REPÚBLICA FEDERATIVA DO BRASIL
MINISTÉRIO DA INDÚSTRIA, COMÉRCIO EXTERIOR E SERVIÇOS
INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

CARTA PATENTE Nº PI 0212339-8

O INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL concede a presente PATENTE DE INVENÇÃO, que outorga ao seu titular a propriedade da invenção caracterizada neste título, em todo o território nacional, garantindo os direitos dela decorrentes, previstos na legislação em vigor.

(21) Número do Depósito: PI 0212339-8

(22) Data do Depósito: 05/09/2002

(43) Data da Publicação do Pedido: 13/03/2003

(51) Classificação Internacional: C12N 15/09; C12N 15/29; C12N 15/82; A01H 5/00; A01H 5/10

(30) Prioridade Unionista: US 60/316,975 de 05/09/2001

(54) Título: MÉTODO DE TRANSFORMAÇÃO DE UMA PLANTA, MOLÉCULA DE ÁCIDO NUCLEICO RECOMBINANTE, VETOR E CÉLULA

(73) Titular: MONSANTO TECHNOLOGY LLC, Sociedade Norte-Americana. Endereço: 800 North Lindbergh Boulevard, St. Louis, MO 63167, ESTADOS UNIDOS DA AMÉRICA(US)

(72) Inventor: QI WANG; PATRICE DUBOIS

Código de Controle: A03CCBF5C8B1C31A AB2BA7585D24DE8D

Prazo de Validade: 10 (dez) anos contados a partir de 11/09/2018, observadas as condições legais

Expedida em: 11/09/2018

Assinado digitalmente por:
Liane Elizabeth Caldeira Lage
Diretora de Patentes, Programas de Computador e Topografias de Circuitos Integrados

Relatório Descritivo da Patente de Invenção para **"MÉTODO DE TRANSFORMAÇÃO DE UMA PLANTA, MOLÉCULA DE ÁCIDO NUCLEICO RECOMBINANTE, VETOR E CÉLULA".**

A presente invenção refere-se ao campo da genética de plantas. Mais especificamente, a presente invenção refere-se à expressão genética específica para semente.

Sementes propiciam uma importante fonte de proteína da dieta para seres humanos e animais domésticos.

Contudo, o teor proteico das sementes é frequentemente incompleto. Por exemplo, muitas proteínas de sementes são deficientes em um ou mais aminoácidos essenciais. Esta deficiência pode ser superada por modificação genética das proteínas nativas ou não-nativas para terem uma composição nutricional mais completa de aminoácidos (ou algum outro aspecto desejável) e para superexpressar as proteínas modificadas nas plantas transgênicas. Alternativamente, um ou mais genes podem ser introduzidos em uma planta de cultivo para manipular seu trajeto metabólico e modificar o teor de aminoácido livre.

Essas abordagens são úteis na produção de colheitas apresentando características agronômicas importantes e propriedades nutricionais e farmacêuticas.

Apesar da disponibilidade de muitas ferramentas moleculares, a modificação genética das sementes é frequentemente restrita por um acúmulo insuficiente da proteína engenheirada. Muitos processos intracelulares podem causar impacto no acúmulo de proteína global, icluindo transcrição, tradução, conjunto da proteína e duplicação, metilação, fosforilação, transporte e proteólise. A intervenção em um ou mais desses processos pode aumentar a quantidade de proteína produzida em sementes geneticamente engenheiradas.

A introdução de um gene pode ocasionar efeito prejudicial no crescimento e desenvolvimento da planta. Sob tais circunstâncias, a expressão do gene pode precisar ser limitada ao tecido alvo desejado. Por exemplo, pode ser necessário expressar um gene de desregulação de aminoácido

em um modo específico para semente para evitar um fenótipo indesejado que pode afetar o rendimento ou outras características agronômicas.

A parte do promotor de um gene desempenha um papel fundamental no controle da expressão genética. Ao longo da região do promotor, o mecanismo de transcrição é montado, e inicia-se a transcrição. A iniciacão da transcrição no promotor pode ser regulada de vários modos. Por exemplo, um promotor pode ser induzido pela presença de um dado composto, expressar um gene apenas em um tecido específico, ou essencialmente, por toda a planta. Assim, a transcrição de uma seqüência de codificação pode ser modificada por ligação operativa da seqüência de codificação a promotores com diferentes características reguladoras.

Sumário da Invenção

A presente invenção inclui promotores capazes de gerar transcrição específica para semente e métodos de modificar, produzir e utilizar os mesmos. A invenção também providencia composições, células hospedeiras transformadas, plantas transgênicas e sementes contendo os promotores de alta expressão, e métodos para preparar e usar os mesmos.

A presente invenção inclui e providencia uma planta transformada contendo uma molécula de ácido nucléico compreendendo na direção 5' a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo que consiste em SEQ ID Nos: 11, 12, 13 e 14 e seus complementos e seqüências de ácido nucléico que hibridizam sob condições rigorosas para quaisquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos. A presente invenção inclui e providencia uma planta transformada contendo uma molécula de ácido nucléico compreendendo na direção 5'a 3': um promotor com uma seqüência de ácido nucléico, que apresenta uma identidade para uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14, e seus complementos em mais do que 90%, operavelmente ligada a uma seqüência de ácido nucléico estrutural onde o promotor é heterólogo com relação à seqüência de ácido nucléico estrutural.

A presente invenção inclui e providencia um método de transformar uma planta de soja compreendendo: providenciar uma molécula de

ácido nucléico compreendendo na direção 5' a 3': um promotor com uma seqüência de ácido nucléico selecionada grupo consistido de SEQ ID NOs: 11, 12, 13, e 14 e os complementos destas, operavelmente ligada a uma seqüência de ácido nucléico estrutural, e transformação da planta com a molécula nucléica.

A presente invenção inclui e providencia um método de transformação de uma planta de soja compreendendo: providenciar uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada dentre o grupo que consiste em SEQ ID NOs: 11, 12, 13 e 14, e seus complementos e seqüência de ácido nucléico que hibridiza sob condições rigorosas para quaisquer SEQ ID Nos: 11, 12, 13, e 14 e os complementos destas em mais do que 90%, operavelmente ligada a uma seqüência de ácido nucléico estrutural, onde o promotor é heterólogo com relação à seqüência de ácido nucléico estrutural.

A presente invenção providencia um método de expressão uma molécula de ácido nucléico estrutural em uma semente compreendendo: crescimento de uma planta transformada contendo uma molécula de ácido nucléico compreendendo na direção 5' a 3': um promotor com uma seqüência de ácido nucléico selecionado grupo que consiste em SEQ ID NOs: 11, 12, 13 e 14 e os complementos desta, e seqüências de ácido nucléico que hibridizam sob condições rigorosas para quaisquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operavelmente ligada à molécula de ácido nucléico estrutural, onde a planta transformada produz a semente e a molécula de ácido nucléico estrutural é transcrita na semente; e isolamento da semente.

A presente invenção providencia um método de obtenção de uma semente incrementada em um produto de um gene estrutural compreendendo: crescimento de uma planta transformada contendo uma molécula de ácido nucléico compreendendo na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos e seqüências de ácido nucléico que hibridizam sob condições rigorosas para quaisquer SEQ ID NOs: 11, 12,

13 e 14 e seus complementos, operavelmente ligadas à molécula de ácido nucléico estrutural, onde a planta transforma produz a semente e a molécula de ácido nucléico estrutural é transcrita na semente, e isolamento da semente da planta transformada.

A presente invenção providencia um método de obtenção de uma farinha grossa incrementada em um produto de um gene estrutural compreendendo: crescimento de uma planta transformada contendo uma molécula de ácido nucléico compreendendo na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos e seqüências de ácido nucléico que hibridizam sob condições rigorosas para quaisquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operavelmente ligadas à molécula de ácido nucléico estrutural, onde a planta transformada produz a semente e a molécula de ácido nucléico estrutural é transcrita na semente, e preparação da farinha grossa compreendendo tal planta transformada ou parte desta.

A presente invenção providencia um método de obtenção de um suprimento incrementado em um produto de um gene estrutural compreendendo: crescimento de uma planta transformada contendo uma molécula de ácido nucléico compreendendo na direção 5'a 3': um promotor com uma seqüência de ácido nucléico hibridiza sob condições rigorosas com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operavelmente ligadas à molécula de ácido nucléico estrutural, onde a planta transformada produz a semente e a molécula de ácido nucléico estrutural é transcrita na semente, e preparação da farinha grossa compreendendo tal planta transformada ou parte desta.

A presente invenção providencia um método de obtenção de um óleo incrementado em um produto de um gene estrutural compreendendo: crescimento de uma planta transformada contendo uma molécula de ácido nucléico compreendendo na direção 5'a 3': um promotor com uma seqüência de ácido nucléico que hibridiza sob condições rigorosas com uma seqüência

de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operavelmente ligadas à molécula de ácido nucléico estrutural, onde a planta transformada produz a semente e a molécula de ácido nucléico estrutural é transcrita na semente, e isolamento do óleo.

A presente invenção inclui e providencia uma molécula de ácido nucléico substancialmente purificada compreendendo seqüências de ácido nucléico selecionadas dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia um vetor compreendendo uma seqüência de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia uma célula contendo um vetor compreendendo uma seqüência de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia uma molécula de ácido nucléico capaz de hibridização específica sob condições rigorosas a uma molécula de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia um vetor compreendendo uma molécula de ácido nucléico capaz de hibridização específica sob condições rigorosas a uma molécula de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia uma célula contendo um vetor compreendendo uma molécula de ácido nucléico capaz de hibridização específica sob condições rigorosas a uma molécula de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia uma seqüência de ácido nucléico substancialmente purificada que apresenta uma identidade para uma seqüência de ácido nucléico selecionada do grupo que consiste em

SEQ ID NOs: 11, 12, 13 e 14 e seus complementos em mais do que 90%.

A presente invenção inclui e providencia um fragmento de ácido nucléico compreendendo pelo menos 20 nucleotídeos consecutivos de uma seqüência de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia um fragmento de ácido nucléico compreendendo pelo menos 30 nucleotídeos consecutivos de uma seqüência de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia um fragmento de ácido nucléico compreendendo pelo menos 50 nucleotídeos consecutivos de uma seqüência de ácido nucléico selecionada dentre o grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção providencia um óleo produzido de uma ou mais sementes de uma planta transformada contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, e seqüências de ácido nucléico que hibridizam sob condições rigorosas para qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia óleo produzido de uma ou mais sementes de uma planta transformada contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos em mais de 90% e seqüências de ácido nucléico que hibridizam sob condições rigorosas a qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operavelmente ligada a uma seqüência de ácido nucléico estrutural onde o promotor é heterólogo com relação à seqüência de ácido nucléico estrutural.

A presente invenção providencia uma semente gerada de uma planta transformada contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico

selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, e seqüências de ácido nucléico que hibridizam sob condições rigorosas para qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia uma semente gerada por uma planta transformada contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos em mais de 90% e seqüências de ácido nucléico que hibridizam sob condições rigorosas a qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operativamente ligadas a uma seqüência de ácido nucléico estrutural onde o promotor é heterólogo com relação à seqüência de ácido nucléico estrutural.

A presente invenção providencia um suplemento alimentar compreendendo uma planta transformada ou parte desta, contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, e seqüências de ácido nucléico que hibridizam sob condições rigorosas para qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia suplemento alimentar compreendendo uma planta transformada ou parte desta, contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos em mais de 90% e seqüências de ácido nucléico que hibridizam sob condições rigorosas a qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operativamente ligadas a uma seqüência de ácido nucléico estrutural onde o promotor é heterólogo com relação à seqüência de ácido nucléico estrutural.

A presente invenção inclui e providencia uma farinha grossa compreendendo material de planta de uma planta transformada contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um pro-

motor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, e seqüências de ácido nucléico que hibridizam sob condições rigorosas para qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos.

A presente invenção inclui e providencia uma farinha grossa compreendendo material de planta de uma transformada contendo uma molécula de ácido nucléico que compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do grupo consistindo em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos em mais de 90% e seqüências de ácido nucléico que hibridizam sob condições rigorosas a qualquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operativamente ligadas a uma seqüência de ácido nucléico estrutural onde o promotor é heterólogo com relação à seqüência de ácido nucléico estrutural.

A presente invenção inclui e providencia um reservatório de sementes, onde pelo menos 25% de tais sementes compreende na direção 5'a 3': um promotor com uma seqüência de ácido nucléico selecionada do gruo que consiste em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos em mais de 90% e seqüências de ácido nucléico que hibridizam sob condições rigorosas para quaisquer SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, operativamente ligadas a uma seqüência de ácido nucléico estrutural onde o promotor é heterólogo com relação à seqüência de ácido nucléico estrutural.

Breve Descrição das Figuras

A figura 1 apresenta um alinhamento da seqüência de duas seqüências de cDNA 7S alfa.

A figura 2 apresenta a seqüência do p-GEM-T 1T.

A figura 2a é um mapa de plasmídeo para o vetor pMON55548.

A figura 3 é um mapa de plasmídeo para o vetor pMON8677.

A figura 3a é um mapa de plasmídeo para o vetor pMON55546.

A figura 4 é um mapa de plasmídeo para o vetor pMON55547.

A figura 5 é um mapa de plasmídeo para o vetor pMON55554.

A figura 6 é um mapa de plasmídeo para o vetor pMON55542.

A figura 7 é uma comparação entre o promotor 7Sα e promotor tARC5.

A figura 8 é um mapa de plasmídeo para o vetor pMON55553.

A figura 9 é um mapa de plasmídeo para o vetor pMON55541.

Breve Descrição das Seqüências

SEQ ID NO: 1 é uma seqüência do promotor 7Sα de soja (*Glycine max* L.).

SEQ ID NO: 2 é uma seqüência do promotor 7Sα de soja (*Glycine max* L.).

SEQ ID NO: 3 é uma seqüência do promotor 7Sα de soja (*Glycine max* L.).

SEQ ID NO: 4 é uma seqüência do promotor 7Sα de soja (*Glycine max* L.).

SEQ ID NO: 5 é um iniciador adaptador (*Hind*III-*Pst*I) para ampliação do gene GlcA da soja.

SEQ ID NO: 6 é um iniciador primário 7Sα para ampliação do gene GlcA da soja.

SEQ ID NO: 7 é um iniciador 7Sα aninhado (*Bgl*II/*Nco*I) para ampliação do gene GlcA da soja.

SEQ ID NO: 8 é um iniciador 7Sα (*Nco*I/*Bgl*II) projetado para ampliação da região do iniciador do gene GlcA da soja.

SEQ ID NO: 9 é um iniciador adaptador do gene 7Sα.

SEQ ID NO: 10 é um iniciador específico para gene 7Sα.

SEQ ID NO: 11 é uma seqüência do promotor 7Sα para soja (*Glycine max* L.).

SEQ ID NO: 12 é uma seqüência do promotor 7Sα para soja (*Glycine max* L.).

SEQ ID NO: 13 é uma seqüência do promotor 7Sα para soja (*Glycine max* L.).

SEQ ID NO: 14 é uma seqüência do promotor 7Sα para soja (*Glycine max* L.).

Definições

As seguintes definições são providenciadas como um auxílio no entendimento da descrição detalhada da presente invenção.

As frases "seqüência de codificação", "seqüência estrutural" e "seqüência de ácido nucléico estrutural" referem-se a uma estrutura física compreendendo uma disposição ordenada de ácidos nucléicos. Os ácidos nucléicos são dispostos em uma série de tripletos de ácido nucléico que formam, cada um, um códon. Cada códon codifica para um aminoácido específico. Assim, a seqüência de codificação, seqüência estrutural e seqüência de ácido nucléico estrutural codificam uma série de aminoácidos formando uma proteína, polipeptídeo, ou seqüência peptídica. A seqüência de codificação, seqüência estrutural e seqüência de ácido nucléico estrutural pode estar contida dentro de uma molécula de ácido nucléico maior, vetor ou similares. Além disso, a disposição ordenada de ácidos nucléicos nessas seqüências pode ser ilustrada na forma de uma listagem de seqüência, figura, tabela, meio eletrônico ou similares.

As frases "seqüência de DNA", "seqüência de ácido nucléico", e "molécula de ácido nucléico" referem-se a uma estrutura física compreendendo uma disposição ordenada de ácidos nucléicos. A seqüência de DNA ou seqüência de ácido nucléico ode estar contida dentro de uma molécula de ácido nucléico maior, vetor ou similares. Além disso, a disposição ordenada de ácidos nucléicos nessas seqüências pode ser ilustrada na forma de uma listagem de seqüência figura, tabela, meio eletrônico ou similares.

O termo "expressão" refere-se à transcrição de um gene para produzir o correspondente mRNA e tradução deste mRNA para produzir o correspondente produto genético (ou seja, um peptídeo, polipeptídeo ou proteína).

O termo "expressão de RNA anti-sentido" refere-se à transcrição de um DNA para produzir uma primeira molécula de RNA capaz de hibridização para uma segunda molécula de RNA. A formação do híbrido RNA-RNA inibe a tradução da segunda molécula de RNA na produção de um produto genético.

"Homologia" refere-se ao nível de similaridade entre duas ou

mais seqüências de ácido nucléico ou aminoácido em termos de percentual de identidade de posição (ou seja, similaridade ou identidade de seqüência). Homologia também refere-se ao conceito de propriedades funcionais semelhantes dentre diferentes ácidos nucléicos ou proteínas.

A frase "heteróloga" refere-se à relação entre duas ou mais seqüências de ácido nucléico ou proteína que são derivadas de diferentes fontes. Por exemplo, um promotor é heterólogo com relação a uma seqüência de codificação, se tal combinação não é normalmente encontrada na natureza. Além disso, uma dada seqüência pode ser "heteróloga" com relação a uma célula ou organismos para o(a) qual ela é inserida (ou seja, não ocorre naturalmente naquela célula ou organismo particular).

"Hibridização" refere-se à capacidade de uma fita de ácido nucléico unir-se com uma fita complementar via pareamento de bases. A hibridização ocorre quando as seqüências de ácido nucléico complementares nas duas fitas de ácido nucléico contatam-se entre si sob condições adequadas.

