

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2010.12.10	(73) Titular(es): IOWA STATE UNIVERSITY RESEARCH FOUNDATION, INC. 310 LAB OF MECHANICS AMES, IA 50011-2131 US REGENTS OF THE UNIVERSITY OF MINNESOTA US
(30) Prioridade(s): 2009.12.10 US 285324 P 2010.06.07 US 352108 P 2010.07.22 US 366685 P	
(43) Data de publicação do pedido: 2012.10.17	
(45) Data e BPI da concessão: 2014.10.22 024/2015	(72) Inventor(es): (74) Mandatário: JOSÉ EDUARDO LOPES VIEIRA DE SAMPAIO R DO SALITRE 195 RC DTO 1250-199 LISBOA PT

(54) Epígrafe: **MODIFICAÇÃO DO ADN MODIFICADA PELO EFECTOR TAL**

(57) Resumo:

A PRESENTE INVENÇÃO TEM POR OBJECTO MATERIAIS E PROCESSOS RELACIONADOS COM O TRATAMENTO DE GENES (POR EXEMPLO, TRATAMENTO DE GENES COM NUCLEASES EFECTORAS SEMELHANTES A UM ACTIVADOR DE TRANSCRIÇÃO; \pm TALEN \pm).

DESCRIÇÃO

MODIFICAÇÃO DO ADN MODIFICADA PELO EFECTOR TAL

DOMÍNIO TÉCNICO

A presente invenção tem por objecto processos para o tratamento de genes e, particularmente, processos que incluem a utilização de sequências efectoras semelhantes a um activador de transcrição (TAL).

ANTECEDENTES

A capacidade para modificar cromossomas através de recombinação homóloga (recombinação genética) tem sido um objectivo procurado ao longo do tempo pelos biólogos. Por exemplo, nas plantas, a recombinação genética pode ajudar e discernir a função dos genes da planta, abrindo novas possibilidades para a melhoria das culturas. Por exemplo, com a recombinação genética é possível realizar a cirurgia genética necessária para as vias metabólicas reorquestradas criarem colheitas de elevado valor, incluindo sementes com perfis alterados de óleos ou de hidratos de carbono, alimentos com melhores qualidades nutricionais ou plantas com uma maior resistência às doenças e à pressão. Nos animais (por exemplo, nos mamíferos), a recombinação genética pode ser utilizada para o tratamento de doenças. Por exemplo, pode-se utilizar a recombinação genética para fazer correcções por engenharia em genes que são defeituosos devido a vários tipos de mutações. Tem sido difícil encontrar processos eficazes para essa recombinação genética.

SUMÁRIO

Os efectores TAL de bactérias patogénicas para plantas do género *Xanthomonas* desempenham papéis importantes na doença ou provocam a defesa por meio da ligação do ADN hospedeiro e dos genes hospedeiros específicos de efectores de activação (ver, por exemplo, Gu et al. (2005) *Nature* 435: 1122; Yang et al. (2006) *Proc. Natl. Acad. Sci. EUA* 103: 10503; Kay et al. (2007) *Science* 318: 648; Sugio et al. (2007) *Proc. Natl. Acad. Sci. EUA* 104: 10720; e Römer et al. (2007) *Science* 318: 645). A especificidade depende do número de variáveis efectoras de repetições imperfeitas, normalmente de 34 aminoácidos (Schornack et al. (2006) *J. Plant Physiol.* 163: 256). Os polimorfismos estão presentes principalmente nas posições 12 e 13 repetidas, que são referidas aqui como di-resíduos variáveis repetidos (DVR).

A presente invenção tem por base, em parte, o facto dos DVR dos efectores TAL corresponderem aos nucleótidos nos seus sítios-alvo, de uma forma directa e linear, um DVR com um nucleótido, com alguma degenerescência e sem um aparente contexto de dependência. Esta verificação surpreendente representa um novo mecanismo para o reconhecimento de ADN de proteínas, que permite a previsão do sítio-alvo de novos efectores TAL específicos do alvo. Tal como aqui descrito, estas proteínas podem ser úteis na investigação e na biotecnologia como nucleases-alvo quiméricas, que podem facilitar a recombinação homóloga em engenharia do genoma (por exemplo, para adicionar ou reforçar traços úteis para os biocombustíveis ou para biorrenováveis em plantas). Estas proteínas também podem ser úteis, por exemplo, como factores de transcrição e, especialmente, para aplicações terapêuticas que requerem um nível muito elevado de especificidade, tal

como agentes terapêuticos contra agentes patogénicos (por exemplo, vírus), como exemplos não limitativos.

Num aspecto, a presente memória descritiva descreve as características de um processo para a modificação do material genético de uma célula, compreendendo (a) providenciar uma célula que contenha uma sequência-alvo de ADN; e (b) introduzir na célula uma enzima que modifica o ADN do efector semelhante ao activador de transcrição (TAL), compreendendo a enzima de modificação do ADN do efector TAL (i) um domínio da enzima de modificação do ADN que pode modificar o ADN de hélice dupla e (ii) um domínio do efector TAL, que compreende várias sequências repetidas do efectora TAL que, em combinação, se ligam a uma sequência específica de nucleótidos na sequência de ADN-alvo, tal como a enzima de modificação do ADN do efector TAL modifica o ADN-alvo dentro ou próximo das sequências nucleotídicas específicas na célula ou na sua progenia. O processo pode ainda compreender providenciar à célula um ácido nucleico que contenha uma sequência homóloga ou pelo menos uma porção da sequência de ADN-alvo, de tal modo que ocorra a recombinação homóloga entre a sequência de ADN-alvo e o ácido nucleico. A célula pode ser uma célula eucariótica, uma célula de mamífero, uma célula de planta ou uma célula procariótica. O ADN-alvo pode ser ADN cromossómico. A introdução pode compreender a transfecção da célula com um vector que codifica a enzima de modificação do ADN do efector TAL, injectando mecanicamente a enzima que modifica o ADN do efector TAL numa célula como uma proteína, libertando a enzima que modifica o ADN do efector TAL na célula como uma proteína, por meio de um sistema de secreção bacteriano do tipo III o introduzindo a enzima que modifica o ADN do efector TAL numa célula como uma proteína, por electroporação. A enzima que modifica o ADN pode ser uma

endonuclease (por exemplo, uma endonuclease de restrição do tipo II, tal como *FokI*).

O domínio efector TAL que se liga à sequência específica de nucleótidos dentro do ADN-alvo pode compreender 10 ou mais repetições da ligação de ADN e, preferencialmente, 15 ou mais repetições da ligação do ADN. Cada repetição da ligação do ADN pode incluir um di-resíduo variável de repetição (DVR) que determina o reconhecimento de um par de bases na sequência-alvo de ADN, em que o DVR compreende um ou mais de: HD para o reconhecimento de C; NG para o reconhecimento de T; NI para o reconhecimento de A; NN para o reconhecimento de G ou de A; NS para o reconhecimento de A ou de C ou de G ou de T; N* para o reconhecimento de C ou de T, em que * representa um intervalo na segunda posição do DVR; HG para o reconhecimento de T; H* para o reconhecimento de T, em que * representa um intervalo na segunda posição do DVR; IG para o reconhecimento de T; NK para o reconhecimento de G; HA para o reconhecimento de C; ND para o reconhecimento de C; HI para o reconhecimento de C; HN para o reconhecimento de G; NA para o reconhecimento de G; SN para o reconhecimento de G ou de A; e YG para o reconhecimento de T. Cada repetição de ligação ao ADN pode conter uma DVR que determina o reconhecimento de um par de bases na sequência de ADN-alvo, em que cada repetição da ligação de ADN é responsável pelo reconhecimento de um par de bases na sequência de ADN-alvo e em que o DVR compreende um ou mais de: HA para o reconhecimento de C; ND para o reconhecimento de C; HI para o reconhecimento de C; HN para o reconhecimento de G; NA para o reconhecimento de G; SN para o reconhecimento de G ou de A; YG para o reconhecimento de T; e NK para o reconhecimento de G e um ou mais de: PD para o reconhecimento de C; NG para o reconhecimento de T; NI para o reconhecimento de A; NN para o reconhecimento de G ou de A; NS para o reconhecimento de A ou de C ou de G ou de T; N*

para o reconhecimento de C ou de T, em que * representa um intervalo na segunda posição do DVR; HG para o reconhecimento de T; H* para o reconhecimento de T, em que * representa um intervalo na segunda posição do DVR; e IG para o reconhecimento de T.

Noutro aspecto, a presente invenção tem por objecto um processo para gerar um ácido nucleico que codifica um efector TAL específico para uma sequência seleccionada de nucleótidos, compreendendo: (1) A linearização de um plasmido iniciador com *PspXI*, o plasmido iniciador compreendendo uma sequência de nucleótidos que codifica um primeiro domínio de repetição da ligação de ADN do efector TAL com um di-resíduo variável de repetição DVR específico para o primeiro nucleótido da sequência de nucleótidos seleccionada, em que o primeiro domínio da repetição da ligação do ADN do efector TAL tem um único sítio *PspXI* na sua extremidade 3' e; (2) a ligação ao sítio do plasmido inicial *PspXI* de um módulo de ADN que codifica um ou mais domínios de repetição da ligação do ADN do efector TAL, que tem DVR específico para os nucleótidos seguintes da sequência de nucleótidos seleccionada, em que o módulo do ADN tem extremidades pegajosas de *XhoI*; e (3) a repetição das etapas (1) e (2) até o ácido nucleico que codifica um efector TAL capaz de se ligar à sequência de nucleótidos seleccionada. O processo pode ainda compreender, após a ligação, a determinação da orientação do módulo de ADN no sítio *PspXI*. O processo pode compreender a repetição das etapas (1) e (2) de 1 a 30 vezes.

Noutro aspecto, a presente invenção tem por objecto um processo para gerar um ácido nucleico que codifica uma endonuclease de efector semelhante a activador de transcrição (TALEN), compreendendo (a) a identificação de uma primeira sequência de nucleótidos no genoma de uma célula; e (b) a

síntese de um ácido nucleico que codifica uma TALEN que contém (i) várias repetições da ligação de ADN que, em combinação, se ligam à primeira sequência única de nucleótidos, e (ii) uma endonuclease que gera um corte de hélice dupla numa posição dentro ou adjacente à primeira sequência de nucleótidos, em que cada repetição da ligação de ADN compreende um DVR que determina o reconhecimento de um par de bases no ADN-alvo, em que cada repetição da ligação do ADN é responsável pelo reconhecimento de um par de bases no ADN-alvo e em que a TALEN compreende um ou mais dos seguintes DVR: HD para o reconhecimento de C; NG para o reconhecimento de T; NI para o reconhecimento de A; NN para o reconhecimento de G ou de A; NS para o reconhecimento de A ou de C ou de G ou de T; N* para o reconhecimento de C ou de T; HG para o reconhecimento de T; H* para o reconhecimento de T; IG para o reconhecimento de T; NK para o reconhecimento de G; HA para o reconhecimento de C; ND para o reconhecimento de C; HI para o reconhecimento de C; HN para o reconhecimento de G; NA para o reconhecimento de G; SN para o reconhecimento de G ou de A; e YG para o reconhecimento de T.

A TALEN pode conter um ou mais dos seguintes DVR: HA para o reconhecimento de C; ND para o reconhecimento de C; HI para o reconhecimento de C; HN para o reconhecimento de G; NA para o reconhecimento de G; SN para o reconhecimento de G ou de A; YG para o reconhecimento de T; e NK para o reconhecimento de G, e um ou mais de: HD para o reconhecimento de C; NG para o reconhecimento de T; NI para o reconhecimento de A; NN para o reconhecimento de G ou de A; NS para o reconhecimento de A ou de C ou de G ou de T; N* para o reconhecimento de C ou de T; HG para o reconhecimento de T; H* para o reconhecimento de T; e IG para o reconhecimento de T.

A primeira sequência de nucleótidos pode cumprir pelo menos um dos seguintes critérios: i) ter um mínimo de 15 bases de comprimento e estar orientada de 5' para 3' com um T a preceder imediatamente o sítio na extremidade 5' e; ii) não ter um T na primeira posição (5') ou um A na segunda posição; iii) terminar em T na última posição de (3') e não ter um G a seguir à última posição; e iv) ter uma composição de bases de 0-63 % de A, 11-63 % de C, 0-25 % de G e 2-42 % de T.

O processo pode compreender a identificação de uma primeira sequência de nucleótidos e de uma segunda sequência de nucleótidos no genoma da célula, em que a primeira e a segunda sequências de nucleótidos cumprem pelo menos um dos critérios estabelecidos antes e estão separadas por 15-18 pb. A endonuclease pode gerar um corte de hélice dupla entre a primeira e a segunda sequências de nucleótidos.

Noutra modalidade, a presente invenção tem por objecto uma TALEN que compreende um domínio de endonuclease e um domínio de ligação de ADN do efector TAL específico para um ADN-alvo, em que o domínio de ligação do ADN compreende várias repetições da ligação de ADN, compreendendo cada repetição um DVR que determina o reconhecimento de um par de bases no ADN-alvo, em que cada repetição da ligação de ADN é responsável pelo reconhecimento de um par de bases no ADN-alvo e em que a TALEN compreende um ou mais dos DVR seguintes: HD para o reconhecimento de C; NG para o reconhecimento de T; NI para o reconhecimento de A; NN para o reconhecimento de G ou de A; NS para o reconhecimento de A ou de C ou de G ou de T; N* para o reconhecimento de C ou de T; HG para o reconhecimento de T; H* para o reconhecimento de T; IG para o reconhecimento de T; NK para o reconhecimento de G; HA para o reconhecimento de C; ND para o reconhecimento de C; HI para o reconhecimento de C; HN para o reconhecimento de G;

NA para o reconhecimento de G; SN para o reconhecimento de G ou de A; e YG para o reconhecimento de T. A TALEN pode conter um ou mais dos seguintes DVR: HA para o reconhecimento de C; ND para o reconhecimento de C; HI para o reconhecimento de C; HN para o reconhecimento de G; NA para o reconhecimento de G; SN para o reconhecimento de G ou de A; YG para o reconhecimento de T; e NK para o reconhecimento de G, e um ou mais de: HD para o reconhecimento de C; NG para o reconhecimento de T; NI para o reconhecimento de A; NN para o reconhecimento de G ou de A; NS para o reconhecimento de A ou de C ou de G ou de T; N* para o reconhecimento de C ou de T; HG para o reconhecimento de T; H* para o reconhecimento de T; e IG para o reconhecimento de T. O domínio da endonuclease pode ser de uma endonuclease de restrição do tipo II (por exemplo, *FokI*).

Ainda noutro aspecto, a presente invenção tem por objecto uma TALEN que compreende um domínio de endonuclease e um domínio efector TAL, em que a sequência de aminoácidos da referida TALEN se selecciona no grupo que consiste nas SEQ ID N.º. 33 a SEQ ID N.º. 55, SEQ ID N.º. 72 e SEQ ID N.º. 73.

A presente invenção também tem por objecto um processo para gerar um animal, compreendendo: providenciar uma célula eucariótica que compreenda uma sequência de ADN-alvo na qual se deseja introduzir uma modificação genética; gerar um corte de hélice dupla dentro da sequência de ADN-alvo com uma TALEN que compreende um domínio de endonuclease e um domínio efector TAL que se liga à sequência de ADN-alvo; e gerar um animal a partir da célula ou da sua progenia, em que tenha ocorrido um corte de hélice dupla. O processo pode ainda compreender a introdução nas células de um ácido nucleico exógeno, compreendendo uma sequência homóloga com pelo menos uma porção do ADN-alvo, em que a introdução é feita em

condições que permitem que ocorra a recombinação homóloga entre o ácido nucleico exógeno e a sequência de ADN-alvo na célula ou na sua progenia; e a geração de um animal a partir da célula ou da sua progenia, em que tenha ocorrido a recombinação homóloga. O animal pode ser um mamífero. A modificação genética pode compreender uma substituição, uma inserção ou uma eliminação.

Ainda noutro aspecto, a presente invenção tem por objecto um processo para gerar uma planta, compreendendo providenciar uma célula de planta que contenha uma sequência de ADN-alvo na qual se pretende introduzir uma modificação genética pré-seleccionada; a geração de um corte de hélice dupla dentro da sequência de ADN-alvo com uma TALEN que compreende um domínio de endonuclease e um domínio efector TAL que se liga à sequência de ADN-alvo; e a geração de uma planta a partir da célula ou da sua progenia em que ocorra o corte de hélice dupla. O processo pode ainda compreender a introdução na célula da planta de um ácido nucleico exógeno compreendendo uma sequência homóloga de pelo menos uma porção da sequência de ADN-alvo, em que a introdução é feita em condições que permitem que ocorra a recombinação homóloga entre o ácido nucleico exógeno e a sequência de ADN-alvo na célula ou na sua progenia; e a geração de uma planta a partir da célula ou da sua progenia, em que tenha ocorrido a recombinação homóloga.

Noutro aspecto, a presente invenção tem por objecto um processo para a recombinação genética direccionada numa célula, compreendendo a introdução na célula de um ácido nucleico que codifica uma endonuclease efectora TAL dirigida a uma sequência de ADN-alvo seleccionada; a indução da expressão da endonuclease efectora TAL dentro da célula; e a identificação de uma célula em que a sequência de ADN-alvo

seleccionada exhibe uma mutação. A mutação pode ser seleccionada no grupo que consiste em eliminação de material genético, inserção de material genético e tanto eliminação como inserção de material genético. O processo pode ainda compreender a introdução de um ADN dador na célula. A célula pode ser uma célula de insecto, uma célula de planta, uma célula de peixe ou uma célula de mamífero.

Noutro aspecto, a presente invenção tem por objecto um processo para gerar um efector TAL com uma maior capacidade de recombinação para um ADN-alvo, compreendendo a geração de um ácido nucleico que codifica um efector TAL que compreende o domínio de ligação de ADN com várias repetições da ligação de ADN, em que cada repetição compreende um DVR que determina o reconhecimento de um par de bases no ADN-alvo, em que cada repetição de ligação ao ADN é responsável pelo reconhecimento de um par de bases no ADN-alvo, em que a geração compreende a incorporação de um ácido nucleico que codifica uma variante da sequência de repetição da ligação de ADN de ordem 0, com especificidade para A, C ou G, eliminando assim a necessidade de T na posição -1 do sítio de ligação.

Noutro aspecto, a presente invenção tem por objecto um processo para gerar um efector TAL com uma maior capacidade de recombinação para um ADN-alvo, compreendendo a geração de um ácido nucleico que codifica um efector TAL que compreende o domínio de ligação de ADN com várias repetições da ligação de ADN, em que cada repetição compreende um DVR que determina o reconhecimento de um par de bases no ADN-alvo, em que cada repetição de ligação ao ADN é responsável pelo reconhecimento de um par de bases no ADN-alvo, em que a geração compreende a incorporação de um ou mais ácidos nucleicos que codificam os domínios de ligação do ADN do efector TAL que contém DVR com uma maior especificidade para G e em que os referidos DVR se

seleccionam no grupo que consiste em RN, R*, NG, NH, KN, K*, NA, NT, DN, D*, NL, NM, EN, E*, NV, NC, QN, Q*, NR, NP, HN, H*, NK, NY, SN, S*, ND, NW, TN, T*, NE, NF, YN, Y* e NQ, em que * representa um intervalo na segunda posição do DVR.

A presente invenção tem também por objecto um processo de produção de um polipéptido que reconhece selectivamente pelo menos um par de bases numa sequência de ADN-alvo, compreendendo a síntese de um polipéptido que compreende a repetição, em que o domínio de repetição compreende pelo menos uma unidade de repetição derivada de um effector semelhante ao activador de transcrição (TAL), em que a unidade de repetição compreende uma região hipervariável que determina o reconhecimento de um par de bases na sequência-alvo do ADN, em que a unidade de repetição é responsável pelo reconhecimento de um par de bases na sequência de ADN e em que a região hipervariável compreende um elemento seleccionado no grupo que consiste em (a) HD para o reconhecimento de C/G; (b) NI para o reconhecimento de A/T; (c) NG para o reconhecimento de T/A; (d) NS para o reconhecimento de C/G ou de A/T ou de T/A ou de G/C; (e) NN para o reconhecimento de G/C ou de A/T; (f) IG para o reconhecimento de T/A; (g) N para o reconhecimento de C/G; (h) HG para o reconhecimento de C/G ou de T/A; (i) H para o reconhecimento de T/A; e (j) NK para o reconhecimento de G/C. Além disso, a presente invenção tem por objecto um polipéptido produzido pelo processo anterior e um ADN compreendendo uma sequência de codificação para o polipéptido produzido por esse processo. Também tem por objecto uma cassette de expressão compreendendo um promotor ligado operacionalmente ao ADN mencionado antes e uma célula hospedeira não-humana compreendendo a casse de expressão. Noutro aspecto, a presente invenção tem por objecto um

organismo não-humano transformado, compreendendo a cassete de expressão.

Ainda noutro aspecto, a presente invenção tem por objecto um processo para reconhecer selectivamente um par de bases numa sequência de ADN num polipéptido, compreendendo a construção de um polipéptido que compreende um domínio de repetição, em que o domínio de repetição compreende pelo menos uma unidade de repetição derivada de um efector TAL, em que a unidade de repetição compreende uma unidade de uma região hipervariável que determina o reconhecimento de um par de bases na sequência de ADN, em que a unidade de repetição é responsável pelo reconhecimento de um par de bases na sequência de ADN e em que a região hipervariável compreende um elemento seleccionado no grupo que consiste em (a) HD para o reconhecimento de C/G; (b) NI para o reconhecimento de A/T; (c) NG para o reconhecimento de T/A; (d) NS para o reconhecimento de C/G ou de A/T ou de T/A ou de G/C; (e) NN para o reconhecimento de G/C ou de A/T; (f) IG para o reconhecimento de T/A; (g) N para o reconhecimento de C/G; (h) HG para o reconhecimento de C/G ou de T/A; (i) H para o reconhecimento de T/A; e (j) NK para o reconhecimento de G/C.

A presente invenção também tem por objecto um processo para regular a expressão de um gene-alvo numa célula, em que se providenciam células que contêm um polipéptido, em que o polipéptido compreende um domínio de repetição, em que o domínio de repetição compreende pelo menos uma unidade de repetição derivada de um efector TAL, em que a unidade de repetição compreende a região hipervariável que determina o reconhecimento de um par de bases numa sequência de ADN, em que a unidade de repetição é responsável pelo reconhecimento de um par de bases na sequência de ADN e em que a região hipervariável compreende um elemento seleccionado no grupo

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

Noutro aspecto, a presente invenção tem por objecto um polipéptido que compreende um domínio de repetição, em que o domínio de repetição compreende pelo menos uma unidade de repetição derivado de um efector TAL, em que a unidade de repetição compreende uma região hipervariável que determina o reconhecimento de um par de bases numa sequência de ADN, em que a unidade de repetição é responsável pelo reconhecimento de um par de bases na sequência de ADN e em que a região hipervariável compreende um elemento seleccionado no grupo que consiste em (a) HD para o reconhecimento de C/G; (b) NI para o reconhecimento de A/T; (c) NG para o reconhecimento de T/A; (d) NS para o reconhecimento de C/G ou de A/T ou de T/A ou de G/C; (e) NN para o reconhecimento de G/C ou de A/T; (f) IG para o reconhecimento de T/A; (g) N para o reconhecimento de C/G; (h) HG para o reconhecimento de C/G ou de T/A; (i) H para o reconhecimento de T/A; e (j) NK para o reconhecimento de G/C. A presente invenção também tem por objecto um ADN que compreende uma sequência de codificação para o polipéptido mencionado antes.

Noutro aspecto, a presente invenção tem por objecto um ADN que é modificado para incluir um par de bases localizado numa sequência de ADN de modo que o par de bases possa ser reconhecido especificamente por um polipéptido que compreende um domínio de repetição, em que o domínio de repetição

compreende pelo menos uma unidade de repetição derivada de um efector TAL, em que a unidade de repetição compreende uma região hipervariável que determina o reconhecimento de um par de bases na sequência de ADN, em que a unidade de repetição é responsável pelo reconhecimento de um par de bases na sequência de ADN e, em que, para receber um reconhecimento selectivo e determinado pela região hipervariável, o par de bases selecciona-se no grupo que consiste em (a) C/G para o reconhecimento por HD; (b) A/T para o reconhecimento por NI; (c) T/A para o reconhecimento por NG; (d) CT ou A/T ou T/A ou G/C para o reconhecimento por NS; (e) G/C ou A/T para o reconhecimento por NN; (f) T/A para o reconhecimento por IG; (g) C/G ou T/A para o reconhecimento por N; (h) T/A para o reconhecimento por HG; (i) T/A para o reconhecimento por H; e (j) G/C para o reconhecimento por NK. Também tem por objecto um vector compreendendo o ADN mencionado antes, uma célula hospedeira não-humana compreendendo o ADN e um organismo não-humano transformado, compreendendo o ADN.

Ainda noutro aspecto, a presente invenção tem por objecto um processo para a produção de um ADN que contenha uma sequência de ADN-alvo que é reconhecida selectivamente por um polipéptido que compreende um domínio de repetição, em que o domínio de repetição compreende pelo menos uma unidade de repetição derivado de um efector TAL, em que a unidade de repetição compreende uma região hipervariável que determina o reconhecimento de um par de bases na sequência de ADN-alvo e em que a unidade de repetição é responsável pelo reconhecimento de um par de bases na sequência de ADN-alvo, compreendendo esse processo a síntese de um ADN que compreende um par de bases capaz de ser reconhecido pela unidade de repetição, em que o par de bases se selecciona no grupo que consiste em (a) C/G para o reconhecimento por HD; (b) A/T para o reconhecimento por NI; (c) T/A para o

reconhecimento por NG; (d) CT ou A/T ou T/A ou G/C para o reconhecimento por NS; (e) G/C ou A/T para o reconhecimento por NN; (f) T/A para o reconhecimento por IG; (g) C/G ou T/A para o reconhecimento por N; (h) T/A para o reconhecimento por HG; (i) T/A para o reconhecimento por H; e (j) G/C para o reconhecimento por NK.

Noutro aspecto, a presente invenção tem por objecto um processo para a modificação de material genético de uma célula de planta. O processo pode incluir (a) a introdução na célula da planta de (i) um primeiro ácido nucleico recombinante compreendendo uma sequência de nucleótidos-alvo modificada, em que a sequência de nucleótidos-alvo modificada compreende uma ou mais modificações na sequência de nucleótidos no que respeita a uma sequência de nucleótidos-alvo correspondente presente na célula da planta e em que a sequência de nucleótidos-alvo compreende ainda um sítio de reconhecimento para uma endonuclease efectora TAL específica da sequência (TALEN); e (ii) um segundo ácido nucleico recombinante compreendendo uma sequência de nucleótidos que codifica a endonuclease efectora semelhante ao activador de transcrição (TAL) específica da sequência; (b) para gerar uma planta contendo uma célula da planta; (c) a análise das células, das sementes ou dos tecidos obtidos a partir da planta ou da sua progenia, para recombinação na sequência de nucleótidos-alvo. O processo pode ainda incluir a introdução na célula da planta de (iii) um terceiro ácido nucleico recombinante compreendendo uma célula de nucleótidos que codifica um marcador seleccionável; e a determinação se a planta ou a sua progenia expressa um marcador seleccionável. O processo pode ainda incluir a etapa de rastreio da planta ou da sua progenia para avaliar a ausência do marcador seleccionável. A sequência de nucleótidos que codifica o marcador seleccionável pode ou não estar flanqueada em um ou

ambos os lados por uma sequência semelhante ou idêntica a uma sequência que seja endógena para a célula da planta (por exemplo, uma sequência no sítio de clivagem para uma segunda nuclease específica da sequência). A sequência de nucleótidos que codifica o marcador seleccionável pode ser flanqueada em ambos os lados por sítios de reconhecimento para uma recombinase específica da sequência. O processo pode ainda incluir a etapa de exogamia da planta, com ou sem a etapa de rastreio da progenia da exogamia quanto à ausência do marcador seleccionável. O primeiro e o segundo ácidos nucleicos recombinantes podem ser introduzidos simultaneamente na célula da planta. Um ou ambos os ácidos nucleicos recombinantes podem ser linearizados antes da etapa de introdução. O primeiro e segundo ácidos nucleicos recombinantes podem estar presentes na mesma estrutura.

Noutro aspecto, a presente invenção tem por objecto outro processo para a modificação de material genético de uma célula. O processo pode incluir providenciar uma primeira célula contendo uma sequência de ADN-alvo cromossômica, em que se deseja que ocorra uma recombinação homóloga; providenciar uma TALEN que compreende um domínio de endonuclease que cliva um ADN de hélice dupla e um domínio efector TAL compreendendo várias sequências de repetição do efector TAL que, em combinação, se ligam a uma sequência específica de nucleótidos dentro do ADN-alvo na célula; e o contacto da sequência de ADN-alvo com a TALEN na célula, de tal modo que a TALEN cliva ambas as hélices de uma sequência de nucleótidos dentro ou adjacentes à sequência de ADN-alvo na célula. O processo pode ainda incluir providenciar um ácido nucleico que compreende uma sequência homóloga a pelo menos uma porção do ADN-alvo, de tal modo que ocorra a recombinação homóloga entre a sequência de ADN-alvo e o ácido nucleico. A sequência de ADN-alvo pode ser endógena em

relação à célula. A célula pode ser uma célula de planta, uma célula de mamífero, uma célula de peixe, uma célula de insecto ou linhas de células derivadas destes organismos para culturas *in vitro* ou células primárias retiradas directamente de tecido vivo e estabelecidas para culturas *in vitro*. O contacto pode incluir a transfecção da célula com um vector que compreenda uma sequência que codifica TALEN e que expresse a proteína TALEN na célula, injectando mecanicamente uma proteína TALEN numa célula, administrando uma proteína de endonuclease do efector TAL na célula por meio de um sistema de secreção bacteriano de tipo III ou introduzindo uma proteína TALEN na célula por electroporação. O domínio de endonuclease pode ser de uma endonuclease de restrição do tipo II (por exemplo, *FokI*). O domínio do efector TAL que se liga a uma sequência específica de nucleótidos dentro do ADN-alvo pode incluir 10 ou mais repetições da ligação de ADN, mais preferencialmente, 15 ou mais repetições da ligação de ADN. A célula pode ser de qualquer organismo procariótico ou eucariótico.

Noutro aspecto, a presente invenção tem por objecto um processo para conceber uma sequência específica TALEN capaz de clivar ADN numa localização específica. O processo pode incluir a identificação de uma primeira sequência de nucleótidos cromossómica endógena única, adjacente a uma segunda sequência de nucleótidos, na qual se deseja incluir um corte de hélice dupla; e a concepção de uma sequência específica TALEN compreendendo (a) vários domínios de repetição da ligação de ADN que, em combinação, se ligam a uma primeira sequência de nucleótidos cromossómica endógena única e (b) uma endonuclease que gera um corte de hélice dupla na segunda sequência de nucleótidos.

A presente invenção também tem por objecto uma TALEN que contém um domínio de endonuclease e um domínio de ligação de ADN efector TAL, específico para uma sequência particular de ADN. A TALEN pode ainda incluir um marcador de purificação. O domínio de endonuclease pode ser de uma endonuclease de restrição do tipo II (por exemplo, *FokI*).

Noutro aspecto, a presente invenção tem por objecto um processo para gerar um animal geneticamente modificado, no qual se deseja introduzir um ácido nucleico. O processo pode incluir providenciar uma célula primária que compreende uma sequência de ADN-alvo cromossómico endógeno na qual se deseja introduzir o ácido nucleico; a geração de um corte de hélice dupla dentro da sequência de ADN-alvo cromossómico endógeno com uma TALEN compreendendo um domínio de endonuclease e um domínio efector TAL específica à sequência de ADN-alvo cromossómico endógeno; introduzindo um ácido nucleico exógeno compreendendo uma sequência homóloga a pelo menos uma porção do ADN-alvo cromossómico humano na célula primária, em condições que permitem a ocorrência da recombinação homóloga entre o ácido nucleico exógeno e o ADN-alvo cromossómico endógeno; e gerando um animal a partir da célula primária em que ocorreu a recombinação homóloga. O animal pode ser um mamífero. A sequência homóloga pode ser uma sequência de nucleótidos seleccionada num grupo que consiste numa sequência de nucleótidos que causa disrupção de um gene após a recombinação homóloga, uma sequência de nucleótidos que substitui um gene após a recombinação homóloga, uma sequência de nucleótidos que introduz uma mutação pontual num gene depois da recombinação homóloga e uma sequência de nucleótidos que introduz um sítio regulador após a recombinação homóloga.

Ainda noutro aspecto, a presente invenção tem por objecto um processo para gerar uma planta geneticamente modificada, em que se introduziu um ácido nucleico desejado. O processo pode incluir providenciar uma célula de planta compreendendo uma sequência de ADN-alvo endógeno, em que se deseja introduzir o ácido nucleico; a geração de um corte de hélice dupla dentro da sequência de ADN-alvo endógeno com um TALEN compreendendo um domínio de endonuclease e um domínio efector TAL que se liga à sequência de nucleótidos-alvo endógena; a introdução de um ácido nucleico exógeno compreendendo uma sequência homóloga pelo menos a uma porção do ADN-alvo endógeno na célula da planta, em condições que permitam que ocorra a recombinação homóloga entre o ácido nucleico exógeno e o ADN-alvo endógeno; e para gerar uma planta a partir de uma célula de planta em que tenha ocorrido a recombinação homóloga.

Noutro aspecto, a presente invenção tem por objecto um processo para a recombinação genética direccionada numa célula. O processo pode incluir a introdução na célula de uma molécula de ácido nucleico que codifica uma TALEN recombinante com uma sequência de ADN-alvo seleccionada; induzindo a expressão da TALEN dentro da célula; e identificando uma célula na qual a sequência de ADN-alvo seleccionada exhibe uma mutação. A mutação pode ser seleccionada no grupo que consiste numa eliminação de material genético, uma inserção de material genético e tanto uma eliminação como uma inserção de material genético. O processo pode ainda incluir a introdução de um ADN dador na célula. A célula pode ser uma célula de insecto, uma célula de planta, uma célula de peixe ou uma célula de mamífero.

Ainda noutro aspecto, a presente invenção tem por objecto um processo para gerar um ácido nucleico que codifica

uma TALEN específica da sequência, compreendendo (1) a selecção de um plasmido iniciador que compreende uma sequência de nucleótidos que codifica um primeiro domínio de repetição da ligação do ADN efector TAL com um DVR específico para o primeiro nucleótido de uma sequência de nucleótidos seleccionada, em que o primeiro domínio de repetição da ligação do ADN efector TAL tem um único sítio *PspXI* na sua extremidade 3' e; (2) a linearização do plasmido iniciador com *PspXI*; (3) a ligação no sítio de *PspXI* de uma molécula de ADN que codifica um ou mais domínios de repetição da ligação de ADN do efector TAL que tem DVR específico para os nucleótidos seguintes da sequência seleccionada de nucleótidos, em que o módulo de ADN tem extremidades pegajosas de *XhoI*; e (4) repetição das etapas (2) e (3) até o ácido nucleico codificar uma TALEN capaz de se ligar à sequência de nucleótidos seleccionada. Nalguns casos, o processo pode ainda incluir, após a ligação na etapa (3), a verificação da orientação do módulo de ADN no sítio de *PspXI*.

Salvo indicação em contrário, todos os termos técnicos e científicos têm aqui o mesmo significado que é normalmente atribuído pelo especialista na técnica a que pertence esta invenção. Embora se possam utilizar outros processos e materiais semelhantes ou equivalentes aos aqui descritos para a prática da presente invenção, os processos e os materiais apropriados estão descritos a seguir. Em caso de conflito, a presente memória descritiva, incluindo as definições, fará o controlo. Além disso, os materiais, os processos e os exemplos são apenas ilustrativos e não pretendem ser limitativos.

Estabelece-se aqui os detalhes de uma ou mais modalidades da presente invenção com os desenhos em anexo e a descrição que se segue. Outras características, objectos e

vantagens da presente invenção serão evidentes a partir da descrição e do desenho e das reivindicações.

DESCRIÇÃO DOS DESENHOS

As figuras 1A-1D descrevem o código de reconhecimento do ADN efector TAL. A figura 1A representa um diagrama de um efector genérico TAL que mostra a região de repetição (caixas a branco) e uma sequência de repetição representativa (SEQ ID N^o. 1) com o DVR sublinhado. A figura 1B representa um diagrama que mostra os melhores modelos de alinhamento (alinhamentos com baixa entropia) para várias DVR efectoras TAL e sequências promotoras de genes recombinados (SEQ ID N^o. 2-11). Um asterisco indica uma eliminação no resíduo 13. A figura 1C representa um diagrama que mostra as associações de nucleótidos-DVR nos alinhamentos em B, mais dez alinhamentos obtidos por rastreio de todos os promotores de arroz com 40 efectores adicionais TAL de *X. oryzae*, retendo para cada efector o melhor alinhamento para o qual o gene a jusante foi activado durante a infecção. A figura 1D representa um diagrama que mostra as frequências dos nucleótidos de flaqueio para os 20 sítios-alvo efector TAL. As posições são relativas à extremidade 5' do sítio-alvo; N, comprimento do sítio-alvo. Os logos foram gerados utilizando WebLogo.

As figuras 2A e 2B dão evidências de que *OsHen1* é activado por Tal1c da estirpe de *pv. Oryzicola* de *Xanthomonas oryzae*, BLS256. A figura 2A é uma figura dos resultados semi-quantitativos de RCP-TI, que mostram abundância relativa dos transcriptos de *OsHen1*, com um gene de actina como referência, em folhas de arroz, 24 horas após a inoculação com os mutantes da permuta do marcador BLS256 de M51, M51 comportando o vector vazio

do cosmido (vv) M51 comportando o cosmido pIJF92, que contém *talla*, *tallb* e *tallc* e a estirpe de tipo selvagem (TS). A figura 2B é um esquema baseado no mapeamento da única mutação de permuta do marcador em M51 por meio da recuperação e sequenciação do terminal de um fragmento *XmaI* contendo um marcador. A região do genoma, as coordenadas do fragmento recuperado e as coordenadas do fragmento genómico BLS256 contidas no cosmido pIJF92 também estão ilustradas.

A figura 3 é uma sequência de aminoácidos de referência de AvrBs3 (SEQ ID N.º. 12).

A figura 4 é uma sequência de ácidos nucleicos de referência de AvrBs3 (SEQ ID N.º. 13).

A figura 5 é um mapa do vector de expressão da nuclease TAL.

A figura 6 é um mapa de um plasmido relator do alvo.

A figura 7 representa um diagrama uma arquitectura esquemática das nucleases TAL. Os sítios de reconhecimento do domínio de ligação do ADN TAL estão representados em letras maiúsculas enquanto a sequência espaçadora está indicada em letras minúsculas.

A figura 8 representa a sequência de aminoácidos (SEQ ID N.º. 16) das repetições 17 e semi-repetições em tandem do domínio de reconhecimento AvrBs3. Os aminoácidos hipervariáveis nas posições 12 e 13 estão dentro de uma caixa.

A figura 9 representa um diagrama que mostra um esquema para um ensaio de leveduras para analisar a eficácia TAL.

A figura 10 representa um gráfico que mostra os resultados do ensaio de levedura da nuclease TAL de AvrBs3.

A figura 11 representa um diagrama que mostra a representação esquemática dos módulos de repetição

simples, duplos ou triplos de AsvBs3 e um vector de clonagem.

As figuras 12A e 12B descrevem uma repetição representativa simples do efector TAL (figura 12A), assim como uma repetição representativa truncada (figura 12B), que está presente na extremidade da região de repetição na maior parte dos efectores TAL. Ilustram-se as sequências de nucleótidos e de aminoácidos codificados. NS representa os nucleótidos que codificam as DVR, que estão indicadas como "XX". Dão-se os números para as posições dos aminoácidos. As sequências foram retiradas de *Tal1c*.

A figura 13 representa uma descrição esquemática do gene *Tal1c* e o processo pelo qual a região de repetição foi reduzida a uma única repetição truncada, resultando em pCS487, também ilustrada. M, sítio de *MscI*; S, sítio de *SphI*.

A figura 14 representa uma descrição esquemática da introdução de uma mutação com tradução silenciosa na extremidade da repetição original truncada em pCS487 para criar um sítio *PspXI* e *XhoI*, originando pCS489. As sequências dos codões 18-21 na sequência original (SEQ ID N.º. 21) e na repetição mutada (SEQ ID N.º. 23) também estão ilustradas. A sequência de aminoácidos codificados (SEQ ID N.º. 22) não foi alterada pela mutação. Os nucleótidos mutados estão em itálico.

A figura 15 representa um mapa de pCS488, que é um plasmido resistente a canamicina que codifica apenas as porções dos terminais N e C de *Tal1c*, sem a região de repetição, no vector de entrada do cruzamento pENTR-D (Invitrogen, Carlsbad, CA).

A figura 16 representa um mapa do plasmido iniciador da repetição única, designado por pCS493, que codifica uma repetição tendo DVR NI. Três outros plasmidos,

designados por pCS494, pCS495 e pCS496 eram idênticos, excepto no facto de as DVR estarem codificadas (indicado no lado direito).

A figura 17A descreve sequências de nucleótidos e aminoácidos codificados para um módulo único de repetição com DVR NI. A extremidade coesa compatível 5' de *XhoI*, o sítio *MscI* e a extremidade coesa compatível 3' de *PspXI/XhoI* estão sublinhadas. A DVR e os nucleótidos que a codificam estão a negrito. Construíram-se três outros módulos de repetição que são idênticos aos ilustrados, excepto no facto das sequências de codificação de DVR, que codificam, respectivamente, HD, NI e NG. A figura 17B é um mapa do plasmido do módulo de repetição simples, designado por pCS502, que contém a sequência de codificação da repetição ilustrada na figura 17A. Os plasmidos designados por pCS503, pCS504 e pCS505 também foram gerados e são idênticos a pCS502, excepto para as DVR que codificam (indicado do lado direito).

