser imobilizada em uma matriz e interage com as moléculas de DNA correspondentes que contêm uma sequência compatível. Portanto, as sequências de DNA específicas podem ser isoladas a partir de uma mistura de moléculas de DNA.

10) Uso de domínios de repetição para a clivagem endonucleolítica de DNA

[000192] A especificidade de reconhecimento de DNA do domínio de repetição pode ser fundida a uma endonuclease de restrição adequada para especificamente clivar o DNA. Portanto, a ligação de sequência específica do domínio de repetição leva à localização da proteína de fusão para algumas sequências específicas, de modo que a endonuclease clive especificamente o DNA no local desejado. Por meio do reconhecimento das sequências de DNA alvo, as nucleases não específicas como *FokI* podem ser transformadas em endonucleases específicas análogas para trabalho feito com nucleases de dedo de zinco. Por exemplo, a distância ideal entre os dois sítios alvo do DNA efetor seriam determinados para que fossem necessários para suportar a dimerização de dois domínios *FokI*. Isto seria realizado através da análise de uma coleção de constructos na qual os dois locais de ligação de DNA são separados por sequências de espaçador de diferentes dimensões. O uso dessa abordagem permite que uma pessoa determine as distâncias que permitem que a clivagem de DNA mediada por nuclease ocorra e a análise funcional das nucleases de efetor adicionais que visam as diferentes sequências de DNA. Em uma abordagem alternativa, um dímero *FokI* de cadeia única recém-

desenvolvido (Mino *et al.* (2009) *J Biotechnol* 140:156-161) é empregado. Nesta abordagem dois domínios catalíticos *FokI* são transcricionalmente fundidos a um domínio de repetição único da invenção. Assim, a funcionalidade de uma nuclease correspondente já não depende da dimerização intermolecular de dois domínios *FokI* que estão localizados em duas proteínas diferentes. Este tipo de construção tem sido utilizado com sucesso no contexto de motivos de ligação de DNA a base de dedo de zinco. Além disso, estes métodos permitem cortes muito específicos em apenas algumas posições no complexo de moléculas de DNA. Estes métodos podem, entre outras coisas, ser usados para introduzir as quebras de filamento duplo *in vivo* e seletivamente incorporar o doador de DNA para essas posições. Estes métodos também podem ser usados especificamente para inserir transgenes.

11) Construção de domínios de repetição com ordem de repetição designada personalizada

[000193] Devido à alta similaridade entre as unidades de repetição individuais de um domínio de repetição, a construção de um polipeptídeo de ligação de DNA personalizado como descrito acima pode não ser viável através de métodos que envolvem métodos de clonagem tradicionais. Conforme detalhado neste exemplo, um domínio de repetição com uma ordem de unidade de repetição que corresponde a uma sequência de DNA desejada em um promotor de interesse, tal como o promotor *Bs4* (Fig. 17B, C), é determinado com base no código de reconhecimento da presente invenção. A geração de uma ordem de unidade de repetição 11,5 específica foi realizada usando a clonagem “Golden Gate” (Engler *et al.* (2008) *PLoS ONE* 3:e3647). Como blocos de construção, nós subclonamos o N- e C-terminal de Hax3, bem como as 12 unidades de repetição individuais assemelhando-se às 11,5 unidades de repetição. Cada bloco de construção continha locais *Bsa* I de flanqueamento individuais (Fig. 18) que permitiram

uma montagem ordenada dos fragmentos em um polipeptídeo de efetor personalizado. O efetor (ARTBs4) foi corretamente montado a partir do total de 14 fragmentos em um vetor binário compatível *Bsa* I que permite a expressão mediada por *Agrobacterium* do polipeptídeo do efetor personalizado como uma fusão GFP marcada N-terminalmente nas células de planta (Fig. 18).

12) Uso de efetores como repressores virais

[000194] A especificidade de ligação de nucleotídeos do domínio de repetição pode ser usada para projetar efetores que interrompem a replicação viral nas células. Estes efetores vão expor uma especificidade de ligação de nucleotídeos alvo de sequência de nucleotídeos em origens de replicação viral e outras sequências críticas para a função viral. Nenhum domínio de proteínas adicional precisa ser fundido a essas proteínas de domínio de repetição a fim de bloquear a função viral. Eles agem como repressores clássicos, cobrindo as origens da replicação ou outras sequências chave, incluindo promotores, realçadores, unidades de repetição de terminal longo, e locais de entrada de ribossomo internos, ligando e tornando-os inacessíveis para hospedeiros ou fatores virais, incluindo a polimerase de RNA dependente de RNA codificado viral, proteínas de nucleocapsídeo e integrases, que participam na replicação viral e função. Este tipo de estratégia tem sido usado com sucesso com proteínas de dedo de zinco (Sera (2005) *J. Vir.* 79:2614-2619; Takenaka *et al.* (2007) *Nucl Acids Symposium Series* 51:429-430).

[000195] Resumindo, a presente invenção adicionalmente abrange as moléculas de ácido nucléico isoladas para serem usadas em qualquer um dos métodos da presente invenção, plantas transformadas compreendendo um polinucleotídeo heterólogo estavelmente incorporado em seu genoma e compreendendo a molécula de nucleotídeo descrita acima, de preferência operacionalmente ligada a um elemento promotor e/ou operacionalmente ligado a um gene de interesse. A planta transformada é de preferência uma

monocotiledônea ou uma dicotiledônia. A invenção abrange também as sementes das plantas transformadas. A invenção abrange as células do hospedeiro humanas e não humanas transformadas com qualquer um dos polinucleotídeos da invenção ou os polipeptídeos da invenção. Os promotores usados em combinação com qualquer um dos nucleotídeos e polipeptídeos da invenção são de preferência promotores específicos de tecido, promotores induzíveis por produtos químicos e promotores induzíveis por patógenos.

[000196] Enquanto a presente invenção pode ser usada em sistemas animais e vegetais, uma modalidade preferida opcional refere-se ao uso em sistemas vegetais. O termo planta inclui células de planta, protoplastos de plantas, culturas de tecidos de células de plantas a partir das quais as plantas podem ser regeneradas, calos de planta, aglomerados de plantas e células de plantas que estão intactas em plantas ou partes de plantas, tais como embriões, pólen, óvulos, sementes, folhas, flores, ramos, frutos, raízes, pontas das raízes, anteras e similares. A progênie, variantes e mutantes das plantas regeneradas também estão incluídos no escopo da invenção, visto que essas parte compreendem os polinucleotídeos introduzidos.

Materiais e Métodos

[000197] **Cepas bacterianas e condições de crescimento.** *Escherichia coli* foi cultivada a 37°C em meio *lysogeny broth* (LB) e *Agrobacterium tumefaciens* GV3101 a 30°C em caldo de extrato de levedura (YEB) suplementado com antibióticos apropriados.

[000198] **Material vegetal e as inoculações.** Plantas *Nicotiana benthamiana* foram cultivadas na estufa (dia e noite em temperaturas de 23°C e 19°C, respectivamente) com 16 h de luz e 40-60% de umidade. Folhas maduras de plantas de cinco a sete semanas de idade, foram inoculadas com *Agrobacterium*, utilizando uma seringa sem agulha conforme descrito anteriormente (S1). Plantas inoculadas foram transferidas para uma câmara de crescimento Percival (*Percival Scientific*) com 16 h de luz, 22°C e 18°C de

temperatura, à noite.

[000199] **Construção de efetores artificiais.** A construção de efetores com a região de repetição modificada foi baseada na ligação de fragmentos de restrição *Esp3I* (Fermentas). *Esp3I* corta fora de sua seqüência de reconhecimento e, normalmente, uma vez por repetição. Para construir um vetor de entrada compatível do sistema *Gateway* (*Invitrogen*) para a geração de efetores da invenção, o N- e C-terminais de *hax3* foram amplificados por PCR usando uma polimerase de verificação (*proof reading*) (*HotStar HiFidelity Polymerase Kit; Qiagen*), combinado pela SOE (encaixe pela extensão de sobreposição)-PCR e inserido em pCR8/GW/TOPO resultando em derivados de *hax3* com 1,5 unidades de repetição (pC3SE26; primeira repetição = NI; última metade de repetição = NG). Uma estrutura variante de 1 bp anterior ao códon de iniciação, foi inserida por mutagênese direcionada por sítio, para permitir na estrutura, fusões N-terminal usando recombinação *Gateway* (*Invitrogen*), resultando em pC3SEIF. Unidades de repetição simples foram amplificadas a partir de efetores TAL usando uma ligação de iniciador de direção para a maioria das unidades de repetição, e de iniciadores inverso de repetição específica. Ambos iniciadores incluíram os sítios *Esp3I*, presentes naturalmente. Para evitar a amplificação de mais de uma repetição, o DNA molde foi digerido com *Esp3I* antes da reação de PCR. Os produtos do PCR foram digeridos com *Esp3I* e clonados em pC3SE26, digerido por *Esp3I*, resultando em derivados de Hax3 com 2.5 unidades de repetição, onde uma repetição simples pode ser retirada com *Esp3I* (repetição HD = 5 repetições de Hax3; repetição NI = 11 repetições de Hax3; repetição NG = 4 repetições de Hax4; repetição NN = mutante G13N de 4 repetições de Hax4). A construção do suporte principal do efector ArtHD consiste no N- e C-terminal de Hax3 com a repetição da última metade mutada, em uma repetição HD. A construção resultante foi restringida por *Esp3I* e desfosforilada. As unidades de repetição de codificação de fragmentos de

DNA foram excisadas com *Esp3I* de derivados de pC3SE26 contendo uma única repetição HD, e purificadas por meio de géis de agarose. A ligação foi realizada com um excesso molar de inserção para vetor, para facilitar a ligação de concatemer, e transformada em *E. coli*. O número de unidades de repetição foi determinado em plasmídeos recombinantes utilizando *StuI* e *HincII*. Efetores ArtX1-3 com uma combinação aleatória de tipos, foram gerados isolando-se unidades de repetição de codificação de fragmentos de DNA, como descrito acima, a partir de unidades de repetição NI, HD, NN, e NG simples clonadas (especificidades para A, C, G/A e T, respectivamente). Os fragmentos foram adicionados em quantidades molares iguais, cada um à reação de ligação de concatemer com vetor pC3SEIF. Os efetores contendo os Plasmídeos da invenção com 12.5 unidades de repetição, foram escolhidos para análise posterior. Os efetores foram clonados por recombinação *Gateway* (*Invitrogen*) em pGWB6 (S2) para a expressão de fusões efetoras GFP N-terminal. Sequências de oligonucleotídeos estão disponíveis mediante solicitação. Todas as construções foram sequenciadas.

[000200] **Construção de repórter GUS.** O promotor *Bs4* mínimo foi amplificado por PCR e inserido pENTR/D-TOPO (*Invitrogen*) com caixas de DNA alvo na extremidade 5'(S3; Fig. S5). Derivados do promotor foram clonados em pGWB3 (S2), contendo um gene *uidA* sem promotor.

[000201] **Construção de *hax2*-transgênica *A. thaliana*.** *hax2* foi clonado sob o controle do promotor *alcA* induzível de *Aspergillus nidulans*, em um derivado compatível *Gateway* do vetor T-DNA binário binSRNACatN (*Zeneca Agrochemicals*), contendo o gene regulador, dependente de etanol, *alcR* acionado por 35S-e, um marcador de seleção *nptII*. *AlcR* aciona a indução depende de etanol do promotor *alcA* (S4). T-DNA contendo esses genes foi transformado em *A. thaliana* Col-0 através de *A. tumefaciens* utilizando inoculação *floral dip* (S5). Transformantes foram selecionados, como plantas resistentes à canamicina, em meio estéril.

[000202] **Construção de ARTBs4, um efetor artificial.** Clonagem “Golden Gate” (Engler et al (2008) PLoS ONE 3:e3647) foi usada para montar efetores com 11,5 unidades de repetição especificamente ordenadas. O N- e C-terminal de Hax3 e 12 unidades de repetição individuais semelhante às 11,5 unidades de repetição foram subclonados. Cada bloco de construção continha sítios individuais de acompanhamento *BsaI*, o que permitiu uma montagem ordenada dos fragmentos em um efetor artificial. Para a montagem alvo de efetores com qualquer composição de repetição desejada, o repertório de bloco de construção de unidades de repetição foi ampliado. Para permitir a especificidade alvo para qualquer uma das quatro bases naturais (A, C, G e T) no DNA, quatro tipos diferentes de repetição foram escolhidas, com base nos aminoácidos 12 e o 13, por unidade de repetição. Os quatro tipos de repetição e suas especificidades são: NI = A; HD = C; NG = T, NN = G ou A. Para gerar um kit de montagem universalmente aplicável, quatro unidades correspondentes a cada um dos quatro tipos de unidade de repetição foram clonadas com sítios de acompanhamento *BsaI* para cada uma das 12 posições de repetição. A soma de 48 blocos de construção assemelha-se a uma biblioteca que pode ser usada para montar efetores com 11,5 unidades de repetição com qualquer composição dos quatro tipos de unidade de repetição.

[000203] **Ensaio de β -glucuronidase (GUS).** Para ensaios GUS transitórios, construção de efetor de liberação de cepas de *Agrobacterium* e construção de repórter GUS foram misturadas em 1:1, e inoculadas em folhas de *Nicotiana benthamiana* com um OD₆₀₀ de 0,8. Dois discos de folha (0,9 cm de diâmetro) foram transformados em amostras dois dias após a infiltração (dpi), e a atividade GUS quantitativa foi determinada utilizando 4-metil-umbeliferil- β -D-glicuronídeo (MUG), como descrito anteriormente (SI). Proteínas foram quantificadas através de ensaios de *Bradford* (BioRad). Os dados correspondem a amostras em triplicata de diferentes plantas. Para ensaios GUS qualitativos, discos de folhas foram transformados em amostras

2 dpi, incubadas em solução de coloração X-Gluc (5-bromo-4-cloro-3-indolil- β -D-glicuronídeo) (S3), descoradas em etanol e secas. Experimentos foram realizados pelo menos duas vezes com resultados semelhantes.