A frase "operavelmente ligada" refere-se ao arranjo espacial funcional de duas ou mais regiões de ácido nucléico ou seqüências de ácido nucléico. Por exemplo, uma região de promotor pode ser posicionada em relação a uma seqüência de ácido nucléico tal que a transcrição de uma seqüência de ácido nucléico é direcionada pela região do promotor. Assim, uma região do promotor é "operavelmente ligada" à seqüência de ácido nucléico.

O termo "promotor" ou "região de promotor" refere-se a uma seqüência de ácido nucléico, normalmente encontrada a montante (5') a uma seqüência de codificação, que é capaz de direcionar a transcrição de uma seqüência de ácido nucléico ao mRNA. O promotor ou região do promotor providencia, tipicamente, um sítio de reconhecimento para RNA polimerase e os outros fatores necessários para iniciação adequada da transcrição. Como aqui examinado, um promotor ou região do promotor inclui variações de promotores derivados por inserção ou deleção de regiões reguladoras, submetendo o promotor à mutagênese aleatória ou direcionada ao sítio, etc.

A atividade ou força de um promotor pode ser medida em termos das quantidades de RNA que ele produz ou da quantidade de acúmulo de proteína em uma célula ou tecido relativo a um promotor cuja atividade transcricional foi previamente avaliada.

O termo "vetor recombinante" refere-se a qualquer agente como um plasmídeo, cosmídeo, vírus, seqüência de replicação autônoma, fago, ou seqüência de nucleotídeo de DNA ou RNA de fita simples linear, de fita simples circular, de fita dupla linear, ou de fita dupla circular. O vetor recombinante pode ser derivado de quaisquer fontes, ser capaz de integração genômica ou replicação autônoma.

"Seqüência reguladora" refere-se a uma seqüência de nucleotídeo localizada a montante (5'), dentro ou a jusante (3') para uma seqüência de codificação. A transcrição e expressão da seqüência de codificação é tipicamente impactada pela presença ou ausência da seqüência reguladora.

O termo "substancialmente homólogo" refere-se a duas seqüências que são pelo menos cerca de 90% idênticas em seqüência, como medido pelo programa BestFit aqui descrito (Versão 10; Genetics Computer Group, Inc. University of Wisconsin Biotechnology Center, Madison, WI), usando parâmetros padrão.

O termo "transformação" refere-se à introdução do ácido nucléico em um hospedeiro receptáculo. O termo "hospedeiro" refere-se a células de bactérias, fungos, animais ou células de animais, plantas ou sementes ou quaisquer partes de planta ou de tecidos incluindo células de planta, protoplastos, caule, raízes, tubérculos, sementes, hastes, folhas, plantas germinadas de semente, embriões, e pólen.

Como aqui empregado, o termo "planta transgênica" refere-se a uma planta onde um ácido nucléico introduzido é estavelmente introduzido ao genoma da planta, por exemplo, genomas nucleares ou de plastídeos.

Conforme aqui empregado, o termo "substancialmente purificado" refere-se a uma molécula separada de substancialmente todas as outras moléculas normalmente associadas com ela em sua forma nativa. Mais preferivelmente uma molécula substancialmente purificada e a espécie predo-

minante presente em uma preparação. Uma molécula substancialmente purificada pode ser maior do que cerca de 60% livre, de preferência cerca de 75% livre, mais preferivelmente cerca de 90% livre, e mais preferivelmente cerca de 95% livre de outras moléculas (com exclusão do solvente) presente na mistura natural. O termo "substancialmente purificado" não se destina a abranger moléculas presentes em seu estado nativo.

Descrição Detalhada das Modalidades Preferidas

A presente invenção providencia promotores capazes de transcrever uma seqüência de ácido nucléico estrutural heteróloga em uma semente, e métodos de modificar, produzir e utilizar a mesma. A invenção também providencia composições, células hospedeiras transformadas e plantas contendo promotores específicos para semente, e métodos para preparar e usar as mesmas.

Moléculas de Ácido nucléico

A presente invenção providencia moléculas de ácido nucléico que hibridizam para uma molécula de ácido nucléico com uma seqüência de ácido nucléico selecionada dentre o grupo de SEQ ID NOs: 11, 12, 13 e 14 e seus complementos. A presente invenção também inclui molécula de ácido nucléico que são fragmentos de tais moléculas.

A hibridização de ácido nucléico é uma técnica bem-conhecida aos versados na técnica da manipulação do DNA. A propriedade de hibridização de um dado par de ácidos nucléicos é uma indicação de sua similaridade ou identidade.

Condições de pouco rigor podem ser usadas para selecionar seqüências de ácido nucléico com identidades de seqüência menores para uma seqüência de ácido nucléico alvo. Pode-se querer empregar condições tais como cerca de 0,15M a cerca de 0,9 M de cloreto de sódio, a temperaturas que variam de cerca de 20°C a cerca de 55°C.

Condições de alta rigorosidade podem ser usadas para selecionar seqüências de ácido nucléico com maiores graus de identidade às seqüências de ácido nucléico apresentadas (Sambrook et al., 1989). (Obs: Citações completas para todas as referências são dadas abaixo)

As condições de alto rigor, tipicamente envolvem hibridização de ácido nucléico em cerca de 2X a cerca de 10X SSC (diluído de uma solução de estoque SSC 20X contendo cloreto de sódio 3M e citrato de sódio 0,3M, pH 7,0 em água destilada), cerca de 2,5X a cerca de 5X de solução de Denhardt (diluído de uma solução de estoque 50X contendo 1% (peso/volume) de albumina de soro bovino, 1% (peso/volume) de ficol, e 1% (peso/volume) de polivinilpirrolidona em água destilada), cerca de 10 mg/ml a cerca de 100 mg/ml de esperma de DNA de esperma de peixe, e cerca de 0,02% (peso/volume) a cerca de 0,1% (peso/volume) de SDS, com uma incubação a cerca de 50°C a cerca de 70°C por várias horas até toda a noite. As condições de alto rigor são, de preferência, providenciadas por SSC 6X, solução de Denhardt 5X, DNA de esperma de peixe 100 mg/ml e SDS 0,1% (peso/volume) com uma incubação a 55°C por várias horas.

A hibridização é geralmente seguida por várias etapas de lavagem. As composições de lavagem geralmente compreendem 0,5X a cerca de 10X de SSC, e 0,01% (peso/volume) a cerca de 0,5% (peso/volume) SDS com uma incubação de 15 minutos a cerca de 20°C a cerca de 70°C. De preferência os segmentos de ácido nucléico permanecem hibridizados após lavagem pelo menos uma vez em SSC a 0,1X a 65°C.

As moléculas de ácido nucleico de preferência hibridizam sob condições de baixa ou alta rigorosidade, com uma seqüência de ácido nucléico selecionada do grupo que consiste em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos. As moléculas de ácido nucléico mais preferivelmente, hibridizam sob condições de alta rigorosidade com uma seqüência de ácido nucléico selecionada do grupo que consiste em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos. Moléculas de ácido nucléico exemplares incluem promotores com seqüências de ácido nucléico de SEQ ID NOs: 11, 12, 13 e 14 e seus complementos ou quaisquer fragmentos destas.

Em uma modalidade alternativa, a molécula de ácido nucléico compreende uma seqüência de ácido nucléico com mais de 85% de identidade, e mais preferivelmente mais do que 86, cerca de 87, cerca de 88, cerca de 89, cerca de 90, cerca de 91, cerca de 92, cerca de 93, cerca de 94,

cerca de 95, cerca de 96, cerca de 97, cerca de 98, ou cerca de 99% de identidade para uma seqüência de ácido nucléico selecionada do grupo que consiste em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos e os fragmentos de quaisquer destas.

O percentual de identidade de seqüência é de preferência determinado usando o programa "BestFit" ou "Gap" de Sequence Analysis Software Package® (Versão 10, Genetics Computer Group, Inc., University of Wisconsin Biotechnology Center, Madison, WI). "Gap" utiliza o algoritmo de Needleman e Wunsch (Needleman and Wunsch, 1970) para encontrar o alinhamento de duas seqüências que maximizam o número de pares e minimiza o número de intervalos. "BestFit" executa um ótimo alinhamento do melhor segmento de similaridade entre duas seqüências e insere intervalos para maximizar o número de pares usando o algoritmo de homologia local de Smith e Waterman (Smith and Waterman 1981, Smith *et al.*, 1983). A identidade percentual é mais preferivelmente determinada usando o programa "Best Fit" usando parâmetros padrão.

A presente invenção também providencia fragmentos de molécula de ácido nucléico que hibridizam sob condições de baixa ou alta rigorosidade a moléculas de ácido nucléico com uma seqüência de ácido nucléico selecionada do grupo que consiste em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, fragmentos de moléculas de ácido nucléico e que apresentam mais do que cerca de 80, cerca de 85, cerca de 90, cerca de 95 ou cerca de 99% de identidade de seqüência com uma seqüência de ácido nucléico selecionada do grupo que consiste em SEQ ID NOs: 11, 12, 13 e 14 e seus complementos, ou fragmentos de quaisquer dessas moléculas.

Em uma modalidade, os fragmentos ficam entre cerca de 3000 e cerca de 1000 nucleotídeos consecutivos, cerca de 1800 e cerca de 150 nucleotídeos consecutivos, cerca de 1500 e cerca de 500 nucleotídeos consecutivos, cerca de 1300 e cerca de 250 nucleotídeos consecutivos, cerca de 1000 e cerca de 200 nucleotídeos consecutivos, cerca de 800 e cerca de 150 nucleotídeos consecutivos, cerca de 500 e cerca de 100 nucleotídeos consecutivos, cerca de 300 e cerca de 75 nucleotídeos consecutivos, cerca

de 100 e cerca de 50 nucleotídeos consecutivos, cerca de 50 e cerca de 25 nucleotídeos consecutivos ou cerca de 20 e cerca de 10 nucleotídeos consecutivos ao longo de uma molécula da presente invenção.

Em uma outra modalidade, o fragmento compreende pelo menos cerca de 20, cerca de 30, cerca de 40, cerca de 50, cerca de 60, cerca de 70, cerca de 80, cerca de 90, cerca de 100, cerca de 150, cerca de 200, cerca de 250, cerca de 500, ou cerca de 750 nucleotídeos consecutivos de uma seqüência de ácido nucléico da presente invenção.

Promotores

Em uma modalidade preferida quaisquer moléculas de ácido nucléico apresentadas podem ser promotores. em uma modalidade particularmente preferida, o promotor e um promotor 7Sα. em uma modalidade, o promotor é específico para tecido ou órgão, de preferência específico para semente. em uma modalidade particularmente preferida, o promotor de preferência expressa genes estruturais associados no endosperma ou embrião.

Em um aspecto, um promotor é considerado tecido ou órgão específico caso o nível de um mRNA naquele tecido ou órgão seja expresso a um nível que é pelo menos 10 vezes maior, de preferência pelo menos cerca de 100 vezes maior ou pelo menos cerca de 1.000 vezes maior do que um outro tecido ou órgão. O nível de mRNA pode ser medido, seja em um ponto de tempo único, ou em múltiplos pontos de tempo, e como tal o aumento de vezes pode ser aumento de vezes médio ou um valor extrapolado derivado de valores experimentalmente medidos. Sendo uma comparação de níveis, qualquer método que mede níveis de mRNA pode ser usado. Em um aspecto preferido, o tecido ou órgãos comparados são uma semente ou tecido de semente com uma folha ou tecido de folha. Em mais um aspecto preferido, tecidos ou órgãos múltiplos são comparados. Uma comparação múltipla preferida é uma semente ou tecido de semente comparado com dois, três, quatro ou mais tecidos ou órgãos selecionados dentre o grupo consistindo em tecido floral, ápice floral, pólen, folha, embrião, broto, primeira folha, ápice do broto, raiz, ponta de raiz, tecido vascular e cotilédone. Como aqui empregado, exemplos de órgãos de planta são sementes, folhas, raiz, etc., e

exemplo de tecidos são primodia de folha, ápice de broto, tecido vascular, etc.

A atividade ou resistência de um promotor pode ser medida em termos da quantidade de acúmulo de mRNA ou proteína que ele produz especificamente, em relação à quantidade total de mRNA ou proteína. O promotor, de preferência, expressa uma seqüência de ácido nucléico operavelmente ligada a um nível maior do que cerca de 2,5%, mais preferivelmente maior do que cerca de 5%, cerca de 6, cerca de 7, cerca de 8, ou cerca de 9%, mesmo mais preferivelmente, maior do que cerca de 10, cerca de 11, cerca de 12, cerca de 13, cerca de 14, cerca de 15, cerca de 16, cerca de 17, cerca de 18, ou cerca de 19% e mais preferivelmente maior do que cerca de 20% do mRNA total.

Alternativamente, a atividade ou resistência de um promotor pode ser expressa em relação a um promotor bem-caracterizado (para o qual a atividade transcricional foi previamente avaliada). Por exemplo, um promotor de interesse pode ser operavelmente ligado a uma seqüência repórter (por exemplo, GUS) e introduzido ao tipo de célula específico. Um promotor conhecido pode ser similarmente preparado e introduzido para o mesmo contexto celular. A atividade transcricional do promotor de interesse é a seguir determinada comparando-se a quantidade da expressão do repórter, em relação ao promotor conhecido. O contexto celular é, preferivelmente, soja.

Seqüências de Ácido Nucléico Estruturais

Os promotores da presente invenção podem ser operavelmente ligados a uma segunda seqüência de ácido nucléico que é heteróloga com relação à seqüência de ácido nucléico do promotor. A segunda seqüência de ácido nucléico pode geralmente ser qualquer seqüência de ácido nucléico para a qual se deseja um nível maior de transcrição. A segunda seqüência de ácido nucléico de preferência codifica um polipeptídeo adequado para incorporação à dieta de um ser humano ou animal ou providencia algum caráter agronomicamente importante.

Segundas seqüências de ácido nucléico adequadas incluem,

sem limitação as que codificam proteínas de armazenagem em sementes, enzimas de passo de ácido graxo, enzimas biossintéticas de tocoferol, enzimas biossintética de aminoácido e enzimas de ramificação de amido.

Proteína de armazenagem em semente preferidas incluem zeínas (Patentes U.S. 4.886.878, 4.885.357, 5.215.912, 5.589.616, 5.508.468, 5.939.599, 5.633.436 e 5.990.384, WO 90/01869, WO 91/13993, WO 92/14822, WO 93/08682, WO 94/20628, WO 97/28247, WO 98/26064 e WO 99/40209), proteínas 7S (Patentes U.S. 5.003.045 e 5.576.203) proteína da castanha do Pará (patente U.S. 5.850.024), proteínas livres de fenilalanina (WO 96/17064), albumina (WO 97/35023), β-conglicinina (WO 00/19839), 11S (Patente U.S. 6.107.051), alfa-hordotionina (Patentes U.S. 5.885.802 e 5.885.801) proteínas de armazenagem de semente ("arcelina") (Patente US. 5.270.200) lectinas (Patente U.S. 6.110.891) e glutenina (Patentes U.S. 5.990.389 e 5.914.450).

Enzimas de passo de ácido graxo preferidas incluem tioesterases (Patentes U.S. 5.512.482, 5.530.186, 5.945.585. 5.639.790, 5.807.893. 5.955.650, 5.955.329, 5.759.829, 5.147.792, 5.304.481, 5.298.421, 5.344.771 e 5.760.206) e desaturases (Patentes U.S. 5.689.050, 5.663.068, 5.614.393. 5.856.157, 6.117.677, 6.043.411, 6.194.167, 5.705.391, 5.663.068, 5.552.306, 6.075.183, 6.051.754, 5.689.050, 5.789.220, 5.057.419, 5.654.402, 5.659.645, 6.100.091, 5.760.206, 6.172.106, 5.952.544, 5.866.789, 5.443.974 e 5.093.249). Enzimas biossintéticas de tocoferol preferidas incluem *tyrA, slr173, ATPT2, dxs, dxr, GGPPS, HPPD GMT, MT1, AANT1, slr 1737* e um constructo anti-sentido para dioxigenase de ácido homogentísico (Kridl et al., *Seed Sci. Res.* 1:209:219 (1991), Keegstra, *Cell* 56(2):247-53 (1989), Nawrath, et al., *Proc. Natl. Acad. Sci U.S.A.* 91:12760-12764 (1994); Xia *et al., J. Gen. Microbiol.* 138:1309-1316 (1992), Cianobase http://www.kazusa.or.jp/cyanobase, Lois *et al., Proc. Natl. Acad. Sci.* U.S.A 95(5):2105-2110 (1998); Takahashi *et al. Proc. Natl Acad. Sci U.S.A.* 95(17), 9879-9884 (1998), Norris *et al., Plant Physiol.* 117:1317-1323 (1998), Bartley and Scolnik, *Plant Physiol.* 104:1469-1470 (1994), Smith *et al., Plant J.* 11:83-92 (1997), WO 00/32757, WO 00/10380, Saint Guily *et al.,*

Plant Physiol, 100(2):1069-1071 (1992), Sato *et al., J. DNA Res.*7(1):31-63 (2000)).

Enzimas biossintéticas de aminoácido preferidas incluem antranilato sintase (Patente U.S. 5.965.727 e WO 97/26366, WO 99/11800 e WO 99/49058) triptofano descarboxilase (WO 99/06581) e treonina descarboxilase (Patentes U.S. 5.534.421 e 5.942.660 e WO 95/19442), treonina desaminase (WO 99/02656 e WO 98/55601), diidrodipicolinato sintase (Patente U.S. 5.367.110), lisina cetoglutarato redutase (WO 98/42831) e aspartato quinase (Patentes U.S. 5.367.110, 5.858.749 e 6.040.160).

Enzimas de ramificação do amido preferidas incluem as apresentadas nas Patentes U.S. 6.232.122 e 6.147.279 e WO 97/22703.

Alternativamente, um promotor e segunda seqüência de ácido nucléico podem ser projetados para regular para menos uma específica seqüência de ácido nucléico. Isto é tipicamente realizado por ligação do promotor a uma segunda seqüência de ácido nucléico que é orientada na direção anti-sentido. Um versado na técnica está familiarizado com tal tecnologia anti-sentido. Qualquer seqüência de ácido nucléico pode ser regulada negativamente deste modo.