A figura 18A descreve sequências de nucleótidos e aminoácidos codificados para um único módulo de repetição com DVR NI, em que as substituições dos nucleótidos (em itálico) evitam a reconstituição do sítio *XhoI* na extremidade 5' no seguimento da ligação entre um sítio de *PspXI/XhoI* e destroem o sítio interno *MscI*. A DVR e os nucleótidos que a codificam estão a negrito. Construíram-se três módulos adicionais de repetição que são idênticos aos ilustrados, excepto para as sequências que codificam DVR, que codificam, respectivamente, HD, NI e NG. A figura 18B representa um esquema dos três módulos de repetição unidos sequencialmente por meio dos módulos de ligação de repetições adicionais num plasmido do módulo de repetição único. O sítio de *MscI* na primeira repetição e

o sítio de *PspXI* na extremidade 3' permanecem únicos e todo o módulo é flanqueado por dois sítios *XhoI*.

A figura 19 é uma lista do conjunto completo de um, dois e três plasmídeos de módulos de repetição.

A figura 20 representa um diagrama de fluxo que descreve as etapas num processo que pode ser utilizado para unir qualquer sequência de repetições na "estrutura" de *Tal1c* para gerar um gene efector TAL customizado.

As figuras 21A e 21B são esquemas que descrevem a junção de módulos de repetição na construção de endonucleases TAL que irão recombinar as sequências de nucleótidos ilustradas. Na figura 21A, os módulos de repetição de plasmídeos designados por pCS519, pCS524, pCS537, pCS551, pCS583 e pCS529 são adicionados sequencialmente à sequência do plasmídeo iniciador designado por pCS493, resultando nos plasmídeos designados por pMAT55, pMAT56, pMAT57, pMAT58, pMAT59 e pMAT60. Na figura 21B, os módulos de repetição de plasmídeos designados por pCS530, pCS533, pCS522 e pCS541 são adicionados sequencialmente à sequência no plasmídeo designado por pMAT1, resultando nos plasmídeos designados por pMAT61, pMAT62, pMAT63 e pMAT64.

A figura 22A é uma representação esquemática da proteína efectora TAL. Os fragmentos *BamHI* (indicados por B) fundiram-se com o domínio analítico da endonuclease *FokI* para criar TALEN. N, terminal N; SLN, sinal de localização nuclear; B, sítio de *BamHI*, DA, domínio da activação de ácido. A figura 22B é um gráfico da actividade TALEN construído com os efectores TAL AvrBs3 e PthXo1. Avr-FokI, AvrBs3 TALEN; fusões de Pth-FokI, PthXo1 TALEN, Avr-FokI e Pth-FokI, AvrBs3 e PthXo1 com uma versão de *FokI*, inactiva sob o ponto de vista catalítico (Bitinaite et al. (1998) Proc. Natl. Acad. Sci. EUA 95:10570-10575); NDZ, nuclease de dedos de

zinco contendo o domínio de ligação do ADN Zif268 (Porteus e Baltimore (2003) Science 300:763).

A figura 23 é uma sequência de aminoácidos de referência de PthXo1 (SEQ ID N^o. 31).

A figura 24 é uma sequência de aminoácidos de referência de PthXo1 (SEQ ID N^o. 32).

A figura 25 é um diagrama do vector pFZ85.

A figura 26 ilustra a sequência de aminoácidos de AvrBs3_TALEN (SEQ ID N^o. 33).

A figura 27 ilustra a sequência de aminoácidos de PthXo1_TALEN (SEQ ID N^o. 34).

A figura 28A é um gráfico que relaciona a actividade das TALEN de AvrBs3 e PthXo1 nos alvos com diferentes comprimentos espaçadores. NDZ, nuclease de dedos de zinco derivada de Zif268. A figura 28B é um gráfico da actividade de uma TALEN heterodimérica. A actividade em levedura contendo vectores de expressão PthXo1-FokI e AvrBs3-FokI e um plasmido com um alvo consisto nos sítios de reconhecimento para cada um deles, com uma orientação da cabeça para a cauda, separados por 15 pb que também estão ilustrados (Avr-FokI, Pth-FokI). Também estão ilustrados, como referência, a actividade das TALEN de AvrBs3 (Avr-FokI) e PthXo1 (Pth-FokI) individualmente e Zif268 (NDZ) nos respectivos alvos. Como controlo negativo ensaiou-se uma cultura de levedura apenas com plasmidos do sítio-alvo para Avr-FokI, Pth-FokI, para a actividade de lacZ (indicado como (-)).

A figura 29A é um quadro que mostra as sequências de DVR TALEN individuais customizados e as respectivas sequências de reconhecimento de ADN. A figura 29B é um gráfico da actividade TALEN customizada. (-), controlo negativo apenas com plasmidos do sítio-alvo; NDZ, controlo positivo de nuclease de dedos de zinco.

A figura 30 é uma descrição do nucleótido e das frequências de DVR nos terminais de 20 pares-alvo e efectores TAL.

A figura 31 é uma representação esquemática do sistema de clonagem Golden Gate [Engler et al. (2008) PLoS One 3: e3647; e Engler et al. (2009) PLoS One 4: e5553].

As figuras 32A e 32B descrevem um conjunto de 58 plasmidos para a junção e a clonagem de matrizes que codificam repetições de efectores TAL customizadas, utilizando a abordagem de clonagem Golden Gate, tal como aqui descrita. Tet, gene de resistência a tetraciclina, um marcador para a selecção de plasmidos; spec, gene de resistência a espectinomicina, um marcador para a selecção de plasmidos; amp, gene de resistência a ampicilina, um marcador para a selecção de plasmidos.

A figura 33 representa um esquema de um processo para a junção e a clonagem de matrizes de codificação de repetições dos efectores TAL customizada, por meio da abordagem Golden Gate, utilizando o conjunto de plasmidos ilustrado na figura 32. Para fins de ilustração, mostra-se a junção de uma matriz de repetição arbitrária. spec, gene de resistência a espectinomicina, um marcador para a selecção de plasmidos; amp, gene de resistência a ampicilina, um marcador para a selecção de plasmidos.

As figuras 34A-34U mostram as sequências de aminoácidos das TALEN, geradas tal como descrito aqui no exemplo 9.

A figura 34A, telomerase-TALEN124; A figura 34B, gridlock-TALEN105; A figura 34C, adh1-TALEN58; A figura 34D, adh1-TALEN63; A figura 34E, adh1-TALEN68; A figura 34F, adh1-TALEN73; A figura 34G, adh1-TALEN89; A figura 34H, gridlock-TALEN106; A figura 34I, adh1-TALEN64; A figura 34J, adh1-TALEN69; A figura 34K, adh1-TALEN74; A figura 34L, tt4-TALEN90; A figura 34M, telomerase-

TALEN121; A figura 34N, telomerase-TALEN126; A figura 34O, gridlock-TALEN107; A figura 34P, gridlock-TALEN117; A figura 34Q, telomerase-TALEN131; A figura 34R, telomerase-TALEN 136; A figura 34S, *adh1*-TALEN60; A figura 34T, *tt4*-TALEN85; A figura 34U, gridlock-TALEN102.

A figura 35 representa um gráfico que mostra a actividade TALEN medido por meio do ensaio da levedura, utilizando monómeros customizados TALEN com comprimentos crescentes (9, 10, 12, 13, 15, 16, 17 ou 18 mers). As TALEN foram recombinadas com *Arabidopsis* e genes de peixe-zebra, conforme indicado.

A figura 36A representa um diagrama que mostra duas sequências-alvo diferentes de ADN do gene *Arabidopsis ADH1*, que são recombinadas por dois pares TALEN. A figura 36B representa um gráfico dos dados de um ensaio de levedura para pares funcionais TALEN, que se recombinam com o gene de *Arabidopsis ADH1*.

A figura 37A é uma representação esquemática de um ensaio de nuclease de restrição utilizado para detectar mutações induzidas por TALEN em protoplastos de *Arabidopsis*. A figura 37B mostra as sequências de nove clones de ADN não digerido num ensaio de endonuclease de restrição. Seis dos clones têm mutações introduzidas pela junção não-homóloga da extremidade (JNHE).

A figura 38A mostra sequências com 0 repetições de vários efectores TAL distintos sob o ponto de vista filogenético, *AvrHah1* para *Xanthomonas gardneri*, *AvrBs3* para *X. campestris* pv. *vesicatoria*, *PthXo1* para *X. oryzae* pv. *oryzae*, *PthA* para *X. citri*, e *Tal1c* para *X. oryzae* pv. *oryzicola*. As posições polimórficas estão dentro de caixas. A figura 38B é uma representação esquemática que mostra as repetições de ordem 0 e de ordem 1 de *PthXo1*. A repetição de ordem "0" que precede

imediatamente a primeira repetição mostra 35 % de identidade e tem uma estrutura secundária semelhante à prevista. A DVR da primeira repetição e os resíduos análogos candidatos da repetição de ordem 0 estão sublinhados. *, gap; H, hélice; E, alongada. A estrutura foi prevista utilizando JPred (Cole et al. (2008) Nucl. Acids Res. 36: W197-W201).

A figura 39 mostra uma análise por "Western blot" de uma proteína total isolada de células de rim embrionário humano 293T, transfectada com plasmidos, que codificam as proteínas do efector TALEN recombinadas com V5, AvrBs3, PthXo1 e Tallc, conforme indicado, no seguimento da imuno-deteccção utilizando um anticorpo de rato anti-V5. Mostra-se a actina imuno-marcada como o controlo para a carga equivalente em cada coluna.

A figura 40A mostra a sequência de aminoácidos de HPRT-3254-17 TALEN e a figura FIG. 40B mostra a sequência de aminoácidos de HPRT-3286-20r TALEN.

A figura 41A é uma representação esquemática que mostra o sítio recombinado TALEN no gene cromossómico humano HPRT. Os sítios de ligação para as TALEN de HPRT-3254-17 e HPRT-3286-20r, o sítio *Bpu10I* num espaçador entre estes sítios e os sítios do iniciador para amplificação da região também estão indicados. As coordenadas no fundo são a distância dos pares de base do primeiro nucleótido da sequência de codificação. A figura 41B mostra os resultados da digestão dos produtos de amplificação de RCP feita com *Bpu10I* da região ilustrada na figura 41A, utilizando ADN genómico isolado a partir de células-matriz tratadas e não-tratadas com TALEN. O ADN genómico foi digerido com *Bpu10I* antes da amplificação. Os fragmentos de ADN foram separados por electroforese em gel de agarose e visualizados utilizando brometo de etídio.

DESCRIÇÃO DETALHADA

O presente pedido de patente tem por objecto materiais e processos relacionados com o reconhecimento do ADN específico de sequências mediado por efectores TAL. Tal como se descreve aqui, as sequências de aminoácidos primárias dos efectores TAL ditam as sequências de nucleótidos às quais se ligam. Os requerentes verificaram que existem relações directas entre as sequências de aminoácidos dos efectores TAL e as suas sequências-alvo de ADN, permitindo prever um sítio de recombinação para os efectores TAL e também permitindo uma customização dos efectores TAL para se ligarem as sequências particulares de nucleótidos. Esta previsão e customização podem ser utilizadas para uma variedade de finalidades. Num exemplo, as sequências particulares de efectores TAL podem fundir-se com sequências de endonucleases, permitindo que a endonuclease se recombine com sequências específicas de ADN e o subsequente corte do ADN nas sequências recombinadas ou próximo delas. Os cortes (isto é, as rupturas das hélices duplas) no ADN podem aumentar drasticamente a frequência da recombinação homóloga. Assim, em combinação com estruturas de ADN que comportam as sequências que têm um elevado grau de semelhança de sequências com uma sequência particular do ADN de recombinação, pode utilizar-se TALEN para facilitar a mutagénese dirigida ao sítio em genomas complexos, isto é, "knock out" ou alteração da função dos genes ou para adicionar genes ou outras sequências com grande precisão e com uma elevada eficácia.

Assim, estão incluídos no âmbito aqui descrito, *inter alia*, materiais e processos para a produção de organismos geneticamente modificados (incluindo, sem limitações, plantas, fungos, *Drosophila*, nematodes, peixe-zebra, murganhos, outros mamíferos e seres humanos). Estes processos

podem incluir, por exemplo, a transfecção de uma célula com vários ácidos nucleicos recombinantes. Por exemplo, pode-se transformar uma célula (por exemplo, uma célula eucariótica) com uma primeira estrutura de ácido nucleico recombinante contendo uma sequência de nucleótidos dadora que inclui alterações relativamente a uma sequência de nucleótidos recombinante correspondente encontrada dentro da célula e uma segunda estrutura de ácidos nucleicos recombinantes que codifica uma nuclease TAL. Nalgumas modalidades, a célula também pode ser transformada com uma terceira estrutura de ácidos nucleicos recombinantes que codifica um marcador seleccionável. Uma sequência de ácidos nucleicos de uma estrutura de ácidos nucleicos doadores pode assim tornar-se incorporada no genoma da célula transformada, tal como se descreve aqui. Por exemplo, células de plantas produzidas utilizando os processos aqui descritos podem crescer de modo a produzirem plantas com a sequência de nucleótidos doadores alterada, incorporada nos seus genomas. Podem utilizar-se sementes dessas plantas para produzir plantas com um fenótipo TAL como, por exemplo, uma característica de crescimento alterada (por exemplo, uma maior resistência ou tolerância a várias condições bióticas e abióticas), uma aparência alterada (por exemplo, a cor ou a altura alteradas) ou uma composição alterada (por exemplo, níveis maiores ou menores de carbono, azoto, óleos, proteínas, hidratos de carbono (por exemplo, açúcar ou amido), aminoácidos, ácidos gordos ou metabólitos secundários) em relação às plantas não modificadas.

Polinucleótidos e polipéptidos

Providenciam-se aqui ácidos nucleicos e polipéptidos isolados. As expressões "ácido nucleico" e "polinucleótido" são utilizadas intermutavelmente e referem-se tanto a ARN

como a ADN, incluindo ADNc, ADN genómico, ADN sintético (por exemplo, sintetizado quimicamente) e ADN (ou ARN) contendo análogos de ácidos nucleicos. Os polinucleótidos podem ter qualquer estrutura tridimensional. Um ácido nucleico pode ser de hélice dupla ou de hélice simples (isto é, uma hélice paralela ou uma hélice simples anti-paralela). Exemplos não limitativos de polinucleótidos incluem genes, fragmentos de genes, exões, intrões, ARN mensageiro (ARNm), ARN de transferência, ARN ribossómico, ribozimas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plasmídeos, vectores, ADN isolado de qualquer sequência, ARN isolado de qualquer sequência, sindas de ácidos nucleicos e iniciadores, assim como, análogos de ácidos nucleicos.

Os polipéptidos da presente invenção (tais como, enzimas que modificam o ADN de efectores TAL, como exemplo não limitativo), podem ser introduzidos numa célula utilizando um vector que codifica ou referidos polipéptidos, por exemplo ou como polipéptidos *per se*, utilizando vectores de administração associados ou combinados por quaisquer técnicas de permeabilização celular, tais como, sonoporação ou electroporação ou técnicas derivadas destas.

Tal como se utiliza aqui, "isolado", quando se refere a um ácido nucleico, refere-se a um ácido nucleico que está separado de outros ácidos nucleicos que estão presentes num genoma, por exemplo, um genoma de planta, incluindo ácidos nucleicos que normalmente flanqueiam um ou ambos os lados do ácido nucleico no genoma. O termo "isolado", tal como se utiliza aqui em relação aos ácidos nucleicos, também inclui qualquer sequência de ocorrência não-natural, desde que essas sequências de ocorrência não-natural não se encontrem na natureza e não tenham sequências imediatamente contíguas num genoma de ocorrência natural.

Um ácido nucleico isolado, pode ser, por exemplo, uma molécula de ADN, desde que as sequências de ácidos nucleicos normalmente encontradas imediatamente a flanquear essa molécula de ADN num genoma de ocorrência natural estejam eliminadas ou ausentes. Assim, um ácido nucleico isolado inclui, sem limitação, uma molécula de ADN que existe como uma molécula separada (por exemplo, um ácido nucleico sintetizado quimicamente ou um ADNc ou um fragmento de ADN genómico produzido por RCP ou por um tratamento com endonuclease de restrição) independentemente de outras sequências, assim como, um ADN que está incorporado num vector, um plasmido que se replica autonomamente, um vírus (por exemplo, um para-retrovírus, um retrovírus, um lentivírus, um adenovírus ou um vírus do herpes) ou o ADN genómico de um procarioto ou eucarioto. Além disso, um ácido nucleico isolado pode incluir um ácido nucleico recombinante, tal como uma molécula de ADN que faz parte de um ácido nucleico híbrido ou de fusão. Um ácido nucleico que existe entre centenas a milhões de outros ácidos nucleicos, por exemplo, em bibliotecas de ADNc ou de bibliotecas genómicas ou lâminas de gel contendo um produto de digestão por restrição de ADN genómico não é considerado como um ácido nucleico isolado.

Um ácido nucleico pode ser produzido, por exemplo, por síntese química ou por uma reacção em cadeia de polimerase (RCP). A RCP refere-se a um procedimento ou a uma técnica na qual se amplifica os ácidos nucleicos-alvo. A RCP pode ser utilizada para amplificar sequências específicas a partir de ADN, assim como, de ARN, incluindo sequências de ADN genómico total ou de ARN celular total. Descrevem-se vários processos de RCP, por exemplo, em PCR Primer: A Laboratory Manual, Dieffenbach and Dveksler, eds., Cold Spring Harbor Laboratory Press, 1995. Geralmente, a informação sobre as sequências das

extremidades da região de interesse ou para além dela é utilizada para conceber iniciadores de oligonucleótidos que são idênticos ou semelhantes na sequência em relação a hélices opostas na matriz que vai ser amplificada. Estão disponíveis várias estratégias de RCP por meio das quais podem ser introduzidas modificações de sequências de nucleótidos específicas de sítio num ácido nucleico-matriz.

Os ácidos nucleicos isolados podem ser obtidos por mutagênese. Por exemplo, pode-se mutar uma sequência de ácidos nucleicos dadores utilizando técnicas-padrão, incluindo a mutagênese dirigida a oligonucleótidos e a mutagênese dirigida ao sítio através de RCP. Ver, *Short Protocols in Molecular Biology*, Capítulo 8, Green Publishing Associates and John Wiley & Sons, editado por Ausubel et al., 1992.

O termo "polipéptido", tal como se utiliza aqui, refere-se a um composto de duas ou mais subunidades de aminoácidos independente da modificação pós-translacional (por exemplo, fosforilação ou glicosilação). As subunidades podem estar ligadas por ligações de péptidos ou outras ligações, tais como, por exemplo, ligações de éster ou de éter. O termo "aminoácido" refere-se a quaisquer aminoácidos naturais e/ou não-naturais ou sintéticos, incluindo os isómeros ópticos D/L.

Por "isolado" ou "purificado", no que se refere a um polipéptido, entende-se que o polipéptido está separado de alguma forma dos seus componentes celulares, com os quais está normalmente ligado na natureza (por exemplo, outros polipéptidos, lípidos, hidratos de carbono e ácidos nucleicos). Um polipéptido purificado pode originar uma banda única maior num gel de poliacrilamina não redutor. Um

polipéptido purificado pode ser, pelo menos, cerca de 75 % puro (por exemplo, pelo menos 80 %, 85 %, 90 %, 95 %, 97 %, 98 %, 99 % ou 100 % puro). Podem obter-se polipéptidos purificados, por exemplo, por extracção de uma fonte natural, por síntese química ou por produção recombinante numa célula hospedeira ou numa planta transgénica e podem ser purificados utilizando, por exemplo, cromatografia por afinidade, imunoprecipitação, cromatografia por exclusão de dimensão e cromatografia de permuta iónica. O grau de purificação pode medir-se utilizando qualquer processo apropriado incluindo, sem limitação, cromatografia em coluna, electroforese em gel de poliacrilamida ou cromatografia líquida de elevada resolução.

Estruturas recombinantes

As estruturas de ácidos nucleicos recombinantes (por exemplo, vectores) são providenciadas aqui. Um "vector" é um replicam, tal como um plasmido, fago ou cosmido, no qual pode ser inserido outro segmento de ADN de modo a provocar a replicação do segmento inserido. Geralmente, um vector é capaz de se replicar quando está associado com os elementos de controlo apropriados. As estruturas de vectores apropriadas incluem, por exemplo, as que são utilizadas por rotina na técnica, tais como plasmidos, vírus, cromossomas artificiais, BAC, YAC ou PAC. O termo "vector" inclui vectores de clonagem e de expressão, assim como, vectores virais e vectores de integração. Um "vector de expressão" é um vector que inclui uma ou mais sequências de controlo da expressão e uma "sequência de controlo da expressão" é uma sequência de ADN que regula e controlo a transcrição e/ou a tradução de outra sequência de ADN. Os vectores de expressão apropriados incluem, sem limitação, plasmidos e vectores virais derivados, por exemplo, de bacteriófagos, baculovírus,

vírus de mosaico do tabaco, vírus do herpes, citomegalovírus, retrovírus, vírus vaccinia, adenovírus e vírus adeno associados. Estão disponíveis comercialmente numerosos vectores e sistemas de expressão em empresas, tais como Novagen (Madison, WI), Clontech (Palo Alto, CA), Stratagene (La Jolla, CA) e Invitrogen/Life Technologies (Carlsbad, CA).

As expressões "região reguladora", "elemento de controlo" e "sequência de controlo de expressão" referem-se a sequências de nucleótidos que influenciam a iniciação e a velocidade da transcrição ou da translação e a estabilidade e a mobilidade do transcripto ou do produto do polipéptido. As regiões reguladoras incluem, sem limitação, sequências promotoras, sequências intensificadoras, elementos de resposta, sítios de reconhecimento da proteína, elementos indutíveis, elementos de controlo do promotor, sequências de ligação de proteínas, regiões de 5' e 3' não traduzidas (RNT), sítios de início transcricional, sequências de terminação, sequências de poliadenilação, intrões e outras regiões reguladoras, que podem residir dentro de sequências de codificação, tais como sinais de secreção, sequências de localização nuclear (SLN) e sítios de clivagem de proteases.

Tal como se utiliza aqui, "ligado operacionalmente" significa incorporado numa estrutura genética de tal modo que as sequências de controlo de expressão controlam eficazmente a expressão de uma sequência de codificação com interesse. Uma sequência de codificação está "operacionalmente ligada" e "sob controlo" das sequências de controlo de expressão numa célula, quando a polimerase de ADN é capaz de transcrever a sequência de codificação no ARN, o qual, se for um ARNm, pode então ser traduzido na proteína codificada pela sequência de codificação. Assim, uma região reguladora pode modular, por exemplo, regular, facilitar ou orientar a transcrição na

célula da planta, na planta ou no tecido da planta, no qual se deseja expressar um ácido nucleico-alvo modificado.

Um promotor é uma sequência de controlo de expressão composta por uma região de uma molécula de ADN, normalmente dentro de 100 nucleótidos a montante do ponto em que começa a transcrição (geralmente próximo do sítio de iniciação para a polimerase II de ARN). Os promotores estão envolvidos no reconhecimento e na ligação da polimerase de ARN e de outras proteínas para iniciar e regular a transcrição. Para colocar uma sequência de codificação sob o controlo de um promotor, normalmente é necessário posicionar o sítio de iniciação da tradução do quadro de leitura translacional do polipéptido entre um e cerca de cinquenta nucleótidos a jusante do promotor. Contudo, um promotor pode ser seleccionado, por exemplo, 5.000 nucleótidos a montante do sítio de iniciação da tradução ou cerca de 2.000 nucleótidos a montante do sítio de iniciação da transcrição. Um promotor normalmente compreende pelo menos um promotor nuclear (básico). Um promotor também pode incluir pelo menos um elemento de controlo, tal como um elemento a montante. Esses elementos incluem regiões de activação a montante (RAM) e, eventualmente, outras sequências de ADN que afectam a transcrição de um polinucleótido, tal como um elemento sintético a montante.

A escolha dos promotores que devem ser incluídos depende de vários factores incluindo, mas não se limitando à eficácia, capacidade de selecção, inductibilidade, nível de expressão desejado e especificidade da célula ou do tecido. Por exemplo, os promotores específicos de tecidos, órgãos e células que conferem apenas transcrição ou que conferem predominantemente num tecido, órgão ou tipo de célula particular, respectivamente, podem ser utilizados. Nalgumas

modalidades, os promotores específicos de tecidos vegetais, tais como células embrionárias, parênquima, meristema básico, feixe vascular, câmbio, floema, córtex, meristema apical do rebento, meristema lateral do rebento, meristema apical da raiz, meristema lateral da raiz, primórdio da folha, mesófilo da folha, ou epiderme das folhas podem ser regiões reguladoras apropriadas. Nalgumas modalidades, os promotores que são essencialmente específicos de sementes ("promotores preferenciais das sementes") podem ser úteis. Os promotores específicos das sementes podem promover a transcrição de um ácido nucleico ligado operacionalmente no endosperma e no tecido cotiledóneo durante o desenvolvimento das sementes. Alternativamente, os promotores constitutivos podem promover a transcrição de um ácido nucleico ligado operacionalmente na maior parte ou em todos os tecidos de uma planta, através do desenvolvimento da planta. Outras classes de promotores incluem, mas não se limitam a promotores indutíveis, tais como promotores que conferem transcrição em resposta a estímulos externos, tais como agentes químicos, estímulos de desenvolvimento ou estímulos ambientais.

Um promotor básico é a sequência mínima necessária para fazer a junção de um complexo de transcrição necessária para a iniciação da transcrição. Os promotores básicos frequentemente incluem um elemento de uma "caixa TATA" que pode estar localizada próximo dos nucleótidos 15 e 35 a montante do sítio de iniciação da transcrição. Os promotores básicos também podem incluir um elemento de uma "caixa CCAAT" box" (normalmente a sequência CCAAT) e/ou uma sequência GGGCG, que pode estar localizada próximo dos nucleótidos 40 e 200, normalmente próximo dos nucleótidos 60 a 120, a montante do sítio de iniciação da transcrição.

Exemplos não limitativos de promotores que podem estar incluídos nas estruturas dos ácidos nucleicos aqui referidos incluem a região de iniciação da transcrição do vírus do mosaico da couve-flor (CaMV) 35S, os promotores 1' ou 2' derivados do ADN-T de *Agrobacterium tumefaciens*, os promotores de um gene específico da folha do milho descrito por Busk ((1997) Plant J. 11:1285-1295), os genes relacionados com kn1 do milho e de outras espécies e as regiões de iniciação da transcrição de vários genes de plantas, tais como o promotor de ubiquitina-1 do milho.

Transcreve-se a região não transcrita (RNT) de 5', mas não é traduzida e permanece entre o sítio do início do transcripto e o codão de iniciação da tradução e pode incluir o nucleótido +1. Pode-se posicionar uma RNT de 3' entre o codão de terminação da tradução e a extremidade do transcripto. As RNT podem ter funções particulares, tais como aumentarem a estabilidade do ARNm mensageiro ou atenuar a tradução. Exemplos de RNT de 3' incluem, mas não se limitam a sinais de poliadenilação e sequências de terminação da transcrição. Uma região de poliadenilação na extremidade 3' de uma região de codificação pode também estar operacionalmente ligada a uma sequência de codificação. A região de poliadenilação pode derivar de um gene natural, de vários outros genes de plantas ou de um ADN-T de *Agrobacterium*.

Os vectores providenciados aqui também podem incluir, por exemplo, origens de replicação e/ou regiões de ligação de estruturas (RLE). Além disso, um vector de expressão pode incluir uma sequência marcadora designada para facilitar a manipulação ou a detecção (por exemplo, a purificação ou a localização) do polipéptido expresso. As sequências marcadoras, tais como as sequências da proteína fluorescente

verde (PFV), da glutathiona-S-transferase (GST), da polihistidina, da c-mic, hemaglutinina ou de um marcador FlagTM (Kodak, New Haven, CT) são normalmente expressas como uma fusão com o polipéptido codificado. Esses marcadores podem ser inseridos em qualquer parte dentro do polipéptido, incluindo em qualquer um dos terminais carboxilo ou amino.

Por "vector de administração" ou "vectores de administração" entende-se qualquer vector de administração que pode ser utilizado na presente invenção para por em contacto ou administrar para dentro das células ou agentes/produtos químicos de compartimentos sub-celulares e moléculas (proteínas ou ácidos nucleicos) necessários na presente invenção. Inclui, mas não se limita a vectores de administração de lipossomas, vectores de administração de vírus, vectores de administração de fármacos, veículos químicos, veículos poliméricos, lipoplexos, poliplexos dendrímeros, microbolhas (agentes de contraste de ultrasons), nanopartículas, emulsões ou outros vectores de transferência apropriados. Estes vectores de administração permitem a administração de moléculas, produtos químicos, macromoléculas (genes ou proteínas) ou outros vectores, tais como plasmidos péptidos desenvolvidos por Diatos. Nestes casos, os vectores de administração são veículos moleculares. Por "vector de administração" ou "vectores de administração" também se entende processos de administração para realizar a transfecção.

Os termos "vector" ou "vectores" referem-se a uma molécula de ácido nucleico capaz de transportar outro ácido nucleico ao qual tenha sido ligada. Um "vector", na presente invenção, inclui, mas não se limita a um vector viral, um plasmido, um vector de ARN ou uma molécula de ADN ou de ARN linear ou circular, que pode consistir em ácidos nucleicos

cromossômicos, não cromossômicos, semi-sintéticos ou sintéticos. Os vectores preferidos são os que são capazes de replicação autónoma (vectores epissômicos) e/ou expressão de ácidos nucleicos aos quais estão ligados (vectores de expressão). Um grande número de vectores apropriados é conhecido dos especialistas nesta técnica e está comercialmente disponível.

Os vectores virais incluem retrovírus, adenovírus, parvovírus (por exemplo, vírus adeno-associados), coronavírus, vírus de ARN de hélice negativa, tais como ortomixovírus (por exemplo, o vírus da gripe), rabdovírus (por exemplo, vírus da raiva e da estomatite vesicular), paramixovírus (por exemplo, vírus do sarampo e vírus Sendai), vírus de ARN de hélice positiva, tais como picornavírus e alfavírus e vírus de ARN de hélice dupla, incluindo adenovírus, vírus do herpes (por exemplo, vírus do herpes simples do tipo 1 e 2, vírus de Epstein-Barr, citomegalovírus) e poxvírus (por exemplo, vírus vaccinia, vírus das aves e vírus dos canários). Outros vírus incluem vírus de Norwalk, togavírus, flavivírus, reovírus, papovavírus, hepadnavírus, e vírus da hepatite, por exemplo. Exemplos de reiovírus incluem: sarcoma da leucose das aves, vírus do tipo C de mamíferos, vírus do tipo B, vírus do tipo D, vírus do grupo HTLV-BLV, lentivírus, spumavírus (Coffin, J. M., *Retroviridae: The viruses and their replication*, In *Fundamental Virology*, terceira edição, B. N. Fields, et al., Eds., Lippincott-Raven Publishers, Filadélfia, 1996).

Por "vector lentiviral" entende-se vectores de lentivírus à base de VIH, que são muito promissores para a administração de genes por causa da sua capacidade relativamente grande de empacotamento, a sua imunogenicidade reduzida e a sua capacidade para transduzir de forma estável e com uma elevada

eficácia uma grande gama de tipos de células diferentes. Os vectores lentivirais são normalmente gerados no seguimento de uma transfecção temporária de três ou mais vírus (empacotamento, envelope e transferência) nas células produtoras. Tal como o VIH, os vectores lentivirais entram na célula-alvo através da interacção de glicoproteínas virais de superfície com receptores na superfície das células. À entrada, o ARN viral sofre uma transcrição inversa que é mediada pelo complexo viral de transcrição inversa. O produto da transcrição inversa é um ADN viral linear, de hélice dupla, que constitui o substrato para a integração viral no ADN das células infectadas. Os referidos vectores lentivirais podem ser "não integradores" ou "integradores".

- Por "vectores lentivirais integradores (ou VL)", entende-se os vectores, como exemplos não limitativos, que são capazes de integrar o genoma de uma célula-alvo.
- Em sentido oposto, por "vectores lentivirais não integradores (ou VLNI)" entende-se vectores eficazes de administração de genes que não integram o genoma de uma célula-alvo através da acção da integrase do vírus.

Um tipo de vector preferido é um epissona, isto é, um ácido nucleico capaz de replicação extra-cromossómica. Os vectores preferidos são os que capazes de replicação autónoma e/ou de expressão dos ácidos nucleicos aos quais estão ligados. Os vectores capazes de direccionar a expressão do gene aos quais estão operacionalmente ligados são referidos aqui como "vectores de expressão". Um vector de acordo com a presente invenção compreende, mas não se limita a um CAL (cromossoma artificial de levedura), um BAC (bacteriano artificial), um vector de baculovírus, um fago, um fagemido,

um cosmido, um vector viral, um plasmido, um vector de ARN ou uma molécula de ADN ou de ARN linear ou circular que pode consistir em ADN cromossómico, não-cromossómico, semi-sintético ou sintético. Em geral, os vectores de expressão com utilidade nas técnicas de ADN recombinante estão muitas vezes sob a forma de "plasmidos", que se referem geralmente a anéis de ADN de hélice dupla circulares e na sua forma de vectores não estão ligados ao cromossoma. Os especialistas nesta técnica conhecem um grande número de vectores apropriados. Os vectores podem conter marcadores seleccionáveis, por exemplo: neomicina-fosfotransferase, histidinol-desidrogenase, di-hidrofolato-reductase, higromicina-fosfotransferase, timidina-cinase do vírus do herpes simples, adenosina-desaminase, glutamina-sintetase e fosforibosil-transferase de hipoxantina-guanina para a cultura de células eucarióticas; TRP1 para *S. cerevisiae*; resistência a tetraciclina, rifampicina ou ampicilina em *E. coli*. Preferencialmente, os referidos vectores são vectores de expressão, em que se coloca uma sequência de codificação de um polipéptido de interesse sob o controlo de elementos de controlo transcricional e translacional apropriados para permitir a produção ou a síntese do referido polipéptido. Assim, o referido polipéptido está contido numa cassette de expressão. Mais particularmente, o vector contém uma origem de replicação, um promotor ligado operacionalmente ao referido polinucleótido de codificação, um sítio de ligação do ribossoma, um sítio de junção do ARN (quando se utiliza ADN genómico), um sítio de poliadenilação e um sítio de terminação da transcrição. Contém ainda elementos intensificadores ou silenciadores. A selecção do promotor dependerá da célula no qual o polipéptido é expresso. Os promotores apropriados incluem promotores específicos do tecido e/ou indutíveis. Exemplos de promotores indutíveis são: o promotor eucariótico de metalotionina, que é induzido

pelos níveis acrescidos de metais pesados, o promotor procariótico de lacZ, que é induzido em resposta a choque térmico de isopropil- β -D-tiogalacto-piranósido (IPTG) e promotores eucarióticos de choque térmico que são induzidos pelo aumento da temperatura. Exemplos de promotores específicos de tecido são a creatina-cinase dos músculos do esqueleto, o antigénio específico da próstata (PSA), uma α -antitripsina-protease, proteínas A e B de um tensioactivo humano (TH), genes de β -caseína e de proteína ácida do soro do leite.

Os promotores indutíveis podem ser induzidos por elementos patogénicos ou condições stressantes, mais preferencialmente, por condições stressantes, como frio, calor, luz UV ou elevadas concentrações iónicas (revisto em Potenza et al. (2004) *In vitro Cell Dev Biol* 40:1-22). O promotor indutível pode ser induzido por produtos químicos [revisto em Moore et al. (2006); Padidam (2003); (Wang et al. (2003); e (Zuo e Chua (2000))].

Os vectores de administração e os vectores que podem estar associados ou combinados em qualquer técnica de permeabilização celular, tal como sonoporação ou electroporação ou derivados destas técnicas.

Deve entender-se que pode estar presente mais do que uma região reguladora num polinucleótido recombinante, por exemplo, intrões, intensificadores, regiões de activação a montante e elementos indutíveis.

As estruturas de ácidos nucleicos recombinantes incluem uma sequência de polinucleótidos inserida num vector apropriado para a transformação de células (por exemplo, células de plantas ou células de animais). Os vectores

recombinantes podem ser produzidos utilizando, por exemplo, técnicas-padrão de ADN recombinante (ver, por exemplo, Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, 2ª ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Uma sequência recombinante de ácido nucleico, tal como aqui descrita, pode integrar-se no genoma de uma célula por via de uma recombinação ilegítima (isto é, aleatória, não-homóloga, não-específica do sítio) ou uma sequência de ácido nucleico recombinante, tal como descrito aqui, pode adaptar-se para integrar-se no genoma de uma célula por via de uma recombinação homóloga. As sequências de ácidos nucleicos, adaptadas para integração por via de recombinação homóloga, são flanqueadas, em ambos os lados, com sequências que são semelhantes ou idênticas às sequências de nucleótidos recombinantes endógenas, que facilitam a integração do ácido nucleico recombinante em sítios particulares no genoma, contendo as sequências endógenas de nucleótidos recombinantes. As sequências de ácidos nucleicos, adaptadas para integração por via de recombinação homóloga, também podem incluir um sítio de reconhecimento para uma nuclease específica da sequência. Alternativamente, o sítio de reconhecimento para uma nuclease específica de uma sequência pode estar localizado no genoma da célula ser transformada. As sequências de ácidos nucleicos dadores, tal como aqui descritas, normalmente são adaptadas para integração por via da recombinação homóloga.

Nalgumas modalidades, um ácido nucleico que codifica um marcador seleccionável também pode ser adaptado para se integrar, por via de recombinação homóloga e assim pode ser flanqueado em ambos os lados com sequências que são semelhantes ou idênticas às sequências endógenas dentro do

genoma da planta (por exemplo, sequências endógenas no sítio de clivagem para uma nuclease específica de uma sequência). Nalguns casos, os ácidos nucleicos contendo sequências de codificação para um marcador seleccionável também podem incluir um sítio de reconhecimento para uma nuclease específica da sequência. Nestas modalidades, o sítio de reconhecimento para a nuclease específica da sequência pode ser o mesmo ou pode ser diferente daquele que está contido dentro da sequência de ácidos nucleicos dadores (isto é, pode ser reconhecido pela mesma nuclease como uma sequência de ácidos nucleicos dadores ou reconhecido por uma nuclease diferente da sequência de ácidos nucleicos dadores).

Nalguns casos, pode adaptar-se uma sequência recombinante de ácidos nucleicos para se integrar no genoma de uma célula por via de uma recombinação específica do sítio. Tal como se utiliza aqui, recombinação "específica do sítio" refere-se à recombinação que ocorre quando uma sequência de ácido nucleico é dirigida para sítios particulares dentro de um genoma não por homologia entre as sequências no ácido nucleico recombinante e as sequências do genoma, mas antes pela acção das enzimas recombinase, que reconhecem sequências específicas de ácidos nucleicos e catalisam a permuta recíproca de hélices de ADN entre estes sítios. A recombinação específica do sítio refere-se assim à clivagem mediada por enzimas e à ligação de duas sequências de nucleótidos definidas. Pode-se utilizar qualquer sistema de recombinação específico do sítio que seja apropriado, incluindo, por exemplo, o sistema Cre-lox ou o sistema FLP-FRT. Nestas modalidades, pode-se introduzir um ácido nucleico que codifica uma enzima recombinase numa célula, para além de uma sequência de nucleótidos dadores e uma sequência que codifica uma nuclease e, nalguns casos, uma sequência de

marcadores seleccionáveis. Ver, por exemplo, a patente norte-americana N.º. 4 959 317.

Endonucleases específicas de sequências

As nucleases específicas de sequências e os ácidos nucleicos recombinantes que codificam as endonucleases específicas de sequências são dados aqui. As endonucleases específicas de sequências podem incluir os domínios de ligação de ADN do efector TAL e domínios de endonuclease. Assim, os ácidos nucleicos que codificam essas endonucleases específicas de sequências podem incluir uma sequência de nucleótidos de um efector TAL específico de sequência, ligado a uma sequência de nucleótidos de uma nuclease.

Os efectores TAL são proteínas de bactérias patogénicas das plantas, que se injectam na célula das plantas por meio do agente patogénico, e assim elas navegam até ao núcleo e funcionam como factores de transcrição para se transformarem em genes específicos das plantas. A principal sequência de aminoácidos de um efector TAL dita a sequência de nucleótidos à qual se liga. Assim, podem prever-se os sítios-alvo para os efectores TAL e os efectores TAL podem também ser tratados por engenharia e gerados com o fim de se ligarem a sequências particulares de nucleótidos, tal como aqui descrito.

As sequências que se fundem com as sequências de ácidos nucleicos que codificam efectores TAL são sequências que codificam uma nuclease ou uma porção de uma nuclease, normalmente um domínio de clivagem não-específica de uma endonuclease de restrição do tipo II, tal como *FokI* (Kim et al. (1996) Proc. Natl. Acad. Sci. EUA 93:1156-1160). Outras endonucleases úteis podem incluir, por exemplo, *HhaI*, *HindIII*, *NotI*, *BbvCI*, *EcoRI*, *BglI* e *AlwI*. O facto de algumas

endonucleases (por exemplo, *FokI*) funcionarem apenas como dímeros pode ser capitalizado no sentido de aumentar a especificidade para um alvo do efector TAL. Por exemplo, nalguns casos, cada monómero de *FokI* pode fundir-se com uma sequência efectora TAL que reconhece uma sequência-alvo de ADN diferente e apenas quando os dois sítios de reconhecimento estão em grande proximidade fazem os monómeros inactivos juntar-se para criar uma enzima funcional. Como é necessária uma ligação de ADN para activar a nuclease, pode-se criar uma enzima de restrição altamente específica do sítio.