[000204] **Expressão de *hax2*, *hax3* e *hax4*.** *hax2*, *hax3* e *hax4* foram expressos em planta sob o controle do promotor 35S do vírus mosaico da couve-flor constitutivo, usando pAGH2, pAGH3 e pAGH4 (S6).

[000205] **Footprinting DNaseI.** Footprinting DNaseI foi realizado conforme descrito (S7), com as seguintes modificações: produtos de PCR marcados com fluorescência de DNA promotor BS3 e BS3-E foram gerados usando plasmídeos pCRBluntII-TOPO::FPBs3 (fragmento promotor BS3 de -211 a +108) e pCRBluntII-TOPO::FPBs3-E (fragmento promotor BS3-E de -224 a +108), respectivamente, como modelo e Phusion DNA polimerase (Finnzymes). Produto de PCR marcados com fluorescência de DNA promotor UPA20-ubm-r16 foi gerado através de plasmídeo pCRBluntII-TOPO::FPU20-ubm-r16 (fragmento promotor UPA20 de -213 a +86 contendo a mutação ubm-r16 (S7) como modelo e Phusion DNA polimerase (Finnzymes). Os Plasmídeos pCRBluntII-TOPO::FPBs3, pCRBluntII-TOPO::FPBs3-E e pCRBluntII-TOPO::FPU20-UBM-r16 foram seqüenciados, usando o Thermo Sequenase Dye Primer Manual Cycle Sequencing Kit (USB) de acordo com instruções do fabricante. Um internal Gene Scan-500LIZ tamanho padrão (Applied Biosystems) foi utilizado para determinar o tamanho do fragmento de DNA.

Exemplo 2: Identificação de uma unidade de repetição

TAL que se liga ao nucleotídeo G

[000206] O domínio de ligação ao DNA de efetores TAL é composto de 34 unidades de repetição aminoácidos arranjados em tandem. As seqüências de aminoácidos das unidades de repetição são na sua maioria conservadas, exceto por dois resíduos adjacentes altamente variáveis (HVRs) nas posições 12 e 13 que definem a especificidade do DNA alvo (Boch et al (2009) Science

326:1509-1512; *Moscou and Bogdanove (2009) Science 326:1501*). A análise funcional identificou grupos HVR que se ligam preferencialmente a A (NI), C (HD), T (NG, IG) ou igualmente bem a G e A (NN) (*Boch et al (2009). Science 326:1509-1512*). Análise bioinformática revelou que HVRs nas dadas interações de efetor de promotor-TAL correspondem especificamente a G (*Moscou and Bogdanove (2009) Science 326:1501*). No entanto, esta análise, foi baseada em um único (HN & NA) ou dois (NK) sítios de interação. Em nossa opinião o número de sítios de interação é baixo demais para tirar conclusões confiáveis sobre a especificidade de HVR. No entanto, estes HVRs podem ser considerados como candidatos adequados que podem mediar a ligação específica à G.

[000207] A fim de esclarecer a especificidade alvo de HVRs com especificidade desconhecida fizemos uso da interação bem caracterizada entre AvrBs3 e a caixa UPA no promotor BS3. Utilizando mutagênese direcionada por sítio nós substituímos o HVR NI na quinta e sexta unidade de repetição por NK, resultando em AvrBs3-NK_{5/6}. No promotor BS3 tipo selvagem os resíduos NI da quinta e da sexta repetição correspondem a nucleotídeos A. Utilizando mutagênese direcionada por sítio nós substituímos os dois nucleotídeos A no promotor BS3 por dois nucleotídeos C, G e T. O promotor BS3 tipo selvagem e os três mutantes promotores foram fundidos a um gene repórter *uidA* e testados via expressão transitória de *Agrobacterium tumefaciens* em combinação com AvrBs3 ou AvrBs3-NK_{5/6} tipo selvagem em folhas de *Nicotiana benthamiana*. Ensaios GUS revelaram que AvrBs3-NK_{5/6} ativou o repórter GUS somente em combinação com o promotor mutante BS3 “GG” enquanto AvrBs3 ativou somente a construção de promotor tipo selvagem Bs3.

[000208] Nossa análise sugere que NK se liga especificamente a G e, portanto, oferece uma opção para gerar mais arranjos de repetição específicos e também especificamente mirar seqüências alvo rica em G.

Exemplo 3: Método de Geração de Efetores *Designer* via clonagem *Golden gate*.

[000209] O domínio de ligação ao DNA de efetores TAL é composto de 34 unidades de repetição de aminoácido arranjados em tandem (REF). As seqüências de aminoácidos das unidades de repetição são na sua maioria conservadas, exceto por dois resíduos adjacentes altamente variáveis (HVRs) nas posições 12 e 13 que definem a especificidade do DNA alvo ((*Boch et al (2009) Science 326:1509-1512*; *Moscou and Bogdanove (2009) Science 326:1501*)). Grupamentos de HVR diferentes ligam-se com diferentes níveis de especificidade a nucleotídeos, C, G ou T individuais. De modo importante, a análise estatística sugere que unidades de repetição arranjadas em tandem não interferem com a especificidade de unidades adjacentes (*Moscou and Bogdanove (2009) science 326:1501*).

[000210] Assim, a montagem modular de unidades de repetição com especificidades pré-caracterizadas são prováveis em oferecer uma maneira eficiente para a geração de módulos de reconhecimento por DNA com especificidade de DNA desejado.

[000211] No entanto, a geração de construções de DNA que codificam domínios de repetição desejados é um desafio devido ao fato de que as unidades de repetição são quase idênticas. No passado, usamos a síntese química para gerar genes efetores que codificam 17,5 unidades de repetição com a composição de HVR desejada. Para maximizar as diferenças entre as unidades de repetição ao nível do DNA exploramos a degenerescência do código genético. A seqüência otimizada por códon das 17,5 unidades de repetição de codificação de seqüência de DNA foram, em contraste com o gene do tipo selvagem efector TAL correspondente, amplificável por PCR e passíveis de mutagênese baseada em PCR. Nossos resultados também demonstram que a síntese química de domínios de repetição de efector geralmente é viável. No entanto, a síntese química não permite a geração

rápida e com custo-benefício de efetores múltiplos com composição de HVR desejado. Além disso, esta abordagem provavelmente não permitirá a geração de domínios de repetição com 20 ou mais unidades de repetição.

[000212] A recém-desenvolvida clonagem “Golden-Gate” oferece uma abordagem alternativa para a geração de arranjos de unidade de repetição de composição desejada. A estratégia é baseada na utilização de enzimas de restrição tipo IIS, que corta fora de sua seqüência de reconhecimento. Vamos trabalhar com enzima *BsaI* tipo IIS, que cria uma extremidade espessa de 4 bp. Devido ao fato de que sítio de reconhecimento e clivagem são separados em enzimas do tipo IIS podemos gerar, por restrição de *BsaI*, em princípio 256 (4⁴) extremidades espessas diferentes, que fornecem a base para ligações de fragmentos múltiplos. Com desenho próprio dos sítios de clivagem, dois ou mais fragmentos cortados por enzimas de restrição do tipo IIS podem ser ligados em um produto sem o sítio de restrição original (*Engler et al (2008) PLoS ONE 3:e3647; Engler et al (2009) PLoS ONE 4:e5553*).

[000213] No entanto, na prática, existem duas limitações para este método. Devido à atividade da exonuclease em algumas reações, extremidades espessas de DNA pendentes de fita simples são reduzidas de 4 a 3 bases, efetivamente tornando o número de extremidades espessas compatíveis, em apenas 16 (2⁴). Em segundo lugar, a eficiência das reações ligação diminui precipitadamente com um grande número de inserções, tal como seria necessário para criar um efector com 17,5 unidades de repetição como tipicamente encontrados em efetores TAL funcionais que ocorrem naturalmente. Para contornar estas limitações, projetamos um processo de ligação de duas etapas que permite a produção eficaz dos efetores de 20, 30, 40 ou mais unidades de repetição.

[000214] A base para o nosso “kit de construção de arranjo de repetição” é um conjunto de “*plasmídeos de inserção*” que contêm unidades repetidas individuais (uma unidade de repetição por plasmídeo), “*vetores*

intermediários” que contêm domínios de repetição que consistem em conjuntos de 10 unidades de repetição, e um “*receptor de vetor*” que contém as regiões de não-repetição N- e C-terminal de um efector TAL. Todas as unidades de repetição são projetadas de tal forma que os locais de reconhecimento de *BsaI* flanqueiam a inserção nos plasmídeos de inserção.

[000215] Para simplificar a explicação da ligação multi-fragmento nós definimos aqui as extremidades diferentes dos genes de unidade de repetição com letras maiúsculas (em vez da sequência pendurada na extremidade pegajosa) e indicamos sua orientação (N- ou C-terminal da unidade de repetição) com N ou C entre colchetes (por exemplo, A [C]). Um plasmídeo de inserção contendo a primeira unidade de repetição de gene é projetado de tal forma que o tratamento cria a terminação *BsaI* A [N] e B [C]. A segunda unidade de repetição do gene tem terminação B [N] e C [C] com clivagem *BsaI*, enquanto a clivagem *BsaI* do plasmídeo de inserção com a terceira unidade de repetição resulta na terminação C [N] e D [C], e assim por diante. Uma vez que somente as terminações compatíveis podem ser fundidas, a terminação B [C] da primeira unidade de repetição do gene se fundia especificamente com a terminação B [N] da segunda unidade de repetição do gene. Da mesma forma a terminação C [C] da segunda unidade de repetição do gene se ligará especificamente com a terminação C [N] da terceira unidade de repetição do gene e assim por diante.

[000216] A digestão *BsaI* libera as unidades de repetição com quatro saliências bp pegajosa que são compatíveis apenas com as unidades repetidas adjacentes projetadas. O local de reconhecimento *BsaI* por si permanece no vetor do plasmídeo de inserção clivado e a inserção liberada não tem local de reconhecimento *BsaI*. As unidades de repetição são unidas na ordem especificada pelas extremidades salientes em uma reação de corte-laqueadura (ligação e clivagem executados simultaneamente). Devido à ação simultânea de *BsaI* e ligase a religação de unidades de repetição dentro do vetor doador

de inserção é evitada uma vez que esta restaura o local de reconhecimento *BsaI*. Por outro lado os produtos de ligação desejados não possuem os locais de reconhecimento *BsaI*. Este projeto experimental torna este procedimento de clonagem altamente eficiente.

[000217] Para gerar efetores que são projetados para reconhecer as seqüências de bases específicas, quatro variantes são feitas para cada posição da unidade de repetição. Essas variantes são unidades de repetição individuais com especificidade de reconhecimento de nucleotídeos específica, (por exemplo, resíduos de HD na posição 12 e 13 para o reconhecimento de uma base C, NI para A e assim por diante). A variante para cada posição é feita com as extremidades pegajosas apropriadas para cada unidade de repetição, por exemplo, terminais A[N] e B[C] para uma unidade de repetição 1, de modo que há quatro plasmídeos de inserção possíveis para a unidade de repetição um, escolhida com base no reconhecimento de DNA desejado. Existem quatro variantes para a unidade de repetição 2, com especificidade de reconhecimento de nucleotídeos diferentes e terminais B[N] e C[C], e assim por diante para cada posição de repetição.

[000218] Ligações são realizadas em duas etapas. Na primeira etapa, 10 unidades de repetição são combinadas em um *vetor intermediário*. Diferentes conjuntos de 10 unidades de repetição podem ser combinados em vetores intermediários. O vetor intermediário 1 contém unidades de repetição 1 a 10, o vetor intermediário 2 contém unidades repetidas 11 a 20 e assim por diante. Na segunda etapa, 10 unidades de repetição montadas separadamente são combinadas em vetores receptores. O *vetor receptor* também contém as áreas não repetidas de N-e C-terminal do efector, tal que um efector completo composta por 10, 20, 30 40 ou outros múltiplos de 10 unidades de repetição é montado no constructo final. O vetor intermediário tem locais *BsaI* na inserção para a introdução de 10 fragmentos de unidade de repetição e também tem locais *Bpil* de planqueamento na seqüência do vetor de

flanqueamento. BpII é outro tipo de enzima IIS com um local de reconhecimento distinto de BsaI. Usando BsaI, as 10 unidades de repetição são primeiro montadas no “vetor intermediário” e usando o BpII 10mers montados são liberados como um fragmento. Este fragmento é ligado em uma reação de ligase de corte BpII com o vetor receptor, que contém locais BpII entre as áreas de não repetição de N- e C-terminal do efector TAL. Neste caso, apenas 2 a 4 inserções são ligadas no vetor receptor. Isto permite tornar cada ligação altamente específica e montar facilmente 40 e mais unidades de repetição.

[000219] O *vetor receptor* no qual o arranjo da unidade de repetição é finalmente clonado, representa um clone de Entrada GATEWAY e, portanto, permite a transferência a base de recombinação do efector em qualquer constructo de expressão desejado. Atualmente, o *vetor receptor* é projetado para gerar um fator de transcrição do tipo TAL. No entanto, com poucas modificações o *vetor receptor* também permite fusões do arranjo de repetição para a endonuclease *FokI* ou outros domínios funcionais desejados.

[000220] Um esquema deste método é fornecido na Figura 19A a D.