O versado na técnica também reconhecerá que um promotor da presente invenção também pode ser projetado para regular para menos uma dada seqüência de ácido nucléico usando o fenômeno da co-supressão. A co-supressão de um gene endógeno usando tanto seqüências de cDNA de comprimento total, como parcial foi descrita (Napoli et al., *The Plant Cell* 2:279-289 (1990)).

Adicionalmente, a expressão de uma dada seqüência de ácido nucléico também pode ser regulada para menos por regiões de não-codificação de tal seqüência de ácido nucléico. O versado na técnica pode rapidamente isolar seqüências contendo DNA genômico que flanqueia a região de codificação de tal seqüência de ácido nucléico, ou íntrons de tal região de codificação, e o uso de regiões de não-codificação para inibição anti-sentido ou co-supressão.

Alvos de tal regulação podem incluir polipeptídeos que têm um

baixo teor de aminoácidos essenciais, ainda que sejam expressos a um nível relativamente alto em um dado tecido. Por exemplo, β-conglicinina e glicinina são expressos abundantemente em sementes, mas são deficientes no sentido nutricional com relação a aminoácidos essenciais. Esta abordagem antisentido pode ser usada também para remover eficazmente outras proteínas indesejáveis, tais como antinutrientes (por exemplo, lectinas), albumina, e alergênicos, de alimento derivado de planta ou para regular para menos enzimas catabólicas envolvidas na degradação dos compostos desejados como aminoácidos essenciais.

Seqüências de Ácido Nucléico Estruturais Modificadas

Os promotores da presente invenção também podem ser operavelmente ligados a uma seqüência de ácido nucléico estrutura modificada que é heteróloga com relação ao promotor. A seqüência de ácido nucléico estrutural pode ser modificada para dar várias características desejáveis. Por exemplo, uma seqüência de ácido nucléico estrutural pode ser modificada para aumentar o teor de aminoácidos essenciais, intensificar a tradução da seqüência de ácido nucléico, alterar modificações pós-tradução (por exemplo, sítios de fosforilação), transportar um produto traduzido para um compartimento dentro ou fora da célula, melhorar a estabilidade da proteína, inserir ou deletar motivos de sinalização de célula, etc.

Em uma modalidade preferida, a seqüência de ácido nucléico estrutural é intensificada para codificar um polipeptídeo com um teor incrementado de pelo menos um, e mais preferivelmente cerca de 2, cerca de 3, ou cerca de 4 aminoácidos essenciais selecionados dentre o grupo consistindo em hisditina, lisina, metionina, e fenilalanina. Aminoácidos não-essenciais também podem ser adicionados conforme necessário, para intensificação estrutural e nutritiva do polipeptídeo. Seqüências de ácido nucléico estrutural particularmente adequadas para tal incremento incluem as que codificam polipeptídeos nativos que são expressos em níveis relativamente altos, têm um teor particularmente baixo de aminoácidos essenciais, ou ambos. Um exemplo desses são as proteínas de armazenagem de semente, como glicinina e β-conglicinina. Outros objetivos adequados incluem arcelina, fa-

seolina, lectina, zeínas e albumina.

Uso de códon em Seqüências de Ácido Nucléico Estruturais

Devido à degeneração do código genético, diferentes códons de nucleotídeo podem ser usados para codificar um dado aminoácido. Uma célula hospedeira freqüentemente apresenta um padrão preferido de uso de códon. Seqüências de ácido nucléico estruturais são de preferência construídas para utilizar o padrão de uso de códon da célula hospedeira particular. Isto de modo geral, melhora a expressão da seqüência de ácido nucléico estrutural em uma célula hospedeira transformada. Quaisquer seqüências de ácido nucléico e aminoácido supradescritas podem ser modificadas para refletir o uso de códon preferido de uma célula hospedeira ou organismo, em que estão contidas. A modificação de uma seqüência de ácido nucléico estrutural para uso ótimo do códon em plantas está descrita na Patente U.S. 5.689.052.

Outras Modificações de Seqüências de Ácido Nucléico Estruturais

Variações adicionais nas seqüências de ácido nucléico estruturais descritas acima, podem codificar proteínas com características equivalentes ou superiores quando comparado com as proteínas, das quais foram engenheiradas. Mutações podem incluir deleções, inserções, truncações, substituições, fusões, rearranjo de seqüências motivo e similares.

Mutações a uma seqüência de ácido nucléico estrutural podem ser introduzidas em, seja de modo específico ou aleatório, ambas as quais são do conhecimento da técnica da biologia molecular. Existem milhares de técnicas de mutagênese direcionada ao sítio, que utilizam tipicamente oligonucleotídeos para introduzir mutações em locais específicos em uma seqüência de ácido nucléico estrutural. Exemplos incluem resgate de fita simples (Kunkel *et al.,* 1985), eliminação de sítio singular (Deng and Nickloff, 1992), proteção de picotamento (Vandeyar *et al.,* 1988) e PCR (Costa *et al.,* 1996). Mutações aleatórias ou não-específicas podem ser geradas por agentes químicos (para um exame geral, *ver* Singer e Kusmierek, 1982) tal como nitrosoguanidina (Cerda-Olmedo *et al.*, 1968; Guerola *et al.,* 1971) e 2-aminopurina (Rogan and Bessman 1970), ou por métodos biológicos tais

como passagem por cepas mutantes (Greener *et al.*, 1997). Métodos adicionais de produção de alterações descritas acima estão descritas por Ausubel *et al.* (1995); Bauer *et al.*, (1985); Craik (1985), Frits Eckstein *et al.* (1982), Sambrook *et al.* (1989), Smith *et al.* (1981) e Osuna *et al.* (1994).

As modificações podem resultar em, seja mudanças conservativas ou não-conservativas na seqüência de ácido nucléico. Alterações conservativas são alterações que não alteram a seqüência de ácido nucléico final da proteína. em uma modalidade preferida, uma proteína tem entre cerca de 5 e cerca de 500 alterações conservativas, mais preferivelmente entre cerca de 10 e cerca de 300 alterações conservativas, mesmo mais preferivelmente entre cerca de 25 e cerca de 150 alterações conservativas, e mais preferivelmente entre cerca de 5 e cerca de 25 alterações conservativas ou entre cerca de 1 e cerca de 5 alterações conservativas.

Alterações não-conservativas incluem adições deleções, e substituições que resultam na seqüência de ácido nucléico alterada. Em uma modalidade preferida a proteína tem entre cerca de 5 e cerca de 500 alterações de aminoácido não-conservativas, mais preferivelmente entre cerca de 10 e cerca de 300 alterações de aminoácido não-conservativas, mesmo mais preferivelmente entre cerca de 25 e cerca de 150 alterações de aminoácido não-conservativas, e mais preferivelmente entre cerca de 5 e cerca de 25 alterações de aminoácido não-conservativas, ou entre cerca de 1 e cerca de 5 alterações não-conservativas.

Modificações podem ser feitas às seqüências de proteína aqui descritas e as seqüências de ácido nucléico que as codificam, que mantêm as desejada propriedades da molécula. A seguir discute-se, com base na alteração da seqüência de aminoácido de uma proteína para criar um equivalente, ou possivelmente uma molécula de segunda geração aperfeiçoada. As alterações de aminoácido podem ser conseguidas alterando-se os códons da seqüência de ácido nucléico estrutural, de acordo com os códons dados na Tabela 1.

Tabela 1: Degeneração do códon de aminoácidos

Aminoácido	Uma letra	Três letras	Códons
Alanina	A	Ala	GCA GCC GCG GCT
Cisteína	C	Cys	TGC TGT
Ácido Aspártico	D	Asp	GAC GAT
Ácido Glutâmico	E	Glu	GAA GAG
Fenilalanina	F	Phe	TTC TTT
Glicina	G	Gly	GGA GGC GGG GGT
Histidina	H	His	CAC CAT
Isoleucina	I	Ile	ATA ATC ATT
Lisina	K	Lys	AAA AAG
Leucina	L	Leu	TTA TTG CTA CTC CTG CTT
Metionina	M	Met	ATG
Asparagina	N	Asn	AAC AAT
Prolina	P	Pro	CCA CCC CCG CCT
Glutamina	Q	Gln	CAA CAG
Argininina	R	Arg	AGA AGG CGA CGC CGG CGT
Serina	S	Ser	AGC AGT TCA TCC TCG TCT
Treonina	T	Thr	ACA ACC ACG ACT
Valina	V	Val	GTA GTC GTG GTT
Triptofano	W	Trp	TGG
Tirosina	Y	Tyr	TAC TAT

Alguns aminoácidos podem ser substituídos por outros aminoácidos em uma seqüência de proteína sem perda apreciável da atividade desejada. Examina-se assim, as várias alterações que podem ser feitas nas seqüências de peptídeo ou seqüências de proteína ou suas correspondentes seqüências de ácido nucléico sem perda apreciável da atividade biológica.

Fazendo-se tais alterações, o índice hidropático dos aminoácidos pode ser considerado. A importância do índice de aminoácido hidropático em conferir função biológica interativa em uma proteína é geralmente entendida na técnica (Kyte e Doolittle, 1982). Aceita-se que o caráter hidropático relativo do aminoácido contribui para a estrutura secundária da proteína resultante, que por sua vez, define a interação da proteína com outras molécula, por exemplo, enzimas, substratos, receptores, DNA, anticorpos antígenos e similar.

Cada aminoácido concedeu um índice hidropático com base em suas características de hidrofobicidade e carga. Estes são: isoleucina (+4,5), valina (+4,2), leucina(+3,8), fenilalanina (+2,8), cisteína/cisteína (+2,5), metionina (+1,9), alanina (+1,8), glicina (-0,4), treonina (-0,7), serina (-0,8), triptofano (-0,9), tirosina (-1,3), prolina (-1,6), histidina (-3,2), glutamato/glutamina/aspartato/asparagina (-3,5), lisina (-3,9), e arginina (-4,5).

Sabe-se na técnica que alguns aminoácidos podem ser substituídos por outros aminoácidos com um índice ou pontuação hidropático similar e ainda resulta em uma proteína com atividade biológica similar, ou seja, ainda obtém uma proteína biologicamente funcional. Fazendo-se tais alterações, a substituição dos aminoácidos cujos índices hidropáticos estão dentro de ±2, é preferida, os dentro de ±1 são mais preferidos, e os dentro de ±0,5 são os mais preferidos.

Também entende-se da técnica que a substituição de aminoácido similares pode ser feita eficazmente com base na hidrofilicidade. A patente U.S. 4.554.101 declara que a maior hidrofilicidade média local de uma proteína, conforme governada pela hidrofilicidade de seus aminoácidos adjacentes correlaciona-se com uma propriedade biológica da proteína. Os seguintes valores de hidrofilicidade, foram cedidos a aminoácidos: arginina/lisina (+3,0), aspartato/glutamato (+3,0±1), serina (+0,3), asparagina/glutamina (+,2), glicina (0), treonina (-0,4), prolina (-0,5±1), alanina/histidina (-0,5), cisteína (-1,0), metionina (-1,3), valina (-1,5), leucina/isoleucina (-1,8), tirosina (-2,3), fenilalanina (-2,5) e triptofano (-3,4).

Fica entendido que um aminoácido pode ser substituído por um outro aminoácido com uma pontuação de hidrofilicidade similar e ainda resultar em uma proteína com atividade biológica similar, ou seja, ainda obter uma proteína biologicamente funcional. Fazendo-se tais alterações, a substituição dos aminoácidos cujos índices hidropáticos estão dentro de ±2 são preferidos, os dentro de ±1 são mais preferidos, e os dentro de ±0,5 são mais preferidos.

Como indicado acima, substituições de aminoácido são portanto, baseadas na similaridade relativa dos substituintes de cadeia lateral de ami-

noácido, por exemplo, sua hidrofobicidade, hidrofilicidade, carga, tamanho e similar. Substituições exemplares que consideram várias das características precedentes, em consideração são bem-conhecidos daqueles versados na técnica e incluem: arginina e lisina, glutamato e aspartato, serina e treonina, glutamina e asparagina e valina, leucina e isoleucina. Alterações que não se espera que sejam vantajosas podem também ser usadas caso essas proteínas resultantes vierem a melhorar a resistência do rúmen, a resistência maior a degradação proteolítica, ou ambas melhoraram a resistência do rúmen e aumentaram a resistência à degradação proteolítica, em relação ao polipeptídeo não-modificado do qual foram engenheiradas. Alternativamente, alterações, podem ser feitas para melhorar a cinética das enzimas metabólicas.

Em um aspecto preferido, a proteína modificada é selecionada de proteínas de armazenagem de semente, enzimas do passo do ácido graxo, enzimas biossintéticas de tocoferol, enzimas biossintéticas de aminoácido e enzimas de quebra do amido.

Vetores Recombinantes

Quaisquer dos promotores e seqüências de ácido nucléico estruturais descritos acima podem ser providenciados em um vetor recombinante. Um vetor recombinante tipicamente, compreende, em uma orientação 5' a 3' um promotor para direcionar a transcrição de uma seqüência de ácido nucléico estrutura e uma seqüência de ácido nucléico estrutura. Promotores adequados e seqüências de ácido nucléico estruturais incluem as aqui descritas. O vetor recombinante pode ainda compreender um terminador transcricional 3' um sinal de poliadenilação 3, outras seqüências de ácido nucléico não-traduzidas, seqüências de ácido nucléico passageiras e visadas, marcadores selecionáveis, intensificadores e operadores conforme desejado.

Meios de preparar vetores recombinantes são do conhecimento da técnica. Métodos para produzir vetores recombinantes particularmente adequados à transformação de planta estão descritos nas Patentes U.S. 4.971.908, 4.940.835, 4.769.061 e 4.757.011. Esses tipos de vetores tam-

bém foram pesquisados (Rodriguez *et al.*, 1988, Glick *et al.*, 1993).

Vetores típicos úteis para expressão de ácidos nucléicos em plantas superiores são bem-conhecidos na técnica e incluem vetores derivados de plasmídeo indutor de tumor (Ti) de *Agrobacterium tumefaciens* (Rogers et al., 1987). Outros vetores recombinantes úteis para trasnformação de planta incluindo o vetor de controle de transferência pCaMVCN também foram descritos (Fromm *et al.*, 1985).

Em uma modalidade, promotores 7Sα múltiplos são operavelmente ligados em um constructo simples a qualquer combinação de genes estruturais. Em uma modalidade preferida, qualquer combinação de uma, duas, três, quatro, cinco ou seis ou mais das SEQ ID NOs: 11, 12, 13 e 14 pode ser operativamente ligadas em um constructo simples a qualquer combinação de genes estruturais.

Promotores Adicionais no Vetor Recombinante

Um ou mais promotores adicionais também podem ser providenciados no vetor recombinante. Esses promotores podem ser operavelmente ligados, por exemplo, sem limitação a quaisquer seqüências de ácido nucléico estruturais supradescritas. Alternativamente, os promotores podem ser operavelmente ligados a outras seqüências de ácido nucléico, tais como as codificando peptídeos transitórios, proteínas marcadoras selecionáveis ou seqüências anti-sentido.

Esses promotores adicionais podem ser selecionados com base no tipo de célula para a qual o vetor será inserido. Ainda, os promotores que funcionam em bactérias, levedura e plantas são todos do conhecimento da técnica. Os promotores adicionais também podem ser selecionados com base em seus aspetos reguladores. Exemplos de tais aspectos incluem intensificação da atividade transcricional, capacidade de indução, especificidade para tecido, e especificidade para o estágio de desenvolvimento. Nas plantas, os promotores que são induzidos, de origem viral ou sintética, essencialmente ativos, temporariamente regulados, e especialmente regulados, foram descritos (Poszkowski et al., 1989, Odell et al., 1985, Chau et al., 1989).

Promotores constitutivos freqüentemente usados incluem o promotor CaMV 35S (Odell et al., 1985), o promotor CaMV 35S intensificado, o promotor do Vírus do Mosaico da Escrofulária (FMV) (Richins et al., 1987), o promotor manopina sintase (*mas*), o promotor nopalina sintase (*nos*), e o promotor octopina sintase (*ocs*).

Promotores estimuláveis úteis incluem promotores induzidos por ácido salicílico ou ácidos poliacrílicos (PR-1, William et al., 1992), induzido por aplicação de protetores (herbicidas benzenossulfonamida substituídos; Hershey and Stoner, 1991), promotores de choque térmico (Ou-Lee et al., 1986, Ainley et al., 1990), um promotor induzido por nitrato derivado da seqüência de ácido nucléico estrutural nitrito redutase do espinafre(Back et al. 1991), promotores induzidos por hormônio (Yamaguchi-Shinozaki et al., 1990, Kares et al., 1990), e promotores induzidos por luz associados com a subunidade pequena de RuBP carboxilase e famílias LHCP (Kughlemeier et al., 1989, Feinbaum et al., 1991, Weisshaar et al., 1991, Lam e Chua 1990 Castresana et al., 1988, Schulze-Lefert et al., 1989).

Exemplos de promotores específicos para tecido ou órgão úteis incluem β-conglicinina, (Doyle et al., 1986, Slighton e Beachy, 1987), e outros promotores específicos para semente (Knutzon et al. 1992, Bustos et al. 1991, Lam e Chua, 1991). Promotores funcionais de planta úteis para expressão preferencial em plastídeo de semente incluem os de proteínas de armazenagem de planta e de proteínas envolvidas na biossíntese de ácido graxo em sementes oleaginosas. Exemplos de tais promotores incluem as regiões reguladoras 5' de tais seqüências de ácido nucléico estruturais como napin (Kridl et al., 1991), faseolina, zeína, inibidor de tripsina de soja, ACP, estearoil-ACP desaturase e oleosina. A regulação específica para semente está ainda debatida em EP 0 255 378.