Uma TALEN específica da sequência, tal como se providencia aqui, pode reconhecer uma sequência particular dentro de uma sequência de nucleótidos-alvo pré-seleccionada presente numa célula. Assim, nalgumas modalidades, pode-se rastrear uma sequência-alvo de nucleótidos quanto aos sítios de reconhecimento de nuclease e uma nuclease particular pode ser seleccionada com base na sequência-alvo. Noutros casos, uma TALEN pode ser tratada por engenharia para atingir uma sequência celular particular. Pode-se inserir uma sequência de nucleótidos que codifica a TALEN desejada em qualquer vector de expressão apropriado e pode-se ligá-la a uma ou mais sequências de expressão de controlo. Por exemplo, uma nuclease que codifica uma sequência pode estar operacionalmente ligada a uma sequência promotora que irá conduzir à expressão constitutiva da endonuclease nas espécies de plantas a serem transformadas. Alternativamente, uma sequência de codificação de endonuclease pode estar operacionalmente ligada a uma sequência promotora que poderá levar à expressão condicional (por exemplo, a expressão em certas condições nutricionais). Por exemplo, um promotor do vírus em mosaico da couve-flor 35S pode ser utilizado para a expressão constitutiva. Outros promotores constitutivos

incluem, sem limitação, o motor de nopalina-sintase, o promotor de ubiquitina e o promotor de actina. Nalgumas modalidades, pode-se utilizar um promotor induzido por um estrogénio artificial para a expressão condicional e podem conseguir-se elevados níveis de transcrição quando uma planta é exposta a estrogénio. Outros promotores condicionais que podem ser utilizados incluem, por exemplo, promotores genéticos de choque térmico indutível pelo calor e promotores regulados pela luz, tais como os que derivam de genes que codificam a subunidade grande de ribulose-bisfosfato-carboxilase.

Com fins terapêuticos, o ADN efector TAL que modifica a enzima da presente invenção e um excipiente aceitável sob o ponto de vista farmacêutico são administrados numa quantidade eficaz sob o ponto de vista terapêutico. Diz-se que essa combinação é administrada numa "quantidade eficaz sob o ponto de vista terapêutico" se a quantidade administrada for significativa sob o ponto de vista fisiológico. Um agente é significativo sob o ponto de vista fisiológico se da sua presença resulta uma alteração detectável da fisiologia do receptor. No presente contexto, o agente é significativo sob o ponto de vista fisiológico se a sua presença resulta num decréscimo e na gravidade de um ou mais sintomas de doença-alvo e numa correcção do genoma da lesão ou da anomalia. Os vectores que contêm ADN recombinante e/ou ácidos nucleicos que codificam um ADN efector TAL que modifica enzimas podem ser introduzidos numa célula por uma variedade de processos (por exemplo, injeção, ingestão directa, bombardeamento de projecteis, lipossomas, electroporação). O ADN efector TAL que modifica as enzimas pode ser expresso estavelmente e de forma temporária nas células utilizando vectores de expressão. As técnicas de expressão em células eucarióticas são bem conhecidas dos especialistas nesta técnica. (Ver,

Current Protocols in Human Genetics: Capítulo 12 "Vectors For Gene Therapy" e Capítulo 13 "Delivery Systems for Gene Therapy").

Noutro aspecto da presente invenção, o ADN do efector TAL que modifica a enzima é praticamente não-imunogénico, isto é, provoca muito pouca ou nenhuma resposta imunológica adversa. Pode-se utilizar uma variedade de processos para melhorar ou eliminar as reacções imunológicas prejudiciais deste tipo, de acordo com a presente invenção. Numa modalidade preferida, o ADN efector TAL que modifica a enzima está praticamente isento de N-formil-metionina. Outra maneira de evitar reacções imunológicas indesejáveis consiste em conjugar o ADN efector TAL que codifica a enzima com polietileno-glicol ("PEG") ou polipropileno-glicol ("PPG") (preferencialmente, com um peso molecular (PM) médio de 500 a 20.000 daltons). A conjugação com PEG ou PPG, tal como descrito por Davis et al. (patente norte-americana 4.179.337) por exemplo, pode providenciar conjugados de ADN efector TAL que modifica as enzimas, solúveis em água, activos sob o ponto de vista fisiológico, não-imunogénico, com actividade antiviral. Há processos semelhantes que também utilizam um copolímero de polietileno-polipropileno-glicol, que já foram descritos por Saifer et al. (patente norte-americana 5 006 333).

Vectores dadores

A presente invenção também tem por objecto ácidos nucleicos recombinantes incluindo sequências de ácidos nucleicos dadores. Uma sequência de nucleótidos dadores pode incluir uma variante da sequência com uma ou mais modificações (isto é, substituições, eliminações ou inserções), no que se refere a uma sequência-alvo de

nucleótidos pré-selecionada encontrada endogenamente dentro do genoma de uma célula a ser transformada (também referida aqui como "uma sequência de nucleótidos recombinantes modificados"). A variante da sequência dentro do ácido nucleico dador normalmente está flanqueada em ambos os lados com sequências que são semelhantes ou idênticas à sequência de nucleótidos recombinante endógena dentro da célula. As sequências de flanqueio podem ter qualquer comprimento apropriado e normalmente têm pelo menos 50 nucleótidos de comprimento (por exemplo, pelo menos 50 nucleótidos, pelo menos 75 nucleótidos, pelo menos 100 nucleótidos, pelo menos 200 nucleótidos, pelo menos 250 nucleótidos, pelo menos 300 nucleótidos, pelo menos 500 nucleótidos, pelo menos 750 nucleótidos, pelo menos 1000 nucleótidos, de cerca de 50 a cerca de 5 000 nucleótidos, de cerca de 100 a 2.500 nucleótidos, de cerca de 100 a cerca de 1000 nucleótidos, de cerca de 100 a 500 nucleótidos, de cerca de 200 a cerca de 500 nucleótidos ou de cerca de 250 a 400 nucleótidos). Assim, A recombinação homóloga pode ocorrer entre a estrutura de ácido nucleico recombinante dador e o alvo endógeno em ambos os lados da sequência variante, de tal forma que o genoma das células resultante contém a variante da sequência dentro do contexto das sequências endógenas, por exemplo, a partir do mesmo gene. Pode-se gerar uma sequência dadora de nucleótidos para atingir qualquer sequência apropriada dentro de um genoma. Numa planta, por exemplo, uma sequência dadora de nucleótidos pode ser recombinada com um gene biossintético lípido, um gene biossintético de hidratos de carbono, um gene de proteínas que armazenam sementes, um gene de doenças ou de resistência a doenças, um gene tolerante do stresse, um gene tolerante a ácidos ou um gene que produz um elemento anti-nutricional. Além disso, a sequência dadora de nucleótidos contém um sítio de reconhecimento para uma nuclease específica da sequência, tal como aqui descrito.

Marcadores seleccionáveis

Alguns dos processos providenciados na presente invenção a utilização de um terceiro ácido nucleico recombinante que codifica um marcador seleccionável ou rastreável. Uma sequência de nucleótidos que codifica um polipéptido que resulta num traço seleccionável pode ser incorporado num vector de expressão contendo uma ou mais sequências de controlo da expressão. Por exemplo, um vector de expressão pode incluir uma sequência que codifica um marcador seleccionável ligada operacionalmente a uma sequência promotora que vai levar à expressão constitutiva na célula da planta a ser transformada. Os marcadores seleccionáveis apropriados podem incluir, sem limitações, polipéptidos que conferem resistência a um antibiótico, tal como canamicina, G418, bleomicina, ampicilina ou higromicina ou um herbicida, tal como glufosinato, cloro-sulfuron ou fosfinotricina.

Em modalidades para serem utilizadas em plantas, por exemplo, um marcador seleccionável pode conferir resistência a um herbicida que inibe o ponto de crescimento ou o meristema, tal como uma imidazolinona ou uma sulfonilureia. Exemplos de polipéptidos desta categoria para as enzimas ALS e AHAS mutantes estão descritas, por exemplo, nas patentes norte-americanas N.º. 5 767.366 e 5 928 937. As patentes norte-americanas N.º. 4 761 373 e 5 013 659 têm por objecto plantas resistentes a várias herbicidas de imidazolinona ou sulfonamida. A patente norte-americana N.º. 4 975 374 tem por objecto células de plantas e plantas que contém um gene que codifica uma glutamina-sintetase (GS) mutante resistente à inibição por herbicidas que são conhecidas por inibirem GS, por exemplo, fosfinotricina e metionina-sulfoximina. A patente norte-americana N.º. 5 162 602 tem por objecto plantas resistentes à inibição por ciclo-hexanodiona e herbicidas de

ácido ariloxifenoxipropanóico. A resistência é conferida por acetil-co-enzima A-carboxilase alterada (ACCase).

Os polipéptidos para conferirem resistência a glifosato (vendidos sob o nome comercial Roundup®) também são apropriados para serem utilizados nas plantas. Ver, por exemplo, as patentes norte-americanas N°. 4 940835 e 4 769 061. A patente norte-americana N°. 5 554.798 descreve plantas transgênicas de milho resistentes a glifosato, em que a resistência é conferida por uma 5-enolpiruvil-3-fosfochiquimato (EPSP)-sintase. Esses polipéptidos podem conferir resistência a composições herbicidas de glifosato incluindo, sem limitação, sais de glifosato, tais como sal de trimetilsulfônio, sal de isopropilamina, sal de sódio, sal de potássio e sal de amônio. Ver, por exemplo, as patentes norte-americanas N°. 6 451 735 e 6 451 732.

Os polipéptidos para a resistência a compostos de fosfono, tais como, glufosinato-amônio ou fosfinotricina e ácidos piridinoxi ou fenoxi-propiónicos e ciclo-hexonas são também apropriados. Ver, por exemplo, a patente europeia N°. 0 242 246, assim como, as patentes norte-americanas N°. 5 879 903, 5 276 268 e 5 561 236.

Outros herbicidas incluem aqueles que inibem a fotossíntese, tais como triazina e benzonitrilo (nitrilase). Ver, por exemplo, a patente norte-americana N°. 4.810.648. Outros herbicidas incluem ácido 2,2-dicloropropiónico, setoxidim, halogenoxifop, herbicidas de imidazolinona, herbicidas de sulfonilureia, herbicidas de triazol-pirimidina, herbicidas de s-triazina e bromoxinilo. Também são apropriados herbicidas que conferem resistência a uma enzima protox. Ver, por exemplo, a publicação das patentes norte-americanas N°. 20010016956 e 6 084 155.

Nalgumas modalidades, pode-se adaptar um ácido nucleico recombinante que codifica um marcador seleccionável para integrá-lo no genoma de uma célula (por exemplo, uma célula de planta ou uma célula de animal) por recombinação específica do sítio. Por exemplo, uma sequência que codifica um marcador seleccionável pode ser flanqueada por sequências de reconhecimento para uma recombinase, tal como, por exemplo, Cre ou FLP. Noutras modalidades, pode adaptar-se um ácido nucleico recombinante que codifica um marcador seleccionável, para a integração num genoma de planta por recombinação homóloga. Nesses ácidos nucleicos, a sequência que codifica o marcador seleccionável pode ser flanqueada por sequências que são semelhantes ou idênticas às sequências endógenas de nucleótidos encontradas dentro do genoma da célula da planta, em que o ácido nucleico recombinante vai ser introduzido. Pelo menos uma das sequências endógenas pode estar no sítio de clivagem para uma nuclease específica da sequência. O ácido nucleico que codifica o marcador seleccionável também pode conter um sítio de reconhecimento para uma nuclease específica da sequência. A nuclease pode ser a mesma nuclease específica da sequência que foi recombinada com a sequência dadora de nucleótidos ou uma nuclease específica da sequência que é diferente da sequência que foi recombinada com a sequência de nucleótidos dadores. Ainda noutras modalidades, Um ácido nucleico recombinante que codifica um marcador seleccionável pode ser adaptado para integração num genoma de uma célula de planta por uma recombinação ilegítima. Esses ácidos nucleicos normalmente não têm as sequências de flanqueio e os sítios de reconhecimento de nuclease estão contidos dentro dos ácidos nucleicos adaptados para recombinação homóloga ou específica do sítio, tal como se descreve aqui.

Processos

Uma ou mais das estruturas providenciadas pela presente invenção podem ser utilizadas para transformar células e/ou uma enzima que modifica o ADN pode ser introduzida nas células, de tal maneira que se gere um organismo geneticamente modificado (por exemplo, uma planta ou um animal). Assim, os organismos e as células geneticamente modificados que contêm os ácidos nucleicos e/ou os polipéptidos aqui descritos também constituem objecto da presente invenção. Nalgumas modalidades, uma célula transformada tem uma estrutura de ácido nucleico recombinante integrada no seu genoma, isto é, pode ser transformada de forma estável. As células transformadas de forma estável normalmente retêm a sequência, de ácidos nucleicos, introduzida em cada divisão de células. Uma estrutura pode integrar-se de uma forma homóloga, de tal modo que a sequência de nucleótidos endógena em relação à célula transformada é substituída pela estrutura, em que a estrutura contém uma sequência que corresponde à sequência endógena, mas que contém uma ou mais modificações no que respeita à sequência endógena. Deve notar-se que embora uma planta ou um animal que contenha essa sequência endógena modificada possa ser designada por um "organismo geneticamente modificado" (OMG) a sequência endógena modificada não é considerada um transgene. Uma estrutura também pode integrar-se de uma maneira ilegítima, de tal modo que se integra aleatoriamente no genoma da célula transformada.

Alternativamente, uma célula pode ser transformada de forma temporária, de tal forma que a estrutura não se integra em todo o seu genoma. Por exemplo, pode-se introduzir um vector de plasmido contendo uma sequência de codificação TALEN numa célula, de tal modo que a sequência de codificação

TALEN é expressa mas o vector não é integrado de forma estável no genoma. As células transformadas temporariamente normalmente perdem algumas ou todas as estruturas de ácidos nucleicos introduzidas em cada uma das divisões de células, de tal modo que o ácido nucleico introduzido não pode ser detectado nas células-filhas após um número suficiente de divisões de células. Apesar disso, a expressão de sequências de codificação TALEN é suficiente para se conseguir a recombinação homóloga entre uma sequência dadora e uma sequência-alvo endógena. Tanto as células transformadas temporariamente como as células transformadas estavelmente podem ser úteis nos processos aqui descritos.

No que respeita particularmente às células de plantas modificadas geneticamente, as células utilizadas nos processos aqui descritos podem constituir parte ou toda uma planta. Essas plantas podem crescer de uma maneira apropriada para as espécies que se considerarem, quer numa câmara de crescimento, numa estufa ou num campo. As plantas geneticamente modificadas podem ser criadas, conforme desejado, para um fim particular, por exemplo, para introduzir um ácido nucleico recombinante noutras linhas de células, para transferir um ácido nucleico recombinante para outras espécies ou para uma posterior selecção de outros traços desejáveis. Alternativamente, as plantas geneticamente modificadas podem propagar-se vegetativamente para essas espécies que se podem gerir por estas técnicas. A progenia inclui descendentes de uma planta ou de uma linha de plantas particular. A progenia de uma dada planta inclui sementes formadas em F_1 , F_2 , F_3 , F_4 , F_5 , F_6 e plantas da geração subsequente ou sementes formadas em BC_1 , BC_2 , BC_3 e plantas da geração subsequente ou sementes formadas em F_1BC_1 , F_1BC_2 , F_1BC_3 e plantas da geração subsequente. As sementes produzidas por uma planta geneticamente modificada podem crescer e depois

auto-alimentar-se (ou reticularem e auto-alimentar-se) para se obterem sementes homozigóticas para a estrutura de ácidos nucleicos.

As células geneticamente modificadas (por exemplo, células de plantas ou células de animais) podem crescer numa cultura em suspensão ou num tecido ou numa cultura de órgãos, se desejado. Para os objectivos dos processos da presente invenção, podem utilizar-se técnicas de cultura em tecido sólido e/ou líquido. Quando se utiliza um meio sólido, as células podem ser colocadas directamente num meio ou podem ser colocadas numa película de um filtro que é depois colocado em contacto com o meio. Quando se utiliza um meio líquido, as células podem ser colocadas num dispositivo de flotação, por exemplo, uma membrana porosa que contacta com o meio líquido. O meio sólido é normalmente constituído por um meio líquido por meio da adição de agar. Por exemplo, um meio sólido pode ser um meio de Murashige e Skoog (MS) contendo agar e auxina numa concentração apropriada, por exemplo, ácido 2,4-diclorofenoxiacético (2,4-D) e uma citocinina numa concentração apropriada, por exemplo, cinetina.

Uma célula pode ser transformada com uma estrutura do ácido nucleico recombinante ou com várias estruturas (por exemplo, 2, 3, 4 ou 5) de ácidos nucleicos recombinantes. Se se utilizam várias estruturas, elas podem ser transformadas simultaneamente ou sequencialmente. As técnicas para a transformação de uma vasta gama de espécies são conhecidas na técnica. Os polinucleótidos e os vectores recombinantes aqui descritos podem ser introduzidos num genoma de um hospedeiro utilizando qualquer um dos numerosos processos conhecidos incluindo electroporação, micro-injecção e processos biolísticos. Alternativamente, podem combinar-se os polinucleótidos ou os vectores com regiões apropriadas de

flanqueio de ADN-T e introduzindo-se num vector hospedeiro convencional de *Agrobacterium tumefaciens*. Essas técnicas de transformação mediadas por *Agrobacterium tumefaciens* incluindo o desarmamento e a utilização de vectores binários são bem conhecidas na técnica. Outras técnicas de transferência de genes e de transformação incluem a transformação dos protoplastos através de cálcio ou PEG, absorção mediada por electroporação de ADN nu, transfecção mediada por lipossomas, electroporação, transformação mediada por vectores virais e bombardeamento de micro-projecteis (ver, por exemplo, as patentes norte-americanas 5 538 880, 5 204 253, 5 591 616 e 6 329 571). Se se utilizar uma cultura de células de plantas ou tecidos de plantas como o tecido receptor para a transformação, as plantas podem ser regeneradas a partir das culturas transformadas utilizando técnicas conhecidas dos especialistas na matéria.

Nalgumas modalidades, pode-se introduzir, directamente na célula, uma enzima que modifica o ADN (por exemplo, uma TALEN). Por exemplo, pode-se introduzir um polipéptido numa célula por injeção mecânica, por administração por via de um sistema de secreção de bactérias do tipo III, por electroporação ou por transferência mediada por *Agrobacterium*. Ver, por exemplo, Vergunst et al. (2000) Science 290: 979-982 para uma discussão do sistema de transporte VirB/D4 de *Agrobacterium* e a sua utilização para mediar a transferência de um complexo T de nucleoproteína em células de plantas.

Ainda no que respeita às plantas, os polinucleótidos, vectores e polipéptidos aqui descritos podem ser introduzidos num certo número de plantas monocotiledóneas e dicotiledóneas e em sistemas de células de plantas, incluindo dicotiledóneas, tais como açafrão, lucerna, soja, café,

amarante, colza (com um elevado teor de ácido erúxico e canola), amendoim ou girassol, assim como, plantas monocotiledóneas, tais como óleo de palma, açúcar de cana, banana, sorgo, milho, trigo, centeio sevada, aveia, arroz, milho painço ou sorgo. Também são apropriadas gimnospermas, tais como abeto e pinheiro.

Assim, os processos aqui descritos podem ser utilizados com plantas dicotiledóneas, por exemplo, às ordens *Magniolales*, *Illiciales*, *Laurales*, *Piperales*, *Aristolochiales*, *Nymphaeales*, *Ranunculales*, *Papeverales*, *Sarraceniaceae*, *Trochodendrales*, *Hamamelidales*, *Eucomiales*, *Leitneriales*, *Myricales*, *Fagales*, *Casuarinales*, *Caryophyllales*, *Batales*, *Polygonales*, *Plumbaginales*, *Dilleniales*, *Theales*, *Malvales*, *Urticales*, *Lecythidales*, *Violales*, *Salicales*, *Capparales*, *Ericales*, *Diapensales*, *Ebenales*, *Primulales*, *Rosales*, *Fabales*, *Podostemales*, *Haloragales*, *Myrtales*, *Cornales*, *Proteales*, *Santales*, *Rafflesiales*, *Celastrales*, *Euphorbiales*, *Rhamnales*, *Sapindales*, *Juglandales*, *Geraniales*, *Polygalales*, *Umbellales*, *Gentianales*, *Polemoniales*, *Lamiales*, *Plantaginales*, *Scrophulariales*, *Campanulales*, *Rubiales*, *Dipsacales* e *Asterales*. Os processos aqui descritos também podem ser utilizados com plantas monocotiledóneas, tais como as que pertencem às ordens de *Alismatales*, *Hydrocharitales*, *Najadales*, *Triuridales*, *Commelinales*, *Eriocaulales*, *Restionales*, *Poales*, *Juncuales*, *Cyperales*, *Typhales*, *Bromeliales*, *Zingiberales*, *Arecales*, *Cyclanthales*, *Pandanales*, *Arales*, *Lilliales* e *Orchidales* ou com plantas pertencendo a *Gymnospermae*, por exemplo, *Pinales*, *Ginkgoales*, *Cycadales* e *Gnetales*.

Os processos podem ser utilizados para uma vasta gama de espécies de plantas, incluindo plantas do género dicotiledóneo *Atropa*, *Alseodaphne*, *Anacardium*, *Arachis*,

Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Piston, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca, Vitis e Vigna; do género monocotiledóneo *Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum e Zea;* ou do género gimnosperma *Abies, Cunninghamia, Picea, Pinus e Pseudotsuga.*

Pode-se identificar uma célula, calo, tecido ou planta transformada e isolá-la seleccionando ou rastreando células tratadas por engenharia no que respeita a traços ou actividades particulares, por exemplo, as codificadas por genes marcadores ou genes de resistência a antibióticos. Essas metodologias de rastreio e de selecção são bem conhecidas dos especialistas na matéria. Além disso, podem utilizar-se processos físicos e bioquímicos para identificar transformantes. Estes processos incluem análise de Southern ou amplificação por RCP para a detecção de um polinucleótido; análise de "Northern blot", protecção de RNase de S1, extensão de iniciador ou amplificação por RCP-TI para a detecção de transcriptos de ARN; os ensaios enzimáticos, para a detecção da actividade de enzimas ou de ribozimas dos polipéptidos e polinucleótidos; e electroforese em gel de proteínas, "Western blot", imuno-precipitação e imuno-ensaios ligados a enzimas para detectar polipéptidos. Também se podem utilizar outras técnicas, tais como hibridação *in situ*,

coloração de enzimas e imuno-coloração para detectar a presença ou a expressão de polipéptidos e/ou polinucleótidos. Os processos para a realização de todas as técnicas referenciadas são bem conhecidas. Os polinucleótidos que são incorporados de forma estável em células de plantas podem ser introduzidos em outras plantas utilizando, por exemplo, técnicas normalizadas de melhoramento genético.

No contexto da presente invenção, "células eucarióticas" referem-se a células de fungos, leveduras, plantas ou animais ou a linhas de células derivadas dos organismos listados a seguir e estabelecidas para uma cultura *in vitro*. Mais preferencialmente, os fungos podem ser dos géneros *Aspergillus*, *Penicillium*, *Acremonium*, *Trichoderma*, *Chrysosporium*, *Mortierella*, *Kluyveromyces* ou *Pichia*. Mais preferencialmente, os fungos podem ser das espécies *Aspergillus niger*, *Aspergillus nidulans*, *Aspergillus oryzae*, *Aspergillus terreus*, *Penicillium chrysogenum*, *Penicillium citrinum*, *Acremonium chrysogenum*, *Trichoderma reesei*, *Mortierella alpine*, *Chrysosporium lucknowense*, *Kluyveromyces lactis*, *Pichia pastoris* ou *Pichia ciferrii*.

Na presente invenção, a planta pode ser do género *Arabidopsis*, *Nicotiana*, *Solanum*, *Lactuca*, *Brassica*, *Oryza*, *Asparagus*, *Pisum*, *Medicago*, *Zea*, *Hordeum*, *Secale*, *Tritium*, *Capsicum*, *Cucumis*, *Cucurbita*, *Citrullis*, *Citrus* ou *Sorghum*. Mais preferencialmente, a planta pode ser das espécies *Arabidopsis thaliana*, *Nicotiana tabaccum*, *Solanum lycopersicum*, *Solanum tuberosum*, *Solanum melongena*, *Solanum esculentum*, *Lactuca saliva*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Oryza glaberrima*, *Oryza sativa*, *Asparagus officinalis*, *Pisum sativum*, *Medicago sativa*, *Zea mays*, *Hordeum vulgare*, *Secale cereal*, *Tritium aestivum*, *Triticum durum*, *Capsicum sativus*, *Cucurbita pepo*, *Citrullus*

lanatus, Cucumis melo, Cirrus aurantifolia, Citrus maxima, Citrus media ou *Citrus reticulata*.

Na presente invenção, a célula de animal pode ser do género *Homo, Rattus, Mus, Sus, Bos, Danio, Canis, Felis, Equus, Salmo, Oncorhynchus, Galls, Meleagris, Drosophila* ou *Caenorhabditis*; mais preferencialmente, a célula de animal pode ser das espécies *Homo sapiens, Rattus norvegicus, Mus musculus, Sus scrofa, Bos taurus, Danio rerio, Canis lupus, Felis catus, Equus caballus, Oncorhynchus mykiss, Gallus gallus* ou *Meleagris gallopavo*; a célula de animal pode ser uma célula de peixe retirada de *Salmo salar*, peixe teleósteos ou espécies de peixe-zebra como exemplos não limitativos. A célula de animal na presente invenção também pode ser uma célula de insecto de *Drosophila melanogaster* como exemplo não limitativo; a célula de animal pode também ser uma célula de verme de *Caenorhabditis elegans*, como um exemplo não limitativo.

Na presente invenção, a célula pode ser uma célula de planta, uma célula de mamífero, uma célula de peixe, uma célula de insecto ou linhas de células derivadas destes organismos para culturas *in vitro* de células primárias retiradas directamente do tecido vivo e estabelecidas para cultura *in vitro*. Como exemplos não limitativos de linhas de células podem seleccionar-se no grupo que consiste em células de OHC-K1; células HEK293; células Caco2; células U2-OS; células NIH 3T3; células NSO; células SP2; células OHC-S; células DG44; células K-562, células U-937; células MRC5; células IMR90; células Jurkat; células HepG2; células HeLa; células HT-1080; células HCT-116; células Hu-h7; células Huvec; células Molt 4.

Todas estas linhas de células podem ser modificadas pelo processo da presente invenção, para se obter modelos de linhas de células para produzir, expressar, quantificar, detectar, estudar, um gene ou uma proteína de interesse. Estes modelos também podem ser utilizados para rastrear moléculas activas sob o ponto de vista biológico com interesse na investigação e produção em vários domínios, tais como na química, nos biocombustíveis, em agentes terapêuticos e na agronomia, citando apenas exemplos não limitativos.

A presente invenção também tem por objecto processos para o aproveitamento de domínios de ligação do ADN específicos de sequências dentro de efectores TAL para, por exemplo, alterar o material genético dentro das células, regular a expressão de genes e recombinar sequências patogénicas, por exemplo, em terapias antivirais. Por exemplo, nalgumas modalidades, a presente invenção tem por objecto processos para modificar o material genético celular. Nalgumas modalidades, os processos incluem a introdução de um polipéptido contendo um domínio de ligação de ADN de efector TAL ou um ácido nucleico que codifica esse péptido, numa célula. O domínio de ligação de ADN do efector TAL pode fundir-se com toda ou com uma porção de uma enzima modificadora de ADN (por exemplo, uma endonuclease). Nalgumas modalidades, os processos incluem a introdução de dois ou mais ácidos nucleicos recombinantes na célula. Um primeiro ácido nucleico recombinante contém uma sequência dadora de nucleótidos que inclui uma ou mais modificações (isto é, substituições, eliminações ou inserções), no que respeita a uma sequência de nucleótidos recombinante pré-seleccionada correspondente, encontrada na célula. A sequência dadora de nucleótidos pode sofrer uma recombinação homóloga com a sequência-alvo endógena de nucleótidos, de tal modo que a sequência endógena ou uma porção dessa sequência é

substituída por uma sequência dadora ou por uma porção dela. A sequência de nucleótidos recombinados normalmente inclui um sítio de reconhecimento para uma TALEN específica da sequência. Nalguns casos uma sequência-alvo de nucleótidos pode incluir sítios de reconhecimento para dois ou mais TALEN distintas (por exemplo, duas sequências-alvo opostas que são distintas, de tal maneira que se podem utilizar TALEN com uma especificidade distinta da ligação às sequências de ADN). Nesses casos, a especificidade da clivagem de ADN pode ser aumentada em comparação com os casos em que apenas uma sequência-alvo (ou várias cópias da mesma sequência-alvo) é utilizada.

Um segundo ácido nucleico recombinante contém uma sequência de nucleótidos que codifica uma sequência específica TALEN, que se liga a um sítio de reconhecimento na sequência-alvo de nucleótidos. Nalguns casos, a sequência dadora de nucleótidos e a sequência de nucleótidos que codifica a nuclease específica da sequência podem estar contidas na mesma estrutura de ácidos nucleicos. Alternativamente, a sequência dadora de nucleótidos e a sequência de codificação TALEN podem estar contidas em estruturas separadas ou o polipéptido TALEN pode ser produzido e introduzido directamente numa célula.

Nalgumas modalidades, pode-se utilizar um terceiro ácido nucleico recombinante contendo uma sequência de nucleótidos que codifica um marcador seleccionável. O segundo e o terceiro ácidos nucleicos recombinantes podem sofrer recombinação com sequências endógenas e assim integrar-se no genoma da célula. Estes eventos de recombinação podem ser ilegítimos (isto é, aleatórios) ou podem ocorrer através de recombinação homóloga ou através de recombinação específica do sítio. Os ácidos nucleicos recombinantes podem ser

transformados simultaneamente ou sequencialmente dentro da célula e podem ser linearizados antes da transformação.

Quando a célula é uma célula de planta, os processos aqui providenciados podem ainda incluir etapas, tais como a geração de uma planta contendo a célula transformada, a geração de progenia da planta, a selecção ou o rastreio de plantas que expressam o marcador seleccionável (se estiver incluído), a geração de progenia das plantas seleccionadas e o ensaio das plantas (por exemplo, tecido, sementes, células precursoras ou plantas completas) ou a progenia das plantas para recombinação da sequência-alvo de nucleótidos. Nalguns casos, os processos podem incluir a exogamia das plantas seleccionadas para eliminar o marcador seleccionável e/ou o rastreio das plantas seleccionadas ou resultantes da exogamia quando à ausência da nuclease específica da sequência.

Nalgumas modalidades, a presente invenção tem por objecto processos para modificar o material genético de uma célula, por exemplo, uma célula procariótica, uma célula de animal ou uma célula de planta. Os processos podem incluir a introdução na célula de um primeiro ácido nucleico recombinante contendo uma sequência-alvo de nucleótidos modificada, que inclui uma ou mais modificações na sequência de nucleótidos no que respeita a uma sequência-alvo de nucleótidos correspondente presente na célula, assim como, um sítio de reconhecimento para uma TALEN específica da sequência e um segundo ácido nucleico recombinante contendo uma sequência de nucleótidos que codifica a TALEN específica da sequência. Quando a célula é uma célula de plantas, pode-se gerar uma planta contendo a célula e células, sementes ou tecidos obtidos a partir da planta (ou da sua progenia), podem ser analisados para recombinação na sequência-alvo de nucleótidos. O primeiro e o segundo ácidos nucleicos

recombinantes podem ser transformados simultaneamente ou em série numa célula e um ou ambos podem ser linearizados antes da transformação. Nalguns casos, o primeiro e o segundo ácidos nucleicos recombinantes podem estar presentes na mesma estrutura.

Nalguns casos, o processo também pode incluir a introdução na célula de um terceiro ácido nucleico recombinante contendo uma sequência de nucleótidos que codificam um marcador seleccionável e a determinação se a célula ou um organismo gerado a partir da célula ou a sua progenia expressam um marcador seleccionável. O processo ainda pode incluir o rastreio da célula, do organismo ou da sua progenia quanto à ausência do marcador seleccionável. A sequência de nucleótidos que codifica o marcador seleccionável pode ou não estar flanqueada em ambos os lados por sequências de nucleótidos que são semelhantes ou idênticas às sequências de nucleótidos endógenas à célula no sítio de clivagem para uma segunda nuclease específica da sequência ou por sítios de reconhecimento para uma recombinase específica da sequência. Nalguns casos, o processo também pode incluir a etapa de exogamia do organismo. A progenia resultante da exogamia pode ser rastreada quanto à ausência do marcador seleccionável.

A presente invenção também tem por objecto processos para a modificação do material genético numa célula (por exemplo, uma célula de planta ou uma célula de animal), compreendendo providenciar uma célula contendo uma sequência-alvo de ADN, por exemplo, uma sequência cromossómica, mitocôndrica ou cloroplasto, em que se deseja que ocorra a recombinação homóloga, providenciando uma TALEN que contém um domínio de enzima que modifica o ADN (por exemplo, um domínio de endonuclease) e um domínio efector TAL com várias

repetições de efectores TAL que, em combinação, se ligam a uma sequência de nucleótidos específico dentro da sequência-alvo de ADN, providenciando um ácido nucleico contendo uma sequência homóloga com pelo menos uma porção do ADN-alvo e fazendo contactar a sequência-alvo de ADN na célula com a endonuclease TAL, de tal modo que ambas as hélices de uma sequência de nucleótidos dentro ou adjacente à sequência-alvo de ADN na célula são clivadas. Essa clivagem pode aumentar a frequência de recombinação homóloga na sequência-alvo de ADN. A sequência-alvo de ADN pode ser endógena à célula. Os processos podem incluir a introdução na célula de um vector contendo um ADNc que codifica a endonuclease TAL e que expressa uma proteína de endonuclease TAL na célula. Nalguns casos, a própria proteína da endonuclease TAL pode ser introduzida na célula, por exemplo, por injeção mecânica, por administração por via de um sistema de secreção de bactérias do tipo III, por electroporação ou por transferência mediada por *Agrobacterium*.

Os processos aqui descritos podem ser utilizados em várias situações. Na agricultura, por exemplo, os processos aqui descritos são úteis para facilitar a recombinação homóloga num sítio-alvo, que pode ser utilizado para eliminar um transgene previamente integrado (por exemplo, um transgene de resistência a herbicidas) de uma linha de plantas, de uma variedade ou de híbridos. Os processos aqui descritos também podem ser utilizados para modificar um gene endógeno de tal modo que a enzima codificada pelo gene confere resistência a herbicidas, por exemplo, modificação de um gene endógeno de 5-enolpiruvil-chiquimato-3-fosfato (EPSP)-sintase, de tal modo que a enzima modificada confere resistência a herbicidas de glifosato. Como outro exemplo, os processos aqui descritos são úteis para facilitar a recombinação homóloga em regiões reguladoras para um ou mais genes endógenos numa planta ou

numa via metabólica de mamíferos (por exemplo, biossíntese de ácidos gordos), de tal modo que a expressão desses genes é modificada de uma forma desejada. Os processos aqui descritos são úteis para facilitar a recombinação homóloga num animal (por exemplo, num rato ou num murganho) em um ou mais genes endógenos de interesse envolvidos, embora os exemplos não sejam limitativos em vias de sinalização metabólicas e internas, tais como as que codificam marcadores das superfícies das células, genes identificados como estando ligados a uma doença particular e quaisquer genes conhecidos como sendo responsáveis por um fenótipo particular de uma célula de animal.

A presente invenção também tem por objecto processos para conceber efectores TAL específicos de sequências capazes de interagir com sequências particulares de ADN (por exemplo, TALEN capaz de clivar o ADN em localizações específicas). Os processos podem incluir a identificação de uma sequência-alvo de nucleótidos (por exemplo, uma sequência cromossómica endógena, uma sequência de ADN mitocôndrico ou uma sequência de ADN cloroplastos) na qual se deseja ter uma ligação do efector TAL (por exemplo, uma sequência adjacente a uma segunda sequência de nucleótidos na qual se deseja introduzir um corte da hélice dupla) e concebendo uma sequência específica do efector TAL, que contenha várias repetições da ligação de ADN que, em combinação, se ligam à sequência-alvo. Tal como aqui descrito, os efectores TAL incluem um certo número de repetições imperfeitas que determinam a especificidade com que interagem com o ADN. Cada repetição liga-se a uma única base consoante a sequência particular de di-aminoácidos nos resíduos 12 e 13 da repetição. Assim, por engenharia genética podem atingir-se as repetições dentro de um efector TAL (por exemplo, utilizando técnicas normalizadas ou as técnicas aqui descritas), atingindo-se então sítios

particulares de ADN. Esses efectores TAL obtidos por engenharia genética podem ser utilizados, por exemplo, como factores de transcrição dirigidos a sequências particulares de ADN. Na figura 1A, ilustra-se um diagrama de um efector genérico TAL, com a região de repetição indicada pelas caixas abertas e a DVR na sequência representativa das repetições (SEQ ID N.º. 1) a sublinhado.

Exemplos de DVR e os seus correspondentes nucleótidos-alvo estão ilustrados no quadro 1A (Ver, também, a publicação da PCT N.º. WO2010/079430).

Quadro 1A

DVR	Nucleótido
HD	C
NG	T
NI	A
NN	G ou A
NS	A ou C ou G
N*	C ou T
HG	T
H*	T
IG	T
* Indica um intervalo na sequência da repetição correspondendo a uma falta do resíduo de aminoácido numa segunda posição do DVR.	

No quadro 1B ilustra-se outros DVR e os correspondentes nucleótidos-alvo.

Quadro 1B

DVR	Nucleótido
HA	C
ND	C
NK	G
HI	C
HN	G
NA	G
SN	G ou A
YG	T

Quando se deseja ter uma clivagem de ADN específica da sequência, pode-se conceber, por exemplo, uma TALEN específica da sequência que contenha (a) vários domínios de repetição da ligação de ADN que, em combinação, se ligam à sequência endógena de nucleótidos cromossômicos e (b) uma endonuclease que gera um corte de hélice dupla na segunda sequência de nucleótidos. Essa clivagem de ADN específica da sequência pode ser útil para aumentar a recombinação homóloga, tal como aqui descrito. Outras utilizações para as TALEN incluem, por exemplo, a sua utilização terapêutica contra vírus. As TALEN podem ser tratados por engenharia genética para recombinar sequências virais particulares, clivando o ADN viral e reduzindo ou abolindo a virulência.

Os materiais e os processos aqui providenciados podem ser utilizados para modificar a sequência de um gene particular de uma forma orientada. Um gene pode conter várias sequências nas quais se pode recombinar um efector TAL por engenharia genética. Tal como aqui descrito, contudo, algumas sequências-alvo podem ser recombinadas de forma mais eficaz. Por exemplo, tal como se mostra no exemplo 9, as sequências que têm características particulares podem ser recombinadas

de forma mais eficaz pelos efectores TAL. Assim, os processos aqui providenciados podem incluir a identificação de sequências-alvo que cumprem critérios particulares. Estes critérios incluem sequências que: i) têm um comprimento mínimo de 15 bases e uma orientação de 5' para 3', com um T a preceder imediatamente o sítio na extremidade 5' e; ii) não têm um T na primeira posição (5') ou um A na segunda posição; iii) terminam em T na última posição (3') e não têm um G próximo da última posição; e iv) têm uma composição de base de 0-63 % de A, 11-63 % de C, 0-25 % de G e 2-42 % de T.

Dado que as TALEN tal como aqui descrito geralmente funcionam como dímeros, algumas modalidades dos processos aqui providenciados podem incluir a identificação de uma primeira sequência genómica de nucleótidos e uma segunda sequência genómica de nucleótidos numa célula, em que a primeira e a segunda sequência de nucleótidos cumprem pelo menos um dos critérios estabelecidos antes e estão separadas por 15-18 pb. Nalguns casos, um polipéptido TALEN pode ligar-se a cada uma das sequências de nucleótidos e a endonuclease contida na TALEN pode clivar dentro do espaçador de 15-18 pb.

A presente invenção também tem por objecto processos para gerar animais geneticamente modificados nos quais se deseja introduzir um ácido nucleico. Esses processos podem incluir a obtenção de uma célula que contenha uma sequência de ADN recombinante cromossómico endógeno, em que se deseja introduzir o ácido nucleico, introduzindo na célula uma TALEN para gerar um corte da hélice dupla dentro da sequência de ADN recombinante cromossómico endógeno, introduzindo na célula um ácido nucleico exógeno contendo uma sequência homóloga a pelo menos uma porção do ADN recombinante cromossómico endógeno, em que a introdução é feita em condições que permitem que ocorra a recombinação homóloga

entre o ácido nucleico exógeno e o ADN-alvo cromossômico endógeno e gerando um animal a partir da célula primária, em que tenha ocorrido a recombinação homóloga. O ácido nucleico homólogo pode incluir, por exemplo, uma sequência de nucleótidos que vai causar a interrupção de um gene após a recombinação homóloga, uma sequência de nucleótidos que substitui um gene após a recombinação homóloga, uma sequência de nucleótidos que introduz uma mutação pontual num genes após a recombinação homóloga ou uma sequência de nucleótidos que introduz um sítio regulador após a recombinação homóloga.