Exemplo 4: Produção e Teste das Nucleases Específicas de DNA Alvo

[000221] As proteínas de fusão compreendendo um domínio de repetição da invenção que reconhece uma sequência de DNA alvo e uma nuclease *FokI* (“nucleases do tipo TAL”) são produzidas como descrito por qualquer uma dos métodos divulgados aqui ou conhecidas na técnica. As proteínas de fusão são testadas para a atividade de nuclease pela incubação com DNA alvo correspondente. O local alvo de DNA do domínio de repetição é clonado em vários locais de clonagem de um vetor de plasmídeo (por exemplo, bluescript). Como controles negativos, seja um “vetor vazio” que não contém nenhum local alvo de nuclease TAL ou locais alvos clonados com mutações são usados. Antes do tratamento do substrato de DNA com a

nuclease do tipo TAL, o vetor é linearizado por tratamento com uma endonuclease padrão adequada que cliva a estrutura principal do vetor. Este vetor linearizado é incubado com proteínas de fusão de nuclease *FokI* de domínio de repetição gerado *in vitro* e os produtos analisados por eletroforese em gel de agarose. A detecção de dois fragmentos de DNA em eletroforese de gel é indicativa para a clivagem mediada por nuclease específica. Por outro lado, os controles negativos que não contêm um local alvo que é reconhecido pelo domínio de repetição não são afetados pelo tratamento com a proteína de fusão de nuclease *FokI* de domínio de repetição. Sistemas acionados por DNA, livres de células para a expressão de gene *in vitro* e síntese de proteínas são usados para gerar as proteínas de fusão da nuclease *FokI* de domínio de repetição (ex., Sistema de Expressão de Proteína de Alto Rendimento T7; Promega). Para utilizar tais sistemas, as seqüências de nucleotídeos da proteína de fusão de nuclease *FokI* são clonadas em frente de uma polimerase RNA T7. Tais proteínas de fusão que são produzidas através da transcrição e tradução *in vitro* são usadas em ensaios de clivagem de DNA sem purificação adicional.

[000222] O artigo “um” e “uma” são usados aqui para se referir a um ou mais de um (ou seja, a pelo menos um) do objeto gramatical do artigo. A título de exemplo, “um elemento” significa um ou mais elementos.

[000223] Ao longo da especificação a palavra “compreendendo”, ou variações como “compreende” ou “composto por”, será entendida como implicando na inclusão de um elemento declarado, número inteiro ou etapa, ou grupo de elementos, números inteiros ou etapas, mas não a exclusão de qualquer outro elemento, número inteiro ou etapa, ou grupo de elementos, números inteiros ou etapas.

[000224] Todas as publicações e pedidos de patentes mencionados na especificação são indicativos do nível daqueles versados na técnica à qual pertence esta invenção. Todas as publicações e pedidos de patentes citados

neste pedido são aqui incorporados por referência, na mesma medida, como se cada publicação individual ou pedido de patente estivesse específica e individualmente indicada para ser incorporada a título de referência. Além disso, cada um dos seguintes pedidos de patentes fica aqui como referência incorporado em sua totalidade: DE 10 2009 004 659.3 depositado em 12 de janeiro de 2009, EP 09165328 depositado em 13 de julho de 2009, e US 61/225.043 depositado em 13 de julho de 2009.

[000225] Embora a invenção anterior tenha sido descrita em algum detalhe a título de ilustração e exemplo para fins de clareza de entendimento, será óbvio que certas mudanças e modificações podem ser praticadas no escopo das reivindicações anexas.

REIVINDICAÇÕES

1. Método *in vitro* para produzir um polipeptídeo que reconhece seletivamente uma sequência de DNA alvo predeterminada, caracterizado pelo fato de compreender sintetizar um polipeptídeo compreendendo um domínio repetido de um efector semelhante a ativador de transcrição (TAL) considerando a sequência de pares de bases presentes na sequência de DNA alvo predeterminada e combinando, na ordem apropriada, unidades de repetição, e em que o domínio repetido compreende unidades de repetição derivadas de um efector TAL, em que as unidades de repetição compreendem regiões hipervariáveis que coletivamente determinam o reconhecimento da sequência de DNA alvo predeterminada, e

em que a região hipervariável compreende um membro selecionado a partir do grupo consistindo em:

- (a) HD para reconhecimento de C/G;
- (b) NI para reconhecimento de A/T;
- (c) NG para reconhecimento de T/A;
- (d) NS para reconhecimento de C/G ou A/T ou T/A ou G/C;
- (e) NN para reconhecimento de G/C ou A/T;
- (f) IG para reconhecimento de T/A;
- (g) N para reconhecimento de C/G;
- (h) HG para reconhecimento de C/G ou T/A; e
- (i) H para reconhecimento de T/A.

2. Método, de acordo com a reivindicação 1, caracterizado pelo fato de que uma região hipervariável corresponde aos aminoácidos 12 e 13 na unidade de repetição.

3. Método, de acordo com a reivindicação 1 ou 2, caracterizado pelo fato de que o domínio repetido compreende 1,5 a 40,5 unidades de repetição.

4. Método, de acordo com a reivindicação 1 ou 2, caracterizado

pelo fato de que o domínio de repetição compreende 11,5 a 33,5 unidades de repetição.

5. Método, de acordo com qualquer uma das reivindicações 1 a 4, caracterizado pelo fato de que um polipeptídeo ainda compreende pelo menos um domínio adicional que é operacionalmente ligado ao domínio repetido.

6. Método, de acordo com a reivindicação 5, caracterizado pelo fato de que um domínio adicional compreende uma proteína bacteriana, viral, fúngica, de oomiceto, humana, animal, vegetal, ou artificial, ou parte da mesma.

7. Método, de acordo com a reivindicação 6, caracterizado pelo fato de que um domínio adicional compreende uma proteína ou parte funcional ou domínio da mesma, que é capaz de modificar o DNA ou RNA.

8. Método, de acordo com a reivindicação 6, caracterizado pelo fato de que um domínio adicional compreende uma proteína ou parte funcional ou domínio da mesma selecionada a partir do grupo consistindo em: um ativador de transcrição, um repressor de transcrição, uma proteína mediadora de resistência, uma nuclease, uma topoisomerase, uma ligase, uma integrase, uma recombinase, uma resolvase, uma metilase, uma acetilase, uma demetilase, e uma deacetilase.

9. Método, de acordo com qualquer uma das reivindicações 1 a 8, caracterizado pelo fato de que o domínio repetido do polipeptídeo é sintetizado pela expressão de uma sequência de DNA que codifica o polipeptídeo e onde a sequência de DNA que codifica o polipeptídeo é montada pelo pré-montagem das unidades de repetição em um ou mais vetores alvo que podem subsequentemente ser montados em um vetor final compreendendo uma sequência de DNA que codifica o polipeptídeo.

10. Método, de acordo com qualquer uma das reivindicações 1 a 9, caracterizado pelo fato de que a unidade de repetição compreende 30 a 40

aminoácidos.

11. Método, de acordo com a reivindicação 10, caracterizado pelo fato de que a unidade de repetição compreende 33, 34, 35 ou 39 aminoácidos.

12. Método, de acordo com qualquer uma das reivindicações 1 a 11, caracterizado pelo fato de que um polipeptídeo reconhece pelo menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, ou 20 pares de base na sequência de DNA alvo predeterminada.

13. Método, de acordo com a reivindicação 12, caracterizado pelo fato de que um polipeptídeo reconhece todos os pares de base na sequência de DNA alvo predeterminada.

14. Método, de acordo com a reivindicação 12, caracterizado pelo fato de que um polipeptídeo é capaz de se ligar à sequência de DNA alvo predeterminada.

15. Método *in vitro* para reconhecer seletivamente uma sequência de DNA alvo predeterminada em uma célula eucariótica por um polipeptídeo, caracterizado pelo fato de compreender introduzir na célula eucariótica um polipeptídeo compreendendo:

(a) um domínio capaz de clivar DNA, e

(b) um domínio repetido de TAL, em que o domínio repetido compreende unidades de repetição derivadas de um efector TAL, em que as unidades de repetição compreendem regiões hipervariáveis que determinam coletivamente o reconhecimento da sequência de DNA alvo predeterminada, em que o domínio repetido é projetado considerando a sequência de pares de base presente na sequência de DNA alvo predeterminada e combinando, na ordem apropriada, unidades de repetição; e

em que o polipeptídeo reconhece especificamente a sequência de DNA alvo, e

em que a região hipervariável compreende um membro

selecionado a partir do grupo consistindo em:

- (a) HD para reconhecimento de C/G;
- (b) NI para reconhecimento de A/T;
- (c) NG para reconhecimento de T/A;
- (d) NS para reconhecimento de C/G ou A/T ou T/A ou G/C;
- (e) NN para reconhecimento de G/C ou A/T;
- (f) IG para reconhecimento de T/A;
- (g) N para reconhecimento de C/G;
- (h) HG para reconhecimento de C/G ou T/A; e
- (i) H para reconhecimento de T/A.

16. Método, de acordo com a reivindicação 15, caracterizado pelo fato de que uma região hipervariável corresponde aos aminoácidos 12 e 13 na unidade de repetição.

17. Método, de acordo a reivindicação 15 ou 16, caracterizado pelo fato de que o domínio repetido compreende 1,5 a 40,5 unidades de repetição.

18. Método, de acordo com a reivindicação 15 ou 16, caracterizado pelo fato de que o domínio repetido compreende 11,5 a 33,5 unidades de repetição.

19. Método, de acordo com qualquer uma das reivindicações 15 a 18, caracterizado pelo fato de que a unidade de repetição compreende 30 a 40 aminoácidos.

20. Método, de acordo com a reivindicação 19, caracterizado pelo fato de que a unidade de repetição compreende 33, 34, 35 ou 39 aminoácidos.

21. Método, de acordo com qualquer uma das reivindicações 15 a 20, caracterizado pelo fato de que uma região N-terminal de um domínio de repetição confere um reconhecimento de especificidade para um T/A 5' da especificidade de reconhecimento da unidade de repetição.

22. Método para modular expressão de um gene alvo em uma célula *in vitro*, caracterizado pelo fato de que são fornecidas células que contém um polipeptídeo, em que o polipeptídeo compreende um domínio repetido de efector TAL, em que o domínio repetido compreende unidades de repetição derivadas de um efector TAL, em que as unidades de repetição compreendem regiões hipervariáveis que determinam coletivamente o reconhecimento de uma sequência de DNA predeterminada, em que o domínio repetido é projetado considerando a sequência de pares de base presente na sequência de DNA alvo predeterminada e combinando, na ordem apropriada, unidades de repetição;

em que a célula é humana ou animal, e

em que a região hipervariável compreende um membro selecionado a partir do grupo consistindo em:

- (a) HD para reconhecimento de C/G;
- (b) NI para reconhecimento de A/T;
- (c) NG para reconhecimento de T/A;
- (d) NS para reconhecimento de C/G ou A/T ou T/A ou G/C;
- (e) NN para reconhecimento de G/C ou A/T;
- (f) IG para reconhecimento de T/A;
- (g) N para reconhecimento de C/G;
- (h) HG para reconhecimento de C/G ou T/A; e
- (i) H para reconhecimento de T/A.

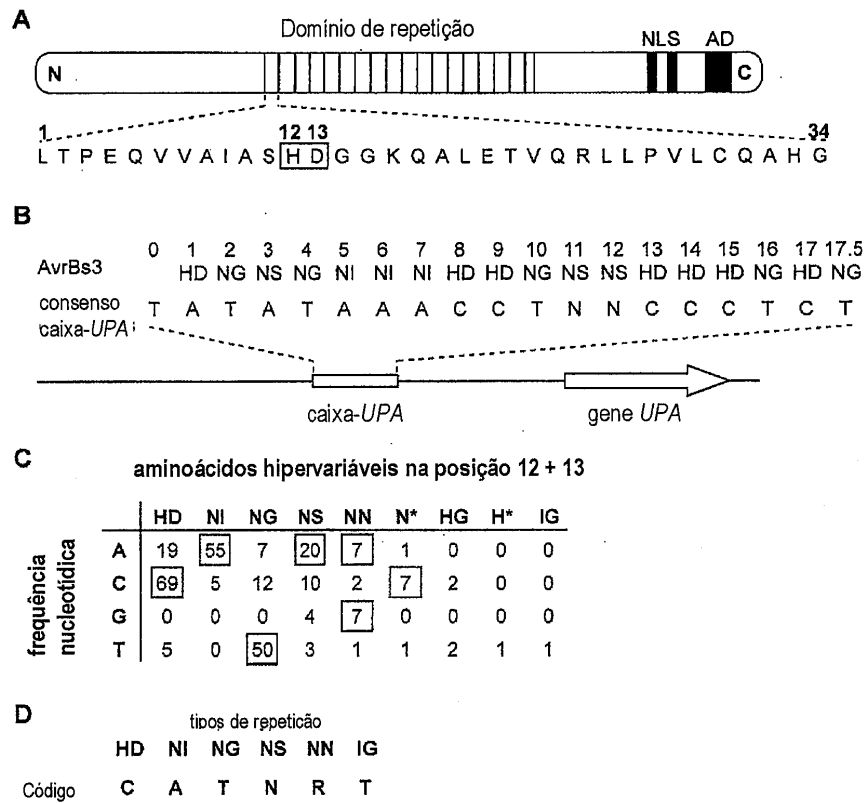


Figura 1

A

Hax2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	21.5
		NN	IG	NG	NI	NG	NG	HD	NG	HD	NI	HD	NI	HD	NG	HD	NG	HD	HD	NG	NG	NI	NG
caixa-Hax2	T	G	T	T	A	T	T	C	T	C	A	C	A	C	T	C	T	C	C	T	T	A	T
Hax3	0	1	2	3	4	5	6	7	8	9	10	11	11.5										
		NI	HD	NI	HD	HD	HD	NS	NS	NS	HD	NI	NG										
caixa-Hax3	T	A	C	A	C	C	C	A	A	A	C	A	T										
Hax4	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	14.5							
		NI	HD	HD	NG	NS	NS	NI	HD	NG	NI	NS	NI	NG	NI	NG							
caixa-Hax4	T	A	C	C	T	A	A	A	C	T	A	A	A	T	A	T							