Um outro promotor específico para semente exemplar é um promotor de lectina. A proteína lectina nas sementes da soja é codificada por uma única seqüência de ácido nucléico estrutural (le1) que é somente expressa durante a maturação da semente. Uma seqüência de ácido nucléico estrutura de lectina e promotor específico para semente foram caracteriza-

dos e usados para direcionar a expressão específica para semente em plantas de tabaco transgênicas (Vodkin et al., 1983, Lindstrom et al., 1990).

Promotores adicionais particularmente preferidos no vetor recombinante incluem promotores nopalina sintase (*nos*), manopina sintase (*mas*) e octopina sintase (*ocs*), que são portados em plasmídeos indutores de tumor de *Agrobacterium tumefaciens,* os promotores do vírus do mosaico da couve-flor (CaMV) 19S e 35S, o promotor 35S de CaMV intensificado, o promotor 35S do Vírus do Mosaico da Escrofulária (FMV), o promotor sensível à luz da pequena subunidade ribulose-1,5-bifosfato carboxilase (ssRUBISCO), o promotor EIF-4A do tabaco (Mandel et al., 1995), sintetase 1 de sacarose de milho (Yang e Russell 1990), desidrogenase 1 do álcool de milho (Vogel et al., 1989) complexo coletor de luz (Simpson 1986), proteína de choque térmico do milho (Odell et al., 1985), o promotor de quitinase de *Arabidopsis* (Samac et al., 1991), os promotores de (Proteína de transferência de lipídeo) LTP do brócolis (Pyee et al., 1995), chalcona isomerase de petúnia (Van Tunen et al., 1988), proteína rica em glicina do feijão 1 (Keller et al., 1989), patatina da batata (Wenzler et al., 1989), o promotor de ubiquitina do milho (Christensen et al., 1992), e o promotor de actina do arroz (McElroy et al., 1990).

Um promotor adicional e de preferência seletivo para semente, seletivo para tecido, constitutivo ou indutor. O promotor é mais preferivelmente nopalina sintase (*nos*), octopina sintase (*ocs*), manopina sintase (*mas*), vírus do mosaico da couve-flor 19S e 35S (CaMV19S, CaMV35S), CaMV (eCaMV) intensificado, ribulose 1,5-bifosfato carboxilase (ssRUBISCO), vírus do mosaico da escrofulária (FMV), AS4 derivado de CaMV, POX1 do trigo, EIF-4 do tabaco, proteína lectina (Le1) ou promotor RC2 do arroz.

<u>Vetores Recombinantes com Seqüências de ácido nucléico estruturais adicionais</u>

O vetor recombinante também pode conter uma ou mais seqüências de ácido nucléico estruturais adicionais. Essas seqüências de ácido nucléico estruturais podem ser geralmente, quaisquer seqüências adequada para uso em um vetor recombinante. Tais seqüências de ácido nu-

cléico estruturais incluem, sem limitação, quaisquer seqüências de ácido nucléico estruturais, e formas modificada destas, descritas acima. As seqüências de ácido nucléico estruturais adicionais também podem ser operavelmente ligadas a quaisquer dos promotores supradescritos. A uma ou mais seqüências de ácido nucléico estruturais podem ser cada qual operavelmente ligadas a promotores separados. Alternativamente, as seqüências de ácido nucléico estruturais podem ser operavelmente ligadas a um único promotor (ou seja, um operon simples).

As seqüências de ácido nucléico estruturais adicionais incluem, sem limitação àquelas que codificam as proteínas armazenadoras de semente, enzimas do passo de ácido graxo, enzimas biossintéticas de tocoferol, enzimas biossintéticas de aminoácido e enzimas de quebra do amido.

Proteínas armazenadoras de semente preferidas incluem zeínas (Patentes U.S. 4.886.878, 4.885.357, 5.215.912, 5.589,616, 5.508.468, 5.939.599, 5.633.436 e 5.990.384, WO 90/01869, WO 91/13993, WO 92/14822, WO 93/08682, WO 94/20628, WO 97/28247, WO 98/26064 e WO 99/40209), proteínas 7S (Patentes U.S. 5.003.045 e 5.576.203) proteína da castanha do Pará (Patente U.S. 5.850.024), proteínas isentas de fenilalanina (WO 96/17064), albumina (WO 97/35023, β-conglicinina (WO 00/19839), 11S (Patente U.S. 6.107.051), alfa-hordotionina (Patentes U.S. 5.885.802 e 5.885.801) proteínas de armazenagem de semente de arcelina (Patente U.S. 5.270.200) lectinas (Patente U.S. 6.110.891) e glutenina (Patentes U.S. 5.990.389 e 5.914.450).

Enzimas do passo de ácido graxo preferidas incluem tioesterases (Patentes U.S. 5.512.482, 5.530.186, 5.945.585. 5639.790, 5.807.893, 5.955.650, 5.955.329, 5.759.829, 5.147.792, 5.304.481, 5.298.421, 5.344.771 e 5.760.206) e desaturases (Patentes U.S. 5.689.050, 5.663.068, 5.614.393, 5.856.157, 6.117.677, 6.043.411, 6.194.167, 5.705.391, 5.663.068, 5.552.306, 6.075.183, 6.051.754, 5.689.050, 5.789.220, 5.057.419, 5.654.402, 5.659.645, 6.100.091, 5.760.206, 6.172.106, 5.952.544, 5.866.789, 5.443.974 e 5.093.249).

Enzimas biossintéticas de tocoferol preferidas incluem *tyrA*,

slr1736, ATPT2, dxs, dxr, GGPPS, HPPD, GMT, MT1, AANT1, sir 1737 e um constructo anti-sentido para dioxigenase ácida homogentísica (Kridl *et al., Seed Sci Res.* 1:209:219 (1991), Keegstra *Cell* 56(2):247-53 (1989), Nawath *et al., Proc. Natl. Acad, Sci U.S.A* 91:12760-12764 (1994); Xia *et al., J. Gen. Microbiol.* 138:1309-1316 (1992), Cianobase http://www.kazusa.or.jp/cianobase, Lois *et al., Proc. Natl. Acad. Sic U.S.A* 95(5):2105-2110 (1998), Takahashi *et al., Proc. Natl. Acad. Sci U.S.A* 95(17), 9879-9884 (1998), Norris *et al., Plant Physiol.* 117:1317-1323 (1998), Bartley and Scolnik *Plant Physiol.* 104:1469-1470 (1994), Smith *et al., Plant J.* 11:83-92 (1997), WO 00/32757, WO 00/10380, Saint Guily *et al., Plant Physiol.,* 100(2):1069-1071 (1992), Sato *et al., J. DNA Res* 7(1):31-63 (2000)).

Enzimas biossintéticas de aminoácido preferidas incluem antranilato sintase (Patente U.S. 5.965.727 e WO 97/26366, WO 99/11800 e WO 99/49058) triptofano descarboxilase (WO 99/06581) e treonina descarboxilase (Patentes U.S. 5.534.421 e 5.942.660 e WO 95/19442), treonina desaminase (WO 99/02656 e WO 98/55601) e aspartato quinase (Patentes U.S. 5.367.110, 5.858.749 e 6.040.160).

Enzimas de quebra do amido preferidas incluem as apresentadas nas Patentes U.S. 6.232.122 e 6.147.279 e WO 97/22703.

Alternativamente, a segunda seqüência de ácido nucléico estrutural pode ser projetada para regular para menos uma dada seqüência de ácido nucléico. Isto é tipicamente realizado por ligação operável do segundo aminoácido estrutural, em uma orientação anti-sentido, com um promotor. Um versado na técnica estará familiarizado com tal tecnologia anti-sentido. Qualquer seqüência de ácido nucléico pode ser negativamente regulada deste modo. Seqüências de ácido nucléico alvo preferíveis contêm um baixo teor de aminoácidos essenciais, ainda que sejam expressos a níveis relativamente altos em particular em tecidos. Por exemplo, β-conglicinina e glicinina são expressos abundantemente em sementes, mas são deficientes do ponto de vista nutricional com relação a aminoácidos essenciais. Esta abordagem anti-sentido pode ser usada também para remover eficazmente outras proteínas indesejáveis, tais como antinutrientes (por exemplo, lectinas),

albumina e alergênicos, de produtos alimentícios derivados de planta, ou para regular para menos enzimas catabólicas envolvidas na degradação dos compostos desejados tal como aminoácidos essenciais.

Marcadores Selecionáveis

Um vetor ou constructo também pode incluir um marcador selecionável. Marcadores selecionáveis podem ser usados também para selecionar plantas ou células de planta que contém o material genético exógeno. Exemplos destes incluem, sem estar limitado a; um gene *neo* (Potrykus et al., 1985), que codifica para gene resistente à canamicina e pode ser selecionado para uso de canamicina, RptII, G418, hpt, etc., um gene bar que codifica para resistência a bialafos, um gene EPSP sintase mutante (Hinchee et al., 1988, Reynaerts et al., 1988), *aad*A (Jones et al., 1987) que codifica resistência a glifosato, um gene nitrilase que confere resistência a bromoxinila (Stalker et al., 1988), um gene acetolactato sintase mutante (ALS) que confere resistência a imidazolinona ou sulfoniluréia (Pedido de Patente Europeu 154.204, 1985), ALS (D'Halluin et al., 1992) e um gene DHFR resistentes a metotrexato (Thillet et al., 1988), O marcador selecionável é de preferência, GUS, proteína fluorescente verde (GFP), neomicina fosfotransferase II (*nptII*), luciferase (LUX), uma seqüência de codificação resistente a antibiótico, ou uma seqüência de codificação resistente a herbicida (por exemplo, glifosato). O marcador selecionável é mais preferivelmente, um marcador resistente a canamicina, higromicina ou herbicida.

Um vetor ou constructo também pode incluir um marcador examinador. Marcadores examinadores podem ser usados para monitorar a expressão. Marcadores examinadores exemplares incluem um gene β-glicuronidase ou *uid*A (GUS) que codifica uma enzima para qual vários substratos cromogênicos são conhecidos (Jefferson, 1987), um gene R-local, que codifica um produto que regula a produção de pigmentos antocianina (cor vermelha) em tecidos de planta (Dellaporta et al., 1988), um gene β-lactamase (Sutcliffe et al. 1978), um gene que codifica uma enzima para a qual vários substratos cromogênicos são conhecidos (por exemplo, PADAC, uma cefalosporina cromogênica), um gene luciferase (Ow et al., 1986), um gene *xy*/E,

(Zukowsky et al., 1983), que codifica uma catecol dioxigenase que pode converter catecóis cromogênicos, um gene α-amilase (Ikatu et al., 1990), um gene tirosinase (Katz et al., 1983), que codifica uma enzima capaz de oxidar tirosina em DOPA e dopaquinona que por sua vez, condensa em melanina, uma α-galactosidase, que irá converter um substrato α-galactose cromogênico.

Incluído nos termos "genes marcadores selecionáveis ou examinadores" também estão genes que codificam um marcador de secreção, cuja secreção pode ser detectada como um meio de identificar ou selecionar células transformadas. Exemplos incluem marcadores que codificam um antígeno de secreção que podem ser identificados por interação com anticorpo, ou mesmo enzimas secretoras que podem ser detectadas cataliticamente. Proteínas de secreção enquadram-se em uma série de classes, incluindo proteínas difusíveis pequenas que são detectáveis (por exemplo, por ELISA), enzimas de pequena atividade que são detectáveis em solução extracelular (por exemplo, α-amilase, β-lactamase, fosfinotricina transferase), ou proteínas que são inseridas ou encerradas na parede celular (tal como proteínas que incluem uma seqüência líder como a encontrada na unidade de expressão da extensão do tabaco PR-S). Outros possíveis genes marcadores selecionáveis e/ou examinadores se evidenciarão aos versados na técnica.

Outros Elementos no Vetor Recombinante

Várias seqüências reguladoras 5' e 3' não-traduzidas de configuração *cis* podem ser incluídas no vetor de ácido nucléico recombinante. Quaisquer seqüências reguladoras podem ser providas em um vetor recombinante com outras seqüências reguladoras. Tais combinações podem ser destinadas ou modificadas para produzir aspectos reguladores desejáveis.

Uma região 3' não-traduzida, propicia tipicamente, um sinal de terminação trancricional e um sinal de poliadenilação que funciona em plantas para ocasionar a adição de nucleotídeos adenilato na extremidade 3' do mRNA. Estes podem ser obtidos a partir de regiões 3' da seqüência de codificação (*nos*) de nopalina sintase, uma seqüência de codificação de proteína

de armazenagem 7Sα'da soja, a seqüência de codificação 5 de arcelina, a seqüência de codificação de albumina e a seqüência de codificação ssRUBISCO E9 da ervilha. Tipicamente, seqüências de ácido nucléico localizada umas poucas centenas de pares de base a jusante do sítio de poliadenilação servem para terminar transcrição. Essas regiões são necessárias para eficiente poliadenilação do mRNA transcrito.

Intensificadores de tradução também podem ser incorporados como parte do vetor recombinante. Assim o vetor recombinante pode de preferência conter uma ou mais seqüências líder não-traduzidas 5' que servem para melhorar a expressão da seqüência de ácido nucléico. Tais seqüências intensificadoras podem ser desejáveis para aumentar ou alterar a eficiência de tradução do mRNA resultante. Seqüências de ácido nucléico 5' preferidas incluem dSSU 5', PetHSP70 5' e GmHSP17.9 5' (Patente U.S. 5.362.865).

O vetor recombinante pode ainda compreender uma seqüência de ácido nucléico codificando um peptídeo transitório. Este peptídeo pode ser útil para direcionar uma proteína ao espaço extracelular, um plastídeo ou a algum outro compartimento dentro ou fora da célula (ver, por exemplo, EP 0 218 571, Patentes U.S. 4.940.835, 5.610.041, 5.618.988 e 6.107.060).

A seqüência de ácido nucléico estrutural no vetor recombinante pode compreender íntrons. Os íntrons podem ser heterólogos com relação à seqüência de ácido nucléico estrutural. Íntrons preferidos incluem o íntron de actina do arroz e o íntron HSP70 do milho.

Proteínas de Fusão

Quaisquer seqüências de ácido nucléico estruturais supradescritas, e suas formas modificadas, podem ser ligadas com seqüências de ácido nucléico adicionais para codificar proteínas de fusão. A seqüência de ácido nucléico adicional de preferência codifica pelo menos um aminoácido, peptídeo ou proteína. Muitas formas de combinação de fusão são possíveis.

Por exemplo, a proteína de fusão pode providenciar um epítopo "rotulado" para facilitar a detecção da proteína de fusão, como GST, GFP,

FLAG ou polyHIS. Tais fusões, de preferência codificam ente cerca de 1 e cerca de 50 aminoácidos, mais preferivelmente entre cerca de 5 e cerca de 30 aminoácidos adicionais, ainda mais preferivelmente entre cerca de 5 e cerca de 20 aminoácidos.

Alternativamente, a fusão pode providenciar funções reguladoras enzimáticas, sinalização de célula ou transporte intercelular. Por exemplo, uma seqüência codificando um peptídeo transitório de plastídeo pode ser adicionada para direcionar uma proteína de fusão aos cloroplastos dentro de sementes. Tais parceiros de fusão de preferência codificam entre cerca de e cerca de 1000 aminoácidos adicionais, mais preferivelmente ente cerca de 5 e cerca de 500 aminoácidos adicionais e ainda mais preferivelmente, entre cerca de 10 e cerca de 250 aminoácidos.

Análise de Seqüência

Na presente invenção, a similaridade ou identidade de seqüência é de preferência determinada usando os programas "BestFit" ou "Gap" de Sequence Analysis Sofware Package® (Versão 10, Genectics Computes Group, Inc., University of Wisconsin Biotechnology Center, Madison, WI). "Gap" utiliza o algoritmo de Needleman e Wunsch (Needleman and Wunsch, 1970) para encontrar o alinhamento de duas seqüências que maximizem o número de pares e minimize o número de folgas. "BestFit" executa um ótimo alinhamento do melhor segmento de similaridade entre duas seqüências. Alinhamentos ótimos são encontrados por inserção de folgas a fim de maximizar o número de pares usando o algoritmo de homologia local de Smith and Waterman (Smith and Waterman, 1981, Smith *et al.,* 1983).

A Sequence Analysis Sofware Package descrita acima contém uma série de outras ferramentas de análise de seqüência para identificar homólogos das seqüências de nucleotídeos e de aminoácido presentemente apresentadas. Por exemplo, o programa "BLAST" pesquisa para seqüências similares a uma seqüência de pesquisa (seja de peptídeo ou ácido nucléico) em uma base de dados específica (por exemplo, base de dados da seqüência mantida no National Center for Biotechnology Information (NCBI) em Bethesda, Maryland, (USA), "FastA"(Lipman and Pearson, 1985, e também

Pearson e Lipman, 1988, Pearson, 1990) realizam uma pesquisa de Pearson and Lipman para similaridade entre uma seqüência de pesquisa e um grupo de seqüência do mesmo tipo (ácido nucléico ou proteína). "TfastA" realiza uma busca Pearson e Lipman para similaridade entre uma seqüência de pesquisa de proteína e qualquer grupo de seqüências de nucleotídeo (ele traduz as seqüências de nucleotídeos em todos os seis quadros de leitura antes de realizar a comparação); "FastX" executa uma busca Pearson and Lipman para similaridade entre uma seqüência de pesquisa de nucleotídeos e um grupo de seqüência de proteínas levando em consideração desvios do quadro. "TfastX" executa uma busca Pearson and Lipman para similaridade entre uma seqüência de pesquisa de proteína e qualquer grupo de seqüências de nucleotídeos levando em consideração desvios do quadro (ele traduz ambas as fitas da seqüência de ácido nucléico antes de realizar a comparação).

Sondas e Iniciadores

Seqüências de ácido nucléico curtas com capacidade em hibridizar especificamente para seqüências de ácido nucléico complementares podem ser produzidas e utilizadas na presente invenção. Tais moléculas de ácido nucléico curtas podem ser usadas como sondas na identificação da presença de uma seqüência de ácido nucléico complementar em uma dada amostra. Assim, construindo-se uma sonda de ácido nucléico isto é complementar a uma pequena parte de uma seqüência de ácido nucléico particular, a presença daquela seqüência de ácido nucléico pode ser detectada e avaliada.