Os processos aqui providenciados também podem ser utilizados para gerar plantas geneticamente modificadas, em que se introduz um ácido nucleico desejado. Esses processos podem incluir a obtenção de uma célula de uma planta contendo uma sequência de ADN recombinante endógeno, em que se deseja introduzir o ácido nucleico, introduzindo uma TALEN para gerar um corte de hélice dupla dentro da sequência de ADN recombinante endógeno, introduzindo na célula da planta um ácido nucleico exógeno contendo uma sequência homóloga pelo menos uma porção do ADN-alvo endógeno, em que a introdução é feita em condições que permitam a ocorrência da recombinação homóloga entre o ácido nucleico exógeno e o ADN-alvo endógeno e gerando uma planta a partir de uma célula de planta em que tenha ocorrido a recombinação homóloga.

O ADN nas células geradas pelos processos de recombinação homóloga facilitados por TALEN aqui providenciados, é modificado, em comparação com as células que não sofreram esses processos e as células que contêm o ADN modificado são referidas como "geneticamente modificadas". Deve notar-se, contudo, que os organismos que contêm essas células não podem ser considerados, para fins reguladores, como OGM, dado que essa modificação envolve uma

recombinação homóloga e não uma integração aleatória de um transgene. Assim, a utilização dos processos facilitados por TALEN aqui descritos para gerar modificações genéticas pode ser vantajosa pelo facto de, por exemplo, os procedimentos-padrão reguladores em conjunto com o tempo e o custo que lhe estão associados poderem ser evitados.

Outros processos de recombinação genética direccionada, tal como os providenciados aqui, podem incluir a introdução numa célula (por exemplo, uma célula de planta, uma célula de insecto, uma célula de peixe teleósteeo ou uma célula de animal) de uma molécula de ácido nucleico que codifica uma TALEN recombinada com uma sequência-alvo de ADN seleccionada, induzindo uma expressão da TALEN dentro da célula e identificando uma célula recombinante em que a sequência-alvo de ADN seleccionada exhibe uma (por exemplo, uma eliminação de material genético, uma inserção de material genético ou tanto uma eliminação como uma inserção de material genético). Também pode ser introduzido um ADN dador na célula.

Nalgumas modalidades, pode utilizar-se uma TALEN monomérica. As TALEN, tal como aqui descritos, funcionam normalmente como dímeros através de um sítio de reconhecimento bipartido com um espaçador, de tal modo que os dois domínios efectores TAL estão fundidos com um domínio catalítico de uma enzima de restrição *FokI*, e em que os sítios de reconhecimento de ADN para cada TALEN resultante estejam separados por uma sequência espaçadora e ligam, cada monómero TALEN a sítios de reconhecido, que permitem que *FokI* dimerize e crie uma quebra da hélice dupla dentro do espaçador (ver, por exemplo, Moscou e Bogdanove (2009) *Science* 326:1501). As TALEN monoméricas também podem ser construídas, contudo, de tal maneira que efectores TAL simples se fundem com uma nuclease que não exige a

dimerização para funcionar. Uma dessas nucleases é, por exemplo, uma variante de cadeia simples de *FokI*, em que os dois monómeros estão expressos como polipéptidos simples (Minczuk et al. (2008) *Nucleic Acids Res.* 36:3926-3938). Outras nucleases monoméricas de ocorrência natural ou originadas por engenharia genética também podem desempenhar este papel. O domínio de reconhecimento de ADN utilizado para uma TALEN monomérica pode derivado de um efector TAL de ocorrência natural. Alternativamente, o domínio de reconhecimento de ADN pode ser tratado por engenharia genética para reconhecer um alvo específico de ADN. As TALEN de cadeia simples resultantes da engenharia podem ser mais fáceis para construir e organizar, já que apenas exigem um domínio de reconhecimento de ADN tratado por engenharia.

Nalgumas modalidades, pode-se gerar uma nuclease dimérica específica de uma sequência de ADN utilizando dois domínios diferentes de ligação de ADN (por exemplo, um domínio de ligação do efector TAL e um domínio de ligação de outro tipo de molécula. Tal como afirmado antes, as TALEN aqui descritas normalmente funcionam como dímeros de um sítio bipartido de reconhecimento com um espaçador. Esta arquitectura da nuclease também pode ser utilizada para nucleases específicas do alvo, geradas a partir de, por exemplo, um monómero TALEN e um monómero de nuclease de dedos de zinco. Nesses casos, os sítios de reconhecimento de ADN para a TALEN e os monómeros de nuclease de dedos de zinco podem ser separados por um espaçador com o comprimento apropriado. A ligação dos dois monómeros pode fazer com que *FokI* dimerize e crie uma quebra da hélice dupla dentro da sequência espaçadora. Os domínios de ligação do ADN, diferentes dos dedos de zinco, tais como os homeo-domínios, as repetições de mib ou os fechos de leucina, também se podem

fundir com *FokI* e servir como um parceira para que um monómero TALEN crie uma nuclease funcional.

Nalgumas modalidades, pode-se utilizar um efector TAL para atingir outros domínios da proteína (por exemplo, domínios da proteína que não são nucleases) para sequências de nucleótidos específicas. Por exemplo, pode-se ligar um efector TAL a um domínio de proteína, sem limitações, a partir de uma enzima que interage com o ADN (por exemplo, metilase, uma topoisomerase, uma integrase, uma transposase ou uma ligase), activadores ou repressores de transcrição ou uma proteína que interage ou que modifica outras proteínas, tais como histonas. As aplicações dessas fusões de efectores TAL incluem, por exemplo, a criação ou a modificação de elementos reguladores epigenéticos, fazendo inserções, eliminações ou reparações no ADN, específicas do sítio, controlando a expressão de genes e modificando a estrutura da cromatina.

Nalgumas modalidades, o espaçador da sequência-alvo pode ser seleccionado ou pode variar de modo a regular a especificidade e a actividade TALEN. Os resultados que aqui se apresentam para TALEN, que funcionam como dímeros através de um sítio de reconhecimento bipartido com um espaçador demonstram que as TALEN podem funcionar numa gama de comprimentos de espaçadores e que a actividade das TALEN varia com o comprimento do espaçador. Ver, por exemplo, o exemplo 6 que se segue. A flexibilidade do comprimento do espaçador indica que o comprimento do espaçador pode ser escolhido para sequências-alvo particulares (por exemplo, num genoma) com uma elevada especificidade. Além disso, a variação da actividade observada para diferentes comprimentos do espaçador indica que o comprimento do espaçador pode ser

escolhido de forma a conseguir o nível desejado da actividade TALEN.

Nalgumas modalidades, a actividade TALEN pode ser regulada fazendo variar o número e a composição das repetições dentro dos domínios de ligação de ADN. Tal como descrito aqui no exemplo 7, por exemplo, uma TALEN à base de PthXoI mostrou uma actividade maior do que uma TALEN à base de AvrBs3. PthXoI difere de AvrBs3 tanto no número como na composição dos DVR das suas repetições. Além disso, os sítios de reconhecimento de ADN de ocorrência natural para estas proteínas diferem na sua divergência das sequências de reconhecimento respectivas, previstas com base no código do ADN do efector TAL descrito por Moscou e Bogdanove (*supra*). Além disso, algumas TALEN vulgares com o mesmo comprimento (12 DVR) mas com uma diferente composição de DVR, diferem na sua actividade e uma TALEN vulgar com 13 DVR tem uma actividade superior a uma TALEN vulgar com 12 DVR. Assim, não só as TALEN podem ser tratadas por engenharia para reconhecer uma sequência de ADN com interesse, mas (1) o número de repetições pode variar com a actividade modulada, (2) diferentes sítios de ligação podem ser seleccionados para se conseguirem diferentes níveis de actividade e (3) a composição dos DVR e o seu ajustamento ao sítio-alvo (de acordo com o código) podem variar para regular a actividade TALEN.

Quando a TALEN está numa forma heterodimérica, por exemplo, com dois monómeros diferentes incluindo cada um deles um domínio efector TAL e um domínio catalítica da nuclease *FokI*, os DVR podem encontrar-se num número equivalente em cada um dos dois domínios efectores TAL ou cada domínio pode exibir números diferentes de DVR. Por exemplo, se se utilizar um total de 22 DVR para ligar um ADN

numa TALEN heterodimérica particular, pode-se encontrar 11 repetições em cada um dos dois domínios efectores TAL; alternativamente, pode-se encontrar 10 repetições num dos dois domínios efectores TAL e 12 no outro. A presente invenção também engloba TALEN com um domínio da enzima de modificação do ADN que funciona como um monómero. Neste caso, todos os DVR podem ser encontrados num único domínio de efector TAL, que está fundido com a enzima monomérica. Neste caso, para se ter uma ligação eficaz, o número de DVR deve ser equivalente ao número total de DVR que seriam encontrados numa TALEN dimérica equivalente. Por exemplo, em vez de ter 10 repetições em dois domínios efectores TAL diferentes (como no caso de uma TALEN dimérica), ter-se-ia 20 repetições num único domínio de efector TAL (como no caso de uma TALEN monomérica).

Num outro aspecto da presente invenção, o número total de repetições dentro da TALEN dimérica ou monomérica é pelo menos de 14. Ainda noutro aspecto da presente invenção, o número total de repetições dentro da TALEN dimérica ou monomérica é de pelo menos de 20. Ainda noutro aspecto da presente invenção, o número total de repetições dentro da TALEN dimérica ou monomérica é de pelo menos de 24. Ainda noutro aspecto da presente invenção, o número total de repetições dentro da TALEN dimérica ou monomérica é de pelo menos de 30.

O presente pedido de patente de invenção também tem por objecto processos para gerar proteínas efectoras TAL com uma capacidade maior de recombinação para um ADN-alvo. Os processos podem incluir, por exemplo, a geração de um ácido nucleico que codifica um efector TAL que tem um domínio de ligação de ADN com várias repetições da ligação do ADN, contendo cada repetição um DVR que determina o reconhecimento

de um par de bases no ADN-alvo, em que cada repetição da ligação do ADN é responsável pelo reconhecimento de um par de bases no ADN-alvo. Tal como se descreve no exemplo 12 a seguir, o relaxamento do requisito para o T na posição -1 do sítio de ligação pode aumentar a capacidade de recombinação para proteínas efectoras TAL tratadas por engenharia. Assim, a geração de um efector TAL que codifica um ácido nucleico pode incluir a incorporação de um ácido nucleico que codifica uma variante da sequência de repetição da ligação de ADN de ordem 0, com especificidade para A, C ou G, eliminando assim a necessidade de T na posição -1 do sítio de ligação.

Além disso, a presente invenção tem por objecto processos para gerar efectores TAL com uma maior capacidade de recombinação para um ADN-alvo. Esses processos podem incluir a geração de um ácido nucleico que codifica um efector TAL que compreende o domínio de ligação de ADN com várias repetições da ligação de ADN, contendo cada repetição um DVR que determina o reconhecimento de um par de bases no ADN-alvo. Tal como se descreve no exemplo 12 a seguir, a especificidade de NN (o DVR mais comum que reconhece G) parece ser geralmente fraca e pode variar com o contexto, mas alguns DVR podem ter uma maior especificidade para G. Assim, a presente invenção tem por objecto processos que incluem a utilização de DVR alternativos que podem ter uma especificidade mais robusta para G. Por exemplo, pode-se utilizar um ou mais DVR seleccionados no grupo que consiste em RN, R*, NG, NH, KN, K*, NA, NT, DN, D*, NL, NM, EN, E*, NV, NC, QN, Q*, NR, NP, HN, H*, NK, NY, SN, S*, ND, NW, TN, T*, NE, NF, YN, Y* e NQ, em que o asterisco indica um intervalo na segunda posição do DVR.

Artigos fabricados

A presente invenção também tem por objecto materiais para procedimentos que contêm, por exemplo, moléculas de ácidos nucleicos que codificam TALEN, polipéptidos TALEN, composições que contêm essas moléculas de ácidos nucleicos ou esses polipéptidos ou linhas de células tratadas por engenharia de endonucleases TAL. Podem utilizar-se esses itens, por exemplo, como instrumentos de pesquisa ou para actividades terapêuticas.

Nalgumas modalidades, um material para procedimento pode incluir sementes de plantas geradas utilizando os processos da presente invenção. As sementes podem ser condicionadas utilizando meios conhecidos na técnica e embaladas utilizando material de embalagem bem conhecido na técnica para preparar um artigo manufacturado. Uma embalagem de sementes pode ter um rótulo, por exemplo, uma marca ou um rótulo aposto no material de embalagem, um rótulo impresso num material de embalagem ou um rótulo inserido dentro da embalagem. O rótulo pode indicar que as sementes contidas dentro da embalagem podem produzir uma colheita de plantas modificadas geneticamente e pode descrever os traços que foram alterados por modificação genética, em relação às plantas não modificadas.

Outras definições

- Resíduos ou subunidades de aminoácidos numa sequência de um polipéptido são designados aqui de acordo com o código de uma letra em que, por exemplo, Q significa Gln ou resíduo de glutamina, R significa Arg ou resíduo de arginina e D significa Asp ou resíduo de ácido aspártico.

- Substituição de aminoácidos significa a substituição de um resíduo de aminoácido por outro, por exemplo, a substituição de um resíduo de arginina com um resíduo de glutamina numa sequência de péptidos, constituindo assim uma substituição de aminoácidos.
- Os nucleótidos são designados da seguinte forma: utiliza-se um código de uma letra para designar a base de um nucleósido: a representa adenina, t representa timina, c representa citosina e g representa guanina. Para os nucleótidos degenerados, r representa g ou a (nucleótidos de purina), k representa g ou t, s representa g ou c, w representa a ou t, m representa a ou c, y representa t ou c (nucleótidos de pirimidina), d representa g, a ou t, v representa g, a ou c, b representa g, t ou c, h representa a, t ou c e n representa g, a, t ou c.
- A expressão "enzima de modificação do ADN" refere-se a qualquer proteína que é capaz de modificar o material genético de uma célula, seja qual for o nível de modificação do ADN (clivagem, interacção covalente, interacção mediada por água...). Proteínas que interagem com o ADN (por exemplo, uma metilase, uma topoisomerase, uma integrase, uma transposase ou uma ligase), activadores ou repressores de transcrição, outras proteínas, tais como histonas, e nucleases é suposto estarem incluídas no significado de "enzima de modificação do ADN". Quando contido numa enzima de modificação do ADN de um efector TAL, a enzima de modificação do ADN é referida como um domínio da enzima de modificação do ADN.
- O termo "nuclease" entende-se como incluindo exonucleases e endonucleases.
- O termo "endonuclease" refere-se a qualquer enzima de tipo selvagem ou uma variante da enzima capaz de

catalisar a hidrólise (clivagem) de ligações entre ácidos nucleicos dentro de uma molécula de ADN ou de ARN, preferencialmente, uma molécula de ADN. Exemplos não limitativos de endonucleases incluem endonucleases de restrição do tipo II, tais como *FokI*, *HhaI*, *HindIII*, *NotI*, *BbvCI*, *EcoRI*, *BglI* e *AlwI*. As endonucleases compreendem também endonucleases de corte raro quando têm normalmente um sítio de reconhecimento de polinucleótidos de cerca de 12-45 pares de bases (pb) de comprimento, mais preferencialmente, de 14-45 pb. As endonucleases de corte raro aumentam significativamente HR por meio da indução de quebras da hélice dupla do ADN (QHD), num local definido (Rouet, Smih et al., 1994; Rouet, Smih et al., 1994; Choulika, Perrin et al., 1995; Pingoud e Silva 2007). As endonucleases de corte raro podem ser, por exemplo, uma endonuclease autoguiada (Paques e Duchateau 2007), uma nuclease quimérica de dedos de zinco (NDZ) resultante na fusão de domínios em dedos de zinco tratados por engenharia num domínio catalítico numa enzima de restrição, tal como *FokI* (Porteus e Carroll 2005) ou uma endonuclease química (Eisenschmidt, Lanio et al., 2005; Arimondo, Thomas et al., 2006; Simon, Cannata et al., 2008). Nas endonucleases químicas, conjuga-se um elemento de clivagem química ou peptídico quer com um polímero de ácidos nucleicos, quer com outro ADN que reconhece uma sequência-alvo específica, recombinao assim a actividade de clivagem com uma sequência específica. As endonucleases químicas também englobam as endonucleases sintéticas como conjugados de ortofenantrolina, uma molécula de clivagem de ADN e oligonucleótidos que formam triplex (OFT), conhecidos por ligarem sequências específicas de ADN (Kalish e

Glazer 2005). Essas endonucleases químicas estão englobadas no termo "endonuclease", de acordo com a presente invenção. Exemplos dessas endonuclease incluem *I-Sce I*, *I-Chu I*, *I-Cre I*, *I-Csm I*, *PI-Sce I*, *PI-Tli I*, *PI-Mtu I*, *I-Ceu I*, *I-Sce II*, *I-Sce III*, *HO*, *PI-Civ I*, *PI-Ctr I*, *PI-Aae I*, *PI-Bsu I*, *PI-Dha I*, *PI-Dra I*, *PI-Mav I*, *PI-Mch I*, *PI-Mfu I*, *PI-Mfl I*, *PI-Mga I*, *PI-Mgo I*, *PI-Min I*, *PI-Mka I*, *PI-Mle I*, *PI-Mma I*, *PI-Msh I*, *PI-Msm I*, *PI-Mth I*, *PI-Mtu I*, *PI-Mxe I*, *PI-Npu I*, *PI-Pfu I*, *PI-Rma I*, *PI-Spb I*, *PI-Ssp I*, *PI-Fac I*, *PI-Mja I*, *PI-Pho I*, *PI-Tag I*, *PI-Thy I*, *PI-Tko I*, *PI-Tsp I*, *I-MsoI*.

As endonucleases de acordo com a presente invenção podem fazer parte de uma endonuclease efectora semelhante a um activador de transcrição (TAL) e (TALEN).

- Por "TALEN" entende-se uma proteína que contém um domínio de ligação de um efector semelhante a um activador de transcrição (TAL) e um domínio de endonuclease, sendo que a fusão de ambos os domínios resulta numa "TALEN monomérica". Algumas TALEN monoméricas podem ser funcionais per se e outras exigem a dimerização com outras TALEN monoméricas. A dimerização pode resultar numa TALEN homodimérica quando ambas as TALEN monoméricas são idênticas ou pode resultar numa TALEN heterodimérica quando as TALEN monoméricas são diferentes. Duas TALEN monoméricas são diferentes quando, por exemplo, o número dos seus DVR é diferente e/ou quando o teor (isto é, a sequência de aminoácidos) de pelo menos um DVR é diferente. Por "enzima de modificação do efector TAL" entende-se uma proteína que compreende um domínio de ligação do efector semelhante a um

activador de transcrição e um domínio de enzima que modifica o ADN.

Por "variante" entende-se uma "variante" de proteína, isto é, uma proteína que não existe naturalmente na natureza e que se obtém por engenharia genética ou por mutagénese aleatória, isto é, uma proteína obtida por engenharia. Esta variante de proteína pode ser obtida, por exemplo, por substituição de pelo menos um resíduo na sequência de aminoácidos de uma proteína selvagem de ocorrência natural por um aminoácido diferente. As referidas substituições podem ser introduzidas, por exemplo, por mutagénese dirigida ao sítio e/ou por mutagénese aleatória.

Por "célula" ou "células" entende-se qualquer célula viva procariótica ou eucariótica, linhas de células derivadas destes organismos para culturas *in vitro*, células primárias de animais ou com origem em plantas.

Por "célula primária" ou "células primárias" entende-se células retiradas directamente de tecidos vivos (isto é, material de biópsia) e estabelecidas para crescimento *in vitro*, que tenham sofrido muito poucas duplicações de população e que sejam por isso mais representativas dos principais componentes e características funcionais dos tecidos dos quais derivaram, em comparação com linhas de células tumorigénicas contínuas ou imortalizadas artificialmente. Estas células representam assim um modelo mais válido em relação para o estado *in vivo* ao qual se referem.

- Por "homóloga" entende-se uma sequência com uma identidade suficiente com uma outra que leve à recombinação homóloga entre as sequências, mais

particularmente, pelo menos 95 % de identidade, preferencialmente, 97 % de identidade e, mais preferencialmente, 99 %.

- "Identidade" refere-se à identidade de sequências entre duas moléculas de ácidos nucleicos ou polipéptidos. A identidade pode ser determinada por comparação de uma posição em cada sequência que pode ser alinhada para fins comparativos. Quando uma posição na sequência comparada está ocupada pela mesma base, então as moléculas são idênticas nessa posição. Um grau de semelhança ou identidade entre sequências de ácidos nucleicos ou de aminoácidos é a função de um número de nucleótidos idênticos ou emparelhados nas posições partilhadas pelas sequências de ácidos nucleicos. Podem utilizar-se vários algoritmos de alinhamento e/ou programas para calcular a identidade entre duas sequências, incluindo FASTA, ou BLAST que estão disponíveis como parte do pacote de análise de sequências GCG (University of Wisconsin, Madison, WI) e podem ser usados, por exemplo, com os parâmetros iniciais.
- Por "mutação" entende-se a substituição, eliminação, inserção de um ou mais nucleótidos/aminoácidos numa sequência de polinucleótidos (ADNc, gene) ou de polipéptidos. A referida mutação pode afectar a sequência de codificação de um gene ou a sua sequência reguladora. Também pode afectar a estrutura da sequência genómica ou a estrutura/estabilidade do ARNm codificado.
- Por "gene" entende-se a unidade básica de hereditariedade, consistindo num segmento de ADN arranjado de uma forma linear ao longo de um cromossoma, que codifica para uma proteína específico ou um segmento de proteína. Um gene normalmente

inclui um promotor, uma região não traduzida 5', uma ou mais sequências de codificação (exões), eventualmente intrões, uma região 3' não traduzida. O gene pode ainda conter um terminador, intensificadores e/ou silenciadores.

- A expressão "gene de interesse" refere-se a qualquer sequência de nucleótidos que codifica um produto genético conhecido putativo.
- Tal como se utiliza aqui, o termo "locus" representa a localização física específica de uma sequência de ADN (por exemplo, de um gene) num cromossoma. O termo "locus" normalmente refere-se à localização física específica de uma sequência-alvo num cromossoma.
- Por "proteína de fusão" entende-se o resultado de um processo bem conhecido na técnica que consiste na junção de dois ou mais genes que originalmente codificam para proteínas separadas, resultando na tradução do referido "gene de fusão" num único polipéptido com propriedades funcionais derivadas de cada uma das proteínas originais.
- Por "domínio catalítico" entende-se o domínio da proteína ou o módulo da proteína de uma enzima contendo o sítio activo da referida enzima; por sítio activo entende-se a parte da referida enzima na qual ocorre a catálise do substrato. As enzimas e também os seus domínios catalíticos são classificados e designados de acordo com a reacção que catalisam. O número da comissão das enzimas (número CE) é um esquema de classificação numérica para as enzimas, com base nas reacções químicas que elas catalisam (World Wide Web em chem.qmul.ac.uk/iubmb/enzyme/). No âmbito da presente invenção, pode-se utilizar qualquer domínio catalítico como um parceiro e pode fundir-se com um domínio efector TAL para gerar uma

proteína quimérica de fusão que resulta numa enzima de modificação do ADN de efector TAL. Exemplos não limitativos desses domínios catalíticos podem ser os de MmeI, EsaSSII, CstMI, NucA, EndA de *Escherichia coli*, NucM, EndA de *Streptococcus pneumonia*, SNase de *Staphylococcus aureus*, SNase de *Staphylococcus hyicus*, SNase de *Shigella flexneri*, *Bacillus subtilis* yncB, fago T7 de enterobactérias de endodesoxiribonuclease I, EndoG de bovino, mutS de proteínas de reparação desemparelhadas do ADN de ttSmr, domínio de clivagem de Metnase.

A prática da presente invenção, utilizará, salvo indicação em contrário, técnicas convencionais de biologia celular, cultura de células, biologia molecular, biologia transgénica, microbiologia, ADN recombinante e imunologia, que estão dentro das competências desta técnica. Essas técnicas estão explicadas completamente na literatura. Ver, por exemplo, Current Protocols in Molecular Biology (Ausubel, 2000, Wiley and son Inc, Library of Congress, EUA); Molecular Cloning: A Laboratory Manual, 3ª edição, (Sambrook et al., 2001, Cold Spring Harbor, Nova Iorque: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); patente de invenção norte-americana N.º. 4 683 195; Nucleic Acid Hybridization (Harries e Higgins eds. 1984); Transcription and Translation (Hames e Higgins eds. 1984); Culture of Animal Cells (Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells and Enzymes (IRL Press, 1986); Perbal, A Practical Guide to Molecular Cloning (1984); the series, Methods in Enzymology (Abelson and Simon, eds.-in-chief, Academic Press, Inc., Nova Iorque), especificamente, Vols. 154 e 155 (Wu et al. eds.) e Vol. 185, "Gene Expression Technology" (Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (Miller and Calos eds., 1987, Cold Spring

Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer e Walker, eds., Academic Press, London, 1987); Handbook of Experimental Immunology, Vols. I-IV (Weir e Blackwell, eds., 1986); e Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

A descrição da presente invenção descrita antes providencia uma maneira e um processo para produzir e utilizar de tal maneira que qualquer especialista na matéria fica habilitado a produzir e a utilizar a mesma, sendo que estes conhecimentos são providenciados, em particular, para os assuntos referidos nas reivindicações em anexo, que fazem parte da descrição original.

Tal como se utilizaram antes, as expressões "seleccionado no grupo que consiste em", "escolhido a partir de" e similares incluem misturas dos materiais especificados.

Quando se estabelece aqui um limite numérico ou um intervalo, os pontos finais estão incluídos. Também, todos os valores e os sub-intervalos dentro de um limite ou intervalo numérico estão especificamente incluídos se estiverem explicitamente escritos.

A descrição anterior foi apresentada para permitir a um especialista na matéria fazer uso da presente invenção e é providenciada no contexto de uma aplicação particular e dos seus requisitos. Assim, a presente invenção não pretende estar limitada pelas modalidades ilustrativas, mas é para ser tomada em consideração no seu âmbito mais lato consistindo com os princípios e as características aqui descritos.

Tendo descrito a presente invenção de uma forma geral, pode ter-se um melhor entendimento com referência a alguns exemplos específicos que vão ser dados aqui para apenas para fins ilustrativos, sendo a presente invenção descrita nos exemplos que se seguem, que não limitam o âmbito da presente invenção descrita nas reivindicações, salvo indicação específica.

EXEMPLOS

Exemplo 1 - Um código governa o reconhecimento do ADN efector TAL

Para determinar se há uma correspondência de um-para-um linear entre os DVR e os nucleótidos contíguos de um sítio-alvo TAL, fez-se o rastreio da região promotora prevista (isto é, os 1.000 pb precedem imediatamente o sítio de início de tradução anotado) do gene-alvo conhecido para cada um dos dez efectores TAL com a sequência de DVR do efector TAL para alinhamentos que minimizam para alinhamentos que minimizam a entropia (aleatoriamente) em associações de DVR-nucleótidos. A fórmula que se segue foi utilizada para quantificar a entropia, em que R é o conjunto dos DVR para o efector, D é o conjunto de quatro nucleótidos (A, C, G, T), e $f_{i,j}$ representa a frequência observada, em que cada um dos DVR i_{th} se associa com o nucleótido j_{th} :

$$\sum_{i \in R} \sum_{j \in D} \max_j(f_i) - f_{i,j}$$

Em cada promotor estavam presentes vários sítios de baixa entropia. Contudo, para o efector AvrBs3, apenas um mapeou o fragmento do promotor *upa20* de 54 pb identificado previamente como suficiente e necessário para a activação e

coincidiu com a caixa UPA comum aos genes directamente activados por AvrBs3 (Kay et al., *supra*). Também para os efectores PthXo1 e AvrXa27, apenas um sítio de cada um se sobrepôs a um polimorfismo entre os alelos activados e não-activados dos seus respectivos alvos, *Os8N3* e *Xa27*. Através dos alinhamentos, nestes três sítios, as associações de DVR-nucleótidos eram consistentes. De tal modo que os alinhamentos remanescentes foram seleccionados com base nessas associações, resultando exactamente num sítio por cada par-alvo de efectores TAL (figura 1B e quadro 2). Cada sítio é precedido por um T (figura 1D).

Para avaliar a especificidade conferida pelas associações de DVR-nucleótido, gerou-se primeiro uma matriz ponderada com base nas frequências de todas as associações de DVR-nucleótido observadas ao longo de dez alinhamentos de sítios-alvo de efectores TAL com uma entropia mínima (figura 1B). Depois, utilizou-se a matriz ponderada para rastrear a região promotora, as 1.000 pb que precedem o início da tradução, de cada modelo de genes não redundante em arroz, *Oryza sativa* spp. japonica cv. Nipponbare (Osal, Release 6.0, rice.plantbiology.msu.edu) para os melhores emparelhamentos com cinco efectores TAL de *Xanthomonas oryzae* (AvrXa27, PthXo1, PthXo6, PthXo7 e Tallc) patogénica do arroz. Para AvrXa27, a sequência mutante de *Xa27* (número de acesso do GenBank AY986492) foi incluída. Esta sequência a montante não está presente em Nipponbare. Ponderaram-se as frequências de associação observadas a 90 % e os restantes 10 % foram distribuídos igualmente com as frequências de todas as associações possíveis. Classificaram-se os alinhamentos utilizando uma pontuação da matriz ponderada (eixo do y), tomada como o logaritmo negativo da pontuação da frequência derivada das frequências de associação de DVR-nucleótido na figura 1B. Assim, a pontuação mais baixa, é o melhor

emparelhamento. Para PthXo1, PthXo6, PthXo7 e Tallc, os gene-alvo identificado experimentalmente foi o melhor ou praticamente o melhor emparelhado. Os melhores emparelhamentos não precedidos por um T, não estão representados na micro-matriz utilizada para identificar o alvo ou falta-lhe os intrões e a evidência de EST. O rastreio das sequências promotoras inversas do complemento não originaram melhores pontuações do alinhamento do que os sítios para a frente no caso dos alvos conhecidos. Este resultado não implica que os efectores TAL se liguem à hélice positiva, mas indicam que funciona numa orientação paralela relativamente à hélice positiva. O alvo conhecido do quinto effector, AvrXa27, é gene de resistência à doença Xa27 (Gu et al., *supra*). A pontuação pior para este emparelhamento (5 368) pode reflectir uma adaptação sub-óptima do hospedeiro, calibrada ou recente. Os sítios melhor pontuados provavelmente compreendem genes recombinados por AvrXa27 para a patogénese.

Utilizando novamente uma matriz ponderada, obtiveram-se dez alinhamentos adicionais por rastreio de todos os promotores de arroz com 40 efectores TAL adicionais de *X. oryzae* e retendo os melhores alinhamentos para os quais o gene a jusante foi activado durante a infecção, com base nos dados públicos de micro-matrizes (PLEXdb.org, acesso 05/3) (quadro 3). Tal como o conjunto inicial, um T precede cada sítio e nenhum dos sítios de hélice inversa teve melhor pontuação. As frequências de associação de DVR-nucleótidos de um total de 20 alinhamentos estão ilustradas na figura 1C. Constituem um código surpreendentemente simples.

As frequências de DVR-nucleótido no conjunto expandido de 20 alinhamentos de nucleótidos de efectores TAL foram utilizadas para gerar uma nova matriz ponderada e escreveu-se

um descritivo computacional em Python v2.5 (www.python.org). O descritivo pode ser utilizado para rastrear qualquer colecção de sequências de ADN quanto aos emparelhamentos com efector TAL particular, com um factor ponderado que pode ser definido pelo utilizador para as associações de DVR-nucleótido observados versus as não observadas. Ver Moscou e Bogdanove (*supra*).

Há alguma degenerescência neste código. Associações fortes podem representar âncoras que têm em conta a maior parte da afinidade de ligação, sendo que as associações fracas providenciam uma medição da flexibilidade. Alternativamente, podem estar envolvidos efeitos de vizinhança. Esta última possibilidade foi investigada por meio da determinação das sequências de associação de nucleótidos de cada DVR, condicionado pelo facto do DVR de cada lado e comparando-os com o total das frequências observadas - por outras palavras, alinhamento os emparelhamentos de DVR-nucleótidos de acordo coma vizinhança de DVR com a esquerda ou a direita e comparando as frequências relativas de cada par assim classificado com a frequência global para esse par. As frequências das associações de DVR-nucleótido seleccionadas pela vizinhança não se desviam significativamente das frequências totais observadas, sugerindo que as associações estão num contexto independente.

As sequências que flanqueiam os 20 sítios-alvo não revelaram nucleótidos conservados, excepto do T em -1, mas tendem em ser ricas em C no seguimento do sítio e pobres em G em todo o lado (figura 1D). Com poucas excepções, os sítios começam com 60 pb a montante do início transcricional anotado e nenhum está mais próximo do que 87 pb em relação ao início da tradução (quadro 2 e quadro 3). Regras adicionais que

governam as associações de DVR-nucleótidos estão descritas nos exemplos 4 e 5.

Considerando estes resultados, é agora possível a previsão de alvos efectores TAL num genoma e a preparação dos alvos *de novo*. A capacidade para prever sítios irá agilizar a identificação de genes hospedeiros importantes em doenças. A capacidade para gerar alvos promete uma concepção de genes resistentes duráveis, que são responsáveis por efectores TAL conservados ou múltiplos. A adaptação dos efectores TAL para a activação arbitrária de genes ou a recombinação de proteínas fundidas para modificação do ADN é também possível tal como se descreve aqui.

Quadro 2

Características do sítio-alvo previsto para pares-alvo de efetores TAL identificados experimentalmente

Efeitor TAL	Fonte	DVR	Gene-alvo	TCS	Caixa	TATA	TIS
AvrKa27 ¹	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> PX099 ^A	17	Ka27 (arroz)	27	-7	87	
AvrBs3 ²	<i>X. campestris</i> pv. <i>vesicatoria</i>	18	Bs3 (pimenta)	59	1	123	
AvrBs3 ³	<i>X. campestris</i> pv. <i>vesicatoria</i>	18	UPA20 (pimenta)	72	1	150	
AvrBs3Δrep16 ^{4,5}	AvrBs3 modificado	14	Bs3-E (pimenta)	85	1	136	
AvrBs3Δrep109 ⁴	AvrBs3 modificado	15	Bs3 (pimenta)	59	1	123	
AvrHah1 ⁶	<i>X. gardneri</i>	14	Bs3 (pimenta)	59	1	121	
PthXo1 ⁷	<i>X. oryzae</i> pv. <i>oryzae</i> PX099 ^A	24	Os8N3 (rice)	79	46	251	
PthXo6 ⁸	<i>X. oryzae</i> pv. <i>oryzae</i> PX099 ^A	23	OSTFX1 (arroz)	31	-780	136	
PthXo7 ⁸	<i>X. oryzae</i> pv. <i>oryzae</i> PX099 ^A	22	OSTFIAY1 (arroz)	333	44	469	
Tallc	<i>X. oryzae</i> pv. <i>oryzicola</i> BLS256	16	OSHEN1 (arroz)	10	-265	217	

DVR, di-resíduos variáveis repetidos; TCS, sítio inicial transcripcional anotado; TIS, sítio de início da tradução. As localizações são relativas à extremidade 5' do sítio-alvo.

¹ Gu et al., *supra*

² Kay et al. (2007) *Science* 318:648

³ Römer et al. (2007) *Science* 318:645

⁴ Herbers et al. (1992) *Nature* 356:172

Quadro 2
Características do sítio-alvo previsto para pares-alvo de efectores TAL identificados experimentalmente

- ⁵ Römer et al. (2009) Plant Physiol.
- ⁶ Schornack et al. (2008) New Phytologist 179:546
- ⁷ Yang et al. (2006) Proc. Natl. Acad. Sci. EUA 103:10503
- ⁸ Sugio et al. (2007) Proc. Natl. Acad. Sci. EUA

Quadro 3

Alvos candidatos de efectores TAL de *Xanthomonas oryzae* em infeções em arroz activado durante a infeção

Efector	Estirpe	DVR	Locus do arroz	r	TCS	Caixa TALTA	TIS	q	Nº. vezes da alteração
Tal1c	BLS256	16	OsHen1	1	10	-265	217	0,01	3,3
Tal2c	BLS256	27	Os03g03034	15	-16	-145	143	0,01	5,2
Tal2d	BLS256	16	Os04g49194	9	27	n.p.	102	3,9E-07	29,7
Tal3b	BLS256	18	Os05g27590	42	34	-1	104	3,4E-08	8,5
Tal4a	BLS256	26	Os03g37840	1	152	221	363	2,2E-04	2,6
Tal4b	BLS256	14	Os09g32100	72	68	n.p.	271	8,0E-03	3,6
Tal4c	BLS256	23	Os06g37080	18	31	n.p.	151	2,7E-10	17,1
Tal6	BLS256	20	Os07g47790	16	-15	-70	93	3,6E-02	21,6

Quadro 3

Alvos candidatos de efectores TAL de *Xanthomonas oryzae* em infecções em arroz activado durante a infecção

Efector	Estilipe	DVR	Locus do arroz	r	TCS	Caixa	TATA	TIS	q	Nº. vezes da alteração
PthXo1	PX099 ^A	24	Os8N3	1	79	46		251	1,0E-08	84,2
PthXo6	PX099 ^A	23	OsTFX1	2	31	-780		136	3,5E-03	2,8
PthXo7	PX099 ^A	22	OsFTIAY1	7	333	44		469	1,6E-06	4,5
Tal9a	PX099 ^A	20	OsHen1	1	44	-3		93	0,13	8,2
Tal7a/8a	PX099 ^A	18	Os01g68740	2	32	-197		102	1,8E-01	1,7
Tal7b/8b	PX099 ^A	20	Os01g40290	57	-2	-276		206	1,8E-01	1,7

DVR, di-resíduos variáveis repetidos; r, ranking de 58.918 modelos de genes rastreados, com base na pontuação da matriz ponderada de DVR; TCS, sítio inicial transcripcional anotado; n.p., não presente; TIS, sítio de início da tradução. As localizações são relativas à extremidade 5' do sítio-alvo. Os valores de q são dados para uma comparação ao longo de cinco períodos de tempo até 96 horas após a inoculação, replicados quatro vezes; a alteração de vezes dada é às 96 horas (PIEXdb, acesso OS3).

Exemplo 2 - As TALEN podem funcionar em levedura

Preparação do plasmido: A sequência de codificação da proteína do efector TAL, AvrBs3, obteve-se por digestão de um plasmido com *Bam*HI. Excisou-se um fragmento de ADN que codifica principalmente o domínio de repetição com *Sph*I. A sequência de aminoácidos de AvrBs3 pode ser encontrada no GENBANK com o N.º. de acesso P14727 e a SEQ ID N.º. 12 (figura 3) e a sequência de ácidos nucleicos com o N.º. de acesso X16130 e SEQ ID N.º. 13 (figura 4). Na figura 4, os sítios de *Bam*HI e *Sph*I estão a negrito e sublinhados. Os fragmentos de *Bam*HI e *Sph*I de AvrBs3 foram clonados no vector de expressão de nuclease pDW1789_TAL (figura 5) adjacente às sequências que codificam o domínio da nuclease *Fok*I. Para clonar o sítio-alvo de AvrBs3 no plasmido relator alvo, sintetizaram-se dois oligonucleótidos de ADN complementares, contendo dois sítios de reconhecimento de AvrBs3 arranjados numa orientação invertida, com uma sequência espaçadora de 18 pb entre eles, e a síntese foi feita com pendentes de *Bgl*III e *Spe*I nas extremidades 5' e 3', respectivamente. Prepararam-se outros plasmidos relatores que tinham sítios de reconhecimento com um comprimento de espaçador de 6, 9, 12 e 15 pb. Os oligonucleótidos hibridados foram clonados no plasmido repórter, pCP5 (figura 6), que foi digerido com *Bgl*III e *Spe*I.

Ensaio de levedura: Os plasmidos relatores-alvo foram transformados n estirpe de levedura *YPH499* (uma estirpe de *MAT*) e seleccionaram-se os transformantes num meio completamente sintético a que faltava triptofano (SC-W). Os plasmidos de expressão de TALEN foram transformados em *YPH500* (uma estirpe de *MAT* α); e os transformantes foram colocados em placa num meio SC a que faltava histidina (SC-H). As colónias de levedura comportando o plasmido-alvo relator e as colónias que continham os plasmidos de expressão de TALEN

foram postas em cultura, durante a noite, a 30 °C, num meio líquido de SC-W e SC-H, respectivamente. Ajustaram-se as culturas pra a mesma DO₆₀₀ e misturou-se 200 µL de cada em 200 µL de meio YPD. Fez-se a incubação da mistura a 30 °C, durante 4 horas, para permitir que os dois tipos de estirpes de levedura se conjugassem. Centrifugou-se a cultura mista e suspendeu-se novamente em 5 mL de meio SC-W-H, a 30 °C, durante a noite ou até a DO₆₀₀ atingir um intervalo de 0,5-1. Colheram-se as células e realizaram-se ensaios quantitativos de β-galactosidase, tal como descrito (Townsend et al. (2009) Nature 459:442-445).

Resultados: A fusão TAL-*FokI* numa nuclease específica do sítio consistindo num domínio de reconhecimento do ADN TAL e o domínio de clivagem não-específico de ADN *FokI*. O domínio de reconhecimento do ADN TAL pode ser tratado por engenharia para se ligar a diferentes sequências de ADN. Tal como descrito aqui no exemplo 1, a especificidade de reconhecimento do ADN para efectores TAL, uma nova classe de domínio de ligação do ADN já foi decifrada. Em particular, O domínio de ligação de ADN de efectores TAL contém um número variável de tandem, 34 repetições de aminoácidos que podem reconhecer e ligarem-se a sequências específicas de ADN. As sequências de aminoácidos das repetições estão conservadas excepto para dois resíduos adjacentes altamente variáveis nas posições 12 e 13 das repetições. Estas posições, em conjunto, especificam nucleótidos individuais no sítio de ligação do ADN, uma repetição com um nucleótido, A arquitectura das TALEN está ilustrada na figura 7. As TALEN funcionam como dímeros, sendo cada monómero composto por repetições de reconhecimento de ADN TAL tratados por engenharia, fundidos com um domínio não-específico da clivagem da endonuclease *FokI*. As repetições de reconhecimento de ADN podem ser tratadas por engenharia para se ligarem a sequências-alvo de

ADN dentro de um genoma de interesse. Os monómeros de nuclease TAL ligam-se a um dos dois semi-sítios de ADN que estão separados por uma sequência espaçadora. Este espaçamento permite que os monómeros de *FokI* dimerizem e cria uma quebra de ADN de hélice duplas (QHD) na sequência espaçadora entre os semi-sítios.