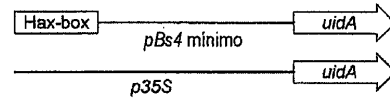
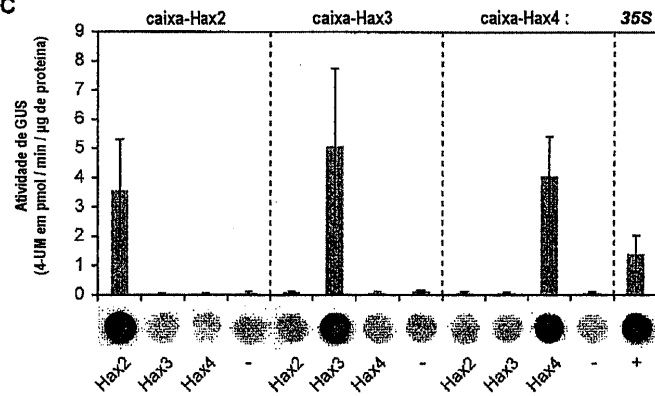
B**C**

Figura 2

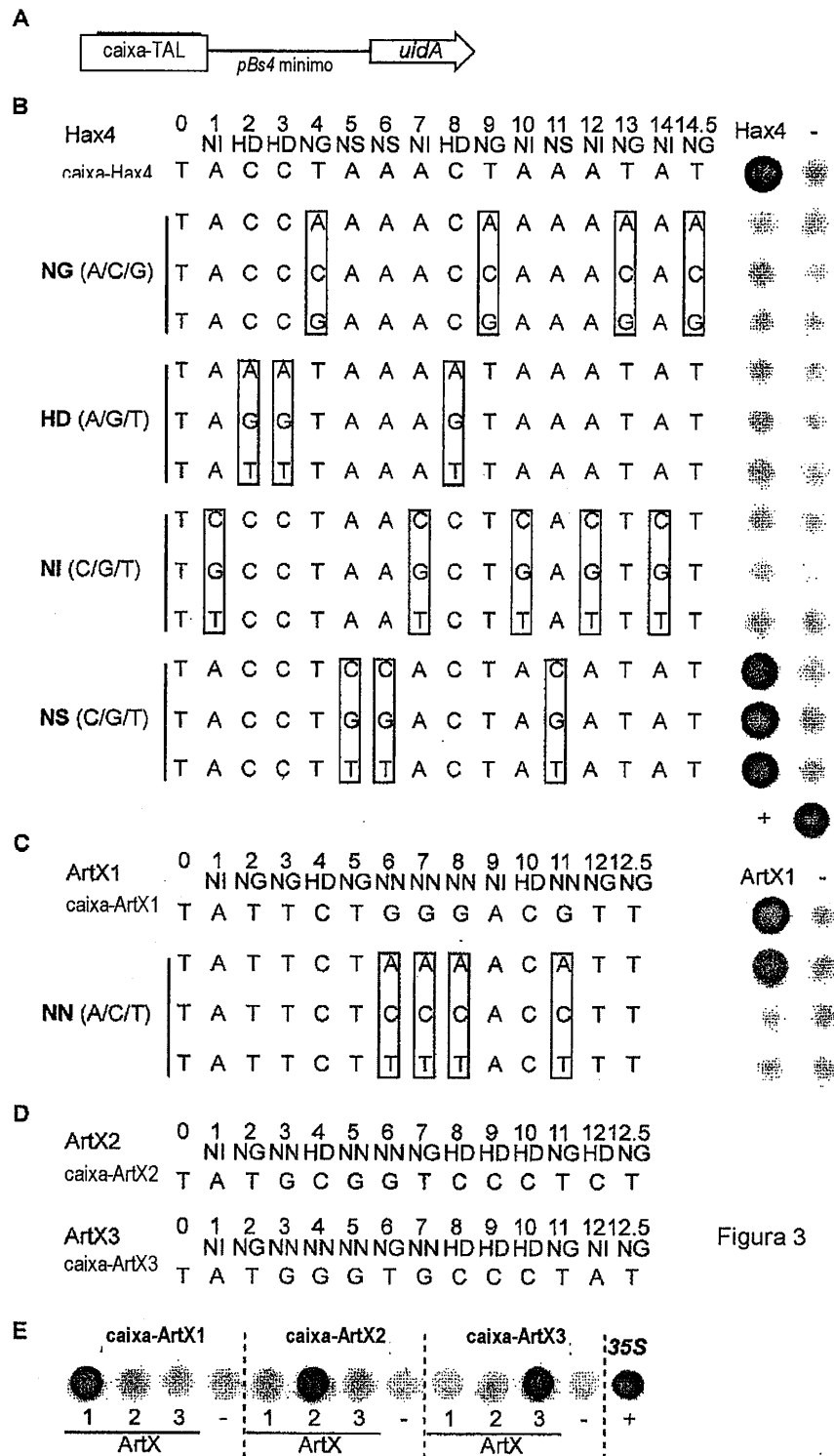


Figura 3

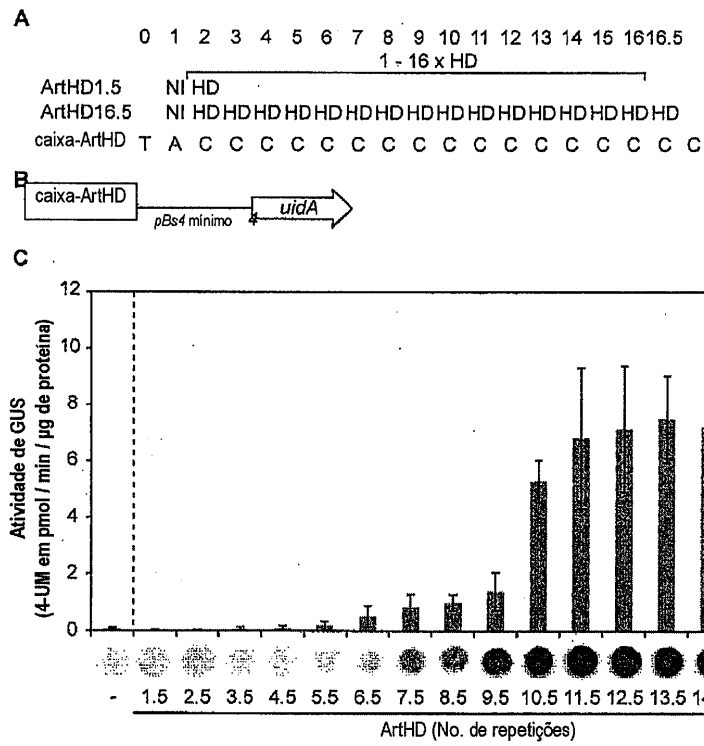


Figura 4

A	caixa-UPA Bs3...	-120	T A T A T A A A C C C T A A C C A T C C	-102	HR (ECW-30R)
	AvrBs3	0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 17.5		+
			HDNGNSNG NI NI NI HDHDNGNSNSHDHDHDNGHDNG		
	AvrBs3Δrep16	0	1 2 3 4 5 6 7 8 9 10 11 12 13 13.5		-
			HDNGNSNG NI NI NI HDHDNGHDNGHDNG		
B	caixa semelhante à UPA Bs3-E	-133	T A T A T A A A C C C T C T C T A T T C	-115	HR (ECW)
	AvrBs3	0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 17.5		-
			HDNGNSNG NI NI NI HDHDNGNSNSHDHDHDNGHDNG		
	AvrBs3Δrep16	0	1 2 3 4 5 6 7 8 9 10 11 12 13 13.5		+
			HDNGNSNG NI NI NI HDHDNGHDNGHDNG		
C	AvrBs3Δrep109	0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 14.5		-
			HDNGNSNG NI NI NI HDHDNGNSNSNGHDNG		
	AvrHah1	0	1 2 3 4 5 6 7 8 9 10 11 12 13 13.5		-
			NN IG NI NI NI HDHDNGNN NI HDHDHDNG		
D	Xa27	-87	T A G A A G A A G A G A C C C A T A	-70	HR
					+
	xa27	-84	T A G A A G A G A C C A A T A G A G	-67	-
	AvrXa27	0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 16.5		
E			NI NN N* NGNSNNNNNN NI NN NI N* HDHD NI NGNG		
F	Xa13	-251	T G C A T C T C C C C C T A C T G T A C A C C A C	-227	Indução gênica
		-223		-199	+
	xa13	0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 23.5		-
			NNHD NI HGHDNGN* HDHD NI NGNG NI HDNGNNNG NI NI NI NI N* NS N*		
	PthXo1				
G	OsTFX1	-136	T A T A A A A G G C C C T C A C C A A C C C A T	-113	+
	PthXo6	0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 22.5		
			NI H* NI NNNNNNNNNNNHD NI HDHGHD NI N* NS NI NI HGHDNSNSNG		
H	OsTFIIAγ1	-469	T A T A A T C C C C A A A T C C C C T C C T C	-447	+
	PthXo7	0	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 21.5		
			NI NG NI NI N* NNHDHDN* NI NI NI NGHDHGNNNSNNHDHDNG N*		