Alternativamente, as seqüências de ácido nucléico curtas, podem ser usadas como iniciadores de oligonucleotídeo para ampliar ou fazer mutação de uma seqüência de ácido nucléico complementar usando tecnologia de PCR. Esses iniciadores podem também facilitar a ampliação de seqüências de ácido nucléico complementares relacionadas (por exemplo, seqüências de ácido nucléico relacionadas de outras espécies).

Seqüências de ácido nucléico curtas podem ser usadas como iniciadores e especificamente como iniciadores PCR. Uma sonda PCR é

uma molécula de ácido nucléico capaz de iniciar uma atividade polimerase enquanto em uma estrutura de dupla fita com um outro ácido nucléico. Vários métodos para determinar a estrutura de iniciadores PCR e técnicas PCR existem na técnica. Buscas geradas por computador usando programas como Iniciador3, STSPipeline ou GeneUP (Pesole et al., 1998) por exemplo, podem ser usados para identificar o potencial de iniciadores PCR.

Quaisquer seqüências de ácido nucléico aqui descritas podem ser usadas como iniciador ou como sonda. O uso desses iniciadores ou sondas podem facilitar em muito, a identificação de plantas transgênicas que contêm os promotores atualmente descritos e seqüências de ácido nucléico estruturais. Tais sondas ou iniciadores podem também ser usados para selecionam cDNA ou bibliotecas genômicas para seqüências de ácido nucléico adicionais relacionadas com ou dividindo homologia com os promotores e seqüências de ácido nucléico estruturais descritos no presente.

Um iniciador ou sonda é geralmente complementar a uma parte de uma seqüência de ácido nucléico que deve ser identificada, ampliada ou passar por mutação e de comprimento suficiente para formar uma molécula dúplex estável e específica para a seqüência com seu complemento. O iniciador ou sonda de preferência é cerca de 10 a cerca de 200 nucleotídeos de comprimento, mais preferivelmente cerca de 10 a cerca de 100 nucleotídeos de comprimento, mesmo mais preferivelmente cerca de 10 a cerca de 50 nucleotídeos de comprimento, e mais preferivelmente cerca de 14 a cerca de 30 nucleotídeos de comprimento.

O iniciador ou sonda pode, por exemplo, sem limitação ser preparado por síntese química direta por PCR (Patentes U.S. 4.683.195 e 4.683.202) ou por excisão do fragmento específico de ácido nucléico de uma molécula de ácido nucléico maior.

Plantas Transgênicas e Células Hospedeiras de Planta Transformada

A invenção também se refere a plantas transgênicas e células hospedeiras transformadas que compreende em uma orientação 5' a 3' um

promotor operavelmente ligado a uma seqüência de ácido nucléico estrutura heteróloga. Outras seqüências de ácido nucléico também podem ser introduzidas na planta ou célula hospedeira juntamente com o promotor e a seqüência de ácido nucléico estrutural. Essas outras seqüências podem incluir terminadores transcricionais 3', sinais de poliadenilação 3', outras seqüências de ácido nucléico não-traduzidas, seqüências transitórias ou alvo, marcadores selecionáveis, intensificadores e operadores. Seqüências de ácido nucléico preferidas da presente invenção incluem vetores recombinantes, seqüências de ácido nucléico estruturais, promotores e outros elementos reguladores descritos acima.

Meios para preparar tais vetores recombinantes são bem-conhecidos na técnica. Por exemplo, métodos de produzir vetores recombinantes particularmente adequados para transformação de planta são descritos nas Patentes U.S. 4.971.908. 4.940.835, 4.769.061 e 4.757.011. Esses vetores também foram revistos (Rodriguez et al., 1988; Glick et al., 1993).

Vetores típicos úteis para expressão de ácidos nucléicos em células e plantas superiores são bem-conhecidos na técnica e incluem vetores derivados de plasmídeo indutor de tumor (Ti) de *Agrobacterium tumefaciens* (Rogers et al., 1987). Outros vetores recombinantes úteis para transformação de planta também foram descritos (Fromm et al., 1985). Elementos de tais vetores recombinantes incluem, sem limitação, os supradescritos.

Uma célula hospedeira transformada pode, geralmente, ser qualquer célula que seja compatível com a presente invenção. Uma planta ou célula hospedeira transformada pode ser, ou derivada de uma planta monocotilédone ou uma planta dicotilédone incluindo, sem limitação a canola, ("crambe"), milho, mostarda, mamona, gergelim, caroço de algodão, semente de linhaça, soja, *Arabidopsis, Phaseolus*, amendoim, alfafa, trigo, arroz, aveia, sorgo, semente de colza, centeio, tritordeum, milheto, capim-do-prado, (forragem) azevém, cana-de-açúcar, arando, papaia, banana, açafrão, óleo de palma, linho, melão almiscarado, maçã, pepino, dendróbio, gladíolo, crisântemo, liliácea, algodão, eucalipto, girassol, *Brassica campestris, Brassica napus,*

turfa, beterraba, café e dioscoréia (Christou, em: *Particle Bombardment for Genetic Engineering of Plants,* Biotechnology Intelligence Unit, Academic Press, San Diego, California (1996)), com canola, milho, *Brassica campestris, Brassica napus,* semente de colza, soja, açafrão, trigo, arroz e girassol preferidos e canola, semente de colza, milho *Brassica campestris, Brassica napus,* soja, girassol, açafrão, óleo de palma e amendoim mais preferidos. Em uma modalidade particularmente preferida, a planta ou célula é ou é derivada de canola. Em uma outra modalidade particularmente preferida, a planta ou célula trata-se de, ou é derivada de *Brassica napus.* Em mais uma modalidade particularmente preferida a planta ou célula é ou é derivada de soja.

A célula ou planta de soja é de preferência uma linhagem de célula de soja superior. Uma "linhagem de elite" é qualquer linhagem que resultou do cultivo e seleção para desempenho agronômico superior. Exemplos de linhagens de elite são linhagens que são comercialmente disponíveis aos fazendeiros ou os que cultivam soja, como HARTZ® variedade H4994, HARTZ® variedade H5218, HARTZ® variedade H5350, HARTZ® variedade H5545, HARTZ® variedade H5050, HARTZ® variedade H5454, HARTZ® variedade H5233, HARTZ® variedade H5488, HARTZ® variedade HLA572, HARTZ® variedade H6200, HARTZ® variedade H6104, HARTZ® variedade H6255, HARTZ® variedade H6586, HARTZ® variedade H6191, HARTZ® variedade H7440, HARTZ® variedade H4452 Roundup Ready®, HARTZ™ variedade H4994 Roundup Ready®, HARTZ® variedade H4988 Roundup Ready®, HARTZ® variedade H5000 Roundup Ready®, HARTZ® variedade H5147 Roundup Ready®, HARTZ® variedade H5247 Roundup Ready®, HARTZ® variedade H5350 Roundup Ready®, HARTZ® variedade H5545 Roundup Ready®, HARTZ® variedade H5855 Roundup Ready®, HARTZ® variedade H5088 Roundup Ready®, HARTZ® variedade H5164 Roundup Ready®, HARTZ® variedade H5361 Roundup Ready®, HARTZ® variedade H5566 Roundup Ready®, HARTZ® variedade H5181 Roundup Ready®, HARTZ® variedade H5889 Roundup Ready®, HARTZ® variedade H5999 Roundup Ready®, HARTZ® variedade H6013 Roundup Ready®, HARTZ® variedade

H6255 Roundup Ready®, HARTZ® variedade H6454 Roundup Ready®, HARTZ® variedade H6686 Roundup Ready®, HARTZ® variedade H7152 Roundup Ready®, HARTZ® variedade H7550 Roundup Ready®, HARTZ® variedade H8001 Roundup Ready® (HARTZ SEED, Stuttgart, Arkansas, U.S.A), A0868, AG0901, A1553, A1900, AG1901, A1923, A2069, AG2101, AG2201, A2247, AG2301, A2304, A2396, AG2401, AG2501, A2506, A2553, AG2701, AG2702, A2704, A2833, A2869, AG2901, AG2902, AG3001, AG3002, A3204, A3237, A3244, AG3301, AG3302, A3404, A3469, AG3502, A3559, AG3601, AG3701, AG3704, AG3750, A3834, AG3901, A3904, A4045, AG4301, A4341, AG4401, AG4501, AG4601, AG4602, A4604, AG4702, AG4901, A4922, AG5401, A5547, AG5602, A5704, AG5801, AG5901, A5944, A5959, AG6101, QR4459 e QP4544 (Asgrow seeds, Des Moines, Iowa, U.S.A), DeKalb variedade CX445 (DeKalb, Illinois).

A invenção também é dirigida a um método de produzir plantas transformada que compreendem, em uma orientação 5'a 3' um promotor operavelmente ligado a uma seqüência de ácido nucléico estrutural heteróloga. Outras seqüências podem também ser introduzidas em plantas juntamente com o promotor e seqüência de ácido nucléico estrutural. Essas outras seqüências podem incluir terminadores transcricionais 3', sinais de poliadenilação 3', outras seqüência não-traduzidas, seqüências transitórias ou alvo, marcadores selecionáveis, intensificadores e operadores. Vetores recombinantes preferidos, seqüências de ácido nucléico estruturais, promotores e outros elementos reguladores incluindo sem limitação aos aqui descritos.

O método geralmente compreende as etapas de selecionar uma planta adequada, transformar a planta com um vetor recombinante e obter a célula hospedeira transformada.

Existem muitos métodos para introduzir ácidos nucléicos em plantas. Métodos adequados incluem infecção bacteriana (por exemplo, *Agrobacterium)*, vetores cromossomais artificiais bacterianos binários, fornecimento direto de ácidos nucléicos (por exemplo, via transformação mediada por PEG, absorção de ácido nucléico medida por dessecação/inibição ele-

troporação, agitação com fibras de carbureto de silício, e aceleração de partículas revestidas de ácido nucléico, etc., (revista em Potrykus et al., 1991)).

A tecnologia para introdução de ácidos nucléicos a células é do conhecimento dos versados na técnica. Métodos podem ser geralmente classificados em quatro categorias: (1) métodos químicos, (Graham e van der Eb, 1973, Zathoukal et al., 1992), (2) métodos físicos como microinjeção (Capecchi, 1980), eletroporação (Wong e Neumann, 1982, Fromm et al., 1985, Patente U.S. 5.384.253), e aceleração de partícula (Johnston and Tang, 1994, Fynan, et al., 1993), (3) vetores virais (Clapp, 1993, Lu et al., 1993, Eglitis e Anderson, 1988), e (4) mecanismos mediados por receptor (Curiel et al., 1992, Wagner et al., 1992). Alternativamente, ácidos nucléicos podem ser diretamente introduzidos no pólen por injeção direta a um órgão reprodutor de planta (Zhou et al., 1983, Hess, 1987, Luo et al., 1988, Pena et al., 1987). Em outro aspecto ácidos nucléicos podem também ser injetados em embriões imaturos (Neuhaus et al., 1987).

A regeneração, desenvolvimento e cultivo de plantas de protoplastos e explantes de planta transformada é ensinada na técnica (Weissbach, e Weissbach, 1988, Horsch et al., 1985). Transfomantes são geralmente cultivados na presença de um meio seletivo que seleciona para células transformadas com sucesso e induz a regeneração dos rebentos de planta (Fraley et al., 1983). Tais rebentos são tipicamente obtidos dentro de dois ou quatro meses.

Os rebentos são a seguir transferidos para um meio de crescimento de raiz apropriadas contendo o agente seletivo e um antibiótico para prevenir crescimento bacteriano. Muitos rebentos desenvolverão raízes. Estes são então transplantados ao solo ou outro meio para permitir o contínuo desenvolvimento das raízes. Os métodos, como esboçado, irão geralmente variar, dependendo da planta particular empregada.

De preferência, as plantas transgênicas regeneradas são autopolinizadas para providenciar plantas transgênicas homozigotas. Alternativamente, o pólen obtido de plantas transgênicas regeneradas podem ser cruzadas com as plantas não-transgênicas, de preferência linhagens con-

sangüíneas de espécies agronomicamente importantes. De modo inverso, pólen de plantas não-transgênicas pode ser usado para polinizar as plantas transgênicas regeneradas.

Uma planta transgênica pode passar ao longo da seqüência de ácido nucléico codificando expressão genética intensificada para sua progênie. A planta transgênica é de preferência homozigótica para o ácido nucléico codificando a expressão genética intensificada e transmite aquela seqüência para toda a sua descendência, como resultado da reprodução sexual. A progênie pode ser crescida de sementes produzidas pela planta transgênica. Essas plantas adicionais podem então ser autopolinizadas para gerar uma linhagem de plantas de geração pura.

A progênil dessas plantas é avaliada, dentre outras coisas, para expressão genética. A expressão genética pode ser detectada por vários métodos comuns como Western blotting, Northern blotting, imunoprecipitação e ELISA.

Plantas ou agentes da presente invenção podem ser utilizados em métodos, por exemplo, sem limitação, na obtenção de uma semente que expressa uma molécula de ácido nucléico estrutural naquela semente, para obter uma semente intensificada em um produto de um gene estrutural, para obter uma farinha melhorada em um produto onde um gene estrutural, para obter suprimento intensificado em um produto de um gene estrutural, e para obter óleo intensificado em um produto de um gene estrutural.

Plantas utilizadas em tais métodos, podem ser processadas. Uma planta ou parte de planta pode ser separada ou isolada de outras partes de planta. Uma parte de planta preferida para este fim é uma semente. Entende-se que mesmo após separação ou isolamento de outras partes de planta, a parte de planta isolada ou separada pode ser contaminada com outras partes de planta. Em um aspecto preferido, a parte de planta separada é maior do que cerca de 50% (p/p) do material separado, mais preferivelmente maior do que cerca de 75% (p/p) do material separado, e mesmo mais preferivelmente mais do que cerca de 90% p/p do material separado. Plantas ou partes de planta da presente invenção geradas por tais métodos

podem ser processadas em produtos usando técnicas conhecidas. Produtos preferido são farinha, suprimento e óleo.

Preparações de Alimento, Farinha, Proteína e Óleo

Qualquer das plantas ou partes desta da presente invenção podem ser processadas para produzir uma preparação de alimento, farinha, proteína ou de óleo. Uma parte de planta particularmente preferida para este fim é uma semente. Em uma modalidade preferida a preparação de alimento, farinha, proteína ou óleo é destinada para animais ruminantes. Métodos de produzir alimento, farinha, proteína e preparações de óleo são do conhecimento da técnica. Ver, por exemplo, Patentes U.S. 4.957.748, 5.100.679, 5.219.596, 5.936.069, 6.005.076, 6.146.669 e 6.156.227. Em uma modalidade preferida, a preparação de proteína é uma preparação de alto teor protéico. Uma tal proteína de alto teor de preferência tem um teor de proteína maior do que cerca de 5% peso/volume, mais preferivelmente cerca de 10% peso/volume, e mesmo mais preferivelmente cerca de 15% peso/volume. Em uma preparação de óleo preferida a preparação de óleo é uma preparação com alto teor de óleo com um teor de óleo derivado de uma planta ou parte desta da presente invenção em mais do que 5% p/v, mais preferivelmente maior do que cerca de 10 p/v, e mais preferivelmente maior do que cerca de 15% p/v. Em uma modalidade preferida a preparação de óleo é um líquido e de um volume maior do que cerca de 1, cerca de 5, cerca de 10, ou cerca de 50 litros. A presente invenção propicia óleo produzido de plantas da presente invenção ou geradas por um método da presente invenção. Tal óleo pode ser um componente menor ou maior do produto resultante. Além disso, tal óleo pode ser combinado com outros óleos. Em uma modalidade preferida o óleo produzido de plantas da presente invenção ou gerado por um método da presente invenção constitui mais do que cerca de 0,5, cerca de 1, cerca de 5, cerca de 10, cerca de 25, cerca de 50, cerca de 75 ou cerca de 90% em volume ou peso do componente de óleo de qualquer produto. Em mais uma modalidade, a preparação de óleo pode ser combinada e pode constituir mais do que cerca de 10, cerca de 25, cerca de 35, cerca de 50, ou cerca de 75% da combinação em volume. Óleo produzido de uma planta da

presente invenção pode ser misturado com um ou mais solvente orgânico ou destilados de petróleo.

Recipientes de semente.

Sementes de plantas podem ser colocadas em um recipiente. Como aqui empregado, um recipiente é qualquer objeto capaz de guardar tais sementes. Um recipiente de preferência contém mais do que cerca de 500, cerca de 1000, cerca de 5000 ou cerca de 25.000 sementes onde pelo menos cerca de 10, cerca de 25, cerca de 50, cerca de 75 ou cerca de 100% das sementes são derivadas de uma planta da presente invenção.

Programas de cultivo

As plantas da presente invenção podem constituir parte de programa de cultivo ou serem geradas de um. A escolha do método de cultivo depende do modo de reprodução da planta, da hereditariedade dos traços característicos que estão sendo melhorados e do tipo de cultivar usado comercialmente (por exemplo, cultivar híbrida F_1, cultivar de linhagem pura, etc.). Abordagens não-limitantes selecionadas para cultivo das plantas das presente invenção são dadas abaixo. Um programa de cultivo pode ser intensificado usando seleção auxiliada por marcador da progênie de qualquer cruzamento. Fica entendido ainda, que qualquer cultivar comercial e não-comercial pode ser utilizada em um programa de cultivo. Fatores tais como, por exemplo, vigor da emergência, vigor vegetativo, tolerância ao estresse, resistência à doença ramificação, florescimento, semeadura, tamanho da semente, densidade da semente, verticalidade, e capacidade de debulho, etc. ditarão, geralmente, a escolha.

Para traços característicos altamente hereditários, uma escolha de plantas individuais superiores avaliada em um único local será eficaz, enquanto para traços característicos com baixa hereditariedade, a seleção deve se basear em valores médios obtidos de avaliações replicadas de famílias de plantas relacionadas. Métodos de seleção populares incluem, normalmente, seleções da ancestralidade, seleções de ancestralidade modificada, seleção de massa, e seleção recorrente. Em uma modalidade preferida um retrocruzamento e programa de cultivo recorrente está garantido.