Para explorar o potencial do domínio de reconhecimento de ADN e do efector TAL, realizaram-se experiências para determinar se os efectores TAL naturais podem funcionar como nucleases quando se fundem com o domínio da nuclease *FokI*. O ensaio à base de levedura foi feito utilizando uma estrutura de expressão de nuclease TAL e uma estrutura relatora do alvo. Tal como ilustrado na figura 5, a estrutura da expressão da nuclease contém um domínio de nuclease *FokI* e um sinal de localização nuclear (SLN) do terminal N, sob o controlo do promotor da levedura *TEF1*. Vários sítios de restrição estão localizados entre o domínio da nuclease *FokI* e um elemento de SLN para facilitar a clonagem de vários efectores TAL. A estrutura repórter-alvo tem um gene repórter de *lacZ* alvo de disrupção com uma duplicação de 125 pb de uma sequência de codificação, tal como se ilustra na figura 6. Os flancos da duplicação, um gene *URA3* e uma sequência-alvo (composta por dois semi-sítios e uma sequência espaçadora) são reconhecidos pelos domínios de ligação de ADN TAL. Se a TALEN se liga e gera quebras de hélice dupla de ADN (QHD) no sítio-alvo, essas quebras, na levedura, são reparadas predominantemente por recombinação homóloga entre as sequências duplicadas de *lacZ*, através da hibridação de hélices simples (Haber (1995) *Bioessays* 17:609). Os resultados da recombinação na reconstituição de um gene funcional *lacZ* e a perda de *URA3* (conferido resistência ao ácido 5-fluoro-orótico). Mediu-se a actividade relativa de

clivagem de TALEN pela determinação da actividade da enzima *lacZ*.

Nestes estudos, um efector TAL natural, AvrBs3, que tinha uma região central de repetição da nuclease, tal como ilustrado na SEQ ID Nº. 16 (figura 8) foi clonado no vector de expressão de nuclease e os sítios-alvo AvrBs3 (dois sítios de ligação arranjados numa orientação invertida), com uma sequência espaçadora de 18 pb, foram clonados no vector-alvo relator. O ensaio da levedura foi feito utilizando o esquema ilustrado na figura 9 e descrito antes. Os resultados mostraram que a actividade de *lacZ* a partir de células de levedura transformadas tanto com o plasmido da nuclease AvrBs3 como com o plasmido-alvo repórter era significativamente mais elevada (15,8 vezes superior) do que a das células de controlo de levedura, que continham apenas o plasmido-alvo relator (figura 10). Não se observou nenhuma actividade com fusões de nucleases feitas apenas com o fragmento *SphI* que codifica predominantemente o domínio de repetição. Isto indicou que as sequências diferentes do domínio de ligação de ADN são necessárias para a actividade de TALEN. Os plasmidos relatores com comprimentos espaçadores de 6 e 9 pb também falharam na exibição de actividade, indicando que o espaço entre os dois sítios de ligação é crítico para permitir que *FokI* dimerize. Estes dados indicam que a nuclease TAL de AvrBs3 pode funcionar como uma nuclease específica do sítio que cliva as suas sequências-alvo cognatas na levedura.

Exemplo 3 - Junção modular de petições de efectores TAL para TALEN costumizadas

Os oligonucleótidos complementares correspondendo a 102 pares de bases de cada quatro repetições de efectores TAL

individuais, especificando cada uma um nucleótido diferente, são sintetizados, hibridados e clonados num vector de clonagem bacteriana de elevado número de cópias, individualmente e em combinação de 2 e 3 repetições em todas as permutações, para originar 4 módulos simples de repetições, 16 duplos e 64 triplos utilizando técnicas-padrão de digestão por restrição e ligação (por exemplo, como ilustrado na figura 11). A sequência de codificação do efector TAL desejada é ligada por introdução dos módulos apropriados, sequencialmente, num vector de clonagem bacteriano de elevado número de cópias pronto para ser atravessado, contendo uma forma truncada do género *Tal1c*, a que falta a região de repetição central, excepto no que respeita à característica da semi-repetição final. Por exemplo, uma sequência de codificação do efector TAL com 18 repetições pode unir-se, por introdução sequencial de 5 módulos triplos e 1 módulo duplo, no vector truncado *Tal1c*.

Exemplo 4 - Um sistema para a junção modular de repetições do efector TAL

Desenvolveram-se plasmidos e processos para gerar genes customizados que codificam o efector TAL. A especificidade funcional dos efectores TAL determina-se pelos DVR nas repetições, tal como aqui descrito; outros polimorfismos nas repetições e noutros sítios das proteínas são raros e inconsequentes no que respeita à especificidade funcional. Assim, os genes customizados de efectores TAL foram gerados por substituição da região de repetição de um gene arbitrário de efector TAL com repetições contendo os DVR desejados. As sequências de repetição fora dos DVR emparelham-se com uma sequência de consenso (ver a seguir). Os fragmentos de ADN que codificam as repetições do efector TAL foram unidos sequencialmente em módulos que codificam uma, duas ou três

repetições e os módulos foram clonados num gene efector TAL a partir do qual as repetições originais foram eliminadas. Cada repetição codificada, com excepção da última (semi) repetição, tinha a sequência LTPDQVVAIASXXGGKQALETVQRLLPV LCQDHG (SEQ ID N.º. 18; figura 12A). A última (semi) repetição tem a sequência LTPDQVVAIASXXGGKQALES (SEQ ID N.º. 20; figura 12B). Em ambas as sequências, "XX" indica a localização do DVR. Os DVR utilizados nas repetições modulares foram NI, HD, NN e NG, que se ligam especificamente a A, C, G e T, respectivamente. Nas experiências que se descrevem a seguir, o gene *Tal1c* da estirpe de *Xanthomonas oryzae* pv. *Oryzicola*, BLS256, com as suas repetições eliminadas, foi utilizado como a "estrutura" para a produção de genes customizados de efectores TAL.

O processo aqui descrito inclui cinco componentes: (1) a geração de plasmidos iniciadores de repetições únicas; (2) a geração de plasmidos de módulos de repetições únicas; (3) a geração de módulos de repetição múltiplos; (4) a geração de um conjunto completo de um, dois e três plasmidos de módulos de repetição; e (5) a junção das sequências customizadas que codificam efectores TAL.

Para gerar plasmido iniciadores de repetição simples, o gene *Tal1c* foi digerido com *MscI* e novamente ligado para eliminar toda a região de repetição excepto para a primeira parte da primeira repetição e a última parte da última, a repetição truncada, resultando no plasmido designado por pCS487 (figura 13). O gene resultante codificou o DVR NI e, tal como a maior parte dos genes efectores TAL, continha dois sítios *SphI* que flanquearam a região de repetição. O gene não continha nenhum sítio *XhoI*.

Em seguida, introduziu-se translaccionalmente uma mutação silenciosa em pCS487 para criar um sítio único de *PspXI*, que engloba um único sítio *XhoI* centrado nos codões 19 e 20. A mutação está descrita na figura 14, que mostra as sequências de nucleótidos originais e alteradas para os codões 18-21 (SEQ ID N°. 21 e SEQ ID N°. 23, respectivamente), codificando ambas as sequências de aminoácidos ALES (SEQ ID N°. 22). O plasmido resultante foi designado por pCS489.

Continuando a mutagénesse, geraram-se três estruturas adicionais com os DVR HD, NN, e NG, para criar os plasmidos designados por pCS490, pCS491 e pCS492, respectivamente. O fragmento *SphI* que engloba a região de repetição modificada foi transferido de pCS489, pCS490, pCS491 e pCS492 para plasmidos resistentes a canamicina designados por pCS488 (figura 15), que codificaram apenas as porções de *Tal1c* do terminal N e do terminal C, sem a região de repetição, no vector de entrada de Gateway pENTR-D (Invitrogen, Carlsbad, CA). Esta transferência resultou em plasmidos iniciadores de repetições simples designados por pCS493 (figura 16), pCS494, pCS495 e pCS496, respectivamente. O sítio *PspXI/XhoI* na repetição truncada permaneceu único nestes plasmidos. O gene efector TAL em pCS488 e cada um dos seus derivados foi precedido das sequências de Shine-Dalgarno e Kozak para uma tradução eficiente em procariotos e eucariotos, respectivamente.

Os plasmidos de módulos de repetição simples foram então preparados. Gerou-se então um plasmido para cada um dos quatro DVR escolhidos (NI, HD, NN e NG). Cada plasmido tinha uma extremidade 5' coesa, compatível, que reconstituiu um *XhoI* mas não um sítio *PspXI* quando ligado num sítio *PspXI* e uma extremidade 3' coesa, compatível, que reconstituiu ambos

os sítios *XhoI* e *PspXI*. Os plasmídeos foram gerados por hibridação de clonagem sintética, de oligonucleótidos complementares com excrescências (figura 17A) no sítio *PspXI/XhoI* de pBluescript SK-, resultando nos plasmídeos designados por pCS502 (figura 17B), pCS503, pCS504 e pCS505, respectivamente. Cada plasmídeo primitiu a introdução de mais repetições na extremidade 3' do módulo de repetição simples no sítio único de *PspXI* reconstituído ou por excisão do módulo de repetição utilizando o sítio *XhoI* reconstituído.

Geraram-se módulos adicionais de repetições simples, um para cada NI, HD, NN e NG. Cada um tinha uma extremidade 5' coesa, compatível, que não reconstituiu um sítio *PspXI* ou *XhoI* quando ligado num sítio site *PspXI*, uma extremidade 3' coesa, compatível, que reconstituiu ambos os sítios *XhoI* e *PspXI* e uma substituição translacional, silenciosa de nucleótidos que destruiu o sítio interno *MscI* (figura 18A). Estes módulos foram gerados por hibridação sintética, com oligonucleótidos complementares com excrescências. A ligação de qualquer um destes módulos adicionais de repetição simples num único sítio *PspXI/XhoI* de um plasmídeo com um módulo de repetição simples (pCS502, pCS503, pCS504 ou pCS505) resultou em nenhum sítio novo de *XhoI* na junção 5', mas resultou na restauração do único sítio *PspXI/XhoI* de 3', resultando assim em plasmídeos que podiam ser linearizados para a introdução de mais repetições adicionais pelo corte com *PspXI*. A reiteração deste processo resultou em módulos contendo várias repetições (figura 18B). Além disso, cada módulo completo de repetições múltiplas podia ser excisado utilizando *XhoI*. Como o sítio *MscI* foi destruído nos módulos adicionais de repetições simples, o sítio *MscI* na repetição inicial permaneceu único e foi útil para verificar a orientação após a subclonagem subsequente do módulo de repetições múltiplas.

Outros módulos de repetições simples foram clonados iterativamente em plasmídeos de módulos de repetições simples para gerar, em conjunto com os plasmídeos dos módulos de repetições simples, um conjunto completo de todos os possíveis módulos de uma, duas e três repetições, para um total de 84 plasmídeos designados desde pCS502 até pCS585 (figura 19). Os módulos contendo mais de três repetições (por exemplo, quatro, cinco, seis, sete, oito, nove, dez ou mais do que dez repetições) são gerados da mesma maneira.

Desviou-se então um processo para unir qualquer sequência de repetições na "estrutura" de *Tal1c* para gerar um gene efector TAL customizado. O processo incluiu as etapas que se seguem, que também estão descritas na figura 20:

- (1) escolhe-se um plasmídeo iniciador de repetições simples com a primeira repetição desejada (pCS493, pCS494, pCS495 ou pCS495, codificando os DVR NI, HD, NN ou NG, respectivamente);
- (2) linearizar o plasmídeo com *PspXI*;
- (3) isolar o módulo para as repetições seguintes a partir do plasmídeo do módulo apropriado (de pCS502 a pCS585) utilizando *XhoI*;
- (4) ligar;
- (5) verificar a orientação por digestão com *MscI* e confirmar a sequência a partir da extremidade 3' e utilizando um iniciador à base de vector; e
- (6) repetir as etapas 2-5 até que todas as repetições estejam unidas.

Exemplo 5 - Biblioteca de plasmidos para a junção modular de TALEN

A junção de repetições de TALEN, tal como aqui descrito (utilizando as etapas descritas na figura 20) resulta em numerosos plasmidos intermédios contendo números crescentes de repetições. Cada um destes plasmidos é armazenado de tal maneira que se gera uma biblioteca de plasmidos para a junção modular de TALEN (pMAT). Por exemplo, as figuras 21A e 21B descrevem a junção de módulos de repetição na construção de endonucleases TAL que irão atingir as sequências de nucleótidos ilustradas. Na figura 21A, os módulos de repetição dos plasmidos designados por pCS519, pCS524, pCS537, pCS551, pCS583 e pCS529 são adicionados sequencialmente à sequência no plasmido iniciador designado por pCS493, resultando nos plasmidos designados por pMAT55, pMAT56, pMAT57, pMAT58, pMAT59 e pMAT60. Na figura 21B, os módulos das repetições de plasmidos designados por pCS530, pCS533, pCS522 e pCS541 são adicionados sequencialmente à sequência no plasmido designado por pMAT1, resultando nos plasmidos designados por pMAT61, pMAT62, pMAT63 e pMAT64.

Exemplo 6 - Geração e análise de TALEN customizadas

O domínio do reconhecimento do ADN TAL foi utilizado para criar TALEN que reconhecem e clivam alvos particulares de ADN (figura 22A), utilizando o sistema descrito nos exemplos 4 e 5. Para avaliar a função de TALEN, adaptou-se um ensaio de leveduras em que a actividade de *lacZ* serve como um indicador de clivagem de ADN (Townsend et al., *supra*). Neste ensaio, um plasmido-alvo e um plasmido de expressão de TALEN são colocados em conjunto na mesma célula por conjugação. O plasmido-alvo tem um gene relator *lacZ* com uma duplicação de 125 pb de sequências de codificação. A duplicação flanqueia

um sítio-alvo reconhecido por um dado TALEN. Quando ocorre uma quebra de ADN de hélice dupla no sítio-alvo, é reparada através de uma hibridação de hélice simples entre as sequências duplicadas, o que cria um gene funcional *lacZ* cuja expressão pode ser medida utilizando ensaios-padrão com β -galactosidase que providencia uma leitura quantificável (figura 22A). Este ensaio tem demonstrado ser um bom ensaio de previsão da capacidade de uma NDZ para criar mutações cromossômicas por meio de JNHE ou para estimular a recombinação homóloga para a edição de um gene em eucariotos superiores (Townsend et al., *supra*; e Zhang et al. (2010) Proc. Natl. Acad. Sci. EUA 107:12028-12033).

Utilizaram-se dois efectores TAL bem caracterizados, AvrBs3 do agente patogénica da pimenta *Xanthomonas campestris* pv. vesicatoria e PthXo1 do agente patogénico do arroz *X. oryzae* pv. *oryzae* (Bonas et al. (1989) Mol. Gen. Genet. 218:127-136; e Yang et al. (2006) Proc. Natl. Acad. Sci. EUA 103:10503-10508). A sequência de aminoácidos de AvrBs3 pode ser encontrada no GENBANK com o N.º. de acesso P 14727 e na SEQ ID N.º. 12 (figura 3) e a sequência de ácidos nucleicos com o N.º. de acesso X16130 e na SEQ ID N.º. 13 (figura 4). A sequência de aminoácidos de PthXo1 pode ser encontrada no GENBANK com o n.º. de acesso ACD58243 e na SEQ ID N.º. 31 (figura 23) e a sequência de ácidos nucleicos com o N.º. de acesso CP000967, o gene ID 6305128 e na SEQ ID N.º. 32 (figura 24). A sequência de aminoácidos de PthXo1, com o n.º. de acesso do GENBANK ACD58243 está truncada no terminal N devido à denotação do codão de partida. A sequência completa está ilustrada na figura 23.

Os domínios de repetição tanto de AvrBs3 como de PthXo1 estão totalmente codificados dentro de um fragmento conservado de *SphI* (figuras 4 e 24). Ambos os genes que

codificam o efector TAL também têm um fragmento de restrição de *Bam*HI que engloba a sequência de codificação para o domínio de repetição e 287 aminoácidos antes e 231 aminoácidos depois (figuras 4 e 24; ver, também, figura 22A). Está ausente do fragmento *Bam*HI o domínio de activação transcricional do efector TAL. Ambos os fragmentos de *Sph*I e os fragmentos de *Bam*HI fundiram-se com um fragmento de ADN que codifica *Fok*I, que está presente no vector de expressão de nuclease pFZ85 (figura 25). As proteínas de fusão entre a nuclease *Fok*I e os fragmentos *Bam*HI codificados por *AvrBs3* e *PthXol* estão indicados nas figuras 26 e 27; SEQ ID N.º. 33 e 34.

Os monómeros de *Fok*I devem dimerizar de modo a clivar, mas o comprimento do espaçador apropriado entre os dois sítios de reconhecimento de ADN não estava claro. Para as NDC, em que a matriz dos dedos de zinco está separada de *Fok*I por um ligante com 4-7 aminoácidos, o espaçador típico entre os dois sítios de reconhecimento é de 5-7 pb (Handel et al. (2009) *Mol. Ther.* 17:104-111). Dado que, por exemplo, 235 aminoácidos separam o domínio de repetição de *Fok*I nas estruturas TALEN de *Bam*HI aqui utilizadas, utilizou-se uma variedade de comprimentos de espaçador tanto para as estruturas de *Bam*HI como de *Sph*I (6, 9, 12, 15 e 18 pb). Como controlo positivo, utilizou-se uma nuclease de dedos de zinco bem caracterizada com um domínio de ligação de ADN derivado do factor de transcrição de murganho *Zif268* (Porteus e Baltimore (2003) *Science* 300:763). Como controlos negativos, os domínios de efector TAL fundiram-se com uma variante de *Fok*I inactiva sob o ponto de vista catalítico ou foram ensaiados contra alvos não-cognatos de ADN.

As células de tipo haplóide contendo quer plasmidos de expressão de TALEN ou plasmidos-alvo em 200 µL de cultura,

durante a noite, foram conjugadas em meio YPD, a 30 °C. Passadas 4 horas, substituiu-se o meio YPD com 5 mL de médio selectivo e incubou-se, durante a noite, a 30 °C. Fez-se a lise das culturas conjugadas, adicionou-se o substrato ONPG e fez-se a leitura da densidade óptica a 415 nm utilizando um leitor de placas de 96 poços (Townsend et al., *supra*). Calculou-se os níveis de β -galactosidase em função da velocidade de clivagem do substrato. Os resultados obtidos com as estruturas-alvo relatoras, que tinham um espaçador de 15 pb a separar os dois sítios de reconhecimento estão ilustrados na figura 22B. Todas as estruturas de expressão de nucleases derivadas do fragmento *SphI*, que codificou principalmente a matriz de repetição falharam em exibir a sua actividade, indicando que as sequências de aminoácidos, para além daquelas que estão na matriz de repetição são necessárias para a função (figura 22B). Observou-se contudo uma actividade robusta tanto para as TALEN de AvrBs3 como de PthXo1 derivadas do fragmento *BamHI* (figura 22B). A actividade de TALEN de PthXo1 aproximou-se do controlo positivo de NDZ. A actividade exige que o domínio funcional *FokI* seja específico para o alvo de ADN reconhecido por uma dada TALEN.

As experiências também levaram ao ensaio de várias distâncias entre os sítios de ligação do efector TAL (variantes com um comprimento de 11 e entre 12 e 30 pb), de modo a identificar os comprimentos dos espaçadores que permitem que *FokI* dimerize mais eficazmente (figura 28A). Ambas as enzimas demonstraram dois comprimentos de espaçadores óptimos - um a 15 pb e o outro quer com 21 pb (AvrBs3) ou 24 pb (PthXo1). Para PthXo1, observou-se uma actividade para todos os comprimentos de espaçadores ensaiados de 13 pb ou mais longos. Alguns comprimentos de espaçadores para AvrBs3

não mostraram actividade, sugerindo contudo que o comprimento do espaçador é crítico para certas TALEN.

As experiências anteriores de ensaio da actividade de TALEN homodiméricas, que se ligam a duas sequências idênticas de reconhecimento colocadas em oposição a cada um dos lados do espaçador. Uma vez que esses sítios palindrômicos provavelmente ocorrem naturalmente em alvos genómicos, realizaram-se as experiências para ensaiar se as TALEN podiam funcionar como heterodímeros. Os sítios de reconhecimento de AvrBs3 e PthXo1 foram colocados numa orientação da cabeça para a cauda em ambos os lados de um espaçador de 15 pb. A actividade das TALEN AvrBs3 e PthXo1 individualmente e de Zif268 nos seus alvos respectivos foi medida como controlos. Como controlo negativo, ensaiou-se uma cultura de levedura apenas com um plasmido do sítio-alvo para o sítio heterodimérico para avaliar a actividade de lacZ. A actividade resultante da TALEN heterodimérica aproximou-se da média das actividades observadas com duas enzimas homodiméricas (figura 28B).

Para ensaiar se os domínios de repetição podem ser unidos às TALEN-alvo com sequências cromossómicas arbitrárias, escolheram-se dois genes que foram previamente recombinados para a mutagénese com NDZ - *ADH1* para *Arabidopsis* e *gridlock* de peixe-zebra (Foley et al. (2009) PLoS One 4:e4348; e Zhang et al., *supra*). Realizou-se uma pesquisa para sequências de 12-13 pb nas regiões de codificação, que eram precedidas por um T de 5' e uma composição de nucleótidos semelhante à dos sítios de ligação dos efectores TAL identificados por Moscou e Bogdanove (*supra*). Em *ADH1* e *gridlock*, esses sítios ocorreram, em média, em 7-9 pb. Selecionara-se quatro sítios de 12 pb em *ADH1* (nas posições 360, 408, 928 e 975 da sequência

cromossômica de genes) e um sítio de 13 pb em *gridlock* (na posição 2.356 da sequência cromossômica de genes; figura 29A). Os domínios de repetição do efector TAL foram preparados de modo a reconhecer estes alvos, utilizando os DVR mais abundantes dos efectores de TAL naturais (NI para A, HD para C, NN para G e NG para T). Para preparar TALEN customizadas, sintetizaram-se repetições com estes DVR individualmente e juntaram-se em módulos de uma, duas ou três repetições, tal como descrito nos exemplos 4 e 5. Estes módulos foram ligados sequencialmente num derivado do gene *Tallc* (Moscou e Bogdanove, *supra*) a partir do qual as repetições originais tinham sido eliminadas e fragmentos de *Bam*HI a partir destes efectores TAL tratados por engenharia, que se fundiram com as sequências que codificam o domínio catalítico de *Fok*I em pFZ85 (figura 25). Criaram-se cinco TALEN customizadas recombinadas com *ADH1* de *Arabidopsis* e o gene *gridlock* do peixe-zebra.

Fez-se o ensaio das TALEN customizadas resultantes num ensaio de levedura como TALEN homodiméricas (o sítio de ligação do ADN idênticos estava duplicado numa orientação inversa em cada um dos lados de um espaçador com 16-18 pb), embora se tenha notado que as TALEN heterodiméricas não devem ser construídas de forma a clivarem directamente alvos de ADN de ocorrência natural. Escolheram-se os comprimentos dos espaçadores com base na distância mais próxima 15 pb da extremidade 3' do sítio candidato da vizinhança próxima (e oposta). Utilizaram-se espaçadores de 16 pb para *ADH1-360-12*, *ADH1-408-12r* e espaçadores de 18 pb para *ADH1-928-12*, *ADH1-975-12r* e *gridlock-2356-13r*. O ensaio de levedura foi realizado tal como se descreveu antes.

Observou-se uma actividade robusta da nucleasa para as TALEN de *ADH1-360-12* e *gridlock-2356-13r* (figura 29B). A

TALEN de *ADH1-928-12* tinha uma actividade modesta que, apesar de tudo, não era insignificante acima dos controlos negativos. Para cada TALEN que deu resultados positivos, a actividade da nuclease era específica do alvo cognato. Estes resultados indicam que se podem criar novas TALEN funcionais juntando domínios de repetição customizados.

Exemplo 7 - Alvos de ocorrência natural e pares de efectores TAL mostram desvios globais e posicionais na composição de nucleótidos e de DVR

Os alvos de 20 pares de bases e os efectores TAL analisados por Moscou e Bogdanove (*supra*) foram avaliados para os desvios da composição global e para os efeitos posicionais em nucleótidos ou frequências de DVR. Observou-se que estes sítios (em hélices positivas) eram geralmente ricos em A e C e pobres em G. a percentagem média de A foi de 31 ± 16 % (desvio-padrão de 1). A percentagem média de C foi de 37 ± 13 %. A percentagem média de G foi de 9 ± 8 % e a percentagem média de T foi de 22 ± 10 %. Dado que os alinhamentos variam nos comprimentos, a análise dos efeitos posicionais foi restringida a cinco posições em cada extremidade. Surpreendentemente, os desvios nas sequências-alvo foram evidentes para A e contra T nas posições 1 e 3 e para T na posição N e, possivelmente, 2. G foi particularmente raro na posição N-1. Este desvio foi reflectido pelos DVR que se emparelham nos efectores, sendo NI o mais comum nas posições 1 e 3, não existindo NG na posição 1, praticamente todos os NG na posição N e raramente NN na posição N-1 (figura 30).

Exemplo 8 - Processo e reagentes para a junção e a clonagem rápida de matrizes de repetição de efectores TAL customizadas

O processo de clonagem Golden Gate [Engler et al. (2008), *supra*; e Engler et al. (2009), *supra*] utiliza a capacidade das endonucleases de restrição do tipo IIS (por exemplo, *BsaI*) para cortar fora dos seus sítios de reconhecimento, para criar extensões customizadas para uma ligação ordenada de fragmentos múltiplos de ADN, feita simultaneamente. Utilizando este processo, pode fundir-se vários fragmentos de ADN numa matriz, numa ordem específica e cloná-los num vector-destino desejado, numa única reacção (figura 31).

Um processo e os reagentes para a junção de matrizes de codificação de repetições de efectores TAL customizadas foram desenvolvidos com base no sistema Golden Gate. Quando os sítios *BsaI* estão posicionados em cada um dos lados da sequência de codificação da repetição de efector TAL, a clivagem liberta um fragmento de repetição flanqueado por pendentes de 4 pb. Como o sítio de clivagem não é específico da sequência, por alternância, os clones de repetição podem ser libertados por uma certa ordem saliências complementares (extremidades pegajosas), permitindo a junção ordenada de matrizes de várias repetições.

Gerou-se uma biblioteca de 58 plasmidos (figuras 32A e 32B) para permitir a junção simultânea de até 10 unidades de repetição em "sub-matrizes" seguido da junção simultânea de uma, duas ou três destas sub-matrizes em conjunto com uma repetição final truncada numa matriz completa customizada. Dez conjuntos alternados de quatro fragmentos, em que cada fragmento está num conjunto que codifica um módulo de repetição com um DVR, diferente dos quatro DVR mais comuns,

HD, NG, NI e NN, foram sintetizados e clonados num vector comportando o gene de resistência à tetraciclina, para um total de 40 plasmidos. Sintetizam-se mais quatro fragmentos que codificaram a repetição do efector TAL truncado de terminal de 20 aminoácidos, codificando cada fragmento um diferente dos quatro DVR mais comuns e clonaram-se num vector diferente comportando o gene de resistência à espectinomicina, para se obter quatro plasmidos, designados como "plasmidos da última repetição" (figura 32A). Todos os fragmentos nos conjuntos alternados são flanqueados por sítios de *BsaI* no vector, de modo que a clivagem com *BsaI* liberta os fragmentos com diferentes extremidades pegajosas que permitem a junção na ordem apropriada; isto é, a saliência na extremidade 3' de um fragmento para o módulo de repetição 1 é complementada apenas na saliência da extremidade 5' do fragmento para o módulo de repetição 2, a saliência na extremidade 3' do módulo de repetição 2 é complementada apenas na saliência da extremidade 5' do módulo de repetição 3 e etc.. Os fragmentos nos últimos plasmidos de repetição estão flanqueados por sítios para uma endonuclease de restrição diferente do tipo IIS, *Esp3I*. Prepararam-se catorze plasmidos adicionais, que se descrevem a seguir, como vectores de destino para receber sub-matrizes que se juntaram.

O primeiro vector de destino, o plasmido pFUS_A foi construído para receber a primeira sub-matriz de 10 repetições para serem associadas numa matriz final de 21 repetições ou menos (contando a repetição final truncada). Construiu-se pFUS_A de tal modo que a clivagem por *BsaI* cria uma saliência num lado complementar à saliência na extremidade 5' do primeiro módulo de repetição e uma saliência no outro lado complementar à saliência na extremidade 3' do 10º módulo de repetição. Para receber uma

sub-matriz de 10 ou menos repetições para serem juntas nessa matriz final, construíram-se plasmido do vector de destino pFUS_B1, pFUS_B2, pFUS_B3, pFUS_B4, pFUS_B5, pFUS_B6, pFUS_B7, pFUS_B8, pFUS_B9 e pFUS_B10 que, quando clivados por *BsaI* tinham saliências respectivamente complementares à saliência da extremidade 5' do primeiro módulo de repetição e a extremidade 3' do módulo de repetição para a correspondente posição numerada (por exemplo, saliência de pFUS_B6 para a extremidade 3' da sub-matriz emparelha com a saliência dos quatro fragmentos de módulos de repetição para a posição 6). As matrizes clonadas nas séries de pFUS_A e de pFUS_B dos plasmidos são flanqueadas pelos sítios *Esp3I* no vector e, quando libertadas por digestão com *Esp3I* as matrizes têm saliências complementares únicas que lhes permitem ligar-se numa certa ordem com um fragmento de repetição final truncado, no vector de destino pTAL, que codifica uma TALEN a que falta a região de repetição. Construiu-se pTAL de tal maneira que a clivagem com *Esp3I* permite a inserção da matriz de repetição na localização correcta e na orientação correcta, em virtude de uma saliência numa extremidade que é complementar da saliência na extremidade 5' da primeira sub-matriz de dez repetições e uma saliência na outra extremidade complementar com a saliência na extremidade 3' do fragmento de repetição final truncado (figura 33).

Os plasmidos finais dos dois vectores de destino pFUS_A30A e pFUS_A30B foram construídos para receber a primeira e a segunda sub-matrizes de dez repetições para serem unidas numa matriz final de 22-31 repetições. Construiu-se pFUS_A30A e pFUS_A30B de tal modo que a digestão com *Esp3I* liberta as matrizes com as saliências complementares apropriadas, de tal modo que as matrizes podem ser ligadas por ordem com uma terceira matriz a partir de um vector pFUS_B e um fragmento de repetição final truncada a

partir de um último plasmido de repetição, libertado do mesmo modo por digestão com *Esp3I* em pTAL (figura 32).

Todos os vectores de destino tinham o gene *lacZ* clonado entre os sítios de endonuclease de restrição do tipo IIS, permitindo um rastreio a azul e branco para os recombinantes. Exceptuando pTAL, que comportam o gene de resistência à ampicilina, todos os vectores de destino comportam um gene de resistência à espectinomicina.

Para construir rapidamente uma repetição de um efector TAL customizado utilizando estes reagentes, estabeleceu-se o processo que se segue. Numa primeira etapa, os plasmidos dos módulos de DVR individuais, apropriados, para as sub-matrizes necessárias de dez ou menos repetições, são misturados em conjunto com o vector de destino apropriado num tubo. Adicionou-se ligase de ADN T4 e endonuclease *BsaI* e faz-se uma incubação da mistura reaccional numa máquina de RCP, durante 10 ciclos de 5 minutos, a 37 °C e 10 minutos, a 16 °C, as respectivas temperaturas óptimas para as duas enzimas. A mistura reaccional é então tratada com a nuclease PLASMID-SAFE™ para hidrolisar todos os fragmentos lineares de ADNhd, de modo a prevenir a clonagem de matrizes incompletas mais curtas por recombinação *in vivo* e depois utilizou-se a mistura para transformar quimicamente as células apropriadas de *E. coli*. Os plasmidos recombinantes resultantes são isolados e confirmou-se as estruturas correctas. Depois, numa segunda etapa, os plasmidos confirmados da primeira etapa misturou-se com os plasmidos apropriados da última repetição e pTAL e realiza-se um ciclo de reacção de digestão e ligação, tal como na primeira etapa. Finalmente, os produtos da reacção são introduzidos em *E. coli* e a estrutura final da matriz, de comprimento completo, é isolada e confirmada. O

protocolo pode ser completado por um especialista dentro de um período de uma semana.

As estruturas de expressão para as TALEN 85, 102 e 117 no quadro 4A, assim como as TALEN HPRT-3254-17 e HPRT-3286-20r, descritas a seguir no exemplo 14, foram feitas utilizando um processo e reagentes descritos neste exemplo.

Subclonaram-se as matrizes de repetição clonadas em pTAL facilmente noutros contextos de genes efectores de TAL utilizando os sítios de endonucleases de restrição conservados de *SphI* que flanqueiam a região de repetição.

Exemplo 9 - Os dos de TALEN customizada mostra um suporte inicial para as "regras" e correlação entre números de DVR e a actividade

O exemplo 6 descreve experiências que levaram à engenharia do domínio de ligação de ADN de TALEN, de tal modo que pode reconhecer sequências únicas de ADN. Tal como descrito, estas TALEN customizadas reconheceram sítios no *ADH1* de *Arabidopsis* e nos genes *gridlock* de peixe-zebra. Outros domínios de ligação do ADN do efector de TAL customizado foram tratados por engenharia para reconhecer não só sítios nestes genes mas também no gene *TT4* de *Arabidopsis* e *telomerase* do peixe-zebra (Foley et al., *supra*; e Zhang et al., *supra*). Estas TALEN customizadas foram feitas utilizando os processos descritos nos exemplos 3, 4 e 8. Na engenharia das TALEN customizadas, as distorções da composição e posicionais observadas foram adoptadas como princípios de concepção ou "regras". Primeiro, uma pesquisa conduziu a sequências das regiões de codificação que eram precedidas por um T de 5' e pelo menos 15 pb de comprimento e que tinham uma composição de nucleótidos consistente com as médias referidas

antes. Especificamente, apenas os sítios com 0-63 % de A, 11-63 % de C, 0-25 % de G e 2-42 % de T foram seleccionados. Em médios, esses sítios ocorreram para cada 7-9 pb. Seleccionaram-se então os sítios que estavam conforme com os desvios posicionais observados descritos antes. A partir deste conjunto, identificaram-se dois pares de sítios de ligação em cada gene que tinham 15-19 pb de comprimento e estavam separados por 15-18 pb, de tal modo que a ligação das TALEN tratadas por engenharia iriam permitir que *FokI* dimerizasse. Os processos de junção modular (exemplos 3 e 4) geraram estruturas de comprimento parcial.

No total, 21 TALEN intermédias e de comprimento completo concebidas para atingir 16 sequências de nucleótidos, cada uma com uma matriz de nove repetições ou mais. As sequências de aminoácidos destas TALEN estão indicadas nas figuras 34A-34U (SEQ ID N.º. 35-55). Estas 21 TALEN foram ensaiadas quanto à sua capacidade para clivar ADN utilizando o ensaio de levedura descrito nos exemplos 2 e 6. Os dados da actividade estão ilustrados na figura 35 e resumidos no quadro 4A.

Alguns das TALEN intermédias, de comprimento parcial, correspondem a alvos que quebram as regras da composição dos nucleótidos e do terminal T. O quadro 4A mostra o comprimento, a conformidade com estas duas regras e a actividade relativa com a de NDZ268 para cada TALEN. Os resultados revelam uma tendência geral que aumentando o comprimento da matriz de DVR aumenta a actividade da TALEN resultante. Isto sugere que há um número mínimo de DVR que é necessário antes de um alvo de ADN poder ser reconhecido *in vivo*. Além disso, a conformidade com as regras parece ser importante. Uma das seis TALEN não exibiu actividade detectável, duas violaram a regra da composição do alvo, duas não terminaram em NG e a outra quebrou ambas as regras (uma

obedeceu a ambas as regras). Três das oito TALEN com actividade inferior a 25 % de NDZ268 violaram uma das regras e uma das quatro TALEN com actividade 25-50 % de NDZ268 não tinha uma sequência de DVR terminando em NG. Notou-se que as TALEN com actividade de 50 % ou mais do que NDZ268 obedeciam a todas as regras e para as TALEN com o mesmo comprimento, aquelas que quebravam as regras geralmente tinham menos actividade do que as matrizes obedientes. A consistente com a tendência geral no que respeita ao comprimento, mesmo para produtos intermédios que não quebram as regras, as correspondentes TALEN de comprimento completo tinha uma actividade mais elevada (quadro 4A e figura 35). A variação do comprimento do espaçador, devido às diferenças do comprimento da TALEN no mesmo alvo pode ter contribuído para esta observação, mas tolera-se algum intervalo de comprimentos do espaçador (Christian et al., *supra*).

Foram evidentes algumas complexidades dos dados. Por exemplo, a actividade variou entre TALEN obedientes do mesmo comprimento, algumas matrizes curtas tinham uma actividade altamente moderada e algumas matrizes longas que eram obedientes tinham pouca ou nenhuma actividade (quadro 4B). Apesar disso, os resultados deram suporte às conclusões de que 1) geralmente um número maior de repetições resulta numa maior actividade e 2) a conformidade com as regras de distorção da composição e posicionais são importantes no que respeita à actividade. Por isso, derivaram-se os seguintes princípios de concepção.

- Os sítios de ligação do efector TAL são concebidos de modo a terem o mínimo 15 bases de comprimento e estarem orientados de 5' para 3' com um T a preceder imediatamente o sítio na extremidade 5', e.

- Um sítio não pode ter um T na primeira posição (5') ou um A na segunda posição.
- Um sítio deve terminar em T (3') e pode não ter um G na posição a seguir à última posição.
- A composição de base do sítio deve estar dentro dos intervalos especificados (média \pm dois desvios-padrão): 0-63 % de A, 11-63 % de C, 0-25 % de G e 2-42 % de T.

Quadro 4A

<i>Actividade, conformidade com as regras e comprimento das TALEN ensaiadas em ensaios de levedura</i>							
Gene	TALEN	Nomes de Christian et al. (supra)	DVR	Activid.	% de GATC	Extr. em NG	Sequência de DVR¹
telomerase	124		9	-	N	Y	HD NN NN NG NG NG NN HD NG
gridlock	105		10	+	N	N	NI HD HD HD HD NG HD NG HD HD
ADH1	58	ADH1-360-12	12	++	Y	N	NI NG HD NI NI NN NI NG NG HD NG HD
ADH1	63	ADH1-408-12r	12	-	Y	N	HD HD HD NI NN NI NI NN NG NI NI NI
ADH1	68	ADH1-928-12	12	+	Y	Y	HD HD NN NN NI NG NN HD NG HD HD NG
ADH1	73	ADH1-975-12r	12	-	N	N	NI NN NI HD NI NI NI HD HD NI HD NI
TT4	89		12	-	Y	N	NN NN HD NI HD NG NN HD NG NI NI HD

Quadro 4A

<i>Actividade, conformidade com as regras e comprimento das TALEN ensaiadas em ensaios de levedura</i>							
Gene	TALEN	Nomes de Christian et al. (<i>supra</i>)	DVR	Activid.	% de GATC	Extr. em NG	Sequência de DVR ¹
gridlock	106	gridlock-2356-13r	13	++	Y	Y	NI HD HD HD HD NG HD NG HD HD NN HD NG
ADH1	64		15	+	Y	Y	HD HD HD NI NN NI NI NN NG NI NI NI HD NI NG
ADH1	69		15	+++	Y	Y	ND ND NN NN NI NG NN HD NG HD HD NG HD NG NG
ADH1	74		15	++	Y	Y	NI NN NI HD NI NI NI HD HD NI HD NI NI HD NG
TT4	90		15	-	Y	Y	NN NN HD NI HD NG NN HD NG NI NI HD HD HD NG
telomerase	121		15	+	Y	Y	HD NG NG NN NG HD HD NN HD NI NG NN NI NG NG
telomerase	126		15	-	N	Y	HD NN NN NG NG NG NN HD NG NI NG HD NN NG NG
gridlock	107		16	++++	Y	Y	NI HD HD HD HD NG HD NG HD HD NN HD NG NG HD NG
gridlock	117		16	++	Y	Y	HD HD HD NN NN NI NI NN

Quadro 4A

<i>Actividade, conformidade com as regras e comprimento das TALEN ensaiadas em ensaios de levedura</i>							
Gene	TALEN	Nomes de Christian et al. (<i>supra</i>)	DVR	Activid.	% de GATC	Extr. em NG	Sequência de DVR¹
							HD HD NN NI HD NN HD NG
telomerase	131		16	+	Y	Y	NI NG NG HD HD HD HD NI HD NN NI NN HD NG HD NG
telomerase	136		17	+	N	Y	NI NN NI HD NI NN NN NI NI NN NG NN NN NI NN HD NG
ADH1	60		18	+++++	Y	Y	NI NG HD NI NI NN NI HG NG HD NG HD NG NG HD NI HD NG
TT4	85		18	+	Y	Y	NI HD NG HD HD NN HD HD NG NN NI NI NN HD NI HD NI NG
gridlock	102		18	+	Y	N	NN NN HD NG HD NI HD HD NG NI HD NI NI HD NN NI HD NI
¹ As sequências-alvo ensaiadas consistem em repetições invertidas das correspondentes sequências de nucleótidos, em que HD, NG, NI e NN correspondem, respectivamente, a C, T, A e G, separadas por uma sequência espaçadora de 16-18 pb.							