Figura 5

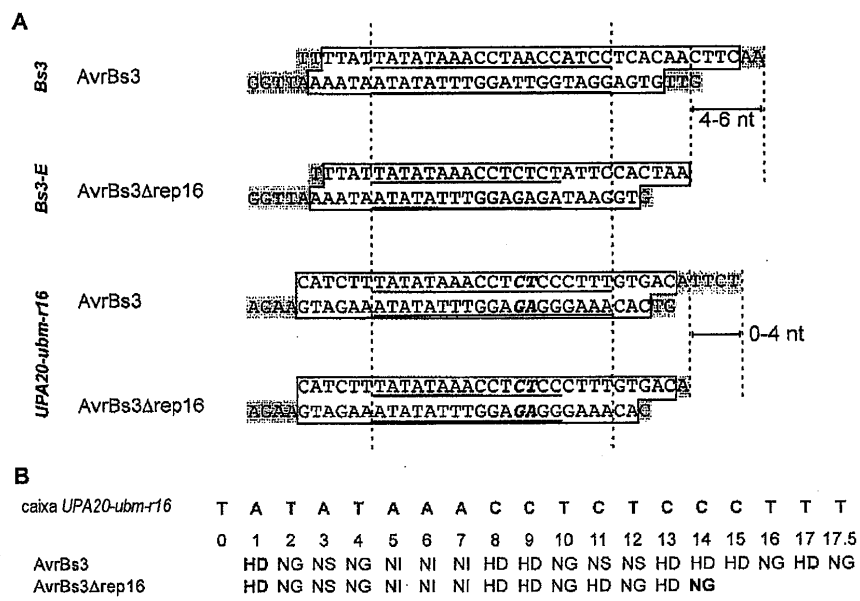


Figura 6

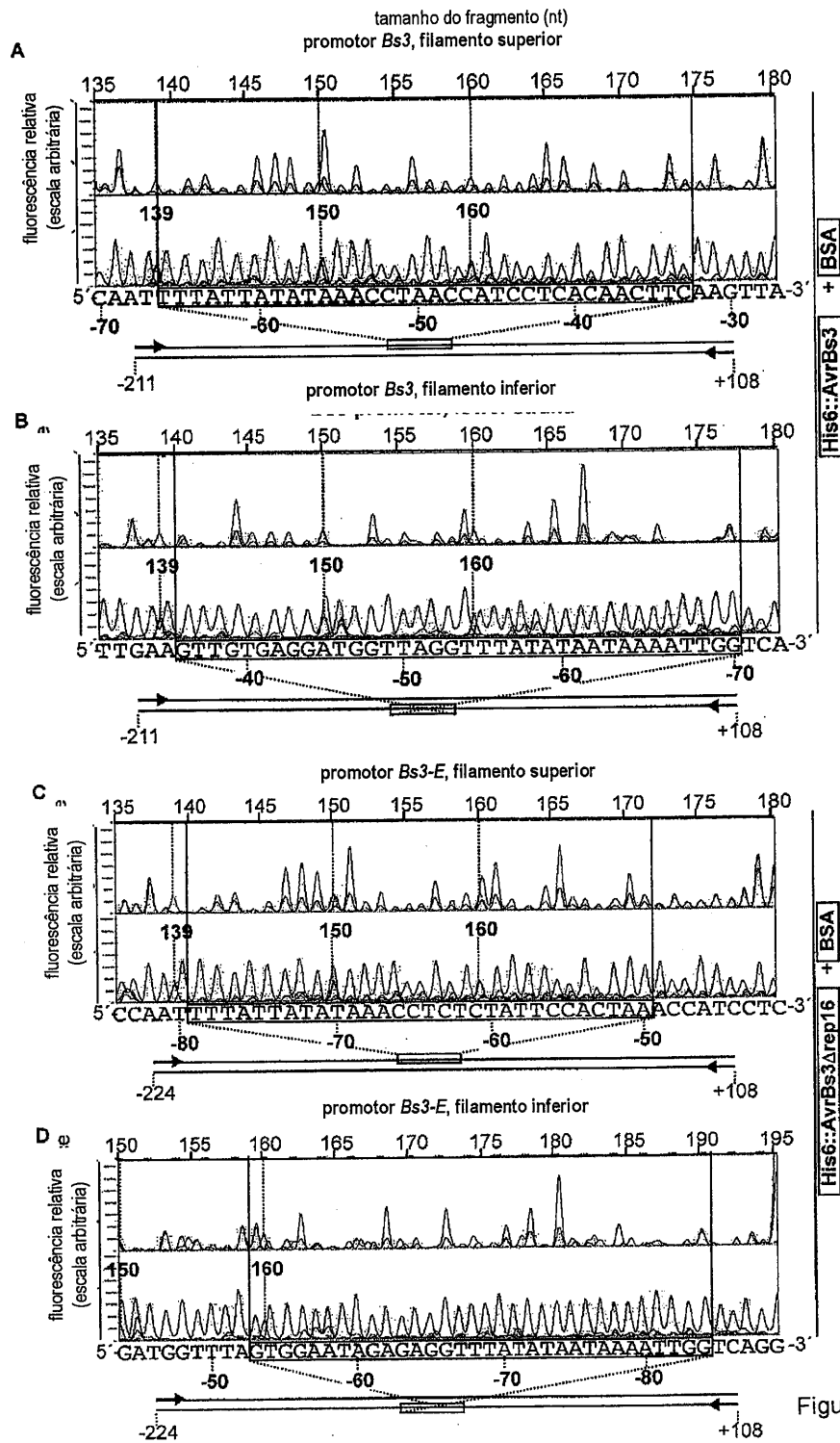


Figura 7

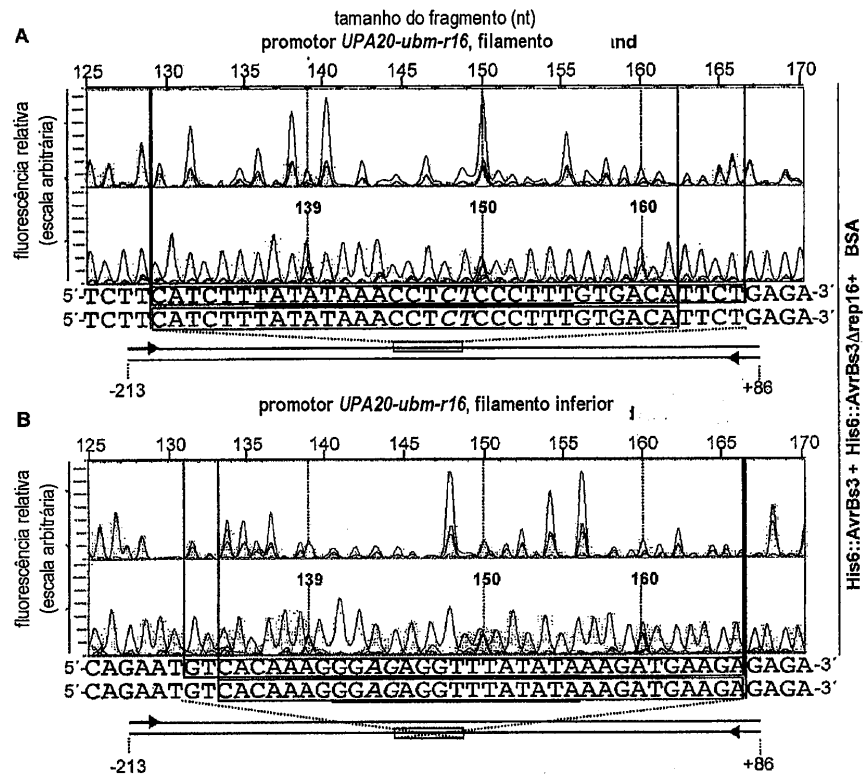


Figura 8

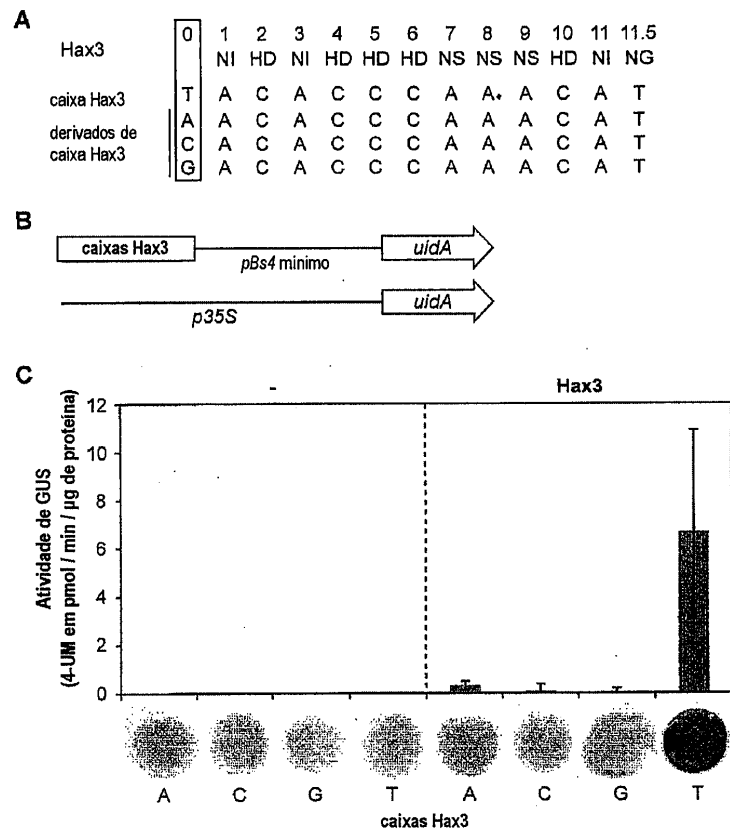


Figura 10

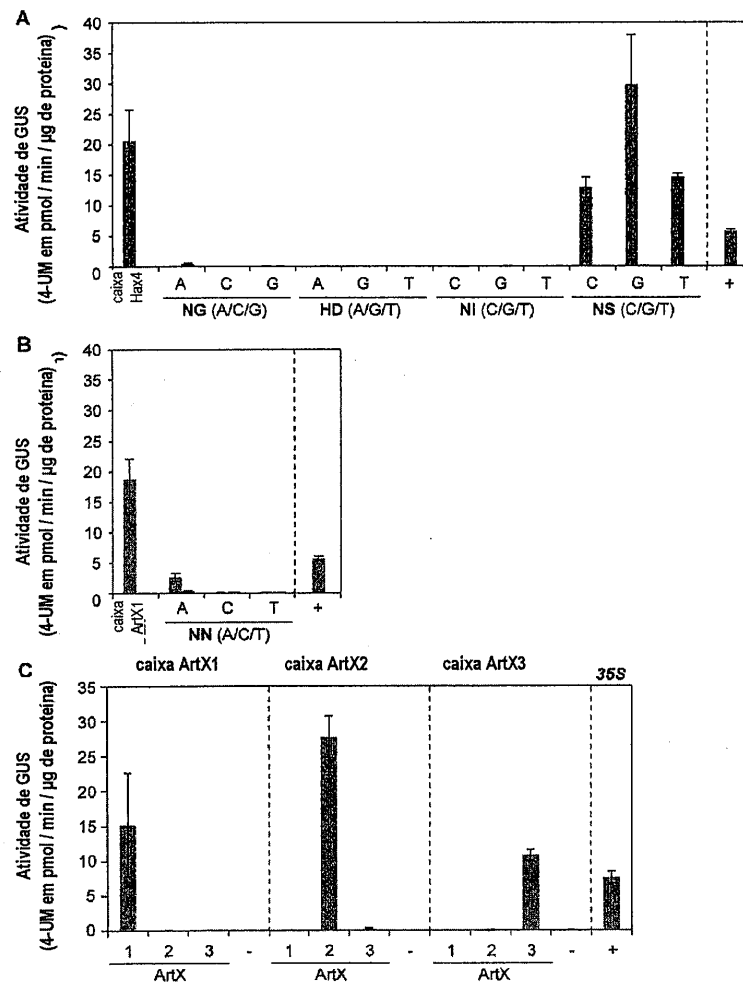


FIGURA 11

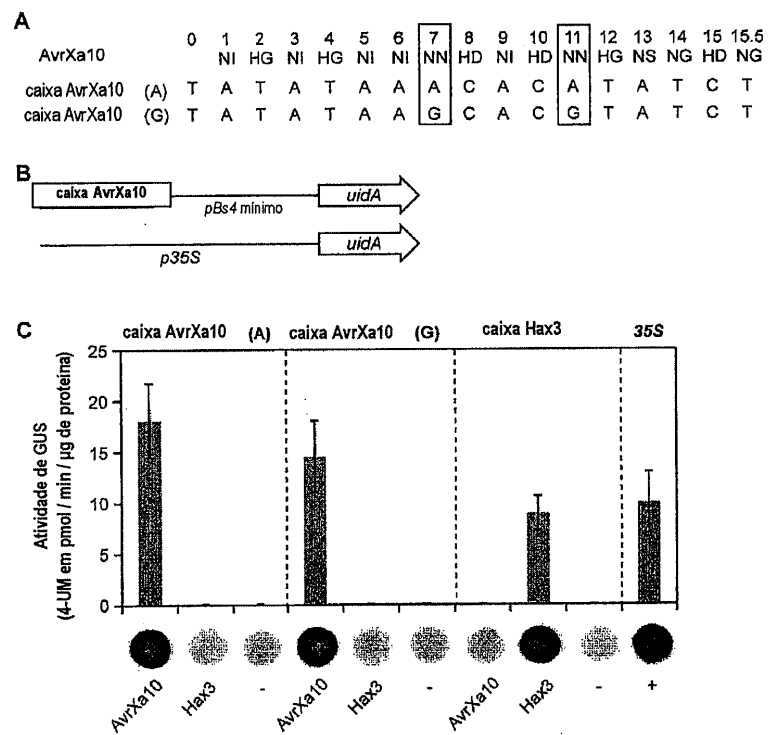


FIGURA 12

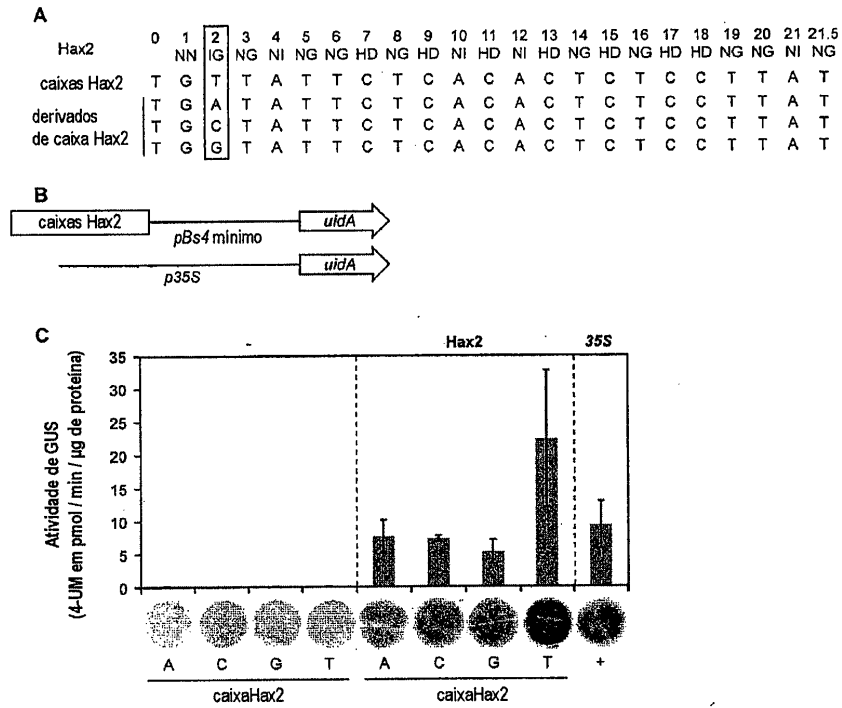


FIGURA 13

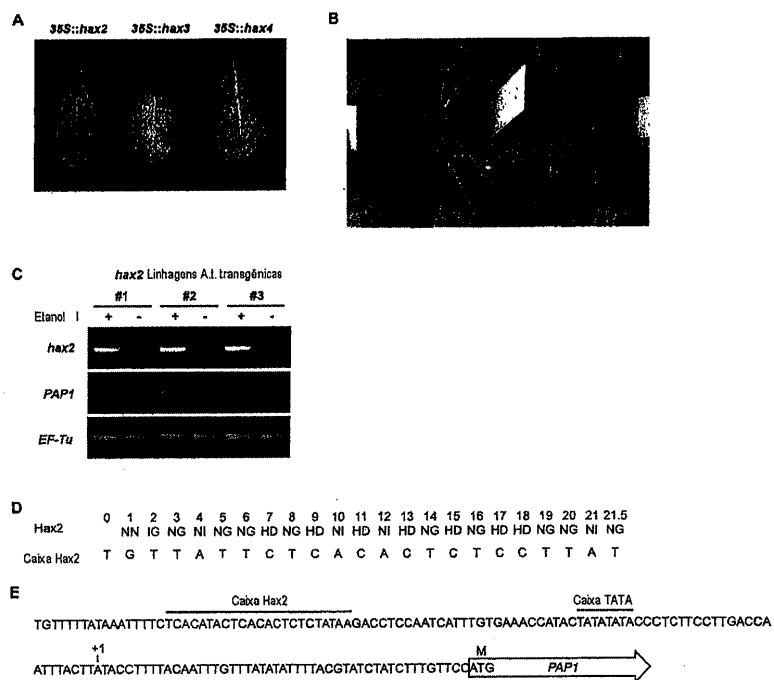


FIGURA 14

Figura 15 A

Tabela 1: Sequências alvo de DNA previstas de efetores TAL

Organismo	Efeitor TAL	Repetições ^b (aminoácidos 12 + 13)	Sequências de DNA previstas ^c	Acesso
Efetores TAL com especificidade mostrada (este trabalho e outros)				
Xcv 71-21	AvrBs3	HD-NG-NS-NG-NI-NI-HD-HD-NG-NS-NS-HD-HD-HD-NG-HD-NG	TCNTAAACCTNNCCCTCT	CAA34257
Xg XV444	AvrFfh1	NN-HG-NI-NI-NI-HD-HD-NG-NN-NI-HD-HD-HD-NG	TRTAAACCTTACCCCT	ABP97430
Xcv 5	Hsx2	NN-HG-NG-NI-NG-NG-HD-NG-HD-NI-HD-NI-HD-NG-HD-NG-HD-NG-NG-NI-NG	TGTTATCTCAGACTCTCCTTAT	AAV43358
Xcv 5	Hsx3	NI-HD-NI-HD-HD-HD-NS-NS-HD-NI-NG	TACACCCNNNCAT	AAV43359
Xcv 5	Hsx4	NI-HD-HD-NG-NS-NS-NI-HD-NG-NI-NS-NI-NG-NI-NG	TACCTNACTANATAT	AAV43360
Xoo PXO86	AvrXa10	NI-HG-NI-HG-NI-NI-NN-HD-NI-HD-NN-HG-NS-NG-HD-NG	TA*AAACACAC*NTCT	AAA92974
Xoo PXO99A	AvrXa27	NI-NN-N*NG-NS-NN-NN-NN-NI-NN-NI-N*HD-HD-NI-NG-NG	TAR*TNRRRARA*CCATT	AAV54168
Xoo PXO99A	PthXo1	NN-HD-NI-HG-HD-NG-N*HD-HD-NI-NG-NG-NI-HD-NG-NN-NG-NI-NI-NI-NI-N*NS-N*	TRCA*CT*CCATTACTRTAAAA*N*	AA546025
Xoo PXO99A	PthXo6	NI-H*NN-NN-NN-NN-NN-NN-HD-NI-HD-HG-HD-NI-N*NS-NI-NI-HG-HD-NS-NS-NG	TA*ARRRRRCAC*CA*AA*CNNT	ABB70183
Xoo PXO99A	PthXo7	NI-NG-NI-NI-N*NN-HD-HD-N*NI-NI-NI-NG-HD-HG-NN-NS-NN-HD-HD-NG-N*	TATAA*RCO*AAATC*RNROCT*	ABB70129
Efetores TAL com especificidade até agora desconhecida				
Xcv 82-8	AvrBs4	NI-NG-NI-NI-NG-NG-NI-NS-NG-NI-NS-NG-HD-HD-NS-HD-NG-NG	TATAATANTANTCCNCTT	CAA48680
Xac NA-1	Ap11	NI-NG-NI-NI-NI-HD-HD-NG-HD-NG-NG-NG-NS-HD-HD-NG-NG	TATAAACCTCTTTTNCCTT	BAA37119
Xac NA-1	Ap12	NI-NG-NI-HD-NI-HD-HD-NG-NG-NG-NG-NG-NI-HD-NG	TATACACCTCTTTTACT	BAA37120
Xac NA-1	Ap13	NI-HD-NI-HD-NI-HD-HD-NG-HD-NG-NI-HD-HD-NI-HD-HD-NG-HD-NG-NI-HD-NG-NG	TACACACCTCTCAACCTCTACTT	BAA37121
Xac 3213	PthA	NI-NG-NI-NS-NI-NG-HD-NG-NG-NG-NG-NG-NG-NS-HD-HD-NG-NG	TATAAATCTCTTTTNCCTT	AAC43587
Xac	PthB	HD-NG-HD-NG-NI-NG-HD-NG-HD-NI-NI-HD-HD-HD-HD-NG-NG-NG	TCCTATCTCAACCCCTTT	AAO72098
Xac C340	PthC	HD-NG-HD-HD-NI-NG-NI-NG-NI-NI-HD-NG-HD-HD-HD-NG-NG-NG	TCCTATATACTCCCTTT	ABO77782
Xac KC21	HsxB3.0	NI-HD-NI-HD-NI-NG-N*NI-N*NI-HD-HD-NI-HD-NG	TACACAT*AAACACT	BAF46269
Xcm	AvrB6	HD-NI-NG-HD-HD-NI-HD-NI-NS-HD-HD-HD-NN-NG	TCATCCACANCCCT	AA800675
Xcm XcmN	PthN	NI-HD-HD-NI-HD-NI-NG-NI-NN-HD-NI-NG-NG-NN	TACCACATARCATT	AA869865
Xam CFBP1851	PTHB	NI-NI-NN-NI-HD-NS-NS-NN-NG-HD-NN-NI-NG	TAARACNNRTCRAT	AAD81494
Xoo IXOH1	AvrXa3	HD-HD-HD-NG-NG-NN-HD-HD-NG	TCCCTTTCCT	AAND1357
Xoo PXO2684	AvrXa7	NI-HG-NI-NI-NS-HD-NN-HD-HD-HD-NS-N*NI-HD-HD-NS-NS-NN-NN-NI-NG-NN-NI-N*NS-N*	TA*AAACGCCCN*CCNNRRATRA*N*	AAF98343
Xoo	PthA1	NN-HD-NS-NG-HD-NN-NG-NI-HD-NS-	TRCNTCKTACNCRCRCRRRRURCT	YP_200770

Figura 15 B

KACC10331		HD-NN-HD-NN-HD-NN-NN-NN-NN-NN-NN-NN NN-NN-NN-HD-NG		NN-NN-NN-NN-NN-NN	
Xac 306	PdbA1	NI-N*-NI-N*-NI-HD-HD-N*-NI-HD-NI- HD-N*-NI-HD-HD-NG	TAA*AACCACACACCTT	NP_644708	
Xac KACC10331	PdbA2	NI-NO-NN-NO-NK-NG-NI-NN-NI-NN- NI-NN-NS-NG-NS-NN-NI-NO-NS-NG	TATRTARARAARNTRNTNT	YP_201652	
Xac 306	PdbA2	NI-HD-NI-HD-NI-HD-ND-NO-HG-ND- NO-NG-NG-NI-NI-NG	TACACACCTCTTTTAAT	NP_644725	
Xac KACC10331	PdbA3	NI-HO-NI-NI-NS-HD-NN-HD-HD-ND- HS-NG-N*-HD-HD-NS-NS-NN-NG	TA-AANRCOCENTOCNNT	YP_201654	
Xac 306	PdbA3	NI-HD-NI-HD-NI-NO-HD-NG-NO-NG- NI-NI-NI-HD-NG	TACACATCTTTAAAAGC	NP_644743	
Xac KACC10331	PdbA4	NI-N*-NI-NS-NN-NG-NN-NS-N*-NS- NN-NS-N*-HD-HG-HD-HD-NS-N*- HD-HG-HD-HD-HD-NG	TA-ANRTRN-NRN-O-COO-N-O-COCTT	YP_200914	