A complexidade de escolha de influências hereditárias do método de cultivo. O cultivo retrocruzado pode ser usado para transferir um ou uns poucos genes favoráveis para um traço característico altamente hereditário para cultivares desejáveis. Esta abordagem foi usada extensivamente para cultivo de cultivares resistentes à doença. Várias técnicas de seleção recorrentes são usadas para melhorar quantitativamente traços característicos adquiridos controlados por numerosos genes. O uso de seleção recorrente em culturas autopolinizadoras depende a facilidade da polinização, da freqüência de híbridos de sucesso para cada polinização, e do número de descendência híbrida de cada cruzamento de sucesso.

Linhagens de cultivo podem ser testadas e comparadas a padrões apropriados em meios representativos de áreas alvo comerciais para duas ou mais gerações. As melhores linhagens são candidatos para novas cultivares comerciais, aquelas ainda deficientes em traços características podem ser usadas como origens para produzir novas populações para seleção adicional.

Um método de identificar uma planta superior é observar seu desempenho relativo a outras plantas experimentais e a uma cultivares padrão amplamente cultivadas. Caso uma simples observação seja inconcludente, observações replicáveis podem dar uma melhor estimativa de seu valor genético. Um criador pode selecionar e cruzar duas ou mais linhagens paternas, seguido por auto-repetição e seleção, produzindo muitas novas combinações genéticas.

O desenvolvimento de novas cultivares requer o desenvolvimento e seleção de variedades, do cruzamento dessas variedades e da seleção de cruzamentos híbridos superiores. A semente híbrida pode ser produzida por cruzamentos manuais entre origens férteis de macho selecionado ou pelo uso de sistemas de esterilidade do macho. Híbridos são selecionados para algumas características genéticas únicas como cor da vagem, cor da flor, produção de semente, cor da pubescência ou resistência a herbicida, que indicam que a semente é de fato um híbrido. Dados adicionais nas linhagens parentais, bem como no fenótipo do híbrido, influência da decisão

do cultivador, se continua com o cruzamento híbrido específico ou não.

O cultivo da linhagem e métodos de cultivo de seleção recorrente podem ser usados para desenvolver cultivares de populações de cultivo. Programas de cultivo combinam traços característicos desejáveis de duas ou mais cultivares ou de várias fontes com base ampla em agrupamento de cultivo dos quais a cultivar é desenvolvida por autodesenvolvimento e seleção dos fenótipos desejados. Novas cultivares podem ser avaliadas para se determinar quais possuem potencial comercial.

O cultivo da linhagem é usado normalmente para aperfeiçoamento de cultivos de autopolinização. Duas origens que possuem traços de complementaridade favoráveis, são cruzadas para produzir um F_1. Uma população F_2 é produzida por individualização de um ou vários F_1's. A seleção dos melhores indivíduos das melhores famílias é realizada. Teste replicado de famílias pode começar na geração F_4 a fim de melhorar a eficácia de seleção para traços com baixa hereditariedade. Em um estágio avançado de endogamia (isto é F_6 e F_7), as melhoras linhagens ou misturas de linhagens fenotipicamente similares são testadas para liberação potencial como novas cultivares.

O cultivo de retrocruzamento foi usado para transferir genes para um aspecto característico simplesmente herdado, altamente transmissível em uma cultivar homozigótica desejável ou linhagem consangüínea, que é a origem recorrente. A fonte da característica a ser transferida é denominada a origem doadora. Espera-se que a planta resultante tenha os atributos da origem recorrente (por exemplo, cultivar) e o desejável traço característico do doador de origem. Após o cruzamento inicial, indivíduos possuindo o fenótipo da origem doadora são selecionados e cruzados repetidamente (retrocruzamento) à origem recorrente. A origem resultante é esperada ter os atributos da origem recorrente (por exemplo, cultivar) e o traço desejável transferido da origem doadora.

O procedimento de geração de semente simples no sentido estrito, refere-se ao plantio de uma população de segregação, colheita de uma amostra de uma semente por planta, e uso da amostra de semente única

para plantar a próxima geração. Quando a população avançou de F_2 até o nível desejado de endogamia, as plantas das quais derivaram-se linhagens, irão cada qual ter aspectos caracteríticos para diferentes indivíduos F_2. O número de plantas em uma população declina a cada geração, devido à falha de algumas sementes em germinar ou algumas plantas em produzir pelo menos uma semente. Como resultado, nem todas as plantas F_2 originalmente amostradas na população serão representadas por uma progênie, quando o avanço da geração está completo.

Em um procedimento de múltiplas sementes, cultivadores, normalmente colhem uma ou mais vagens de cada planta em uma população e debulham-nas junto para formar um volume. Parte do volume é usado para plantar a próxima geração e parte é colocada em estoque. O procedimento foi referido como descendência de semente única modificada ou técnica de vagem em volume.

O procedimento de múltiplas sementes foi usado para economizar trabalho na colheita. É consideravelmente mais rápido debulhar vagens com uma máquina do que remove uma semente de cada vagem manualmente para o procedimento de semente simples. O procedimento de semente múltipla também torna possível o plantio do mesmo número de sementes de uma população de cada geração de cruzamento consangüíneo.

Descrições de outros métodos de cultivo que são comumente usados para diferentes características e colheitas podem ser encontradas em uma de várias literaturas de referência (por exemplo Fehr, *Principles of Cultivar Development* Vol. 1 pp 2-3 (1987)).

Uma planta transgênica da presente invenção também pode ser reproduzida usando apomixia. Apomixia é um método geneticamente controlado de reprodução em plantas onde o embrião é formado sem união de um ovo e um espermatozóide. Existem três tipos básicos de reprodução por apomixia: 1) aposporia onde o embrião se desenvolve de um ovo cromossomalmente não-reduzido em um saco embrionário derivado do núcleo, 2) diplosporia onde o embrião desenvolve-se de um ovo não-reduzido em um saco embrionário derivado da célula-mãe megaspórica e 3) embrionário ad-

ventício onde o embrião desenvolve-se diretamente de uma célula somática. Na maioria das formas de apomixia, a pseudogamia ou fertilização do núcleo polar para produzir endosperma é necessária para viabilidade da semente. Na aposporia uma cultivar nutriz pode ser usada como uma fonte de pólen para a formação do endosperma em sementes. A cultivar nutriz não afeta a genética da cultivar apomíctica aposporosa, visto que o ovo não-reduzido da cultivar desenvolve-se partenogeneticamente, mas torna possível a produção do endosperma. A apomixia é economicamente importante, especialmente em plantas transgênicas, porque ela ocasionam qualquer genótipo, não importando quanto heterozigótico seja o cultivo. Assim, com reprodução apomíctica, plantas transgênicas heterozigotas podem manter sua fidelidade genética por todos os ciclos de vida repetidos. Métodos para produção de plantas apomícticas são conhecidos da técnica. Ver Patente U.S. 5.811.636.

Outros Organismos

Um ácido nucléico da presente invenção pode ser introduzido em qualquer célula ou organismo como célula de mamífero, mamífero, célula de peixe, peixe, célula de pássaro, pássaro, célula de alga, algas, célula de fungo, fungos, ou célula de bactéria. Hospedeiros preferidos e transformantes incluem: células de fungo como *Aspergillus*, leveduras, mamíferos, particularmente bovinos e porcinos, insetos, bactérias e algas. Bactérias preferidas são *E. coli* e *Agrobacterium tumefaciens*.

Métodos para transformar tais células ou organismos são conhecidos da técnica (EP 0 238 023, Yelton et al., 1984: Malardier *et al.* 1989; Becker e Guarente, Ito *et al.*, 1983, Hinnen *et al.*, 1978 e Bennett e LaSure, 1991). Métodos de produzir proteínas de tais organismos são também conhecidos (Kudla et al., 1990, Jarai e Buxton, 1994, Verdier, 1990, MacKenzie et al., 1993, Hartl et al., 1994, Bergeron et al., 1994, Demolder et al., 1994, Craig, 1993, Gething e Sambrook, 1992, Puig e Gilbert, 1994, Wang e Tsou, 1993, Robinson et al. 1994, Enderlin e Ogrydziak, 1994, Fuller et al., 1989, Julius et al., 1984, e Julius et al., 1983).

Exemplos

Os exemplos a seguir são dados e não devem ser interpretados

de modo algum como limitadores do escopo da presente invenção.

Exemplo 1 - Geração de clones de um promotor 7Sα para soja (*Glycine mas* L)

Um promotor 7Sα é obtido de DNA genômico de soja (Asgrow A3244) usando um kit Universal Genome Walker (Clontech Laboratories, Inc., Palo Alto, California) e especificações do fabricante. O procedimento consiste em duas ampliações de PCR, usando um iniciador adaptador e um iniciador específico para gene para cada reação de ampliação. Para identificar uma região de mínima homologia que seria adequada como DNA matriz, as seqüências das regiões de codificações dos genes 7Sα e 7Sα' (GlcX (7Sα') e GlcA (7Sα) são alinhados e duas regiões de não-homologia são identificadas (Figura 1). Com base nas regiões identificadas de não-homologia são preparados os seguintes iniciadores específicos para gene:

Para a ampliação por PCR primária, Glc-A-GW primário" extremidade 3' (complementar à seqüência sublinhada na Figura 1)

5'-CTTCTGATGAGGTGGGCGTGGGAATGGGAA-3'(SEQ ID NO: 6)

Para a ampliação por PCR secundária, GlcA-GW-aninhado (com sítios de clonagem *Gbl*II e *Nco*I adicionados na extremidade 5' do iniciador): extremidade 3'(complementar à seqüência em negrito da Figura 1)

5'-CCATGGAGATCTATCTTGTTCTCATCCTCATCCTCATCCTCATC-3' [SEQ ID NO: 7]

O seguinte iniciador adaptador é usado nesta metodologia:

extremidade 5'primário: seqüência AP1:

5'-GTATACGACTCACTATAGGGC-3' [SEQ ID NO: 9]

AP5 (com sítios de clonagem *Hind*III e *Pst*I adicionados na extremidade 5' d iniciador)

extremidade 5'aninhada

5'-AAGCTTCTGCAGGGTCGACGGCCCGGGCTGGT-3' [SEQ ID NO: 5]

Deste procedimento, três clones são isolados e seqüenciados. Um alinhamento da seqüência ilustrou que todos os três clones representam o mesmo gene 7Sα com clone 8A como o mais extenso. Os clones são identificados como 8A (2,3 kb em comprimento, 7C (1,5 kb em comprimento) e 9C (1,3 kb em comprimento). Esses clones iniciais contém a seqüência

associada com, tanto o promotor e uma parte da região de codificação de um gene 7Sα (uma das subunidades de β-conglicinina). Para confirmar que o clone é um gene 7Sα, a seqüência 3'do clone 8A é alinhada com uma seqüência de cDNA publicada (GenBank Número de Acesso X17698). Um tal alinhamento confirma que o produto PCR é homólogo para a região a montante de 7Sα e não do gene 7Sα'.

Clones 7C, 8A e 9C são então subclonados para providenciar fragmentos menores contendo apenas o promotor e as regiões UTR 5' do gene 7Sα. Reações de ampliação por PCR como supradescrito são realizadas usando os clones 7C, 8A, e 9C como as matrizes. O iniciador AP5 é o iniciador adaptador e o iniciador específico para gene é:

5'-CCATGGAGATCTAAGGAGGTTGCAACGAGCGTGGCAT - 3' [SEQ ID NO: 10]

Este iniciador específico para gene é projetado para introduzir um sítio de restrição *Bgl*II e *Nco*I na extremidade 3' para facilitar clonagem subseqüente. Clones resultantes são seqüenciados usando metodologia padrão e subclonados em um novo vetor fácil pGEM-T (Promega, Madison, WI, US Patente U.S. 4.766.072) gerando o seguinte:

p-GEM-T 1A. inserto de 1,7 Kb derivado do clone original 8A,

p-GEM-T 2A. inserto de 1,3 Kb derivado do clone original 7C, e

P-GEM-T 3A. inserto de 1,1 kb derivado do clone original 9C.

Seqüências de nucleotídeo de pGEM-T 1A, p-GEM-T 2A e p-GEM-T 3A são usadas para prognosticar as correspondentes seqüências de aminoácido. Seqüências de aminoácido prognosticadas são então alinhadas à seqüência de tradução de proteína do produto genético 7Sα (GlcA) (GenBank Número de Acesso X17698 e SWISS-PROT número de Acesso P13916) para identificar um códon iniciador. Seqüências são alinhadas usando software GCG (Genetics Computer Group, Oxford Molecular Group, Inc.) função de alinhamento da seqüência. Os clones 1A, 2A, e 3A são então usados como matrizes para uma outra reação PCR usando o AP5 como o iniciador adaptador e a seguinte seqüência como o iniciador específico para gene:

5'-*CCATGGAGATCT*AGTATCTTAATTATTTATTAAGGTAT - 3' [SEQ ID NO: 8]

Adicionalmente, os sítios *Nco*I e *Bgl*II são adicionados à extremidade 5' do iniciador (mostrado em itálico). Os produtos PCR desta reação são purificados e clonados novamente no vetor fácil pGEM-T resultando em p-GEM-T 1T, inserto de 1724 pares de base derivado do clone 1A, p-GEM-T "novo" 2A, inserto de 1135 pares de base derivados do clone 2A e p-GEM-T "novo"3A, inserto de 918 pares de base derivado do clone 3A.A seqüência de p-GEM-T 1T é mostrada na Figura 2.

Exemplo 2

Este exemplo demonstra que as atividades do promotor das novas variantes 7Sα clonadas são confirmadas em cotilédones de soja transitoriamente transformadas.

Os três clones do exemplo 1 (1T, novo 2A, e novo 3A) são purificados por eletroforese de gel (kit QIAquick Cat. nº 28704, Qiagen Company, Valencia, California) e subclonados em pMON8677 (Figura 3), na posição originalmente ocupada pelo promotor e35S, a montante do gene repórter GUS. Os mapas de plasmídeo resultantes são mostrados na Figura 2a como pMON55548 (1T) Figura 3a como pMON55546 (novo 2A) e Figura 4 como pMON55547 (novo 3A). Dois desses plasmídeos pMON55546 e pMON55548 são usados em ensaios transitórios de cotilédones de soja.

Sementes e plantas de soja (Asgrow A3244) são colhidas 25-28 dias após florescência e tratadas osmoticamente durante a noite a 25°C no escuro em meio de GAMBORG (G5893, Sigma Company. St. Louis Missouri) suplementado com 50 mM de glutamina, 111 mM de maltose, 125 mM de rafinose, 125 mM de manitol e 3 g/l de ágar purificado, pH 5,6. Os resultantes 125 cotilédones são cortados em metades e bombardeados com DNA superespiralado purificado dos constructos do promotor 7sα de pMON55546 e pMON55548 usando tecnologia de pistola de partícula (Maliga et al., 1995, "Methods in Plant Molecular Biology, A Laboratory Course Manual, "Cold Spring Harbor Laboratory Press, página 47). Um constructo luciferase e335S acionado separado é incluído em uma relação molar 1:1 com cada um dos constructos do promotor como um controle de baixa expressão. Tecidos bombardeados são então incubados por 48 horas a 25°C.

Proteínas são extraídas de seis cotilédones de soja bombardeadas usando tampão de extração de 1 ml contendo fosfato de potássio 0,1 M (pH 7,8, 10 mM DTT, EDTA 1 mM, glicerol a 5% e inibidor proteinase (1 comprimido/50 ml, Roche Molecular Biochemicals, Indianapolis, IN), Uma alíquota de 100 µl do extrato de proteína é usada para ensaio Luciferase seguinte a um procedimento "Steady-Glo" pela Promega (Cat. N°. E2510 Promega Corporation Madison, WI). Uma alíquota de 50 µl do extrato de proteína é usada para um protocolo de ensaio GUS padrão com poucas modificações (Maliga et al., 1995, "Methods in Plant Molecular Biology, A Laboratory Course Manual", Cold Spring Harbor Laboratory Press, página 29). Cada amostra é ensaiada duas vezes e o valor médio é usado para análise de dados. A atividade GUS é normalizada usando atividade luciferase e os resultados indicam que todas as variantes dos promotores 7Sα são funcionais em tecidos de cotilédone de soja.

Exemplo 3

Produção de plantas de soja transgênicas contendo o promotor 7Sα

O cassete de expressão em pMON55548 (7Sα-1T-GUS-NOS-3') foi subclonado para gerar o vetor pMON55554, um vetor de transformação em agrobactéria capaz de demonstrar a eficácia de um promotor 7Sα em plantas de soja (Figura 5). Um marcador de seleção resistente ao glifosato (CP4) também está incluído em pMON55554. O vetor é introduzido na cepa bacteriana de *Agrobacterium tumefaciens* ABI e as células transformadas resultantes são usadas para infectar cotilédones de soja (Asgrow A3244). Como um controle, um promotor Arcelnin 5 truncado é usado no lugar do 7Sα-1T para gerar pMON55542 (Figura 6). O vetor é introduzido em uma cepa bacteriana de *Agrobacterium tumefaciens* ABI e as células transformadas resultantes são usadas para infectar cotilédones de soja (Asgrow A3244).

Plantas resistentes ao glifosato são selecionadas após regeneração de plantas de tecidos infectados por *A. tumefaciens*. Sementes maduras de plantas selecionadas são analisadas para atividade GUS. Para testar

atividade GUS, oito sementes de cada evento (linhagem) transgênica são semeadas individualmente. Cerca de 20 mg de tecido de semente triturado é extraído usando 200 µl de tampão de extração contendo fosfato de potássio 0,1 M (pH 7,8), DTT 10 mM, EDTA 1 mM, glicerol a 5%, e inibidor de proteinase (1 comprimido /50 ml, Roche Molecular Biochemicals, Indianapolis, IN). O teor de proteína do extrato é determinado usando Bio-Rad Protein Assay (Bio-Rad nº 61234A) e a atividade GUS é medida usando um protocolo de ensaio GUS padrão com poucas modificações (Maliga et al., 1995, "Methods in Plant Molecular Biology, A Laboratory Course Manual", Cold Spring Harbor Laboratory Press, página 29). A atividade GUS é normalizada contra a concentração de proteína. Cada amostra foi avaliada duas vezes e o valor médio foi usado para análise de dados.