Quadro 4B

Excerto do quadro 4A, ordenado pelo nível de actividade

DVR	Actividade	% de GATC	Extremidades em NG
9	-	n	y
12	-	y	n
12	-	n	n
12	-	y	n
15	-	y	y
15	-	n	y
10	+	n	n
12	+	y	y
15	+	y	y
15	+	y	y
16	+	y	y
17	+	n	y
18	+	y	y
18	+	y	n
12	++	y	n
13	++	y	y
15	++	y	y
16	++	y	y
15	+++	y	y
16	++++	y	y
18	+++++	y	y

Exemplo 10 - Pares de TALEN heterodiméricas que clivam as suas pretendidas sequências-alvo de ocorrência natural num ensaio de leveduras

Os dados nos exemplos 2, 6 e 9 demonstram que as TALEN customizadas podem ser tratadas por engenharia para reconhecer novas sequências-alvo de ADN. Os dados da actividade de levedura para as TALEN customizadas foram reunidos utilizando monómeros individuais de TALEN, que

reconheceram um sítio-alvo homodimérico. Isto é, a sequência-alvo da TALEN foi duplicada numa orientação inversa do outro lado de um espaçador de 15-18 pb. A clivagem das sequências cromossómicas endógenas, contudo, geralmente irá exigir que duas TALEN customizadas diferentes reconheçam duas sequências diferentes em cada um dos lados de um espaçador. Tal como descrito no exemplo 6, esta capacidade foi demonstrada para as TALEN de AvrBs3 e PthXo1 em conjunto com um sítio-alvo polimérico correspondente, num ensaio de levedura. Os requerentes analisaram se duas TALEN diferentes customizadas podiam reconhecer e clivar uma sequência de ADN de ocorrência natural. Utilizando o ensaio da levedura descrito no exemplo 2, as TALEN customizadas, concebidas para clivar duas sequências-alvo diferentes no gene *ADH1* de *Arabidopsis* foram ensaiadas quanto à sua actividade nestes alvos. As sequências de ADN dos sítios-alvo e as correspondentes TALEN estão ilustradas na figura 36A. As sequências de aminoácidos das TALEN estão indicadas na figura 34. A actividade de β -galactosidase obtida no ensaio de levedura está desenhada num gráfico ilustrado na figura 36B. A actividade das TALEN na sua sequência-alvo de ocorrência natural estava significativamente acima dos controlos negativos, indicando que as TALEN podem ser tratadas por engenharia para reconhecer e clivar as sequências-alvo de ADN endógeno.

Exemplo 11 - As TALEN clivam genes naturais em *Arabidopsis* e introduzem mutações por junção da extremidade imprecisa e não-homóloga

Um dos pares activos de *ADH1* de *Arabidopsis* foi ensaiada para determinar se se pode ligar, clivar e mutar ADN cromossómico. Cada uma das TALEN individuais de *ADH1* que compreendia este par (pTALEN 69 e 74) foi clonada num vector de expressão de plantas pFZ14, que coloca as TALEN sob o

controlo de um promotor constitutivo 35S (Zhang et al., *supra*). As estruturas resultantes foram então introduzidas em protoplastos de *Arabidopsis* por electroporação. Passadas 48 horas, isolou-se o ADN genómico e digeriu-se com Tth1111. Um sítio de clivagem de Tth1111 está colocado na sequência espaçadora entre os dois sítios de reconhecimento de TALEN (figura 37A). Seria de esperar que a clivagem do ADN cromossómico pela TALEN introduzisse mutações por uma junção de extremidades imprecisas e não-homólogas (JENH), que resultaria num falhanço da clivagem por Tth1111. Um fragmento de 375 pb que engloba o sítio de reconhecimento de TALEN foi então amplificado por RCP. O produto da RCP foi digerido em função de Tth1111 para eliminar a maior parte do ADN genómico remanescente que não tinha sido modificado por JENH mediada por TALEN. Os produtos da digestão sofreram então um ciclo em gel de agarose. Observou-se um produto de RCP não clivado e esses produtos de RCP não clivados são diagnósticos para a actividade de nuclease (neste caso a actividade de TALEN) na sequência-alvo endógena (Zhang et al., *supra*). Clonou-se o ADN não cortado e analisou-se por sequenciação de ADN. A sequenciação de nove clones independentes revelou que seis completavam mutações introduzidas por JENH (figura 37B). Assim, as TALEN clivam locais cromossómicos endógenos e introduzem quebras da hélice dupla de ADN e mutações.

Exemplo 12 - Aumento da capacidade de recombinação

No núcleo do código do ADN de efector TAL, os quatro DVR mais comuns têm, cada um, uma especificidade aparente de um-para-um para os quatro nucleótidos, com base nas frequências de associação. Isto foi marcadamente assim para HD, NG e NI, mas menos para NN (figura 1C). NN associa-se mais frequentemente com G, mas quase sempre com A e algumas vezes com C ou T. Para um efector TAL reunido aleatoriamente com NN

em quatro localizações numa sequência de 13 DVR, com G em todas as posições correspondentes num alvo artificial originou a melhor actividade (Boch et al. (2009) *Science* 326:1509-1512). A reduziu mas não aboliu a actividade e C e T eliminaram a actividade detectável. Observou-se uma drástica perda de actividade quanto C, T ou A foram substituídos por G justamente na primeira posição no sítio de ligação para os 24 efectores de DVR PthXo1, que é um NN (Romer et al. (2010) *New Phytol.* 187:1048-1057). Isto estava, contudo, em contraste com a observação de que o AvrHah1 muito mais curto (14 DVR) começa com um NN que alinha com A e o efector PthXo6 de 23 DVR, que tem três NN numa fila nas posições 4-6 que alinham cada uma com A, em que ainda ambas estas proteínas são altamente activas (ver, Schornack et al. (2008) *New Phytol.* 179:546-556; e Romer et al., *supra*). Assim, a especificidade de NN para G parece ser geralmente fraca e pode variar com o contexto.

A invariância observada da tímida que precede imediatamente os sítios-alvo do efector TAL é um requisito para vários efectores [Boch et al., *supra*; Romer et al., *supra*; e Romer et al. (2009) *Plant Physiol.* 150:1697-1712]. A sequência de aminoácidos que precede imediatamente a região de repetição em efectores TAL, que está altamente conservada (figura 38A), partilha uma semelhança significativa com a repetição, tanto em sequências de aminoácidos como na prevista estrutura secundária (figura 38B e Bodganove et al. (2010) *Curr. Opin. Plant Biol.* 13:394-401). Pôs-se a hipótese de que esta sequência, designada por repetição de "ordem 0", é a base para o requisito de T na posição -1 do sítio de ligação e que os resíduos na posição análoga de DVR (figura 38B) especificam o nucleótido.

Com base nestas observações, pôs-se a hipótese de que incorporando repetições com uma elevada especificidade para G

e relaxando o requisito para T na posição -1, a capacidade de recombinação para proteínas de efectores TAL tratadas por engenharia pode ser aumentada. As experiências foram iniciadas para ensaiar novos e raros DVR quanto a uma especificidade mais robusta para G do que a que exhibe NN e para substituir os resíduos análogos de DVR da repetição de ordem 0 por DVR comuns.

Novos e raros DVR para uma especificidade robusta para G: Os módulos descritos antes (ver, por exemplo, exemplo 4) utilizaram quatro DVR particulares (NI, HD, NN e NG) para especificar a ligação às quatro bases de nucleótidos (A, C, G e T, respectivamente). As repetições que continham outros DVR também podiam ser úteis e podem ter aumentado a especificidade e/ou a afinidade para as quatro bases em comparação com NI, HD, NN e NG. Para aumentar a especificidade para G, prepararam-se várias repetições que codificam DVR novas e raras. As DVR raras NK, HN e NA associadas com G, sugeriram que N pode ser importante como um dos outros resíduos (figura 1C). Assim, construiu-se um vasto conjunto de derivados que codificam repetições com DVR ilustrados no quadro 5. A coluna da esquerda lista os DVR que têm um aminoácido polar (R, K, D, E, Q, H, S, T ou Y) na posição 12 e N na posição 13. As clonas da direita listam as combinações de N na primeira posição com qualquer um dos outros 17 aminoácidos (G, L, V, R, K, D, E, Q, H, T, M, C, P, Y, W ou F) na segunda posição do DVR. Para ter em conta a possibilidade de uma maior especificidade sem N, também se prepararam repetições com um aminoácido polar (R, K, D, E, Q, H, S, T ou Y) na posição 12, um intervalo (*) na posição 13 (coluna do meio).

Ensaíram-se novos DVR artificiais quanto à função num ensaio quantitativo à base de um gene repórter para avaliar a

actividade da activação transcricional de efectores TAL, tais como GUS ou à base de um repórter duplo de luciferase, um ensaio de expressão transitória mediado por *Agrobacterium* em *Nicotiana benthamiana* ou no ensaio de TALEN à base do repórter *lacZ* em *Saccharomyces cerevisiae*, descrito antes (ver, por exemplo, exemplo 2) Incorporam-se os módulos de repetição contendo os DVR que vão ser analisados num efector TAL ou numa TALEN com níveis mensuráveis e de sub-saturação de actividade e ensaiam-se as proteínas resultantes para avaliar as diferenças na actividade de um conjunto de alvos de ADN com permutações integradas dos quatro nucleótidos nas posições correspondentes. Em particular, começando com as variantes de PthXo1 minimamente activas em *in planta* e ensaios de leveduras e respostas a desemparelhamentos em três repetições adicionadas, ensaiam-se as TALEN contendo cada uma das novas e raras repetições (em árvores homoméricas), *in vivo*, em função de alvos com G em cada uma das posições correspondentes. Para aquelas em que se verificou uma actividade aumentada, repetiram-se os ensaios com alvos permutados com outros nucleótidos nessas posições para acertar a especificidade.

Quadro 5. DVR a serem ensaiados^a

Polar + N	Polar*	N + todos	
RN	R*	NG	NH
KN	K*	NA	NT
DN	D*	NL	NM
EN	E*	NV	NC
QN	Q*	NR	NP
HN	H*	NK	NY
SN	S*	ND	NW
TN	T*	NE	NF
YN	Y*	NQ	

Quadro 5. DVR a serem ensaiados^a

Polar + N	Polar*	N + todos
^a São conhecidas as frequências de associação dos N*, NG e NS. Um, asterisco representa um intervalo correspondente à 2 ^a posição no DVR (isto é, a 13 ^a posição da sequência de consenso da repetição).		

Substituições comuns de DVR em posições análogas de DVR, da repetição de ordem 0, para diminuir a especificidade de T na posição -1: As previsões de estruturas secundárias e o alinhamento das sequências de repetição de ordem 0 e das sequências de consenso da repetição, sugeriram que as posições ocupadas por KR* (o asterisco indica um intervalo), na posição de ordem 0, eram análogas às do DVR e, por isso, foram os resíduos que especificaram T em -1. As variantes de PthXol com substituições de HD, NG, NI e NN para KR e, separadamente, para R* foram produzidas numa "estrutura" de Tallc descrita antes. As actividades destas variantes foram comparadas com o efector de tipo selvagem em ensaios *in planta* e em leveduras, utilizando alvos com nucleótidos correspondentes na posição -1, nomeadamente, C, T, A e G, respectivamente. Prepararam-se variantes adicionais de PthXol que tinham S, o resíduo na posição 11 da sequência de consenso da repetição, substituído por K na posição 11 da repetição de ordem 0. E prepararam-se outras variantes que tinham esta substituição combinada com uma substituição de K, o resíduo na posição 16 da sequência de consenso da repetição, para V na posição 15 da repetição de ordem 0 (quadro 6). Pode-se uma caixa TATA proximal para a actividade do efector TAL. Além disso, PthXol é útil para esta experiência porque, ao contrário de AvrBs3, para o qual o T em -1 parece fazer parte da caixa TATA, a caixa TATA mais

próxima do sítio de ligação de PthXol está 46 pb a jusante e não seria perturbada pelas modificações em -1.

Se as modificações anteriores não resultarem numa maior recombinação para G ou no aumento da capacidade para atingir sequências precedidas por nucleótidos diferentes de T, então fazem-se ensaios de um conjunto mais compreensivo de DVR artificiais para avaliar a especificidade de G e substituições diferentes das dos DVR comuns, para a repetição de ordem 0.

Quadro 6. Estruturas da repetição de ordem 0 preparadas e ensaiadas quanto à especificidade para a recombinação com A, C, G ou T na posição -

1

Sequências de repetição naturais de ordem 0 (específicas T em -1)	Especificação da substituição de T	Especificação da substituição de A	Especificação da substituição de C	Especificação da substituição de G
...KIA*KRGGV... (74) [†]	...KIA*NGGGV... (75)	...KIA*NIIGV... (76)	...KIA*HDGGV... (77)	...KIA*NMGGV... (78)
...KIA*KRGGV... (79)	...KIASNGGGV... (80)	...KIASNIIGV... (81)	...KIASHDGGV... (82)	...KIASNMGGV... (83)
...KIAKR*GGV... (84)	...KIAKNGGGV... (85)	...KIAKNIIGV... (86)	...KIAKHDGGV... (87)	...KIAKNGGGV... (88)
...KIA*KRGGV... (89)	...KIASNGGGK... (90)	...KIASNIIGK... (91)	...KIASHDGGK... (92)	...KIASNMGGK... (93)

Os DPR candidatos e substituídos estão a negro. Outras substituições ou modificações estão sublinhadas. Os asteriscos indicam um intervalo relativamente à sequência de consenso da repetição.

SPQ ID N°.

Exemplo 13 - Novos DVR específicos de nucleótidos previstos

Observou-se que quando os DVR listados nos quadros 1A e 1B foram agrupados pelo segundo resíduo de aminoácido no DVR (isto é, o 13º na repetição total), houve uma correlação quase perfeita desses aminoácidos com os nucleótidos especificados pelo DVR, independentemente do aminoácido na primeira posição do DVR (quadro 7). Assim, os DVR que terminam num intervalo (indicado por um asterisco) especificam C ou T ou T; os DVR que terminam num D especificam C; os DVR que terminam num G especificam T; e os DVR que terminam em N especificam G ou A ou G. Também se observou que os aminoácidos na posição 1 do DVR eram qualquer um de H, I, N, S ou Y. Estas observações sugeriram que se determina a especificidade do DVR pelo resíduo na segunda posição, independente de o resíduo na primeira posição ser H, I, N, S ou Y. Por isso, previram-se as especificidades para várias DVR novos (isto é, ainda não observados) que combinam resíduos observados na segunda posição com resíduos H, I, S, N ou Y na primeira posição. Assim, previu-se que I*, S* e Y* especificam C ou T ou T; ID previu-se que SD e YD especificam C; previu-se que SG especifica T; e previu-se que IN e YN especificam G ou A ou G. Embora houvesse um exemplo de um K na segunda posição, com base na especificidade de NK observada, previu-se que HK, IK, SK e YK especificam G.

Estes novos DVR foram analisados e comparados com DVR existentes quanto à função e a especificidade em ensaios quantitativos do efector TAL e da actividade de TALEN, tal como descrito nos exemplos 2 e 11.

Quadro 7¹

<i>DVR agrupados e rodeados pelo seu segundo resíduo</i>		
1º Resíduo	2º Resíduo	Nucleótido

Quadro 7¹

<i>DVR agrupados e rodeados pelo seu segundo resíduo</i>		
1^o Resíduo	2^o Resíduo	Nucleótido
N	*	C ou T
H	*	T
H	A	C
N	A	G
H	D	C
N	D	C
H	G	T
I	G	T
N	G	T
Y	G	T
N	I	A
H	I	C
N	K	G
H	N	G
S	N	G ou A
N	N	G ou A ¹
N	S	A ou C ou G ¹

¹ Um asterisco indica um intervalo. Os grupos de DVR com as mesmas especificidades estão dentro de linhas espessas.

Exemplo 14 - TALEN comuns clivam alvos endógenos em células de animais e introduzem mutações por junção de extremidades imprecisas não homólogas

Para analisar se as TALEN podem ser utilizadas para a mutagênese direccionada em células de animais, primeiro avaliou-se a expressão dos efectores TAL AvrBs3, PthXo1 e Tallc em células 293T de rim humano embrionário (RHE).

Eliminou-se o codão de paragem dos genes que codificam AvrBs3, PthXo1 e Tal1c e subclonaram-se os genes no vector de expressão de mamífero pcADN3.2/V5-DEST (Invitrogen, Carlsbad, CA) numa matriz, com a sequência a jusante desse vector, que codifica o epitopo V5 para a imuno-deteccção de proteínas. O pcADN3.2/V5-DEST coloca o gene efector TAL sob o controlo do promotor de citomegalovírus (CMV) humano constitutivo. Transfectaram-se as células 297T de RHE utilizando Lipofectamine 2000 (Invitrogen) obtendo-se como resultado individualmente plasmidos, passadas 24 horas, isolou-se o total das proteínas de cada lote de células transfectado e submeteram-se a uma electroforese em gel de poliacrilamida, análise por "western blotting" e imuno-marcação utilizando um anticorpo de murganho anti-V5. Detectaram-se as proteínas marcadas com um conjugado de anticorpo de cabra anti-murganho-peróxidase de rábano-silvestre, utilizando o estojo de quimioluminescência SuperSignal West Pico Chemiluminescent (ThermoScientific, Inc.). Confirmou-se a carga equivalente por imuno-marcação e deteção de actina. Cada proteína do efector TAL foi expressa de forma detectável sem degradação aparente (figura 39).

Em seguida, concebeu-se um par de TALEN, tal como descrito no exemplo 9, para recombinar uma sequência no gene endógeno humano *HPRT* e designou-se por HPRT-3254-17 e HPRT-3286-20r (figura 40A e figura 40B). O plasmido pTALEN141, que codifica HPRT-3254-17 e o plasmido pTALEN142, que codifica HPRT-3286-20r foram preparados utilizando o processo à base de clonagem designado por Golden Gate, utilizando os reagentes descritos no exemplo 8. Os genes de TALEN foram então subclonados no vector de expressão de mamífero pcADN3.1(-) (Invitrogen, Inc.), que os coloca sob o controlo do promotor constitutivo de CMV, originando os plasmidos pTALEN141M e pTALEN142M. Depois, transfectaram-se células

293T de RHE, tanto com pTALEN141M como com pTALEN142M, em conjunto e separadamente com pcADN3.1(-), como controlo negativo. Passadas 72 horas, isolou-se o ADN genómico e digeriu-se com as endonucleases de restrição *Bbu10I*. O sítio de *Bbu10I* existe dentro do espaçador que separa os sítios de ligação de HPRT-3254-17 e HPRT-3286-20r em *HPRT* (figura 41A). No seguimento da digestão com *Bbu10I*, utilizou-se uma RCP para amplificar um fragmento de 244 pb que se estendia no sítio recombinado de TALEN a partir tanto de TALEN tratado como das amostras de controlo. Amplificou-se o fragmento esperado a partir de ambas as amostras, indicando que a digestão com *Bbu10I* do ADN genómico tinha sido incompleta. A seguir à digestão dos produtos de RCP com *Bbu10I*, contudo, resultou uma clivagem completa do produto amplificado a partir das amostras de controlo, mas numa clivagem incompleta do produto, a partir da amostra tratada com TALEN (figura 41B). A presença do produto de RCP resistente à clivagem na amostra tratada com TALEN dá evidências de que o sítio endógeno de *Bbu10I* estava mutado *in vivo*, em resultado da reparação imperfeita por meio da junção de extremidades não homólogas de uma quebra de hélice dupla mediada por TALEN, no alvo pretendido, em *HPRT*. Assim, as TALEN podem ser utilizadas para mutagénese direccionada em células de mamíferos.

Lisboa, 15 de Janeiro de 2015.

REIVINDICAÇÕES

1. Processo para modificar o material genético de uma célula, caracterizado pelo facto de compreender:

(a) providenciar uma célula contendo uma sequência-alvo de ADN; e

(b) introduzir na célula um vector que codifica uma endonuclease efectora semelhante a um activador de transcrição (TAL), compreendendo a nuclease do effector TAL:

(i) um domínio de endonuclease que pode clivar ADN de hélice dupla, em que o domínio da endonuclease é de uma endonuclease de restrição do tipo II e

(ii) um domínio de um effector TAL que compreende várias sequências de repetição de efectores TAL que, em combinação, se ligam a uma sequência específica de nucleótidos na sequência-alvo de ADN,

de tal maneira que a endonuclease do effector TAL cliva a sequência-alvo de ADN dentro ou próximo da sequência de nucleótidos específica na célula ou na sua progenia.

2. Processo, de acordo com a reivindicação 1, caracterizado pelo facto de compreender ainda providenciar à célula um ácido nucleico contendo uma sequência homóloga pelo menos a uma porção da sequência-alvo de ADN, de tal forma que ocorra a recombinação homóloga entre a sequência-alvo de ADN e o ácido nucleico.

3. Processo, de acordo com a reivindicação 1 ou com a reivindicação 2, caracterizado pelo facto de a célula ser uma célula de mamífero.
4. Processo, de acordo com a reivindicação 1 ou com a reivindicação 2, caracterizado pelo facto de a célula ser uma célula de planta.
5. Processo, de acordo com uma qualquer das reivindicações 1 a 4, caracterizado pelo facto de a introdução compreender a transfecção da célula com um vector que codifica a endonuclease do efector TAL.
6. Processo, de acordo com uma qualquer das reivindicações 1 a 5, caracterizado pelo facto da endonuclease de restrição do tipo II ser *FokI*.
7. Processo, de acordo com uma qualquer das reivindicações 1 a 6, caracterizado pelo facto de o domínio do efector TAL que se liga a uma sequência específica de nucleótidos dentro da sequência-alvo de ADN compreender 10 ou mais sequências de repetição do efector TAL, mais preferencialmente, 15 ou mais sequências de repetição do efector TAL.
8. Processo, de acordo com a reivindicação 7, caracterizado pelo facto de a sequência de repetição do efector TAL compreender um di-resíduo variável de repetição (DVR), que determina o reconhecimento de um par de bases na sequência-alvo de ADN, em que cada sequência de repetição do efector TAL é responsável pelo reconhecimento de um par de bases na sequência-alvo de ADN e em que DVR compreende um ou mais de:

HD para o reconhecimento de C;
NG para o reconhecimento de T;
NI para o reconhecimento de A;
NN para o reconhecimento de G;
NS para o reconhecimento de A;
HG para o reconhecimento de T;
IG para o reconhecimento de T;
NK para o reconhecimento de G;
HA para o reconhecimento de C;
ND para o reconhecimento de C;
HI para o reconhecimento de C;
HN para o reconhecimento de G; e
NA para o reconhecimento de G.

9. Processo para gerar um ácido nucleico que codifica uma endonuclease de um efector TAL, processo esse caracterizado pelo facto de compreender:

(a) a identificação de uma primeira sequência única de nucleótidos no genoma de uma célula; e
(b) a síntese de um ácido nucleico que codifica uma endonuclease de efector TAL que compreende (i) várias sequências de repetição do efector TAL que, em combinação, se ligam à primeira sequência única de nucleótidos e (ii) uma endonuclease que gera uma quebra da hélice dupla dentro ou num sítio adjacente à primeira sequência de nucleótidos, em que cada sequência de repetição do efector TAL compreende um DVR que determina o reconhecimento de um par de bases no ADN-alvo, em que cada sequência de repetição do efector TAL é responsável pelo reconhecimento de um par de bases no ADN-alvo e em que a endonuclease do efector TAL compreende um ou mais dos seguintes DVR:

HD para o reconhecimento de C;
NG para o reconhecimento de T;
NI para o reconhecimento de A;
NN para o reconhecimento de G;
NS para o reconhecimento de A;
HG para o reconhecimento de T;
IG para o reconhecimento de T;
NK para o reconhecimento de G;
HA para o reconhecimento de C;
ND para o reconhecimento de C;
HI para o reconhecimento de C;
HN para o reconhecimento de G; e
NA para o reconhecimento de G.

10. Processo, de acordo com a reivindicação 9, caracterizado pelo facto de a primeira sequência de nucleótidos cumprir pelo menos um dos seguintes critérios:

- i) ter um mínimo de 15 bases de comprimento e estar orientada de 5' para 3' com um T a preceder imediatamente o sítio na extremidade 5';
- ii) não ter um T na primeira posição (5') ou um A na segunda posição;
- iii) terminar num T na última posição (3') e não ter um G na posição ao lado da última posição; e
- iv) ter uma composição de bases de 0-63 % de A, 11-63 % de C, 0-25 % de G e 2-42 % de T.

11. Processo de acordo com a reivindicação 9 ou com a reivindicação 10, caracterizado pelo facto de compreender ainda a identificação de uma segunda sequência de nucleótidos no genoma da célula, em que a primeira e a segunda sequências de nucleótidos estão separadas por 15-18 pb.

12. Processo de acordo com uma qualquer das reivindicações 9 a 11, caracterizado pelo facto de a endonuclease gerar uma quebra na hélice dupla entre a primeira e a segunda sequências de nucleótidos.
13. Monómero de endonuclease de efector TAL, caracterizado pelo facto de compreender um domínio de endonuclease e um domínio de ligação do ADN do efector TAL específico para um ADN-alvo, em que o domínio da endonuclease é de uma endonuclease de restrição do tipo II, em que o domínio de ligação do ADN compreende várias sequências de repetição do efector TAL, compreendendo cada sequência de repetição do efector TAL um DVR que determina o reconhecimento de um par de bases no ADN-alvo e em que cada sequência de repetição do efector TAL é responsável pelo reconhecimento de um par de bases no ADN-alvo.
14. Monómero de endonuclease do efector TAL, de acordo com a reivindicação 13, caracterizado pelo facto de a endonuclease de restrição do tipo II ser *FokI*.
15. Monómero de endonuclease do efector TAL, de acordo com a reivindicação 14, caracterizado pelo facto de o referido monómero funcionar como um dímero com outro monómero através de um sítio de reconhecimento bipartido com um espaçador, permitindo que *FokI* dimerize e crie uma quebra da hélice dupla no referido ADN-alvo dentro do referido espaçador.
16. Monómero de endonuclease do efector TAL, de acordo com a reivindicação 13, caracterizado pelo facto de o monómero da endonuclease do efector TAL compreender um ou mais dos seguintes DVR:

HD para o reconhecimento de C;
NG para o reconhecimento de T;
NI para o reconhecimento de A;
NN para o reconhecimento de G;
NS para o reconhecimento de A;
HG para o reconhecimento de T;
IG para o reconhecimento de T;
NK para o reconhecimento de G;
HA para o reconhecimento de C;
ND para o reconhecimento de C;
HI para o reconhecimento de C;
HN para o reconhecimento de G; e
NA para o reconhecimento de G.

17. Monómero da endonuclease do efector TAL, de acordo com a reivindicação 13, caracterizado pelo facto de a sequência de aminoácidos da referida endonuclease do efector TAL ter pelo menos 95 % de identidade com uma sequência seleccionada entre a SEQ ID N.º. 33 a SEQ ID N.º. 55, a SEQ ID N.º. 72 e a SEQ ID N.º. 73.
18. Processo para gerar um animal não humano, caracterizado pelo facto de compreender:
 - i) providenciar uma célula eucariótica contendo uma sequência-alvo de ADN na qual se deseja introduzir uma modificação genética, em que a célula contém um ácido nucleico que codifica uma endonuclease do efector TAL, de acordo com uma qualquer das reivindicações 13 a 17;
 - ii) gerar uma quebra da hélice dupla dentro da sequência-alvo de ADN com a endonuclease do efector TAL; e

iii) gerar um animal não humano a partir da célula ou da respectiva progenia em que ocorreu uma quebra da hélice dupla.

19. Processo, de acordo com a reivindicação 18, caracterizado pelo facto de compreender:

a introdução na célula de um ácido nucleico exógeno que compreende uma sequência homóloga a pelo menos uma porção do ADN-alvo, em que a introdução é feita em condições que permitem que ocorra a recombinação homóloga entre o ácido nucleico exógeno e a sequência-alvo do ADN na célula ou na sua progenia;
e
a geração de um animal não humano a partir da célula ou da respectiva progenia, em que tenha ocorrido a recombinação homóloga.

20. Processo para gerar uma planta, caracterizado pelo facto de compreender:

i) providenciar uma célula de planta contendo uma sequência-alvo de ADN na qual se deseja introduzir uma modificação genética pré-seleccionada, em que a célula da planta contém um ácido nucleico que codifica uma endonuclease do efector TAL, de acordo com uma qualquer das reivindicações 13 a 17;
ii) gerar uma quebra da hélice dupla dentro da sequência-alvo de ADN com a endonuclease do efector TAL; e
iii) gerar uma planta a partir da célula ou da respectiva progenia em que ocorreu uma quebra da hélice dupla.

21. Processo, de acordo com a reivindicação 20, caracterizado pelo facto de compreender ainda:

a introdução de um ácido nucleico exógeno na célula da planta, que compreende uma sequência homóloga a pelo menos uma porção do ADN-alvo, em que a introdução é feita em condições que permitem que ocorra a recombinação homóloga entre o ácido nucleico exógeno e a sequência-alvo do ADN na célula ou na sua progenia; e

a geração de uma planta a partir da célula ou da respectiva progenia, em que tenha ocorrido a recombinação homóloga.

22. Processo para a recombinação genética direccionada numa célula, caracterizada pelo facto de compreender:

i) a introdução na célula de um ácido nucleico que codifica uma endonuclease do efector TAL, de acordo com uma qualquer das reivindicações 13 a 17, em que a referida endonuclease do efector TAL está recombinada com uma sequência-alvo de ADN seleccionada;

ii) indução da expressão da endonuclease do efector TAL dentro da célula; e

iii) identificação de uma célula em que a sequência-alvo de ADN seleccionada exhibe uma mutação.

23. Processo, de acordo com a reivindicação 22, caracterizado pelo facto de a mutação se seleccionar no grupo que consiste em eliminação do material genético, inserção do material genético e tanto eliminação como inserção de material genético.

24. Processo, de acordo com a reivindicação 22, caracterizado pelo facto de compreender ainda a introdução de um ADN dador na célula.
25. Processo, de acordo com uma qualquer das reivindicações 22 a 24, caracterizado pelo facto de a célula ser uma célula de insecto, uma célula de planta, uma célula de peixe ou uma célula de mamífero.
26. Ácido nucleico que codifica uma endonuclease de efector TAL, caracterizado pelo facto de compreender (i) um domínio de endonuclease que pode clivar o ADN de hélice dupla, em que o domínio de endonuclease é de uma endonuclease de restrição do tipo II e (ii) um domínio do efector TAL compreendendo uma pluralidade de sequências de repetição do efector TAL que, em combinação, se ligam a uma sequência-alvo de ADN seleccionada.
27. Ácido nucleico, de acordo com a reivindicação 26, caracterizado pelo facto de o referido domínio de endonuclease ser de *FokI*.
28. Cassete de expressão, caracterizada pelo facto de compreender um promotor ligado operacionalmente ao ácido nucleico, de acordo com a reivindicação 26 ou com a reivindicação 27.
29. Célula hospedeira, caracterizada pelo facto de compreender a cassette de expressão, de acordo com a reivindicação 28.

Lisboa, 15 de Janeiro de 2015.

RESUMO

MODIFICAÇÃO DO ADN MODIFICADA PELO EFECTOR TAL

A presente invenção tem por objecto materiais e processos relacionados com o tratamento de genes (por exemplo, tratamento de genes com nucleases efectoras semelhantes a um activador de transcrição; "TALEN").

Figura 1



Figura 2

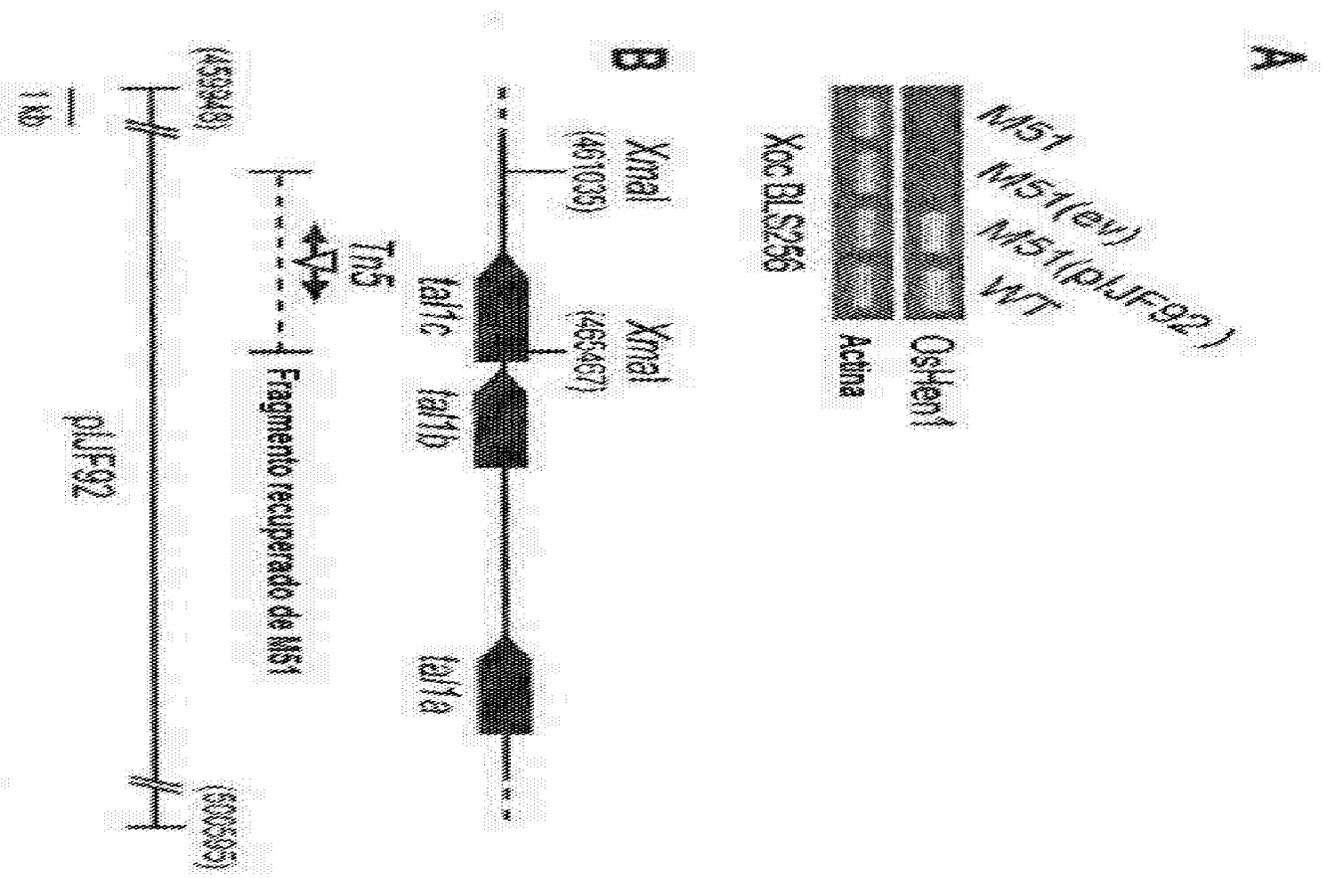


Figura 3

MDPRSRTPSPARELLPPOPDPVOFTADRGVSPFAGSPLDGLPARTMSRTRLPSPPAPSPAFSAOSSFDLLRQFPDPSLFTNTSL
FDLSLPPFGAHTTEAATGSEINDEIVCSGLRADDAPPPTMRYAVTARPPRAKPPARRRAAOPSDASPAQVOLTRTGVSDQDDEKI
KPKRVSTVADHHEALVGHCFTHAHVALSDHPALGTAVKYQDMAALPEATHEAIVGVKOWSGARALEALLTVAGELRGPPL
QLDTGOLLKAKRGGVTAVENVAHNRNALTGAPLNLTEQVVAJSHDGGKQALETVORLLPVLCOAHGILTPOQVVAJASNGGGK
QALETVORLLPVLCOAHGILTPOQVVAJASNGGGKQALETVORLLPVLCOAHGILTEQVVAJASNGGGKQALETVORLLPVLCOAH
GLTPEQVVAJASNGGGKQALETVOALLPVLCOAHGILTEQVVAJASNGGGKQALETVOALLPVLCOAHGILTEQVVAJASNGGGKQ
LETVOALLPVLCOAHGILTEQVVAJASHDGGKQALETVORLLPVLCOAHGILTEQVVAJASHDGGKQALETVORLLPVLCOAHGILT
POQVVAJASNGGGKQALETVORLLPVLCOAHGILTEQVVAJASNGGGKQALETVOALLPVLCOAHGILTEQVVAJASNGGGKQAL
ETVORLLPVLCOAHGILTEQVVAJASHDGGKQALETVORLLPVLCOAHGILTEQVVAJASHDGGKQALETVORLLPVLCOAHGILT
PEQVVAJASHDGGKQALETVORLLPVLCOAHGILTEQVVAJASNGGGKQALETVORLLPVLCOAHGILTEQVVAJASHDGGKQAL
ETVORLLPVLCOAHGILTEQVVAJASNGGGKQALESNVAQLSRPPALALTNQHLVALACLGGRPALDAVKKGLPHAPALKRTN
RRIPERTSHRVADHADVRLGFFQCHSPAQAFDQANTQFGMSRHGLOLFRVIGVTELEARSSTLPPASORWIDRILDASGM
KRAKPSPTSTOTPDQASLMAFADSLERQDLAPPNHEDQDTRASSRKRSDRAVTPSADQSEYRVPEORDAHLPLSWRV
KRPRTSUGGGLPDPGTPVANDLASSTIMREDEDEPFAGADDFPANEELANLWELLPQ (SEQ ID NO:12)