Xac 306	PdbA4	NI-N*-NI-NT-HD-ND-NO-HD-NG- NG-XG-NG-NS-HD-HD-NG-NG	TA-AAACCTCTTTTCCTT	NP_644793	
Xac JX01	PdbXo2	NI-HO-MI-NN-NN-NN-NN-HDI-HD- NS-NS-NS-NS-HD-ND-NG-HD-HD- HG-NG-NG	TA-ARRARCACHNNCRCTCCCTT	AAS46028	
Xac PXO61	PdbXo3	NI-HD-NI-HQ-NI-NI-NI-HD-NN-HD- FD-HD-NG-HD-NO-NI-HD-HD-NN- NS-NI-NN-NN-NG-NN-HD-N*-NS-N*	TA-A*AAACGCCCTTACCRRNRTRON+	AA546027	
Xac PXO99A	TalZa	NI-NG-NN-NG-NK-NG-NI-NN-NI-NN- NI-HD-N*-NS-NG	TATRTARARAOONT	YP_001912778	
Xac PXO99A	TalH	NI-NN-NN-NI-NI-NI-HD-NS-HO-NN- NN-NN-NI-HO-HD	TARRRAACNNRRRAAC	YP_001913182	
Xac PXO99A	TalSa	NI-NS-HD-FG-NS-NN-HD-S*-NG-NN- NN-HD-HD-NG-HD-NG	TANC*NRO*TRRCTCTT	YP_001913457	
Xac PXO99A	TalSa	NI-N*-NI-NS-NN-NG-NN-NS-N*-NS- NN-NS-N*-NI-HG-HD-NI-HD-HD-NG	TA-ANRTRGNRN-N*-CACCT	YP_001913480	
Xac PXO99A	TalTs	NI-HG-NI-NI-NI-NI-HD-NS-NN-NS- NN-HD-NN-NI-HD-NS-SY-G	TA*AAACRNBNRCRACRNT	YP_001913735	
Xac PXO99A	TalTb	NI-HG-NS-HG-HG-HD-NS-NO-HD-NN- NG-HG-NO-HD-HG-HD-HD-NI-NN-NG	TA-N*NCNKTC*TC*CAART	YP_001913759	
Xac PXO99A	TalTa	NI-HG-NI-NI-NN-HD-NS-NN-NS- NN-HD-NN-NI-HD-NN-NS-NG	TA-AAARCNBNNRCRACRNT	YP_001913948	
Xac PXO99A	TalTb	NI-HG-NS-HG-HG-HD-NS-NG-HD-NN- NG-HG-NG-HD-HG-HD-HD-NI-NN-NG	TA-N*NCNKTC*TC*CAART	YP_001913952	
Xac PXO99A	TalTa	HD-HD-HD-NG-N*-NM-HD-HD-N*-NI- NI-NN-HD-HI-ND-HD-NI-HD-NG-NG	TCCTT*CCC*AARCCACACTT	YP_001915089	
Xac PXO99A	TalTc	HD-HD-NN-NN-NG-NG-HD-NS-HG- HD-NI-N*HD-HD-HD-N*-NN-NI-NN- HD-HI-ND-HD-NG-NN-HG-NG	TCCRRTTONCTCCCARCORCT	YP_001915091	
Xac PXO99A	TalTd	NLANN-NI-HG-HG-NN-HG-HD-HG-HD- HD-HD-NG	TARA-N*COOCTT	YP_001915101	
Xac PXO99A	TalTe	NN-HD-NS-NG-HD-ARN-N*-NI-HD-NS- HD-NN-HD-NN-HD-NN-NN-NN-NN- NN-NN-NN-HD-NG	TRCNTCRACNCRCKRRRRRKCT	YP_001915105	
Xac C8	PdCBa	NN-HD-HD-HD-NI-NI-NN-NI-HD-HD- HD-HG-NN-NN-ID-NS-NN-HD-J*- NS-N*	TRCCCAAARCCORRCONON+	ABY60853	
Xac C8	PdCBb	NI-HG-NI-NI-NN-HD-NS-NN-NS- NN-HD-NI-NI-HD-NN-NS-NG	TA-AAARCNBNNRCRACRNT	ACB11564	
Xac	AvrPdh3	NS-HD-NI-AG-NI-HG-NG-HD-NN-NI- NG-HD-NN-NS-NG-NN	TNCATATTTCRATGRNRTR	AAW59492	
Xac	AvrPdh14	NS-HD-NI-NG-NI-HO-NI-MI-NG-NG- HD-NN-NI-NG-HD-NN-NS-NN-NN-NN	TNCATATAATTTCRATCRNRTR	AAW59493	
Xac 12879	avrTw	NI-NG-NI-NI-HD-HD-HD-HD-NG-HD- NG-NS-NI-HD-NG-NI-NS-NG	TATAACACCOCTTACATANT	ACN39605	

AvrBs3

Sequência de proteína:

MDPIRSRTSPARELLPGPQPDGVQPTADRGVSPAGGPLDGLPARRTMSRTRLPSPPAPSPAFSAGSFSDLLRQF
 DPSLENTSLFDSLPPFGAHTTEAATGEWDEVQSLRAADAPPTMRVAVTAARP PRAKPAARRRAAQPSDASPAQ
 VDLRTLGYSQQQKEIKPKVRSSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEATHEAIVGV
 GKQWSGARALEALLTVAGELRGPPQLDTGQLLKIARGGVTAVEAVHAWRNALTGAPLNLTPQQVVAIASHDGKK
 QALETVQRLLPVLCQAHGLTPQQVVAIASNGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNSGGKQALETVQR
 LLPVLCQAHGLTPQQVVAIASNGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNIGGKQALETVQALLPVLCQA
 HGLTPQQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQALLPVLCQAHGLTPQQV
 VAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASHDGKKQALETVQRLLPVLCQAHGLTPQQVVAIASNGG
 GKQALETVQRLLPVLCQAHGLTPQQVVAIASNSGGKQALETVQALLPVLCQAHGLTPQQVVAIASNSGGKQALETV
 QRLLPVLCQAHGLTPQQVVAIASHDGKKQALETVQRLLPVLCQAHGLTPQQVVAIASHDGKKQALETVQRLLPVLC
 QAHGLTPQQVVAIASHDGKKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGKQALETVQRLLPVLCQAHGLTP
 QVVAIASHDGKKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGKQALETVQRLLPVLCQAHGLTPQQVVAIAS
 LGGREPALDAVKGLPHAPALIKRTNRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSRHGLLQL
 FRRVGVTLEARSGLTPPASQRWDRIQASGMKRAKPSPTSTQTPDQASLHAFADSLERDLAPSPMHEGDQTRAS
 SKRSRSDRAVTGPSAQQSFEVRVPEQRDALHLPLSWRVKRPRTSIGGGLPDPGPTPTAADLAASSTVMREQDEDPF
 AGAADDFPAFNEELAWLMELLQ

Acesso:

CAA34257.1

Organismo:

Xanthomonas campestris pv. vesicatoria Cepa 71-21

Description of the AvrBs3 protein sequence:

N-Terminus: amino acid (AA) 1-288

MDPIRSRTSPARELLPGPQPDGVQPTADRGVSPAGGPLDGLPARRTMSRTRLPSPPAPSPAFSAGSFSDLLRQF
 DPSLENTSLFDSLPPFGAHTTEAATGEWDEVQSLRAADAPPTMRVAVTAARP PRAKPAARRRAAQPSDASPAQ
 VDLRTLGYSQQQKEIKPKVRSSTVAQHHEALVGHGFTHAHIVALSQHPAALGTAVVKYQDMIAALPEATHEAIVGV
 GKQWSGARALEALLTVAGELRGPPQLDTGQLLKIARGGVTAVEAVHAWRNALTGAPLN

Repetição 1: AA 289-322

LTPEQVVAIASHDGKKQALETVQRLLPVLCQAHG

Repetição 2: AA 323-356

LTPQQVVAIASNGGKQALETVQRLLPVLCQAHG

Repetição 3: AA 357-390

LTPQQVVAIASNSGGKQALETVQRLLPVLCQAHG

Repetição 4: AA 391-424

LTPEQVVAIASNGGKQALETVQRLLPVLCQAHG

Repetição 5: AA 425-458

LTPEQVVAIASNIGGKQALETVQALLPVLCQAHG

Repetição 6: AA 459-492

LTPEQVVAIASNIGGKQALETVQALLPVLCQAHG

Repetição 7: AA 493-526

LTPEQVVAIASNIGGKQALETVQALLPVLCQAHG

Repetição 8: AA 527-560

LTPEQVVAIASHDGKKQALETVQRLLPVLCQAHG

Repetição 9: AA 561-594

LTPEQVVAIASHDGKKQALETVQRLLPVLCQAHG

Repetição 10: AA 595-628

LTPQQVVAIASNGGKQALETVQRLLPVLCQAHG

Repetição 11: AA 629-662

LTPEQVVAIASNSGGKQALETVQALLPVLCQAHG

FIGURA 16A

Repetição 12: AA 663-696
 LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG

Repetição 13: AA 697-730
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

Repetição 14: AA 731-764
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

Repetição 15: AA 765-798
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

Repetição 16: AA 799-832
 LTPQQVVAIASNGGGRPALETQVRLLPVLCQAHG

Repetição 17: AA 833-866
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

Repetição 17.5: AA 867-886
 LTPQQVVAIASNGGGRPALE

C-terminal: AA 887-1164
 SIVAQLSRPDPALAAITNDHIVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQ
 CHSHPAQAFDDAMTQFGMSRHGLQLFRRVGVTELEARSGLTPPASQRWDRILOASGMKRAKPSPTSTQTPDQASL
 HAFADSLERDLADAPSPMHEGDQTRASSRKRSRSDRAVTGPSAQOSFEVRVPEQRDALHLPLSWRVKRPRTSIGGGL
 PDPGTPTAADLAASSTVMREQDEDPFAGAADFPFANEEELAWLMELLPQ

Hax2

Sequência de proteína:

MDPIRSRTPSPARELLSGPQPDGVQPTADRGVSPAGGPLDGLPARRTMSRTRLSPSPAPSPAFSADSFS
 DLLRQFDPSLFTNTSLFDSLPPFGAHHTAATGEWDEVQSGLRADAPPPTMRVAVTAARPPRAKPAFRRR
 AAQPSDASPAQVDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTTHAHIVALSQHPAALGTAVVKY
 QDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPQLDGTGQLLKIARGGVTAVEAVHAWRN
 ALTGAPLNLTPQVVAIASNNGGKQALETVQRLLPVLCQAPHDLTPQVVAIASIGGKQALETVQRLLP
 VLCQAPHDLTPQVVAIASNNGGKQALETVQRLLPVLCQAPHCLTPQVVAIASNIGGKQALETVQALLP
 VLCQAPHCLTPQVVAIASNNGGKQALETVQRLLPVLCQAPHDLTPQVVAIASNNGGKQALETVQRLLP
 VLCQAPHDLTPQVVAIASHDGGKQALETVQRLLPVLCQAPHDLTPQVVAIASNNGGKQALETVQRLLP
 VLCQAPHDLTPQVVAIASHDGGKQALETVQRLLPVLCQAPHDLTPQVVAIASNIGGKQALETVQALLP
 VLCQAPHCLTPQVVAIASHDGGKQALETVQALLPVLCQAPHDLTPQVVAIASNIGGKQALETVQRLLP
 VLCQAPHDLTPQVVAIASHDGGKQALETVQRLLPVLCQAPHDLTPQVVAIASNNGGKQALETVQRLLP
 VLCQAPHDLTPQVVAIASHDGGKQALETVQRLLPVLCQAPHDLTPQVVAIASNNGGKQALETVQRLLP
 VLCQAPHDLTPQVVAIASHDGGKQALETVQRLLPVLCQAPHDLTPQVVAIASHDGGKQALETVQRLLP
 VLCQAPHDLTPQVVAIASNNGGKQALETVQRLLPVLCQAPHDLTPQVVAIASNNGGKQALETVQALLP
 VLCQAPHDLTPQVVAIASNIGGKQALETVQRLLPVLCQAPHDLTPQVVAIASNNGGKQALESI FAQLS
 RPDPAALAAITNDRLVALACIGGRSALNAVKGLEPNALTLIRANSRIPERTSHLVADHTQVVRVLGFFQC
 HSHPAQAFDEAMTQFGMSRHGLQLFRRVGVTELEARSGLTPPASQRWDRILOASGMKRAKPSPTSTQTP
 DQASLHAFADSLERDLADAPSPMHEGDQTRASSRKRSRSDRAVTGPSAQOSFEVRVPEQRDALHLPLLSWG
 VKRPRTRIGGLDLPDGPMDADLVASSTVWVEQDADPFAGTADDFPAFNEELAWLMELLPH

Acesso:

AA43358.1

Organismo:

Xanthomonas campestris pv. armoraciae Cepa 5

Descrição da sequência de proteína Hax2:

N-terminal: aminoácido (AA) 1-288

MDPIRSRTPSPARELLSGPQPDGVQPTADRGVSPAGGPLDGLPARRTMSRTRLSPSPAPSPAFSADSFSDDLRLQF
 DPSLFTNTSLFDSLPPFGAHHTAATGEWDEVQSGLRADAPPPTMRVAVTAARPPRAKPAFRRRAAQPSDASPAQ
 VDLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTTHAHIVALSQHPAALGTAVVKYQDMIAALPEATHEAIVGV
 GKQWSGARALEALLTVAGELRGPPQLDGTGQLLKIARGGVTAVEAVHAWRNALTGAPLN

Repetição 1: AA 289-323

LTPEQVVAIASNNGGKQALETVQRLLPVLCQAPHD

FIGURA 16B

Repetiç#o 2: AA 324-358
 LTPEQVVAIASIGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 3: AA 359-393
 LTPEQVVAIASNGGGKQALETVQRLLPVLCQAPHC
 Repetiç#o 4: AA 394-428
 LTPEQVVAIASNIGGGKQALETVQALLPVLCQAPHC
 Repetiç#o 5: AA 429-463
 LTPEQVVAIASNGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 6: AA 464-498
 LTPEQVVAIASNGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 7: AA 499-533
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 8: AA 534-568
 LTPEQVVAIASNGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 9: AA 569-603
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 10: AA 604-638
 LTPEQVVAIASNIGGGKQALETVQALLPVLCQAPHC
 Repetiç#o 11: AA 639-673
 LTPEQVVAIASHDGGKQALETVQALLPVLCQAPHD
 Repetiç#o 12: AA 674-708
 LTPEQVVAIASNIGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 13: AA 709-743
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 14: AA 744-778
 LTPEQVVAIASNGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 15: AA 779-813
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 16: AA 814-848
 LTPEQVVAIASNGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 17: AA 849-883
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 18: AA 884-918
 LTPEQVVAIASHDGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 19: AA 919-953
 LTPEQVVAIASNGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 20: AA 954-988
 LTPEQVVAIASNGGGKQALETVQALLPVLCQAPHD
 Repetiç#o 21: AA 989-1023
 LTPEQVVAIASNIGGGKQALETVQRLLPVLCQAPHD
 Repetiç#o 21.5: AA 1024-1043
 LTPEQVVAIASNGGGKQALE

C-terminal: AA 1044-1321
 SIFQLSRPDPALAAALTNDRIVALACIGGRSALNAVKGDLNALTILRRANSRIPERTSHIVADHTQVVRVLGFFQ
 CHSHPAQAFDEAMTQFGMSRHGLLQLFRRVGVTELEARSCTLPASQRWDRILQASGMKRAKPSPTSTQTPDQASL
 HAFADSLERDLDAFSPMHEGDQTRASSRKRSDRAVTGPSAQQSFEVVRPEQRDALHLPLLSWGVKRPRTTRIGGL
 LDPGTPMDADLVASSTVVWEQDADEFAGTADDFPAFNEEELAWLMELLPH

FIGURA 16C

Seqüência de proteína:

Acessão: ;
 AAY43359.1

Organismo:
Xanthomonas campestris pv. armoraciae Cepa 5

N-terminal: aminoácido (AA) 1-288

Repetição 1: AA 289-322
LTPEOVVAIA²⁸⁹SNIGGKOALET³⁰⁰VORLLPVLCOAHG³²²

Repetição 2: AA 323-356
LTPOQVVAIASHDGGKQALETVORLLPVLCOAHG

Repetição 3: AA 357-390
LTPEQVVAIASNIGGKQALETVQALLPVLCOAHG

Repetição 4: AA 391-424
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG

Repetição 5: AA 425-458
LTPOQVVAIASHDGGKQALETVORLLPVLCQAHG

Repetição 6: AA 459-492
LTPQQVVAIASHDGGKQALETVQRLLPVLCAHG

Repetição 7: AA 493-526
LTPQQVVAIASNSGGKQALETVQRLLPVLCAHG

Repetição 8: AA 527-560
LTPQQVVAIASNSGGKQALETVQRLLPVLCAHG

Repetição 9: AA 561-594
LTPQQVVAIASNSGGKQALETVQRLLPVLCQAHG

Repetição 10: AA 595-628
LTPEQVVAIASHDGGKQALETVORLLPVLCOAHG

Repetição 11: AA 629-662
LTPEQVVAIASNIGGKQALETVORLLPVLCOAHG

Repetição 11.5: AA 663-682

FIGURA 16D

SVAQISRPDPALAAIINDHLVACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQ
 CHSHPAQAFDDAMTQFGMSRHGLLQLFRRVGVTELEARSCTLPASQWRDILQASGMKRAKPSPTSTQTPDOASL
 HAFADSEIRDLDLASPMHEGDTQRASSRRKRSRDRAVGTGSAQSQFVRVPEQRDALHLLLSWGKVRPRITRIGL
 LDPGTMDMDLVA SSTWVEODADPAGTADPFGAGTAPFVFNNEELALMELLQPL

Sequência de proteína:

[illegible]

AAAY43360.1

Xanthomonas campestris pv. *armoraciae* Ceba5

N-terminal: aminoácido (AA) 1-288

MDPIRSRTPSPARELLSGPQDGVQPTADRGVSPAGGFLDGLPARRTMSRTRLSPSPAPSPAFSADSFSDDLRLQF
DPSLFNTSLFDSLPPFGAHHTEAATGEWDEVQSLGRAADAPPTMMRVATAARPRAKPAFRRRAQPSDASPAQ
VDLRTLGSYQQQKEIKPKVRSRTVAQHHEALVGHGFTHAIVALSQHAPALGTAVVKYQDMIALPEATHEAIVGV
GKOWSGARQALLTVAGELRGPIPLVDGOLLKTKAGRGVTAVEAVHAWNRLGAPLN

1. AA 289-322
LTPEOVVAIASNIGGKOALETVOAILPVLCOAHG

LTPROOVVAIASHDGGKOALETVOBILPVI,COAHG

1.TPEOVVAIASHDGGKOALETVOBLLPVLCOAHG

LTPOOVVAIASNGGGKOALETVORLLPVLCOAHG

LTPEOVVAIASNSGGKOALETVOALLPVLCOAHG

LTPOOVVAIASNSGGKOALETVOALLPVLCOAHG

LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG

LTPQQVVAIASHDGGKQALETVQRLLPVLCQAHG

Journal of Management Inquiry

FIGURA 16E

LTPQQVVAIASNGGKQALETVQRLLPVLCQAHG

Repetição 10: AA 595-628

LTPEQVVAIASNIGGKQALETVQALLPVLCQAHG

Repetição 11: AA 629-662

LTPQQVVAIASNSGGKQALETVQALLPVLCQAHG

Repetição 12: AA 663-696

LTPEQVVAIASNIGGKQALETVQRLLPVLCQAHG

Repetição 13: AA 697-730

LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG

Repetição 14: AA 731-764

LTPEQVVAIASNIGGKQALETVQALLPVLCQAHG

Repetição 14.5: AA 765-784

LTPEQVVAIASNGGGRPALE

C-terminal: AA 785-1062

SIVAQLSRPDPALAAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQ
CHSHPAQAFDDAMTQFGMSRHLLQLFRRVGVTELEARSGLPPASQRWDRI LQASGMKRAKPSPTSTQTPDQASL
HAFADSLERDLAPSPMHEGDQTRASSRKRSRSDRAVTGPSAQQSFEVVRVPEQRDALHLPLSWRVKRPRTSIGGGL
PDPGTPTAADLAASSTVMREQDEDPFAGAADDFFAFNEEBELAWLMELL PQ