Um evento (linhagem) é rejeitado caso nenhuma das oito sementes tenha atividade GUS detectável. Dentre cada um dos eventos mostrando atividade GUS, a semente com a maior atividade é selecionada. O ensaio de atividade GUS é repetido para sementes selecionadas e os resultados são resumidos na Figura 7. A comparação entre 28 eventos positivos de pMON55554 e 11 eventos positivos de pMON55542 mostram que o promotor 7Sα é, no mínimo tão robusto quanto o promotor Arcelin 5 truncado (Figura 7).

Exemplo 4

Produção de Arabidopsis transgênico contendo o promotor 7Sα ou promotor Arcelin 5.

Expressão de GUS acionada por, seja o promotor 7Sα1T em pMON55553 (Figura 8) ou o tArc5 m pMON55541 (Figura 9) é examinada em *Arabidopsis thaliana*. A expressão de pMON55553 (7Sα1T-GUS-NOS) GUS e pMON55541 (tArc5-GUS-NOS) GUS é examinada em plantas germinadas de semente de uma semana e em uma planta madura. Plantas e plantas germinadas de semente expressando o plasmídeo pMON55553 mostram que GUS está ausente em raízes ou folhas maduras quando o promotor 7Sα1T é usado. A atividade é vista em hipocótilos e cotilédones que é ocasionada pela proteína GUS residual sintetizada durante a embriogênese. Plantas

e plantas germinadas de semente expressando pMON55541 mostram que GUS está presente em raízes e folhas maduras, além dos sinais detectados em hipocótilos e cotilédones quando se usa o promotor tArc5. Os dados ilustram que o promotor 7Sα melhorou a especificidade para semente comparado com o promotor tArc5.

Referências

Ainley et al., *Plant Mol. Biol.*, 14:949,1990.

Ausubel *et al., Current Protocols in Molecular Biology,* John Wiley and Sons, Inc., 1995.

Bartley and Scolnik, *Plant Physiol.* 104:1469-1470, 1994.

Back *et al., Plant Mol. Biol.*, 17:9, 1991.

Bauer *et al., Gene*, 37:73, 1985.

Becker and Guarente, Em: Abelson and Simon (eds). *Guide to Yeast Genetics and Molecular Biology, Methods Enzymol.*, Vol. 194 págs. 182-187, Academic Press, Inc. New York.

Bennett and LaSure (eds.), *More Gene Manipulations in Fungi*, Academic Press, CA, 1991.

Bergeron *et al., TIBS,* 19:124-128, 1994.

Bustos *et al., EMBO J.*, 10:1469-1479, 1991.

Castresana *et al., EMBO J.*, 7:1929-1936, 1988.

Capecchi, *Cell* 22(2):479-488, 1980.

Cerda-Olmedo *et al., J. Mol. Biol,* 33:705-719, 1968.

Chau *et al., Science,* 244:174-181. 1989.

Christensen *et al., Plant Mol. Biol.* 18:675, 689, 1992.

Clapp, *Clin. Perinatol.*, 20(1):155-168, 1993.

Costa *et al., Methods Mol. Biol.*, 57:31-44, 1996.

Craik, *BioTechniques* 3:12-19, 1985.

Craig, *Science,* 260:1902-1903, 1993.

Curiel *et al., Hum. Gen Ther.*, 3(2):147-154, 1992.

Cyanobase, http://www.kazusa.or.jp/cyanobase.

Dellaporta *et al.*, Stadler Symposium, 11:263-282, 1988.

Demolder *et al., J. Biotechnology*, 32:179-189, 1994.

Deng and Nickloff, *Anal. Biochem,* 200:81, 1992.

D'Halluin *et al., Bio/Technology* 10:309-314, 1992.

Doyle *et al., J. Biol. Chem* 261;9228-9238, 1986.

Eglitis and Anderson, *Biotechniques, 6(7):608-614, 1988.*

Enderlin and Ogrydziak, *Yeast* 10:67-79, 1994.

Feinbaum *et al., Mol. Gen. Genet.* 226:449-456, 1991.

Fraley *et al., Proc., Natl. Acad. Sci. U.S.A.,* 80:4803, 1983.

Frits Eckstein *et al., Nucleic Acids Research,* 10:6487-6497, 1982.

Fromm *et al., Proc. Natl. Acad. Sci. USA*, 82(17):5824-5828, 1985.

Fromm *et al., Bio/Technology* 8:833, 1990.

Fuller *et al., Proc. Natl. Acad. Sci. (U.S.A),* 86:1434-1438, 1989.

Fynan *et al., Proc. Natl. Acad. Sci USA* 90(24):11478-11482, 1993.

Gething and Sambrook, *Nature*, 355:33-45, 1992.

Glick *et al.,* Methods in Plant Molecular Biology and Biotechnology, CRC Press, Boca Raton, Fla., 1993.

Graham and Van der Eb, *Virology* 54(2):536-539, 1973.

Greener *et al., Mol. Biotechnol.* 7:189-195, 1997.

Hartl *et al., TIBS,* 19:20-25, 1994.

Hershey and Stoner, *Plant Mol. Biol.* 17:679-690, 1991.

Hess, *Intern Rev. Cytol.,* 107:367,1987.

Hinchee *et al., Bio/Technology,* 6:915-922, 1988.

Hinnen *et al. Proc. Natl. Acad. Sci (U.S.A),* 75:1920, 1978.

Horsch *et al., Science*, 227:1229-1231, 1985.

Ikatu *et al., Bio/Technol.,* 8:241-242, 1990.

Jarai and Buxton, *Current Genetics,* 26:2238-2244, 1994.

Jefferson (I), *Plant Mol. Biol. Rep.,* 5:387-405, 1987.

Jefferson (II) *et al., EMBO J.* 6:3901-3907, 1987.

Johnston and Tang, *Methods Cell Biol.,* 43(A):353-365, 1994.

Jones *et al., Science*, 266:789-793, 1994.

Jones *et al., Mol. Gen. Genet.* 1987.

Julius *et al., Cell*, 32:839-852, 1983.

Julius *et al., Cell,* 37:1075-10-89, 1984.

Kares *et al., Plant Mol. Biol.* 15:905, 1990

Katz *et al., J. Gen. Microbiol.,* 129:2703-2714, 1983.

Keegstra, *Cell,* 56(2):247-253, 1989.

Keller *et al., EMBO L*, 8:1309-1314, 1989.

Knutzon *et al., Proc. Natl. Acad. Sci U.S.A.,* 89:2624-2628, 1992.

Kridl *et al., Seed Sci. Res.,* 1:209, 1991.

Kudla *et al., EMBO*, 9:1355-1364, 1990.

Kuhlemeier *et al., Seeds,* 1:471, 1989.

Kunkel, *Proc. Natl. Acad. Sci. U.S.A.* 82:488-492, 1985.

Kyte and Doolittle, *J. Mol. Biol.,* 157:105-132, 1982.

Lam and Chua, *J. Biol. Chem.* 266:17131-17135, 1990.

Lam and Chua *Science,* 248:471, 1991.

Lipman and Pearson, *Science,* 227:1435-1441, 1985.

Lindstrom, *et al., Developmental Genetics,* 11:160, 1990.

Lois *et al., Proc. Natl. Acad. Sci. U.S.A.* 95(5): 2105-2110, 1998.

Lu *et al., J. Exp. Med.* 178(6): 2089-2096, 1993.

Luo *et al., Plant Mol. Biol. Reporter,* 6:165, 1988.

MacKenzie *et al., Journal of Gen. Microbiol.* 139:2295-2307, 1993.

Malardier *et al., Gene,* 78:147-156, 1989.

Mandel *et al., Plant Mol. Biol.* 29:995-1004, 1995.

McElroy *et al., Seeds*, 2:163-171, 1990.

Nawrath *et al., Proc. Natl. Acad. Sci U.S.A.,* 91:12760-12764, 1994.

Needleman and Wunsch, *Journal of Molecular Biology, 48:443-453, 1970.*

Neuhaus et al., *Theor. Appl. Genet.*, 75:30, 1987.

Odell *et al., Nature,* 313:810, 1985.

Osuna *et al., Critical Reviews In Microbiology* 20:107-116, 1994.

Ou-Lee *et al., Proc. Natl. Acad. Sci U.S.A.,* 83:6815, 1986.

Ow *et al., Science*, 234:856-859, 1986.

Pearson and Lipman, *Proc. Natl. Acad. Sci (U.S.A.),* 85:2444-2448, 1988.

Pearson, "Rapid and Sensitive Sequence Comparison with FASTP and FASTA". En *Methods in Enzymology*, (R. Doolittle, ed.), 183, 63-98, Academic Press, San Diego, California, USA, 1990.

Pearson *Protein Science* 4:1145-1160, 1995.

Pena *et al., Nature* 325:274, 1987.

Poszkowski *et al., EMBO J.,* 3:2719, 1989.

Potrykus *et al., Ann. Rev. Plant Physiol. Plant Mol. Biol.* 42:205, 1991.

Potrykus *et al., Mol. Gen. Genet.* 199:183-188, 1985.

Puig and Gilbert, *J. Biol. Chem.* 269:7764-7771, 1994.

Pyee *et al., Plant J.,* 7:49-59, 1995.

Reynaerts *et al.,* "Selectable and Screenable Markers". En *Plant Molecular Biology Manual,* (Gelvin and Schilperoort, eds.), Kluwer, Dordrecht, 1988.

Richins *et al. Nucleic Acids. Res.* 20:8451, 1987.

Robinson *et al., Bio/Technology*, 1:381-384, 1994.

Rodriguez *et al.,* Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Butterworths, Boston, 1988.

Rogan and Bessman, *J. Bacteriol.,* 103:622-633, 1970.

Rogers *et al., Meth. In Enzymol,* 153-253-277, 1987.

Saint Guily *et al., Plant Physiol.,* 100(2):1069-1071, 1992.

Samac *et al., Seeds* 3:1063-1072, 1991.

Sambrook *et al., Molecular Cloning: A Laboratory Manual* Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 1989.

Sato *et al., J. DNA Res.* 7(1):31-63, 2000.

Schulze-Lefert *et al., EMBO J.*, 8:651, 1989.

Simpson, *Science,* 233:34, 1986.

Singer and Kusmierek *Ann. Rev. Biochem.*, 52:655-693, 1982.

Slighton and Beachy, *Planta*, 172:356, 1987

Smith *et al., Em Genetic Engineering: Principles and Methods,* (Setlow et al., ed.), Plenum Press. NY, 1-32, 1981.

Smith and Waterman, *Advances in Applied Mathematics,* 2:482-489, 1981.

Smith *et al., Nucleic Acids Research,* 11:2205-2220, 1983.

Smith *et al., Plant J.,* 11:83-92, 1997.

Stalker *et al., J. Biol. Chem.* 263:6310-6314, 1988.

Sutcliffe *et al., Proc. Natl. Acad. Sci. (U.S.A.)* 75:3737-3741, 1978.

Takahashi *et al., Proc. Natl. Acad. Sci. U.S.A.,* 951(17):9879-9884, 1998.

Thillet *et al., J. Biol. Chem.* 263:12500-12508, 1988.

Vandeyar *et al., Gene*, 65:129-133, 1988

Van Tunen *et al., EMBO J.* 7:1257, 1988.

Verdier, *Yeast* 6:271-297, 1990

Vodkin *et al., Cell* 34:1023, 1983.

Vogel *et al., J. Cell Biochem., (Suppl)* 13D:312, 1989.

Wagner *et al., Proc. Natl. Acad. Sci. USA,* 89(13):6099-6103, 1992.

Wang and Tsou, *FASEB Journal* 7:1515-1517, 1993.

Weissbach and Weissbach, *Methods for Plant Molecular Biology* (Eds.) Academic Press, Inc. San Diego, CA, 1988.

Weisshaar *et al., EMBO J.* 10:1777-1786, 1991.

Wenzler *et al., Plant Mol. Biol.* 12:41-50, 1989.

Williams *et al., Biotechnology* 10:540-543, 1992.

Wong and Neumann, *Biochem. Biophys. Res. Commun.* 107(2): 584-587, 1982.

Xia *et al., J. Gen. Microbiol.* 138:1309-1316, 1992.

Yang *et l. Proc. Natl. Acad. Sci. USA,* 87:4144-48, 1990.

Yamaguchi-Shinozaki *et al., Plant Mol. Biol.* 15:905, 1990.

Yelton *et al. Proc. Natl. Acad. Sci. (USA),* 81:1470-1474, 1984.

Zhou *et al., Methods in Enzymology* 101:433, 1983.

Zukowsky *et al., Proc. Natl. Acad. Sci. (USA)* 80:1101-1105, 1983.

Patentes US. Nºs 4.683.195, 4.683.202, 4.757.011. 4.769.061, 4.885.357, 4.886.878, 4.940.835, 4.957.748, 4.971.908, 5.057.419, 5.093.249, 5.100.679, 5.147.792, 5.215.912, 5.219.596, 5.270.200, 5.298.421, 5.304.481, 5.344.771, 5.362.865, 5.576.203, 5.508.468, 5.003.045, 5.955.329, 5.367.110, 5.858.749, 6.040.160, 5.610.041, 5.618.988, 6.107.060, 5.811.636, 4.766.072, 5.003.045, 5.576.203, 5.384.253, 5.443.974, 5.512.482, 5.530.186, 5.534.421, 5.552.306, 5.589.616, 5.508.468, 5.614.393, 5.663.068, 5.663.068, 5.633.436, 5.639.790, 5.654.402, 5.659.645, 5.689.050, 5.689.050, 5.689.052, 5.705.391, 5.760.206, 5.759.829, 5.789.220, 5.807.893. 5.850.024, 5.856.157, 5.866.789, 5.885.802, 5.885.801, 5.914.450, 5.942.660, 5.945.585, 5.952.544, 5.955.650, 5.965.727, 5.995.329, 5.990.384, 5.990.389, 5.936.069, 5.939.599. 6.005.076, 6.051.754, 6.075.183, 6.043.411, 6.100.091, 6.107.051, 6.110.891, 6.117.677, 6.194.167, 6.146.669, 6.147.279, 6.156.227, 6.172.106, 6.232.122.

Pedidos Europeus 0 154 204, 0 238 023, 0 255 378.

Pedidos PCT: WO 90/01869, WO 91/13993, WO 92/14822, WO 93/08682, WO 94/20628, WO 95/19442, WO 97/26366, WO 97/28247, WO 97/22703, WO 98/55601, WO 98/26064, WO 96/17064, WO 97/35023, WO 00/19839, WO 99/06581, WO 99/02656, WO 99/40209, WO 99/11800, WO 99/49058, WO 00/32757, WO 00/10380.

REIVINDICAÇÕES

1. Método de transformação de uma planta, caracterizado pelo fato de que compreende:

(a) fornecer uma molécula de ácido nucleico que compreende na direção 5' a 3' um promotor com uma sequência de ácido nucleico de SEQ ID NO: 12 operavelmente ligada a uma sequência de ácido nucleico estrutural, e

(b) transformação de uma planta com a molécula de ácido nucleico, como definida na etapa (a) para produzir uma planta transformada.

2. Método de acordo com a reivindicação 1, caracterizado pelo fato de que a referida planta transformada produz uma semente e a referida sequência de ácido nucleico é transcrita na semente e a referida semente é isolada da planta transformada.

3. Molécula de ácido nucleico recombinante, caracterizada pelo fato de que compreende uma sequência de ácido nucleico de SEQ ID NO: 12, em que a referida molécula de ácido nucleico é operavelmente ligada a uma sequência de ácido nucleico heteróloga.

4. Vetor, caracterizado pelo fato de que compreende a molécula de ácido nucleico recombinante, como definida na reivindicação 3.

5. Célula, caracterizada pelo fato de que compreende o vetor como definido na reivindicação 4, em que a célula é selecionada do grupo consistindo em uma célula bacteriana e uma célula fúngica.