Figura 5

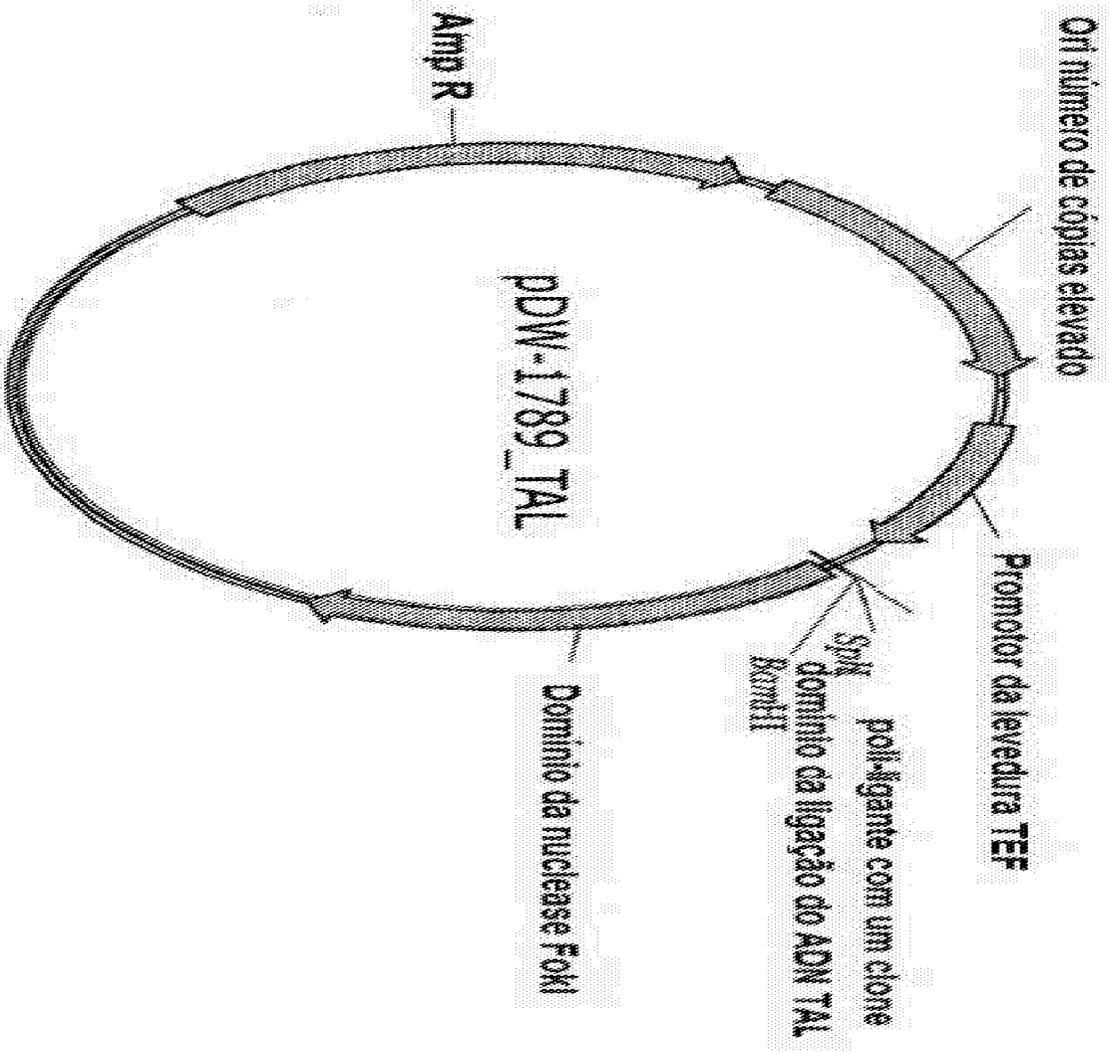


Figura 6

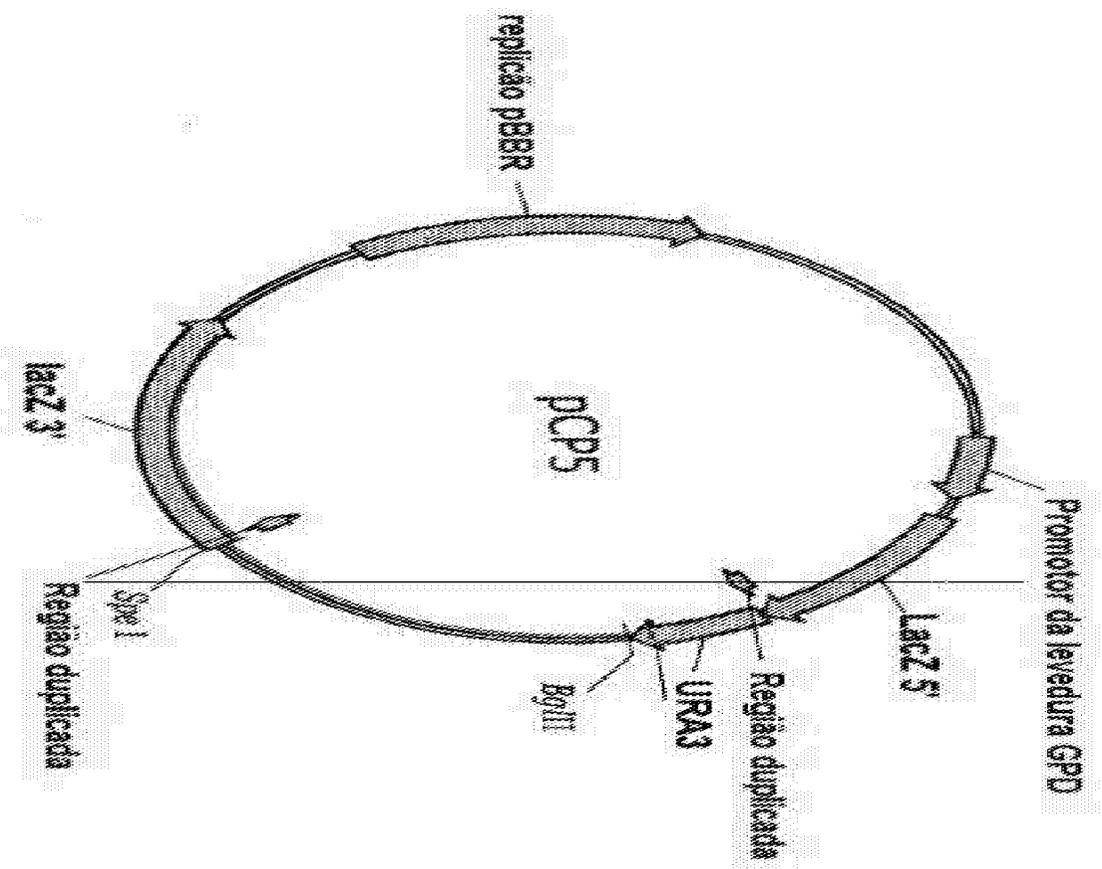


Figura 7

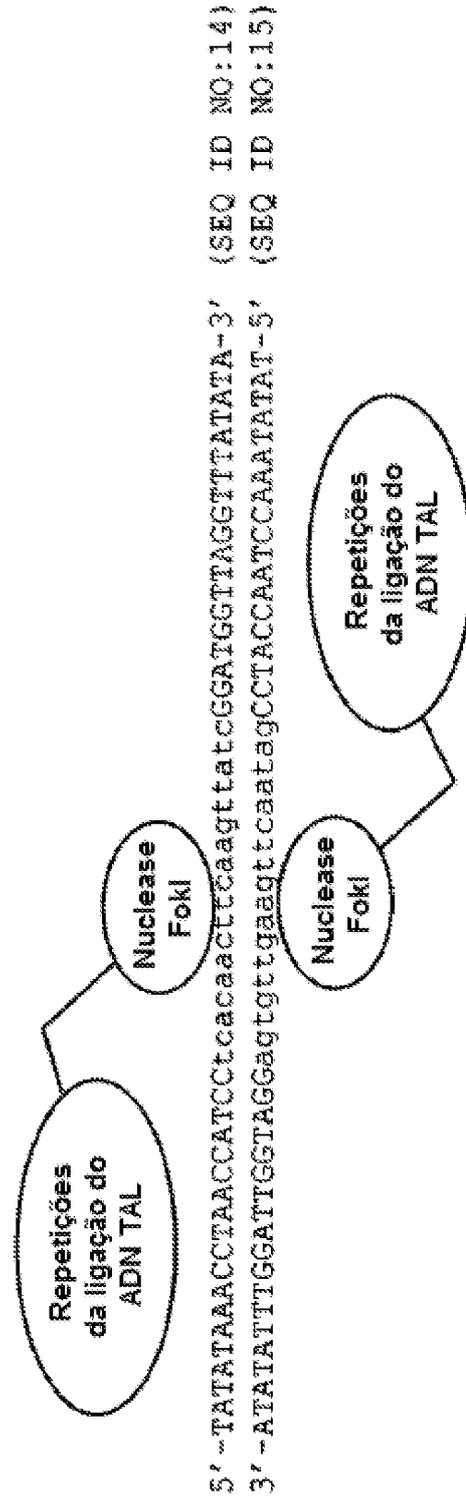


Figura 8

LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNGGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNGGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNI GGKQALETVQALLPVLCQAHG
LTPEQVVAIASNI GGKQALETVQALLPVLCQAHG
LTPEQVVAIASNI GGKQALETVQALLPVLCQAHG
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNGGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNSGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNGGGRPALETVQRLLPVLCQAHG
LTPEQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPEQVVAIASNGGGRPALE (SEQ ID NO:16)

Figura 9

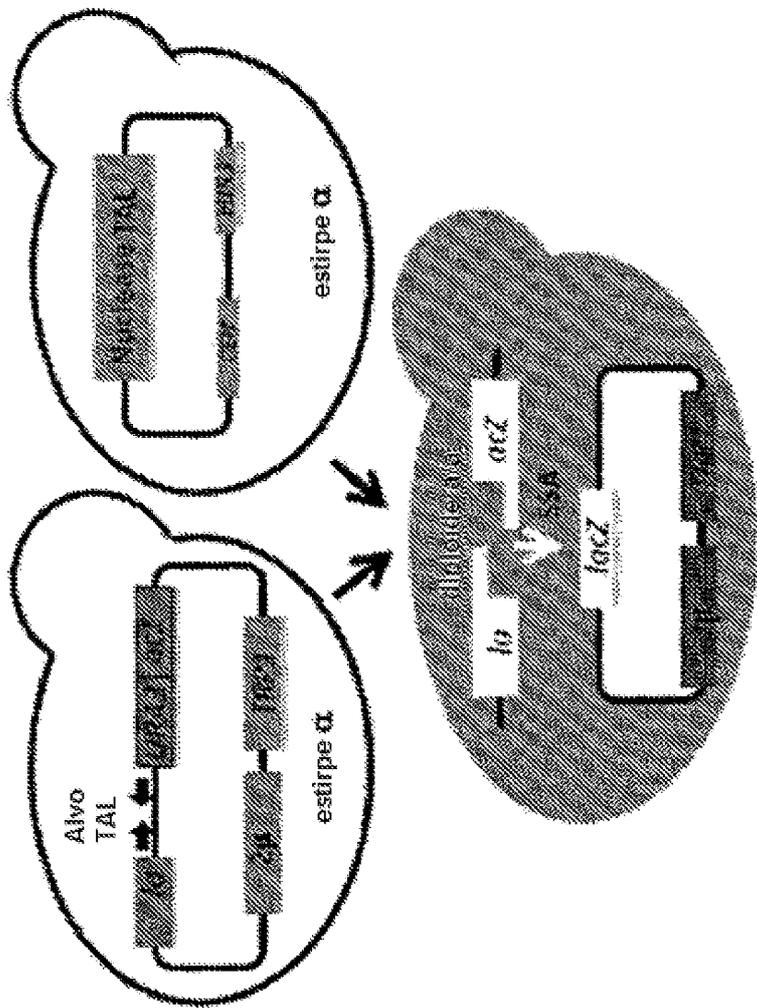


Figura 10

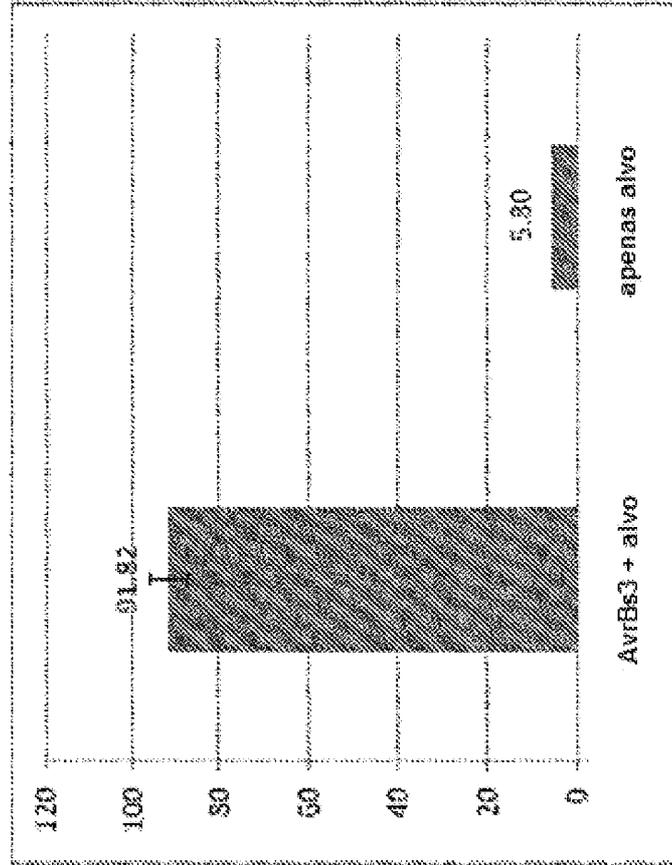


Figura 11

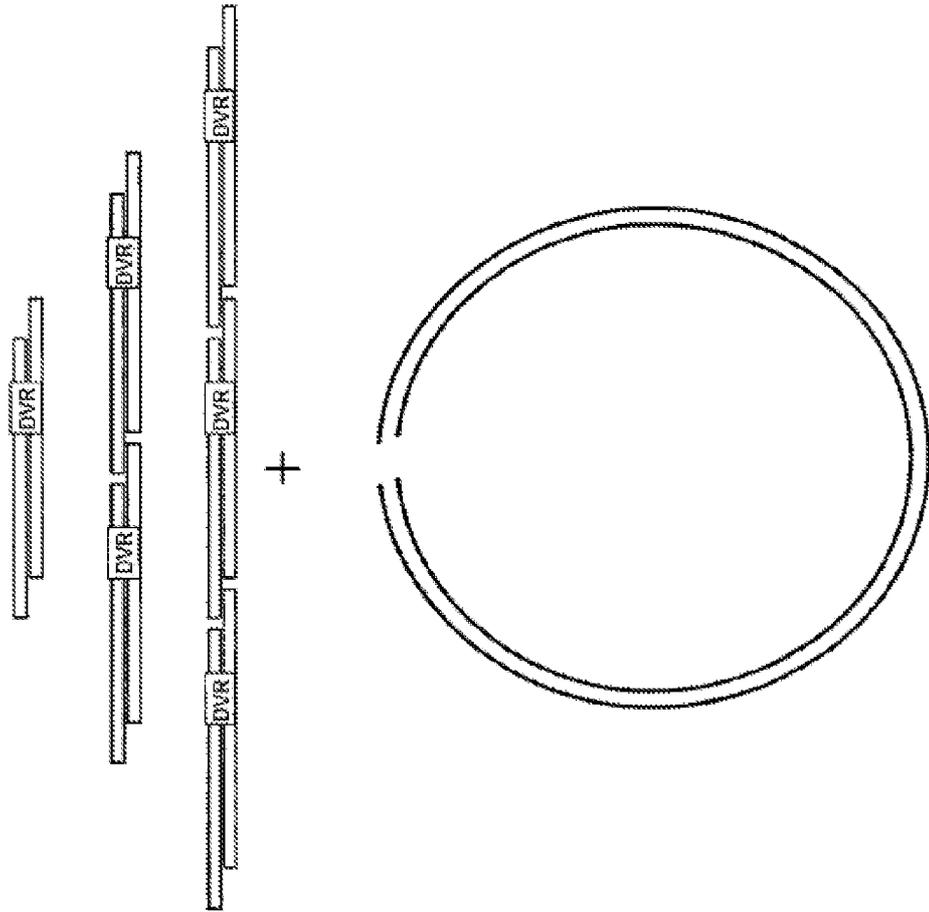


Figura 12

A

MscI

CTGACCCCGGGCAAGGGTTCGTGGCCATTCGGCAGCCTNNNNNSGCGAAGAGAGGAGCTGAGCAGCGTCCAGCCGCTCTTTGGCGGTTCGTGTCTGCGAAGGACCTATGAGC (SEQ ID NO:17)
L T P A Q V V A I A S X X G G K Q A L E T V Q R L L P V L C Q D H G (SEQ ID NO:18)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

B

MscI

CTGACCCCGGGCAAGGGTTCGTGGCCATTCGGCAGCCTNNNNNSGCGAAGAGAGGAGCTGAGCAGCGTCCAGCCGCTCTTTGGCGGTTCGTGTCTGCGAAGGACCTATGAGC (SEQ ID NO:19)
L T P A Q V V A I A S X X G G K Q A L E S (SEQ ID NO:20)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Figura 13

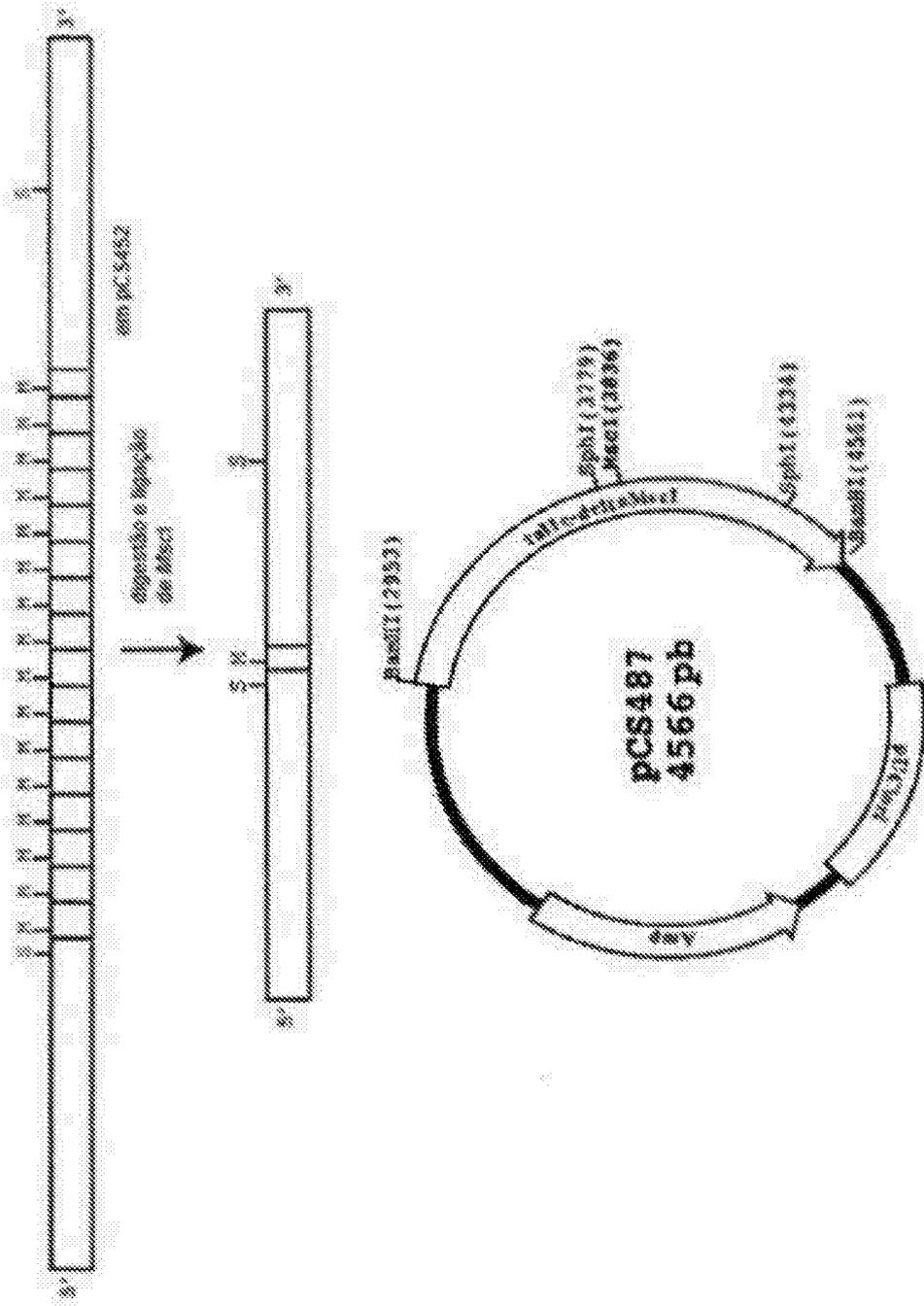


Figura 14

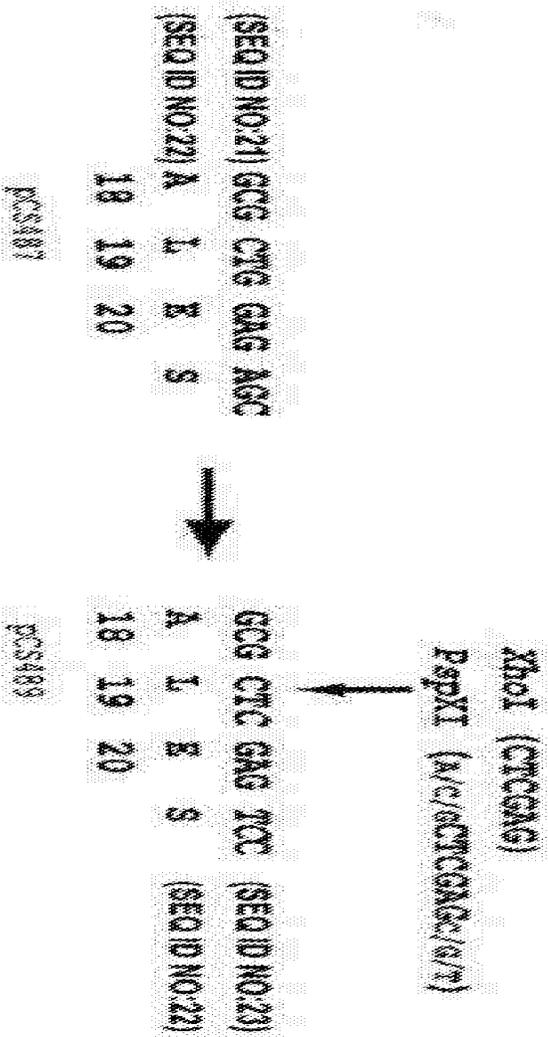


Figura 15

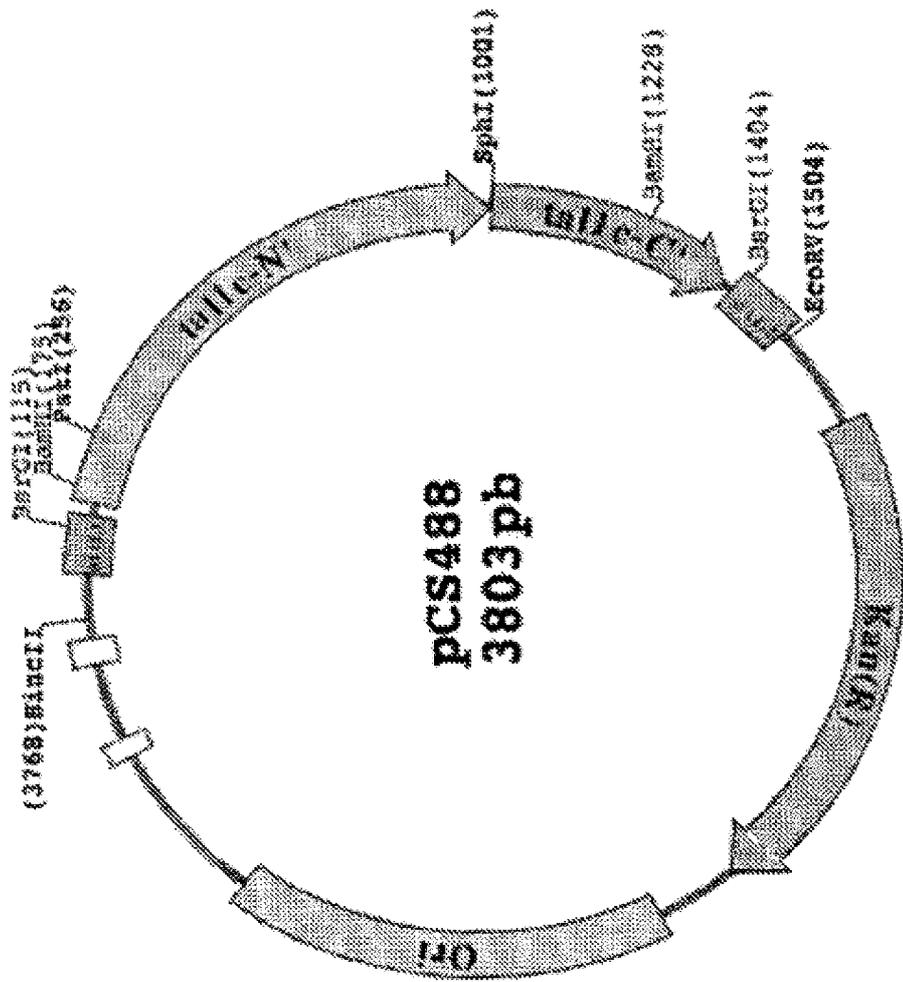
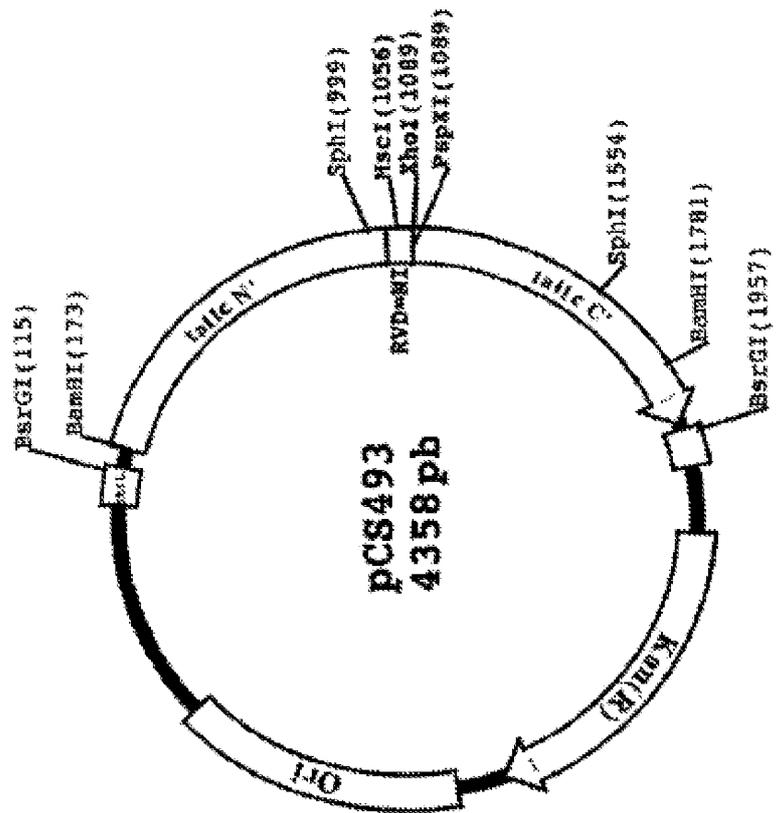


Figura 16



pCS493 DVR=NI
pCS494 DVR=HD
pCS495 DVR=NI
pCS496 DVR=NG

Figura 20

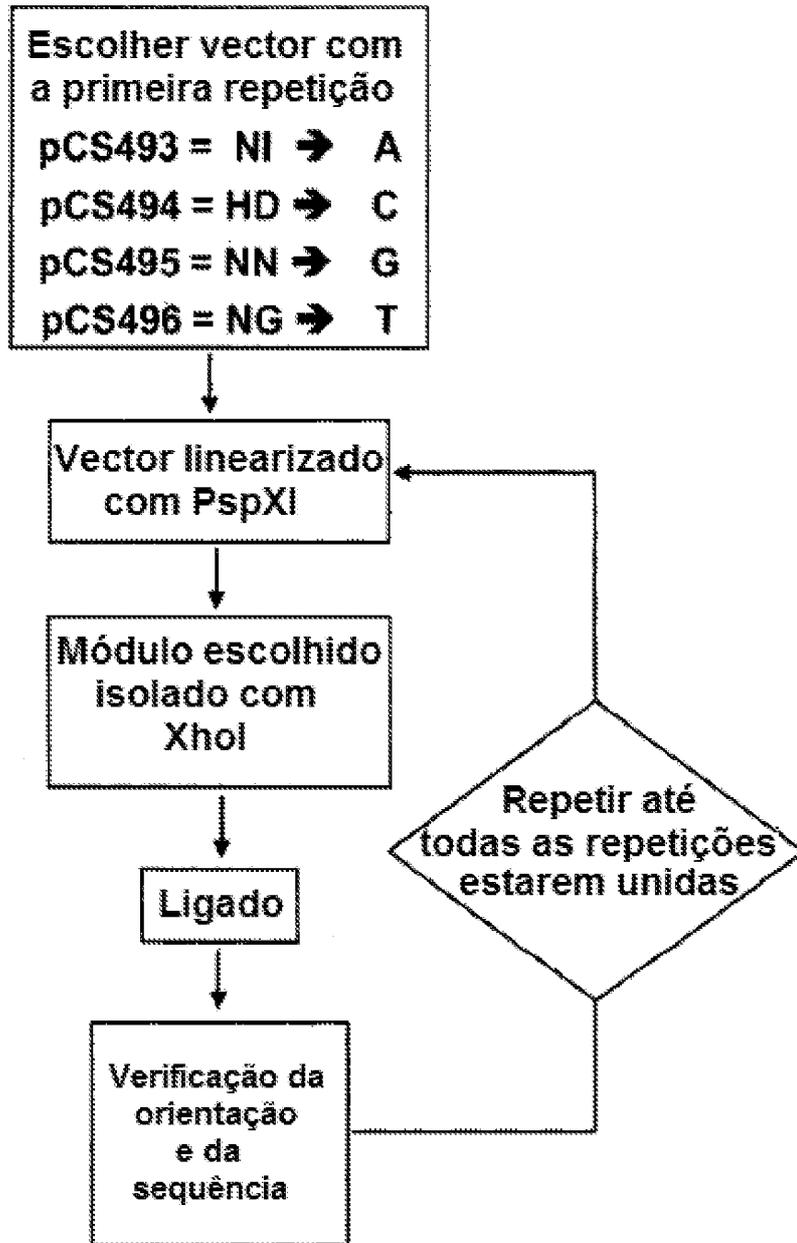


Figura 21A

ADM-ativo de Adh
 D/R de Adh

A T C A N G A T T C T C T T C A C T (REQ ID:MO:29)
 NI NG NO NI NI NN NI NG NG HD NG HD NG NG HD NI NC NG

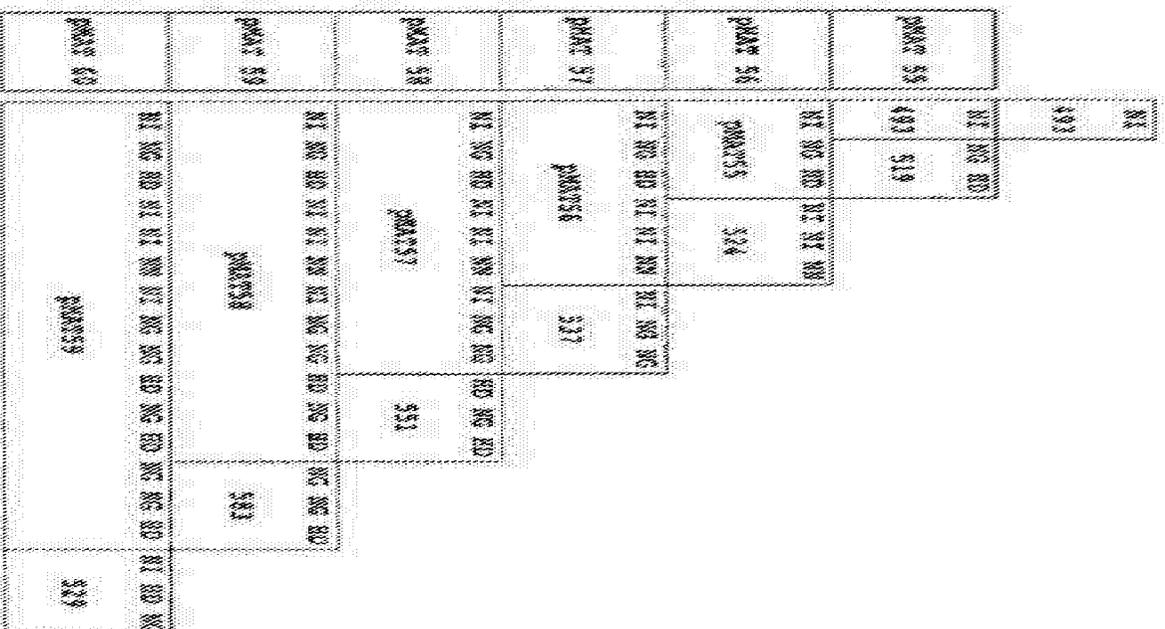


Figura 22

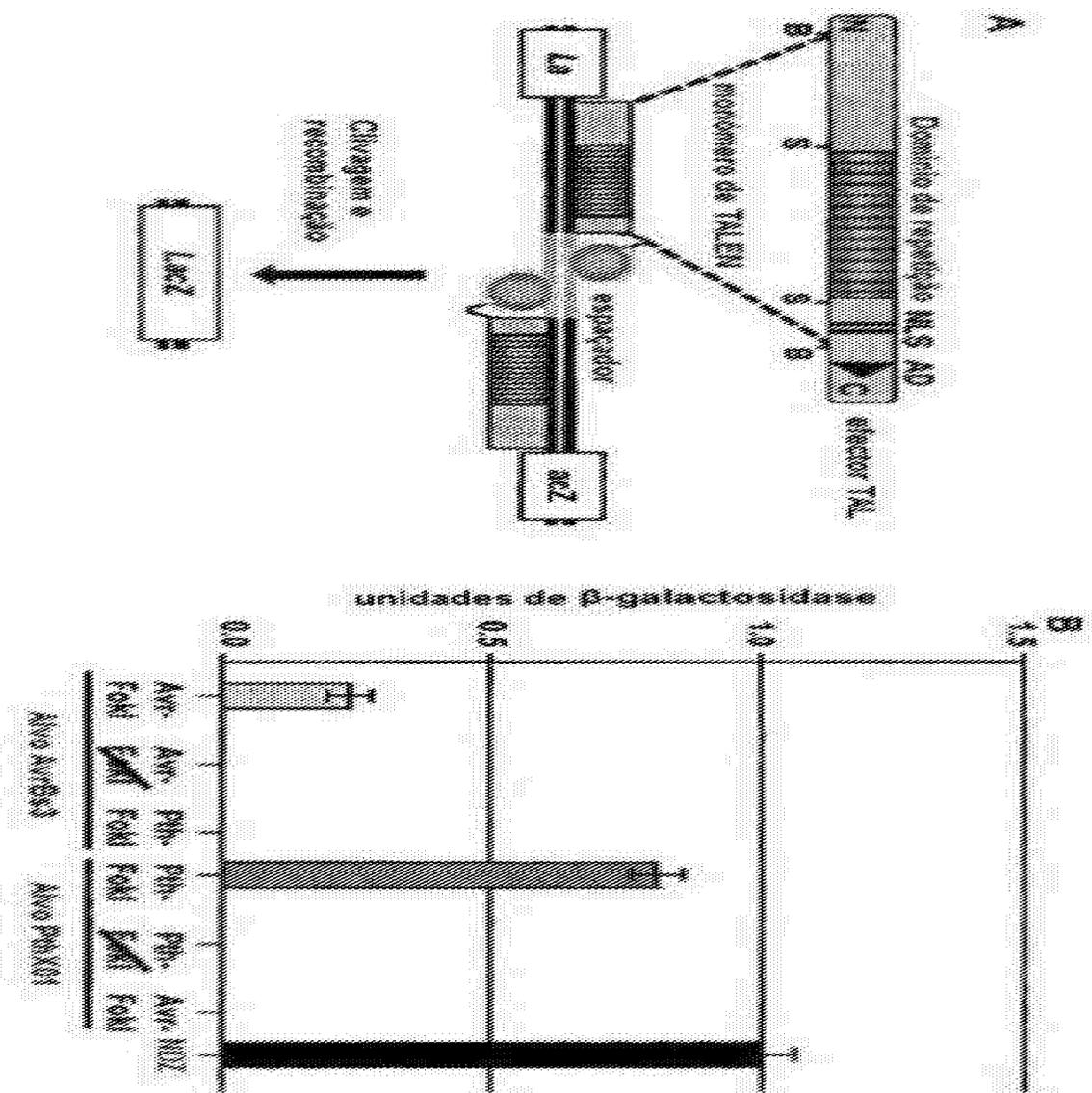


Figura 23

MDPNSRRTSPSARELLPPQPPDORNOPTADRGSGAPPAGGPELDSI.PARRTMSRTRLPQPPAPSPAFSAGSFSOLLROFDPS
LLDTSLLDSKPAVGTPTAAAPAEDEVOSGLRAADDPPTVRYAVTAAQPPRAKPAQRORAQPSQASPAQNDLRTL
GYSODDDEKPKVIGSTVACHHEALVGHFTAAHVALSRHPALGTVAVKYQDMAALFEATHEDVGVKONSSARA
LEALLTVAGELRGPELQDITGALVKAKKGSVTANEVHASRNALTGAPLNLTPAONVAASNGGKQALETVQRLLPILC
QAHGCTPAONVAASHDGKQALETMORLLPILCOAHGCTPDDVVAASNGGKQALETVORLLPILCOAHGCTPDDVVA
IASHGKQALETVORLLPILCOAHGCTPDDVVAASHDGKQALETVORLLPILCOAHGCTPDDVVAASNGGKQALE
TVORLLPILCOAHGCTPDDVVAASNGGKQALETVORLLPILCOAHGCTPDDVVAASHDGKQALETVORLLPILCOH
GLTPAONVAASHDGKQALETVQOLLPILCOAHGCTPDDVVAASNGGKQALATVORLLPILCOAHGCTPDDVVAASN
GGKQALETVORLLPILCOAHGCTPDDVVAASNGGKQALETVORLLPILCOAHGCTONVONVAASNGKQALETVOR
LLPILCOAHGCTPAONVAASHDGKQALETVORLLPILCOAHGCTPDDVVAASNGGKQALETVORLLPILCOAHGCT
QEDVVAASNGGKQALETVORLLPILCOAHGCTPDDVVAASNGGKQALETVORLLPILCOAHGCTPAONVAASNG
GKQALETVORLLPILCODHGCTLADVVAASNGKQALETVORLLPILCOAHGCTODDVAASNGKQALETVORLLP
VICOAHGCTPDDVVAASNGGKQALETVORLLPILCODHGCTLDVONVAASNGKQALETVORLLPILCODHGCTPDDV
VAASNGGKQALETVORLLPILCODHGCTPNDVVAASNGKQALESVIADLSRPPALAAITNDHVALAQLGORPAM
DAVKKGLPHAPELRRVNRPRIGERTSHRVADYAOVRYLEFEOCHSPAYAFDEANTOFQNSRNGLYOLFRRVGVTELE
ARGGTLPASORNDRLDASSKRAKPSPTSACTPDDASLHAFADSIERQLDAPSPNHEGDTIGASSKRSRSDRAV
GFSKHSFETRVPEQRDALLPLSNRWKRRFRTRIGGGLPJPGTPIAQLAASSTWANNEDAAFTKQADDPAFANEEL
ANNLWELLPOSSVIGSTI(SFC ID NO 31)

Figura 25

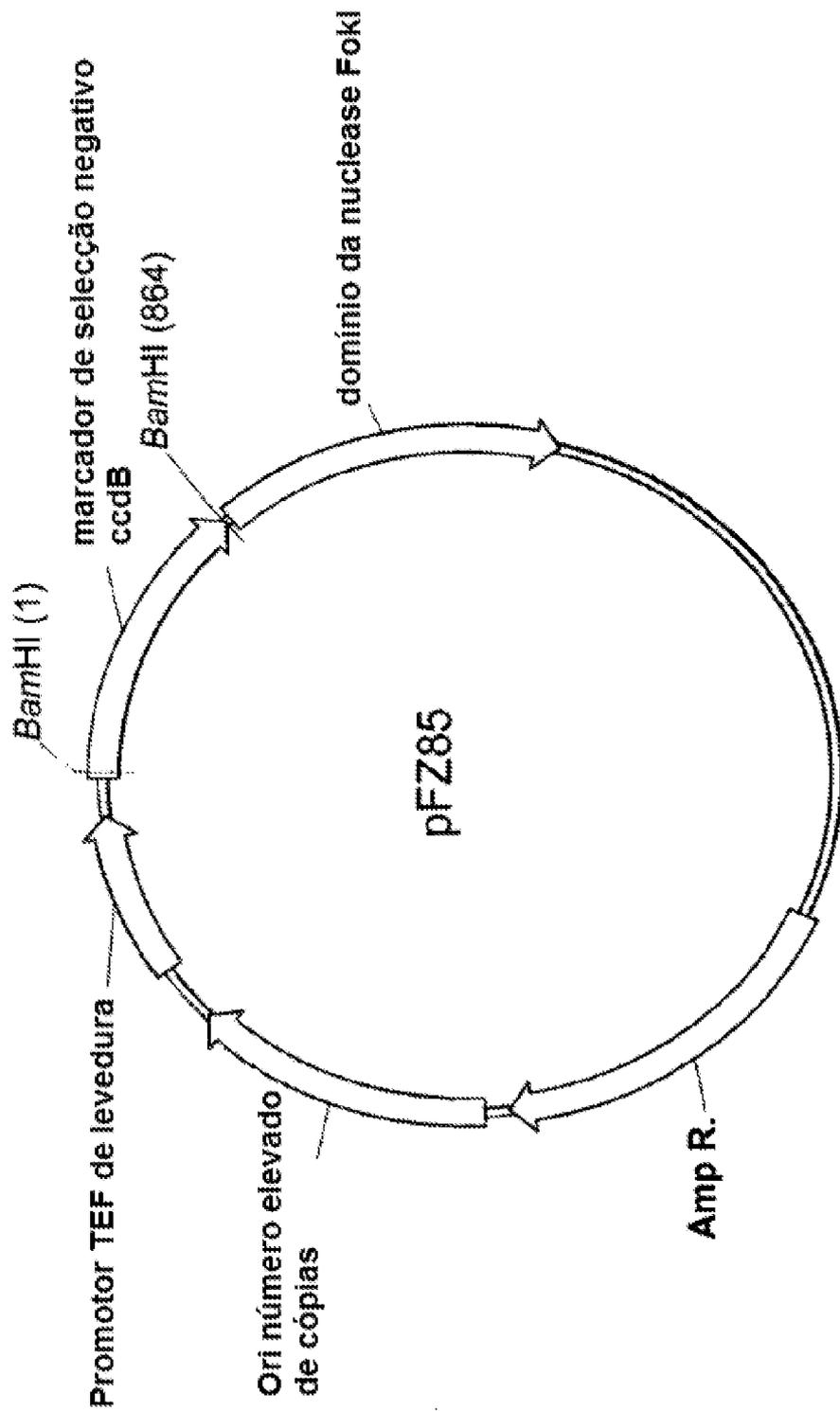


Figura 26

avrBs3_TALEN
MASSPPKKRKRKVSWKDASGWSRMHADPFRSRTPS
PARELLPGPQPDGVQPTADRGVSPFAGGPLDGLP
ARRTMSRTRLPSPPAPSPAPFAGSFSDLLRQDFP
SLFNLSLFDLSPFFGAWTEAATGEWDEVQSGLR
AADAPPPTMRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQOEKIKPKVVRSTVAQHH
EALVGHGFTHAHIVALSQHPAALGTVAVKYQDMI
AALPEATHEAIVGVGKQWNGARALEALLTVAGEL
RGPFLQDGTGQLLKIARCGVTAVEAVHAWRNAL
TGAPLNLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNIGGGKQALETVQALLPV
LCQAHGLTPQOVVAIASNIGGGKQALETVQALLPV
LCQAHGLTPQOVVAIASNIGGGKQALETVQALLPV
LCQAHGLTPQOVVAIASNIGGGKQALETVQALLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGKQALETVQALLPV
LCQAHGLTPQOVVAIASNNGGGKQALETVORLLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGRPALETVORLLPV
LCQAHGLTPQOVVAIASHDGGKQALETVORLLPV
LCQAHGLTPQOVVAIASNNGGGRPALESIVAQLSR
PDPALAAALNDHLVALACLGGRPALDAVKKGLPH
APALIKRTNRRIPERTSHRVADHAQVVRVLGFFQ
CHSHPAQAFDDAMTQFGMSRHGLLQLFRRVGVTE
LEARSGTLPPASQRWDRILOASGMKRAKPSPTST
QTPDQASLHAFADSLERDLDAPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQSTFEVVRVPEQRDALHLP
LSWRVVKRPRTSIGGGLPDPISRSQLVKSELEEK
SELRHKLKVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMORYVEENQTRN
KHINPNEWKVPSSVTEFKFLFVSGHFKGNYKA
QLTRLMHITNCGAVLSVEELIGGEMIKAGTLT
LEEVRRKFNNGEINF (SEQ ID NO:33)