FIGURA 16F

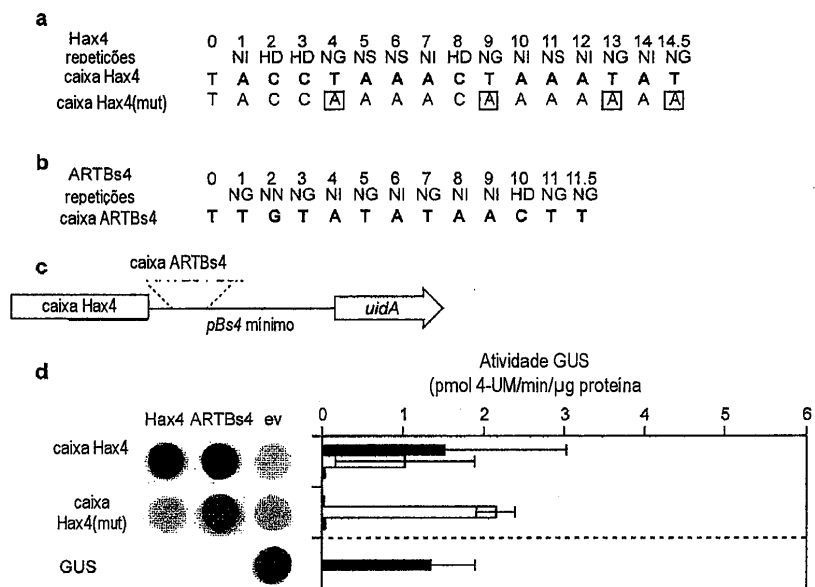


FIGURA 17

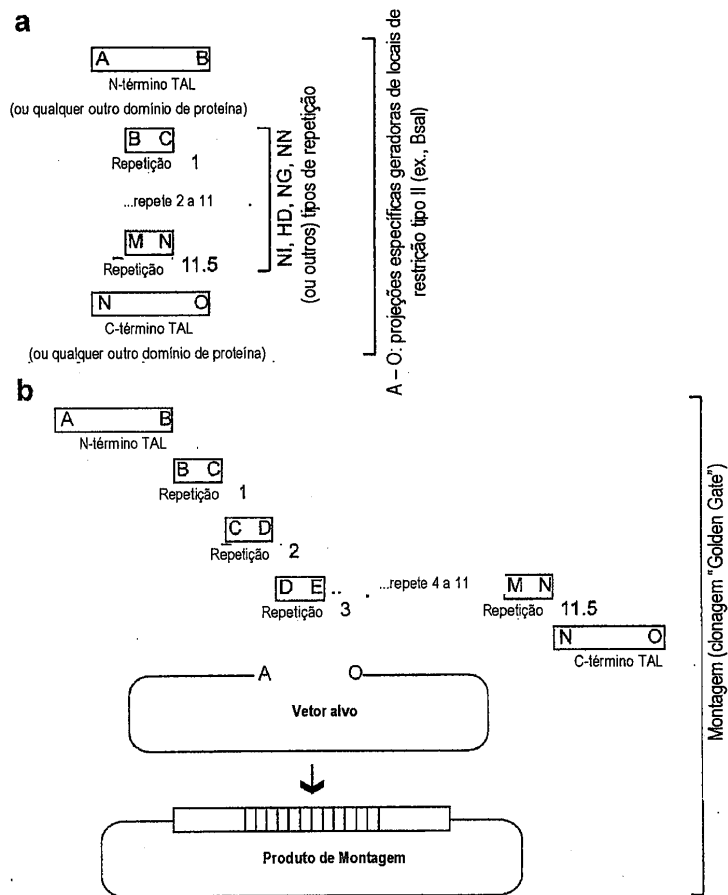


FIGURA 18

FIG. 19A

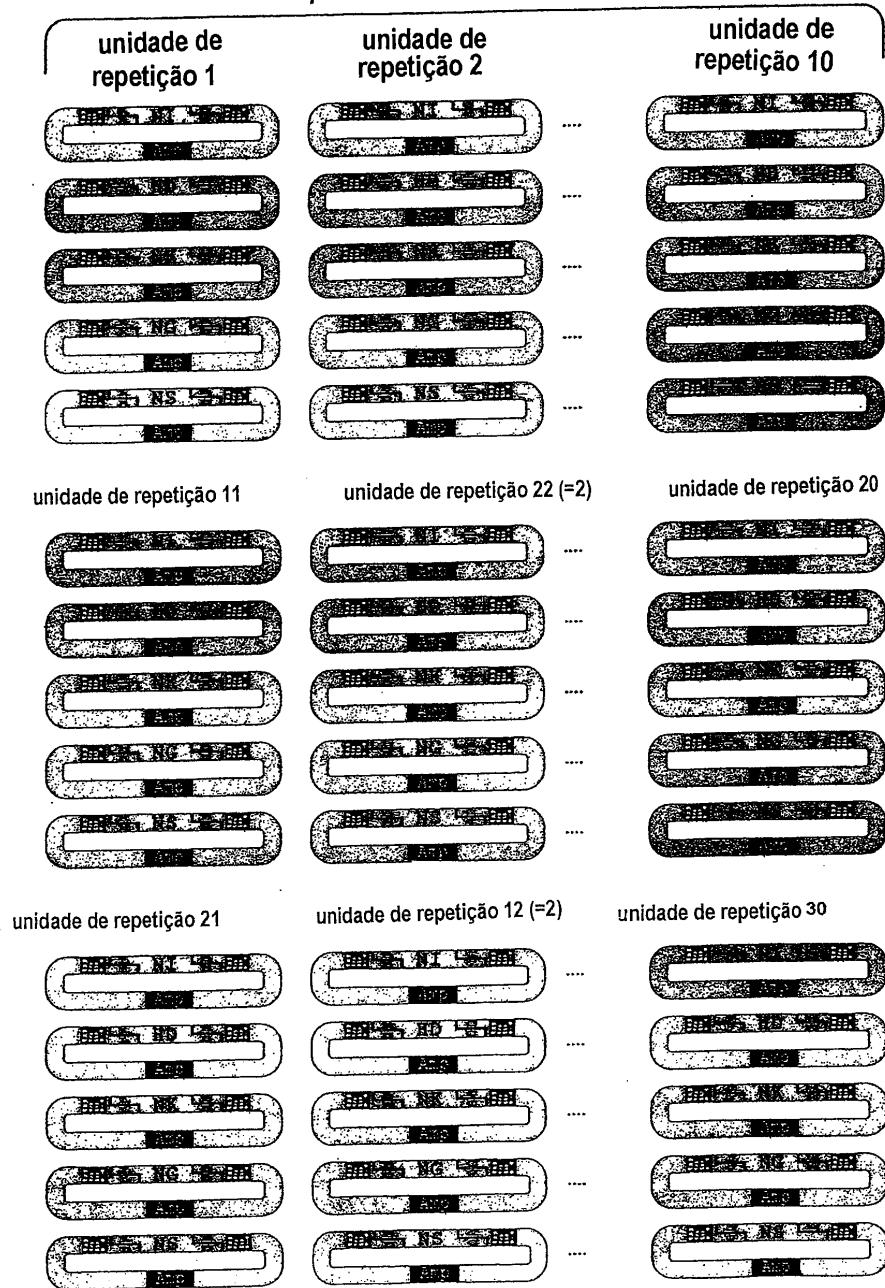
plasmídeos com inserto

FIG 19B
vetores intermediários

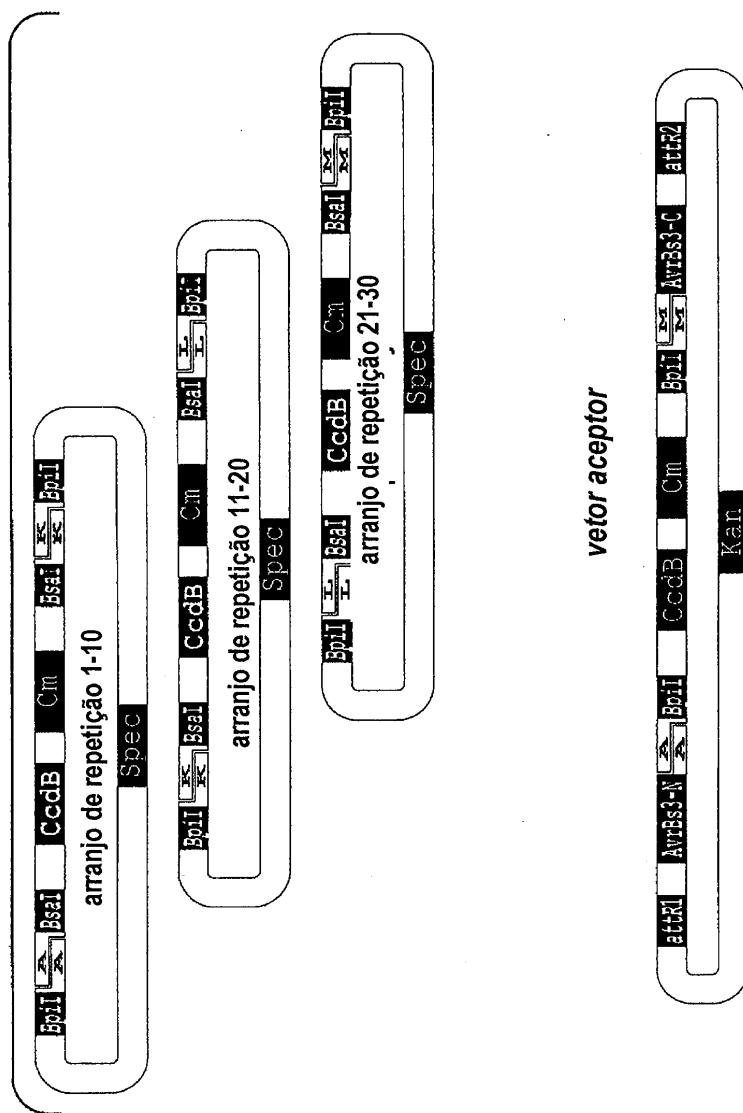


FIG. 19C

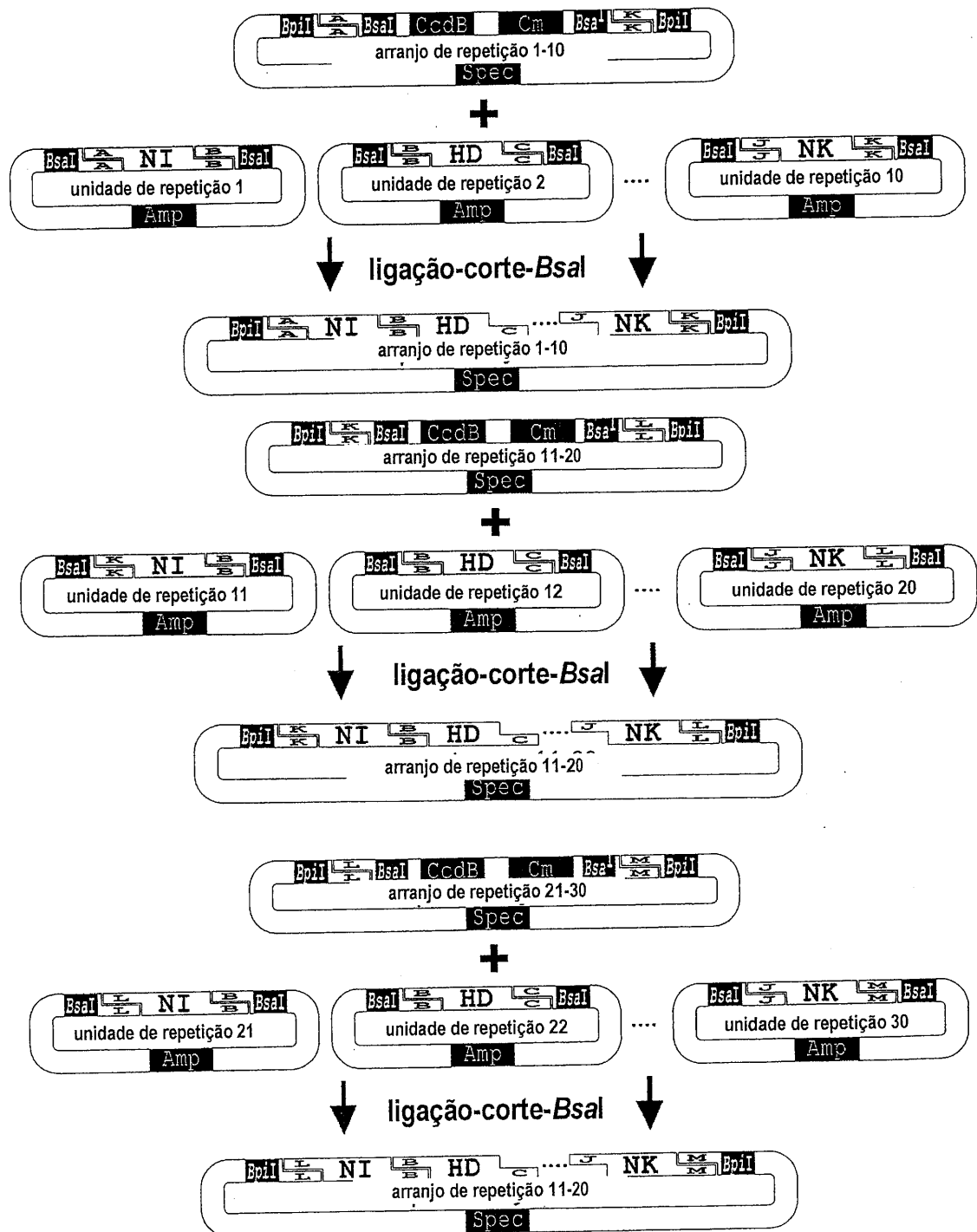


FIG. 19D