Figura 27

pthXol_TALEN
MASSEPKKKRKVSWKDASGWSEPMHADPIRPRRPS
FARELLPGPQPDREVQPTADRGVSAFAGSELDGLP
ARRTVSRTRLPSFPAPSPAFSAGSFSDLLRPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPPTVVRVAVTAARPFRAKFAPRRRAAQPSD
ASFAAQVDLRLTGYSQQQQEKIKPKVRSSTVAQHH
EALVGHGFTHAHIVALSOHPAALGTVAVTYQHI I
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGVVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASNNGGKQALETVORLLPV
LCQAHGLTPAQVVAIASHDGGKQALETMORLLPV
LCQAHGLPPDQVVAIASNIGGKQALETVORLLPV
LCQAHGLTPDQVVAIASHGGGKQALETVORLLPV
LCQAHGLTPDQVVAIASHDGGKQALETVORLLPV
LCQAHGLTPDQVVAIASNNGGKQALETVORLLPV
LCQAHGLTPDQVVAIASNNGGKQALETVORLLPVL
CQAHGLTPDQVVAIASHDGGKQALETVORLLPVL
CQTHGLTPAQVVAIASHDGGKQALETVQQLLPVL
CQAHGLTPDQVVAIASNIGGKQALATVORLLPVL
CQAHGLTPDQVVAIASNNGGKQALETVORLLPVL
CQAHGLTPDQVVAIASNNGGKQALETVORLLPVL
CQAHGLTQVQVVAIASNIGGKQALETVORLLPVL
CQAHGLTPAQVVAIASHDGGKQALETVORLLPVL
CQAHGLTPDQVVAIASNNGGKQALETVORLLPVL
CQAHGLTQEQVVAIASNNGGKQALETVORLLPVL
CQAHGLTPDQVVAIASNNGGKQALETVORLLPVL
CQAHGLTPAQVVAIASNIGGKQALETVORLLPVL
CQDHGLTLAQVVAIASNIGGKQALETVORLLPVL
CQAHGLTQDQVVAIASNIGGKQALETVORLLPVL
CQDHGLTPDQVVAIASNIGGKQALETVORLLPVL
CQDHGLTLDQVVAIASNNGGKQALETVORLLPVL
CQDHGLTPDQVVAIASNNGGKQALETVORLLPVL
CQDHGLTPNQVVAIASNNGGKQALETVORLLPVL
ALAALTNLHVALACLGGRPAMDAVKKGLPHAPE
LIIRVNRRIIGERTSHRVADYAQVVRVLEFFOCHS
HPAYAFDEAMTQFGMSRNLVQLFRRVGVTELEA
RGGTLPPASQRWDRILOASGMKRAKPSPTSAQTP
DQASLHAFADSLERDLDAFSPMHEGDQTRASSRK
RSRSDRAVTGPSAQQAVEVRVPEQRDALHLFLSW
RVKRPRTRIWGGLPDPISRSQLVKSELEEKSEL
RHKLYVPHEYIELIETARNSTQDRILEMKVMEF
FMKVYGYRGKHLGGSRKPDGAIYTVGSPIDYGVI
VDTKAYSGGYNLPIGQADEMORYVEENQTRNKHI
NPNEWKVPSSVTEFKFLFVSGHFKGNKYKAQLT
RLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEE
VRRKFNNGEINF (SEQ ID NO:34)

Figura 28

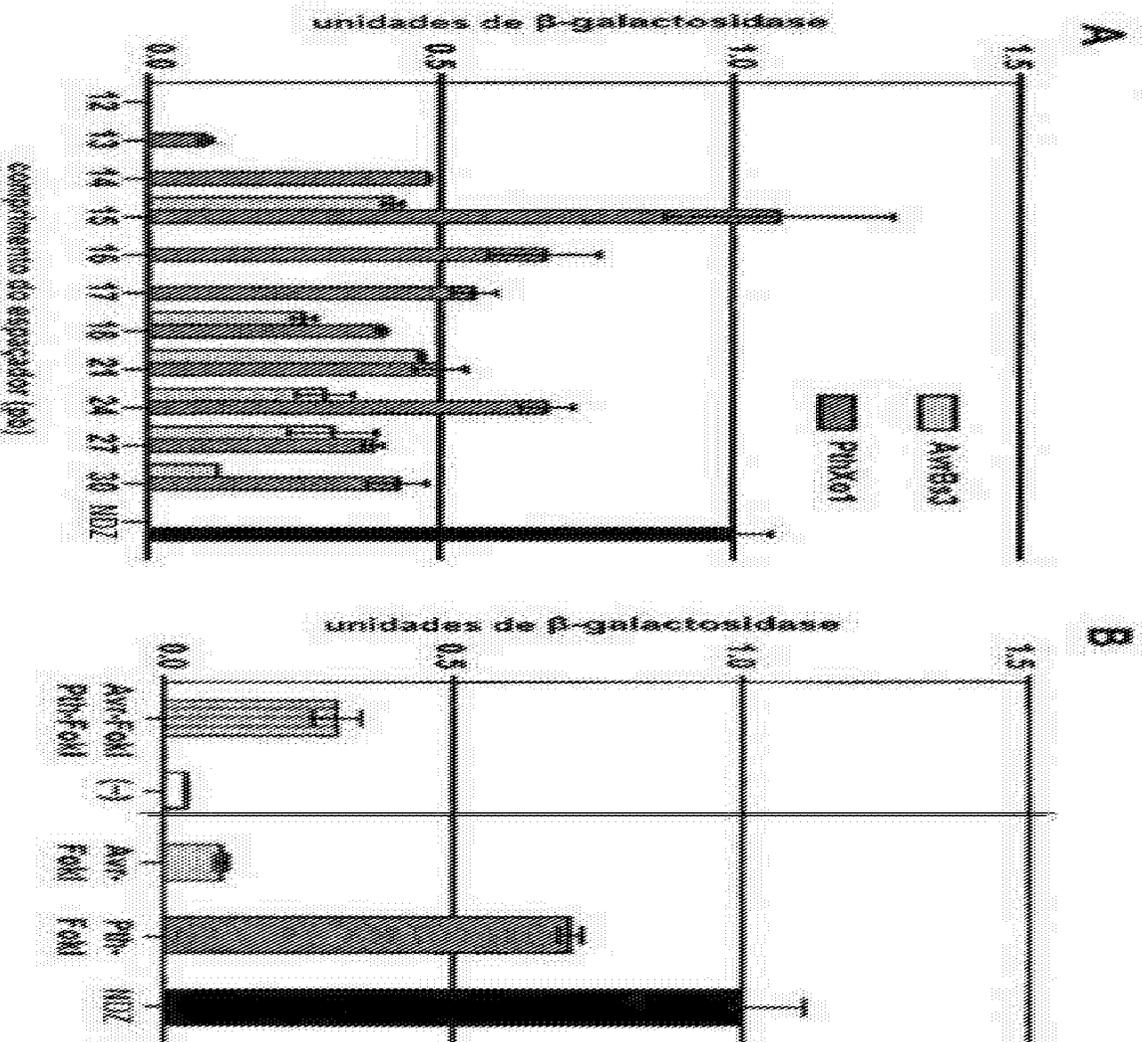


Figura 29A

TALEN customizada	1	2	3	4	5	6	7	8	9	10	11	12	13
ADH1-308-12	NI	NG	HD	NI	NI	NN	NI	NG	NG	HD	NG	HD	HD
	A	T	C	A	A	G	A	T	T	C	T	C	C
ADH1-408-12x	HD	HD	HD	NI	NN	NI	NI	NI	NN	NG	NI	NI	NI
	C	C	C	A	G	A	A	G	T	A	A	A	A
ADH1-928-12	HD	HD	NN	NN	NI	NG	NN	HD	NG	HD	HD	HD	NG
	C	C	G	G	A	T	G	C	T	C	C	C	T
ADH1-975-12x	NI	NN	NI	HD	NI	NI	NI	HD	HD	NI	HD	NI	NI
	A	G	A	C	A	A	A	C	C	A	C	A	A
gndlock-2356-13f	NI	HD	HD	HD	HD	HD	NG	HD	NG	HD	HD	NN	HD
	A	C	C	C	C	C	T	C	T	C	C	G	C

NI=A HD=C NN=G NG=T

Figura 29B

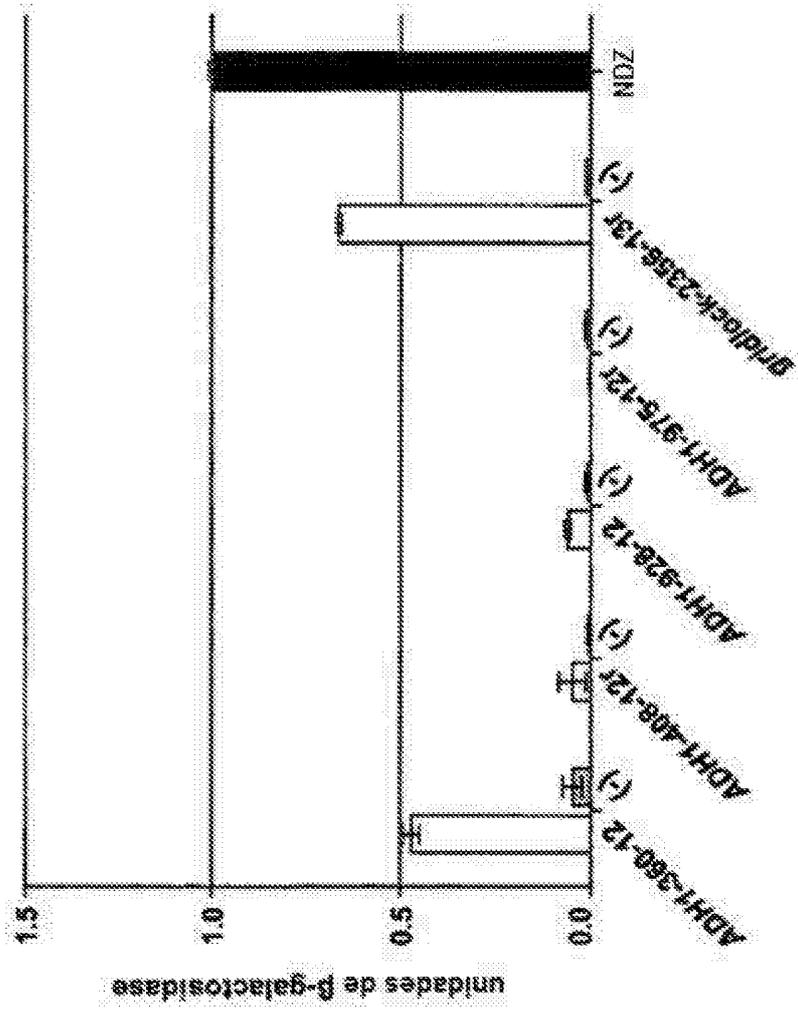


Figura 30

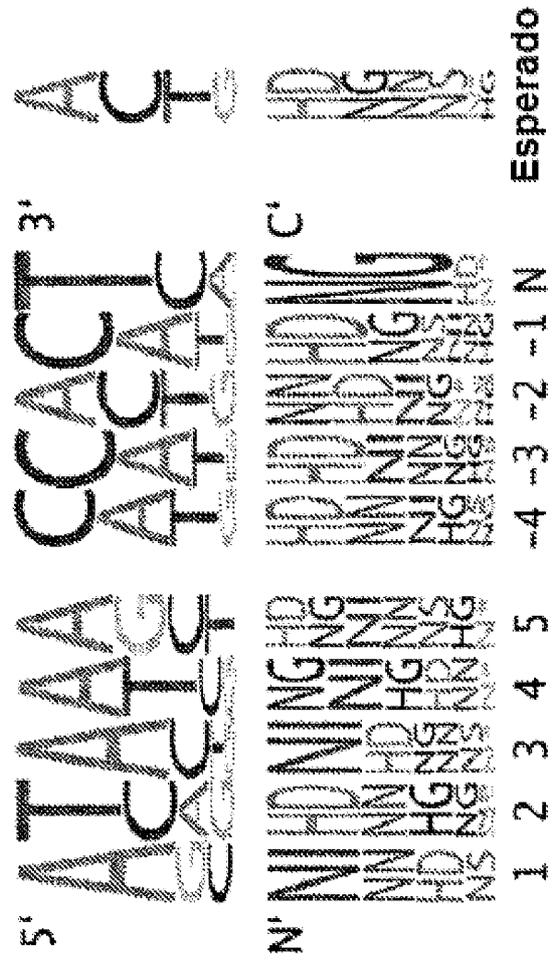
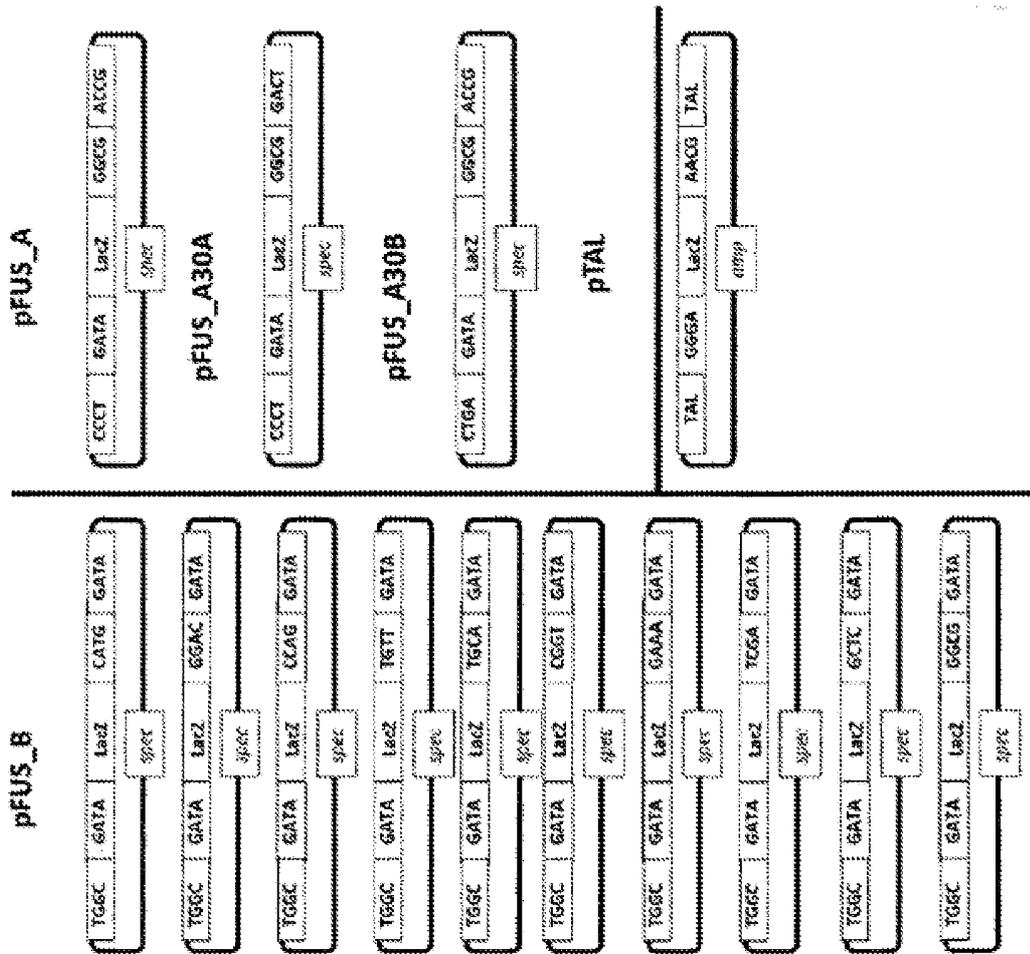


Figura 32B



STRUKTUR DAN TALEN: NI NN HO HO HO NI NI NG HO HO NI NI NG HO NI NI NG HO NI

Figura 33

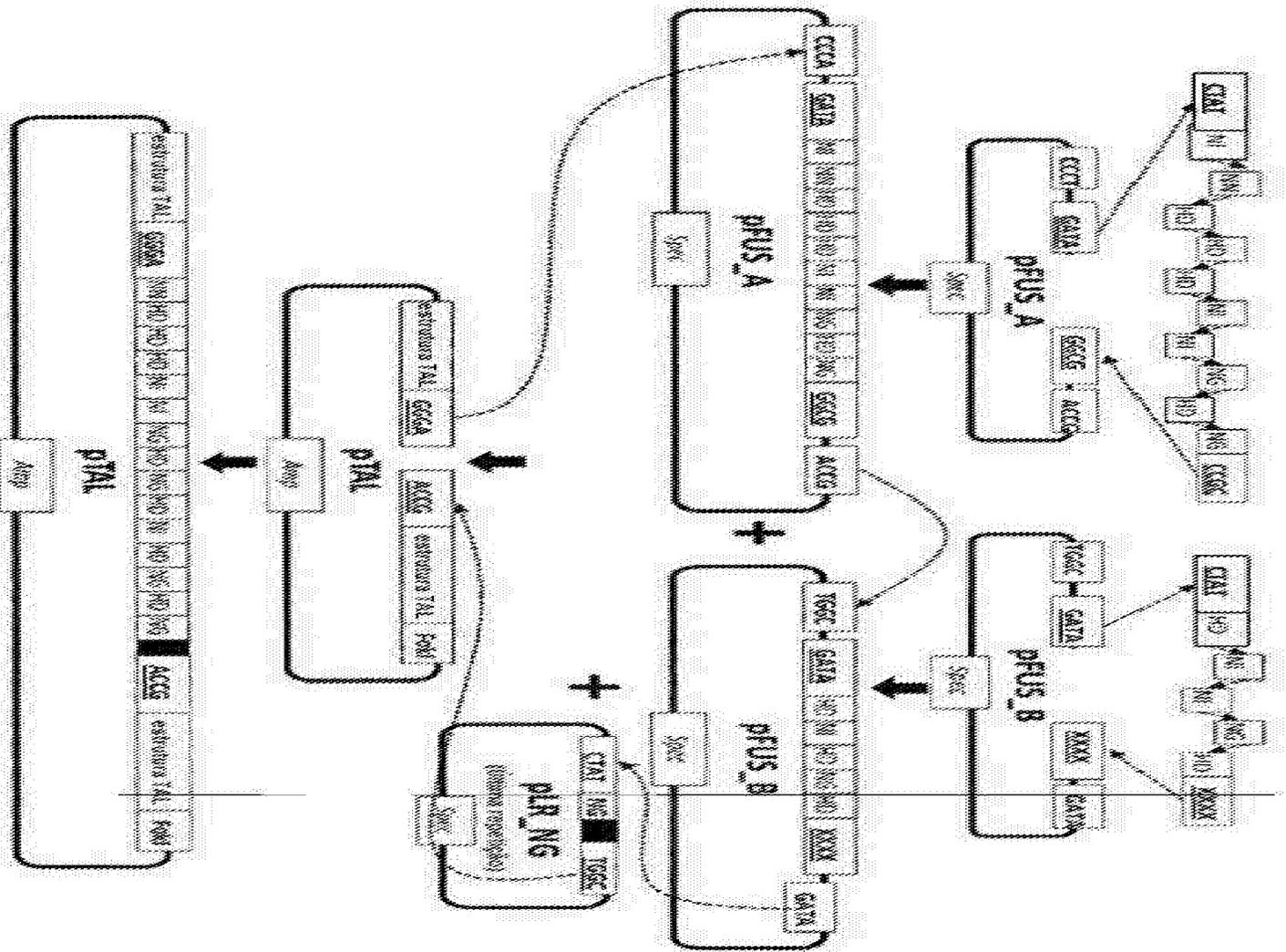


Figura 34A

telomerase-TALEN124_9DVRs
MASSPPKKKRKVSWKDASGWSRMHADPFRPRRPS
PARELLFGPQPDRVQPTADRGVSAAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPTDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPTVVRVAVTAARPPRAKPAFRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSSTVAQHH
EALVGHGFTNAHIVALSQHPAALGTVAVTYQHI I
TALPEATHEDIVGVGKOWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALESIVAQLSR
RDPALAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRILQASGMKRAKPSPTSA
QTPDQTSLHAFADSLERDLDAKSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPISRSQLVKSELEKK
SELRHKLKYVPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMORYVEENQTRN
KHINPNEWKVPSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLLIGGEMIKAGTLT
LEEVRKXFNNGEINF (SEQ ID NO:35)

Figura 34H

gridlock-TALEN106_13DVRs
MASSPPKKKKRVSWKDASGWSRMHADPIRPRRPS
PARELLFGPOPDRVQPTADRGVSAFAGSEPLDGLP
ARRTVSRTLPSFPAPSPAFSAGSFSDLLRPFDP
SLDTSLLDSMPAVGT PHTAAAPAENDEAQSALR
AADDPPFTVRVAVTAARPPRAKPAPRRRAAOPSD
ASPAAQVDLRTLGYEQQQQEKIKPKVRSSTVAQHH
EALVGHGFTHAHIVALSOHPAALGTVAVTYQHI I
TALPEATHEDI VGVGKQWSGARALEALLTDAGEL
RGPFLQLDTGQLVKI AKRGGVTAMEAVHASRNAL
TGAPLNLTFAQVVA IASNI GGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVA IASNGGGKQALETVQRLLPV
RDPALAAALNDHLVALACLGGRFALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRILOASGMKRAKPSPTSA
QTPDQTS LHAFADSLERDLDAFSPMHEGDQTRAS
SRKRSRSDRAVTGFSAQQAVEVRVFEQRDALHLP
LSWRVKRPRTRIWGGLPDFISRSQLVKSELEEKK
SELRHKLKYPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLP I GQADEMORYVEENQTN
KHINPNEWVKVYPSVTEFKFLFVSGHFKGNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:42)

Figura 34I

adh1-TALEN64_15DVRs
MASSEPKKKRKVSWKDASCWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAFAGSPLDGLF
ARRTVSRTRLPSPPAFSPAFSAGSFSDLLRPFDF
SLDTSLLDSMPAVGTPHTAAAPAENDEAQSALR
AADDPPFTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSSTVAQHH
EALVGHGFTHAHTVALSQHPAALGTVAVTYQHII
TALPEATHEDI VGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTFDQVVAIASNIGGKQALETVQRLLPV
RDPALAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRILOASGMKRAKPSPTSA
QTPDOTSLHAFADSLERDLDPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPISRSQLVKSELEEK
SELRHKLKYPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRN
KHINPNEWKVYPSSVTEFKFLFVSGHFKGNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:43)

Figura 34J

adh1-TALEN69_15DVRs
MASSPPKKRKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAFAGSFLDGLP
ARRTVSRTLPSPPAPSPAFSAGSFSDLLRPFDP
SLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDFPPTVRVAVTAARPPRAKPAFRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSSTVAQHH
EALVGHGFTHAHIVALSQHPAALGTVAVTYQHI I
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPFLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGGKQALETVQRLLPV
RDPALAAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVNRRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASORWDRILQASGMKRAKPSPTSA
QTPDQTSLHAFADSLERDLDAFSPMHEGDQTRAS
SRKRSRSDRAVTGFSAQQAVEVVRVFEQRDALHLP
LSWRVKRPRTRIWGGLPDPISRSQLVKSELEEK
SELRHKLKYPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSSGYNLPIGQADEMORYVEENQTRN
KHINPNEWKVPYSSVTEFKFLFVSGHFKNYKA
QLTRLNHITNONGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:44)

Figura 34L

tt4-TALEN90_15DVRs
MASSPPKKKRKXVSWKDASGWSPMHADPIRPRRPS
PARELLPGPQPDRVQPTADRGVSAFAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRPFDP
SLLDTSLLDSMPAVGTPHTAAAPAEWDEAQSALR
AADDPPPTVRVAVTAARFPRAKPAFRRAAQPSD
ASPAAQVDLRTLGYSSQQQOEKIKPKVRSSTVAQHH
EALVGHGFTTHAHIVALSQHPAALGTVAVTYQHII
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGVVTAMEAVHASRNAL
TGAPLNLTPAQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGKQALE SIVAQLSR
RDFALAALTNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRILQASGMKRAKPSPTSA
QTPDQTSLNHAFADSLERDLDAPSPMHEGDQTRAS
SRKRSRSDRAVTGPSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPI SRSQLVKSELEEK
SELRHKLKYPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSSGYNLPIGQADEMQRYVEENQTRN
KHINPNEWKVPSSVTEFKFLFVSGHFKGNYKA
QLTRLNHI TNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:46)

Figure 34M

telomerase-TALEN121_15DVRs
MASSPPKKRKRKVSWKDASGWSRMHADPIRPPRPS
FARELLPGPOFDRVQPTADRGVSAFAGSPLDGLP
ARRTVSRTRLPSFPAPSPAFSAGSFSDLLRPFDP
SLDTSLLDSMPAVGTFHTAAAPAENDEAQSALR
AADDPPPTVRVAVFAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSSTVAQHH
EALVGHGFTHANTVALSQHPAALGTVAVTYQHI I
TALPEATHEDIVGVGKQWSGARALEALLTDAGEL
RGPPLQLDTGQLVKIYAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIANNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIANNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIANNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNIGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIANNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNIGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVVAIASNNGGGKQALESTVAQLSR
RDPALAAALTNHDLVALACLGGRRPALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHPAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTPDQTSLHAFADSLERDLDAFSPMHEGDQTRAS
SRKRSRSRAVTPGSAQQAVEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPISRSOLVKSELEKK
SELRHKLKYPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMORYVEENQTRN
KHINPNEWNKVYPSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:47)

Figura 34N

telomerase-TALEN126_15DVRs
MASSPPKKKRKVSWKDASGWSRMHADPIRPRRPS
PARELLPGPQPDVQPTADRGVSAPAGSPLDGLP
ARRTVSRTRLPSPPAPSPAFSAGSFSDLLRFFDP
SLLDTSLLDSMPAVGTPHTAAAPAENDEAQSALR
AADDPPPTVRVAVTAARPPRAKPAPRRRAAQPSD
ASPAAQVDLRTLGYSSQQQEKIKPKVRSSTVAQHH
EALVGHGFTHAHIVALSOHPAALGTVAVTYQHI I
TALPEATHEDI VGVGKQWNGARALEALLTDAGEL
RGPPLQLDTGQLVKIAKRGGVTAMEAVHASRNAL
TGAPLNLTPAQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNIGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASHDGGKQALETVQRLLPV
LCQDHGLTPDQVVAIANNNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNNGGGKQALESIVAQLSR
RDPALAAALNDHLVALACLGGRPALDAVKKGLPH
APEFIRRVNRRIAERTSHRVADYAHVVRVLEFFQ
CHSHFAHAFDEAMTQFGMSRHGLVQLFRRVGVTE
FEARYGTLPPASQRWDRI LQASGMKRAKPSPTSA
QTFDQTS LHAFADSLERDL DAPS PMHEGDQTRAS
SRKRSRSDRAVTGPSAQQA VEVRVPEQRDALHLP
LSWRVKRPRTRIWGGLPDPISRSQLVKSELEEK
SELRNKLKYPHEYIELIEIARNSTQDRILEMKV
MEFFMKVYGYRGKHLGGSRKPDGAIYTVGSPIDY
GVIVDTKAYSGGYNLPIGQADEMQRVVEENQTRN
KHINPNEWVKVYPSVTEFKFLFVSGHFKNYKA
QLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLT
LEEVRKFNNGEINF (SEQ ID NO:48)

Figura 35

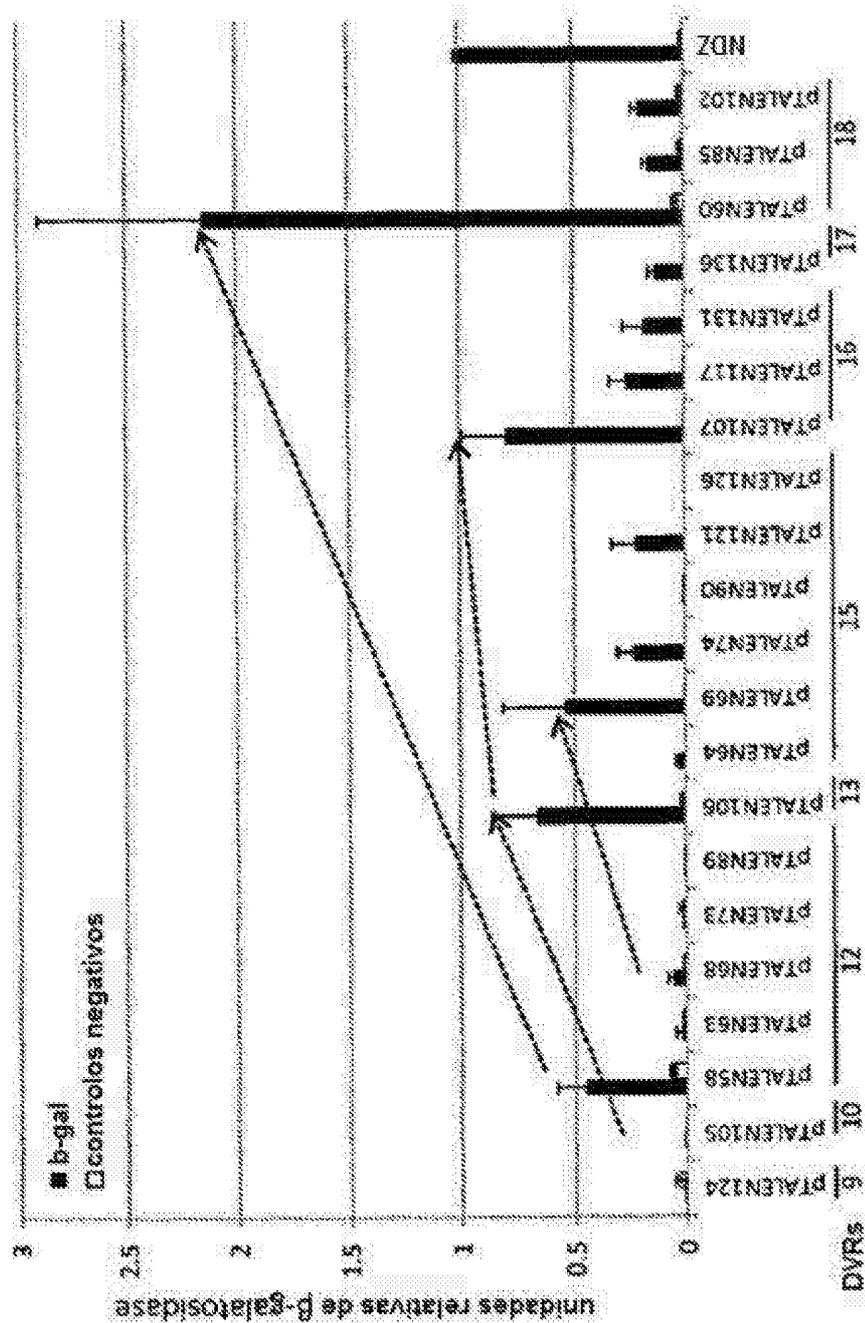


Figura 36

A

	<u>alvo de pTALEN 60</u>	<u>alvo de pTALEN 64</u>
alvo de Adh1	TATCAAGATTCTCTTCACCTTCTCTGTCACACCGATGTTACTTCTGGGA (SEQ ID NO:56)	
	<u>alvo de pTALEN 69</u>	<u>alvo de pTALEN 74</u>
alvo de Adh1	TCCGGATGCTCCTTGGACAAGGCTGTATTGTCAGTTGGTTGCTA (SEQ ID NO:57)	

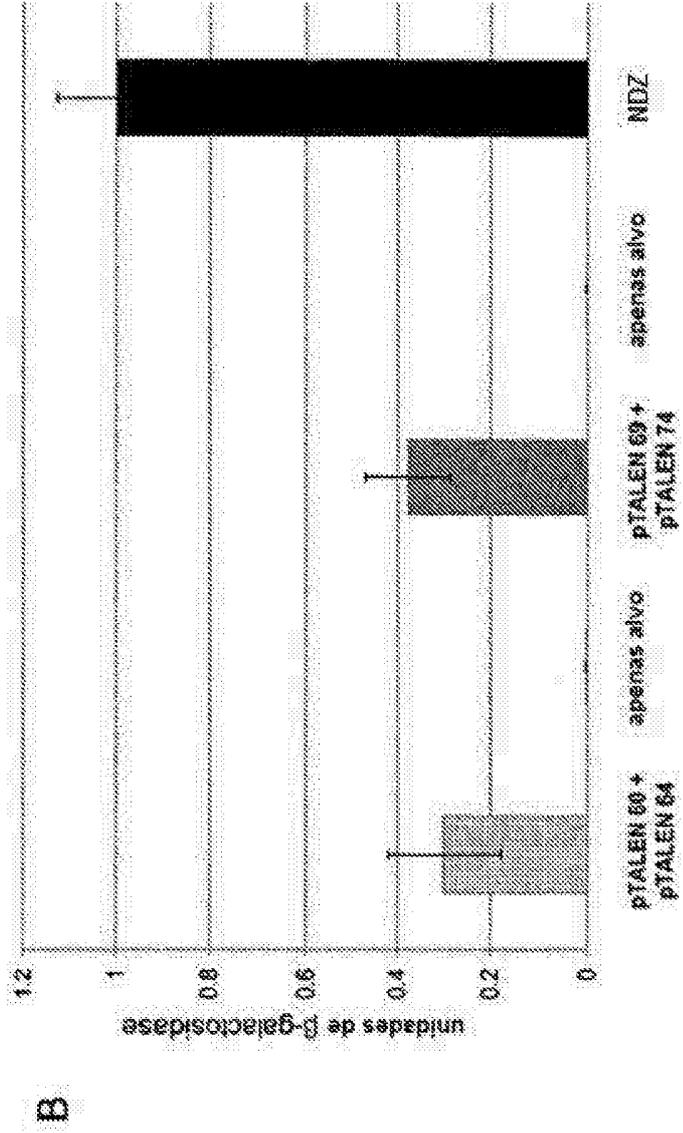
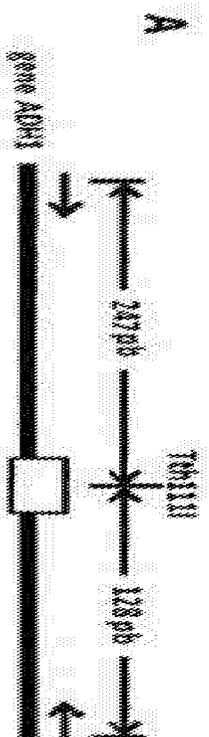


Figura 37



B

ADH1	1	10	20	30	40	48	
Neo modificado	CCGGATGCTCTCTTACAAAGCTGTGATTTGTCAGTTGTCGTTGCT						(SEQ ID NO:58)
	CCGGATGCTCTCTTACAA			TTGTCAGTTGTCGTTGCT			-8 (SEQ ID NO:59)
	CCGGATGCTCTCTTACAAAG			TATTTGTCAGTTGTCGTTGCT			-5 (SEQ ID NO:60)
	CCGGATGCTCTCTTACAA			TTGTCGTTGCT			-15 (SEQ ID NO:61)
	CCGGATGCTCTCTTACAAAG			ATTGTCAGTTGTCGTTGCT			-5 (SEQ ID NO:62)
	CCGGATGCTCTCTTACAAAGCTGTGATTTGTCAGTTGTCGTTGCT						0 (SEQ ID NO:58)
	CCGGATGCTCTCTTACAA			ATTGTCAGTTGTCGTTGCT			-7 (SEQ ID NO:63)
	CCGGATGCTCTCTTACAAAG			TATTTGTCAGTTGTCGTTGCT			-4 (SEQ ID NO:64)
	CCGGATGCTCTCTTACAAAGCTGTGATTTGTCAGTTGTCGTTGCT						0 (SEQ ID NO:58)
	CCGGATGCTCTCTTACAAAGCTGTGATTTGTCAGTTGTCGTTGCT						0 (SEQ ID NO:58)

Figura 38

A

AvrHah1	LDTGQLFKIAKRGVTA NEAVH AMRNALTGAPLN	(SEQ ID NO: 65)
AvrBs3	LDTGQLFKIAKRGVTNEAVH AMRNAL TGAPLN	(SEQ ID NO: 66)
PthXo1	LDTGQIN KIAKRGVTNEAVH ASRNALTGAPLN	(SEQ ID NO: 67)
PthA	LDTGQLFKIAKRGVTNEAVH AMRNAL TGAPLN	(SEQ ID NO: 68)
Tal1c	LDTGQIN KIAKRGVTNEAVH ASRNALTGAPLN	(SEQ ID NO: 69)

B

repetição de ordem 0	LDTGQLVKIA* <u>K</u> RGVTAMEAVHASRNALTGAPLN	(SEQ ID NO: 70)
1ª repetição	LTPAQVVAIASNNGKQALFTVQRLLFVLCQANG	(SEQ ID NO: 71)

HHHHHHHH HHHHHHHHHHHH
 | | | | | | | | | |

HHHHHHHHHHHH HHHHHHHHHHHH
 HEEEEEE HHHHHHHHHHHH

Figura 39

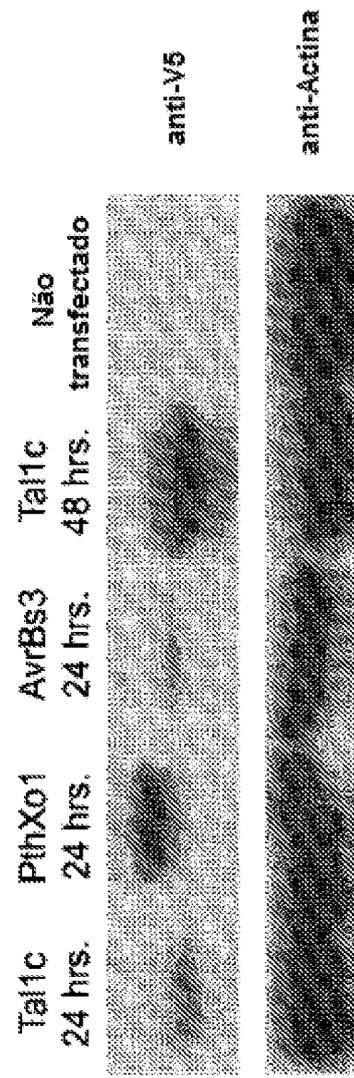


Figura 41A

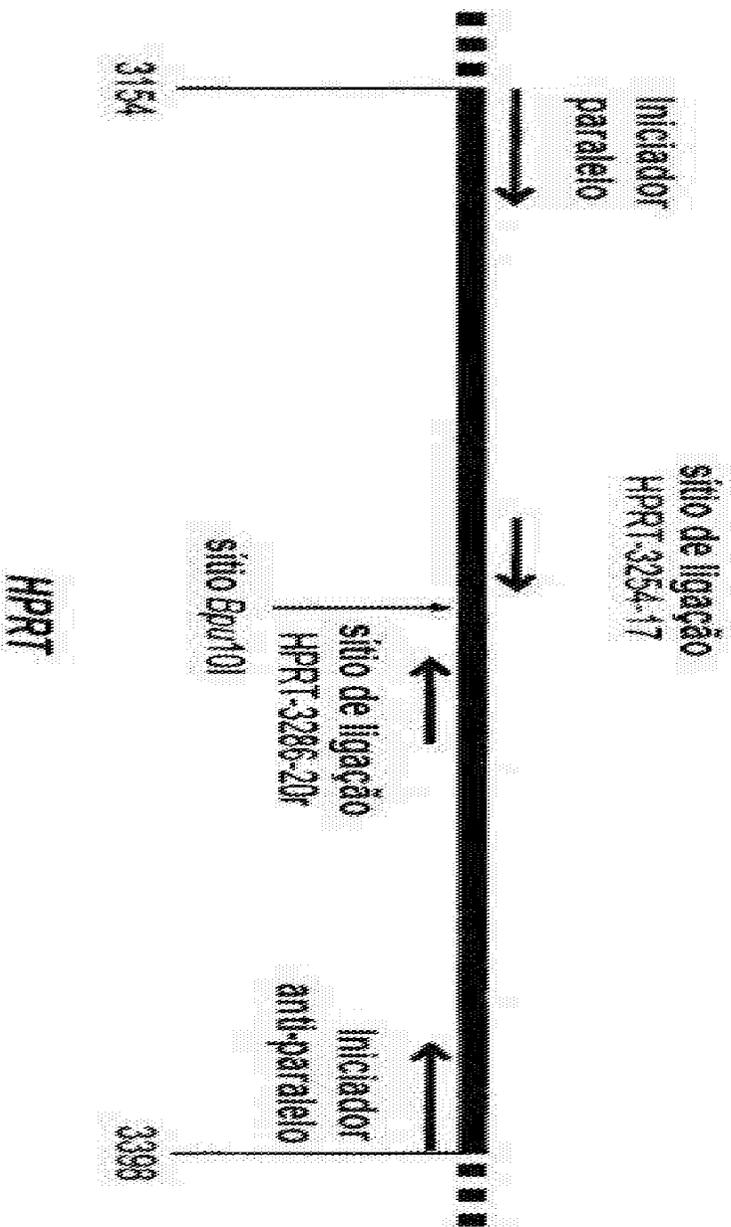


Figura 41B