6. Célula de acordo com a reivindicação 5, caracterizada pelo fato de que a referida célula bacteriana é *Agrobacterium tumefaciens* ou *E. coli.*

PI0212339

GAP de M13759 e X17698;
Topo: M13759 - mRNA de soja M13759 para a subinidade alfa beta-conglicinina
Fundo: mRNA de soja M17698 M13759 para a subinidade alfa beta-conglicinina

```
 951 TTAAGATATACTATGATGAGAGCGCGGTTCCCATTACTGTTGCTGGGAGT 1000
                ||||||||||| |||||||||||||||||||||||| |
   1 ...........ATGATGAGAGCACGGTTCCCATTACTGTTGCTGGGACT 38

1001 TGTTTTCCTAGCATCAGTTTCTGTCTCATTTGGCATTGCGTATTGGGAAA 1050
     ||||||||| || |||||||||||||||||||||||| || |||||||
  39 TGTTTTCCTGGCTTCAGTTTCTGTCTCATTTGGCATTGCTTACTGGGAAA 88

1051 AGCAGAACCCCAGTCACAACAAGTGCCTCCGAAGTTGCAATAGCGAGAAA 1100
     | |||||||||  |||||||||| |||| |||||||||||||||| |
  89 AAGAGAACCCCAAACACAACAAGTGTCTCCAGAGTTGCAATAGCGAGAGA 138

1101 GACTCCTACAGGAACCAAGCATGCCACGCTCGTTGCAACCTCCTTAAGGT 1150
     ||||| ||||||||||||||||||||||||||||||||||||||||||
 139 GACTCGTACAGGAACCAAGCATGCCACGCTCGTTGCAACCTCCTTAAGGT 188

1151 GGAGGAAGAAGAA...GAATGCGAAGAAGGTCAAATTCCACGACCACGAC 1197
     |||| ||||||||   |||   | |||  || |   ||||||||||||
 189 GGAGAAAGAAGAATGTGAAGAAGGTGAAATTCCACGACCACGACCACGAC 238

1198 CACAACACCCGGAGAGGGAACGTCAGCAACACGGTGAGAAGGAGGAAGAC 1247
     |||||||||||||||||||| |||||||| ||||||||||||||||||
 239 CACAACACCCGGAGAGGGAACCTCAGCAACCCGGTGAGAAGGAGGAAGAC 288

1248 GAAGGTGAGCAGCCACGTCCATTCCCATTCCCACGCCCACGCCAACCTCA 1297
     |||| |||||| |||||||| ||||||||||||||||   |||||||
 289 GAAGATGAGCAACCACGTCCAATCCCATTCCCACGCCCA...CAACCTCG 335

1298 TCAAGAGGAAGAGCACGAGCAGAAGGAGGAACACGAATGGCATCGCAAGG 1347
     |||||| ||||||||||||||| |||||||| ||||||| ||||||||
 336 TCAAGAAGAAGAGCACGAGCAGAGAGAGGAACAGGAATGGCCTCGCAAGG 385

1348 AGGAAAAACACGGAGGAAAGGGAAGTGAAGAGGAACAAGATGAACGTGAA 1397
     ||||||||| ||||| |||||||||||||||||| | || |
 386 AGGAAAAACGCGGAGAAAAGGGAAGTGAAGAGGAAGATGAGG..... 427

1398 CACCCACGCCCACACCAACCTCATCAAAAGGAAGAGGAAAAGCACGAATG 1447
                           | |||| |||||| |  | ||| |
 428 ..........................ATGAGGATGAGGAACAAGATGAACG 452

1448 GCAACACAAGCAGGAAAAGCACCAAGGAAAGGAAAGTGAAGAAGAAGAAG 1497
      |||  |   |         | || |    |  | |  |||||   | | |
 453 TCAATTC...CCATTCCCACGCCCACCTCATCAGAAGGAAGAGCGAAACG 499
```

FIG. 1. Alinhamento de duas seqüências de 75 alfa cDNA

Comprimento de seqüência P-Gm-7salfa-1T: 1724 5-jul-01 11:16:17 checagem: 7862

```
   1  AAGCTTCTGC AGGGTCGACG GCCCGGGCTG GTCTGTCTTT TCAATTTTTT
  51  TGGCCACATA TTATTCGGGT TCTGTGACCT TTTCAAAATG ACTGCTATTA
 101  CCTCCTGACC TTGCTATTAC ATCTTGACCA TCACTAGGCA TTTAAAAGTA
 151  TTAGTCATAG TCACATATTA CTACAAAGCG AGATTGATCT CTCTAATCTA
 201  ATGGGTGGGA AAACACTTAT AATATATGAT TCAAGAAAAG AAAGTAAATA
 251  AAACAATTTT ATTATATAAA GACTATTAGG ATAAAAAAAA CCTTAAAAGT
 301  GCTTGGATTT GGACCAGACT TGAATTTTAA TTTAATGATA TTATAATATG
 351  TGAATATATT TTTGAGACAA TTGTAAATTT CAGATAAAAA AATAATGTAA
 401  TTAAAATTGT AATAACTATA TCGTATACTT AATTAATTAT TAAATGTGAC
 451  AAAAAAGATA TACATCAAAA CTTAATGTTT CATGACTTTT TTTTTTAATG
 501  TGTGTCCTAA ATAGAAATTA AAAATAAAAA TTATTATATC CAAATGAAAA
 551  AAACATTTAA TACGTATTAT TTAAGAAATA ACAATATATT TATATTTTAA
 601  TATGTATTCA CATGTAAATT TAAAAACAAA AACAAAATTT CTCTTTTATT
 651  GATTAATTAA AATAATTTTA TAACTACATT TATTTTCTAT TATTATCAAT
 701  TTTCTTCTGT TTTTTTATTT GGCATATATA CCTAGACAAG TCAAAAAATG
 751  ACTATTCTTT AATAATCAAT CATTATTCTT ACATATTGTT CGAACTACGA
 801  GTTATGAAGT GTCAATTGCA CCTTAGTGTT TTGATAGGCC TCCATTTGCC
 851  GCTCATTAAT TAATTTGATA ACAGCCGTAC CGATCAATTA CTTATGCTTC
 901  TTCCATCGTA ATTATATGCA TGTCGGTTCT TTTAATCTTG GTACTCTCGA
 951  ATGCCACCAC AACACTGACT AGTCTCTTGG ATCATGAGAA AAAGCCAAAG
1001  AACAAAAAAG ACAACATAAA GAGTATCCTT TGCAAAAAAA TGTCTAAGTT
1051  CATAAAATAC AAACAAAAAC GCAATCACAC ACAGTGGACC CAAAAGCCAT
1101  GCACAACAAC ACGTACTCAC CAAGGTGCAA TCGTGCTGCC CAAAAACATT
1151  CACCAACTCA ATCCATGATG AGCCCACACA TTTGTTGTTT GTAACCAAAT
1201  CTCAAACGCG GTGTTCTCTT TGGAAAGCAA CCATATCAGC ATATCACACT
1251  ATCTAGTCTC TTGGATCATG CATGCGCAAC CAAAAGACAA CACATAAAGT
1301  ATCCTTTCGA AAGCAATGTC CAAGTCCATC AAATAAAATT GAGACAAAAT
1351  GCAACCTCAC CCCACTTCAC TATCCATGGC TGATCAAGAT CGCCGCGTCC
1401  ATGTAGGTCT AAATGCCATG CACATCAACA CGTACTCAAC ATGCAGCCCA
1451  AATTGCTCAC CATCGCTCAA CACATTTCTT GTTAATTTCT AAGTACACTG
1501  CCTATGCGAC TCTAACTCGA TCACAACCAT CTTCCGTCAC ATCAATTTTG
1551  TTCAATTCAA CACCCGTCAA CTTGCATGCC ACCCCATGCA TGCAAGTTAA
1601  CAAGAGCTAT ATCTCTTCTA TGACTATAAA TACCCGCAAT CTCGGTCCAG
1651  GTTTTCATCA TCGAGAACTA GTTCAATATC CTAGTATACC TTAATAAATA
1701  ATTTAAGATA CTAGATCTCC ATGG
```

FIG. 2. Seqüências de p-GEM-T 1T

PI0212339

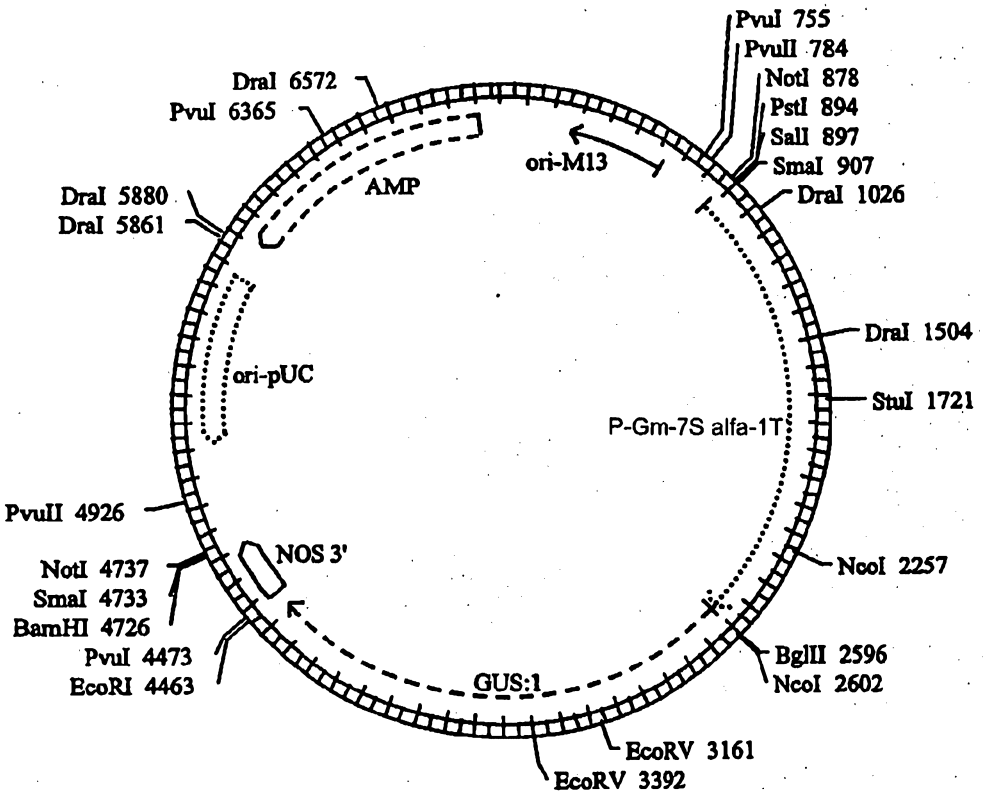


FIG. 2A. Um mapa de plasmídeo de pMON55548

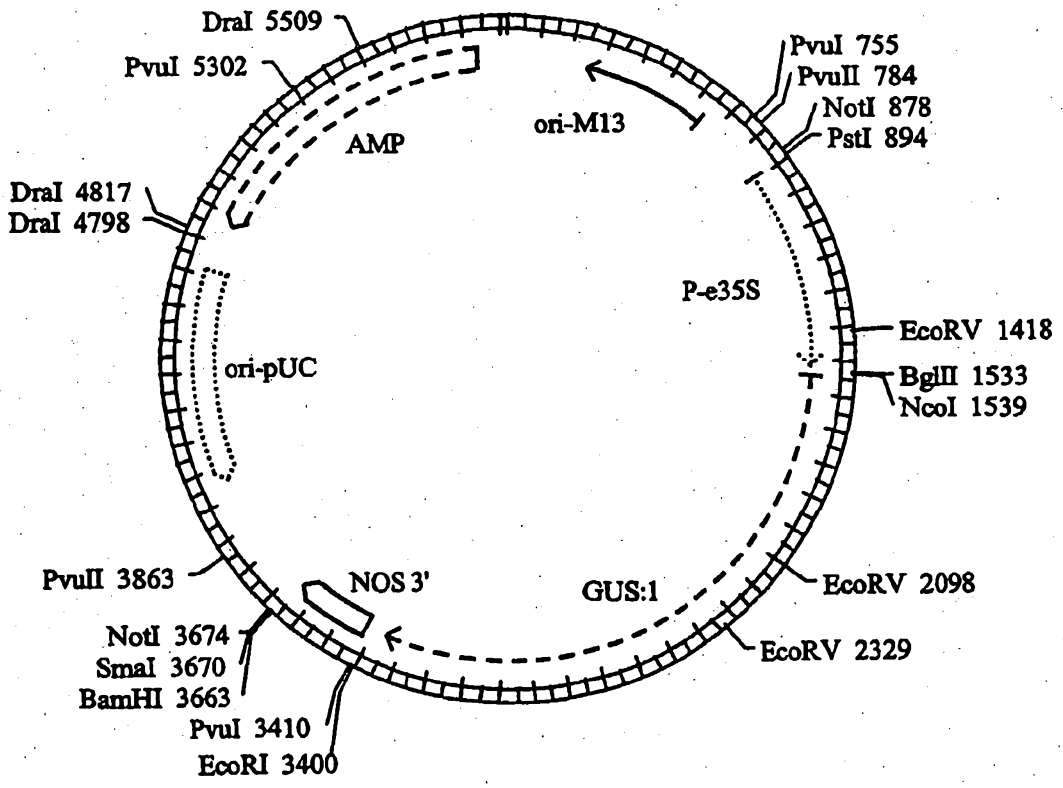


FIG. 3. Um mapa de plasmídeo de pMON8877

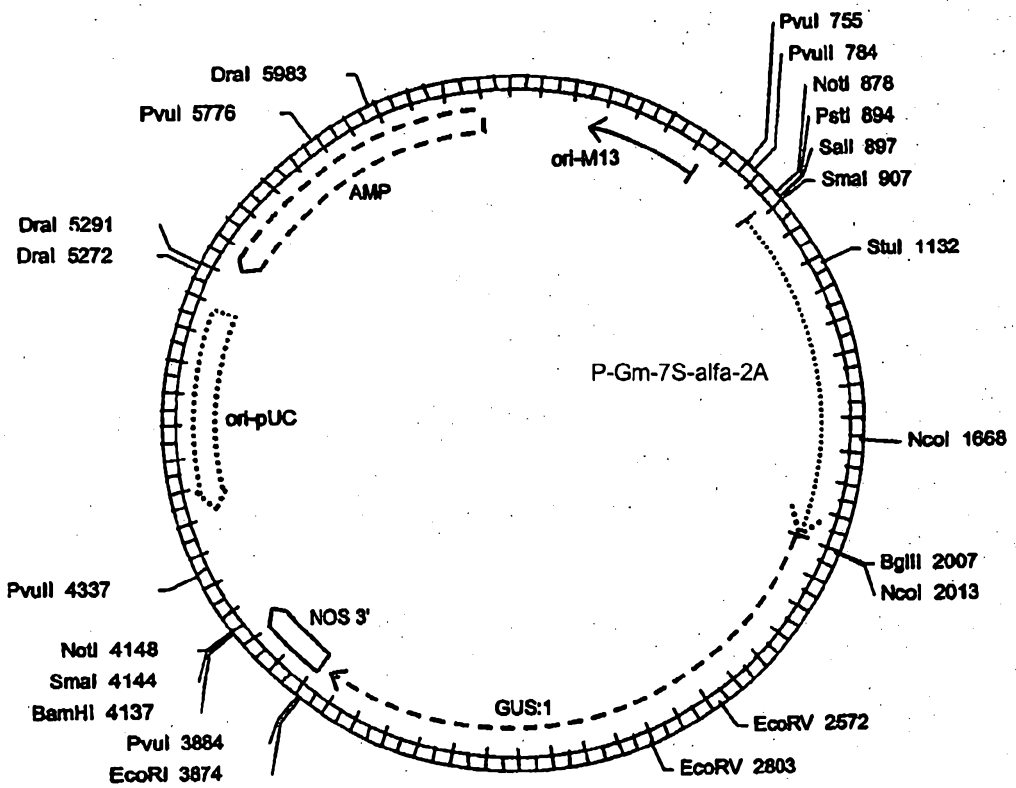


FIG. 3A. Um mapa de plasmídeo para vetor pMON55546

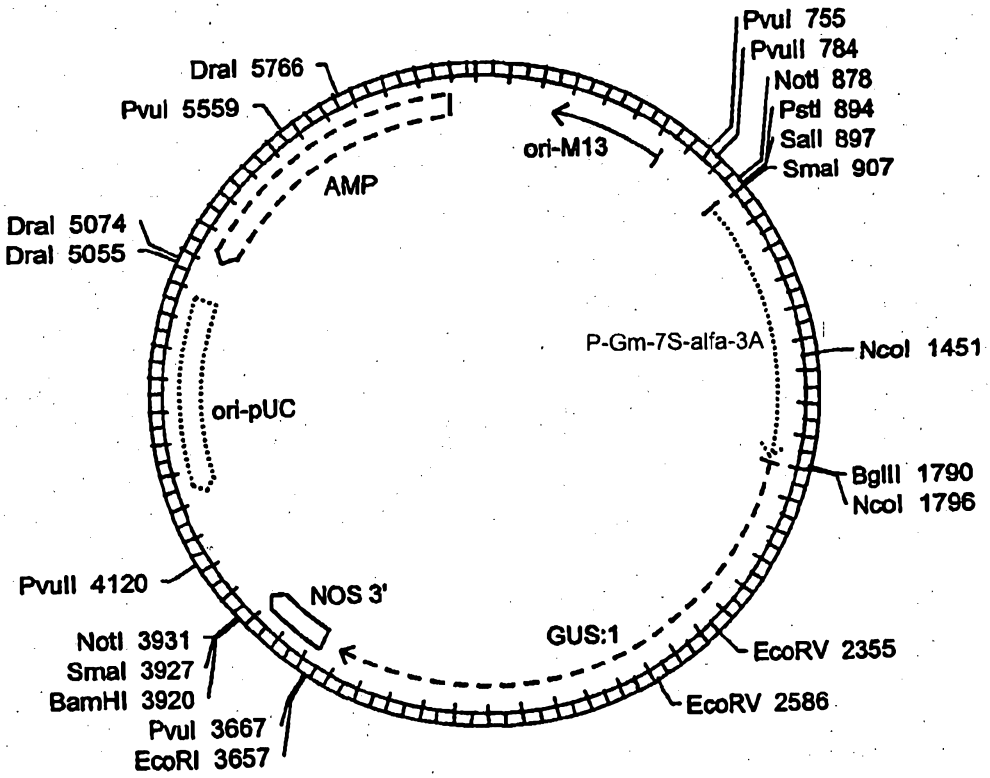


FIG. 4. Mapa de plasmídeo para vetor pMON55547

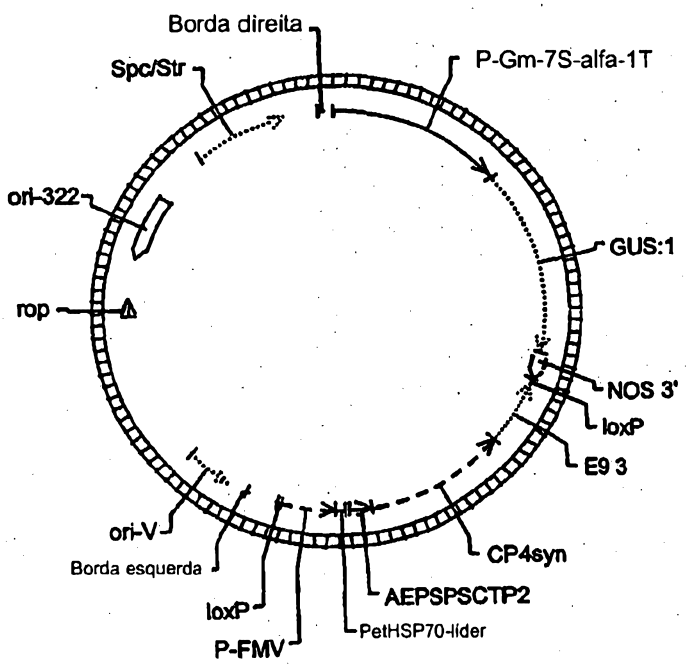


FIG. 5. Um mapa de plasmídeo para vetor pMON55554

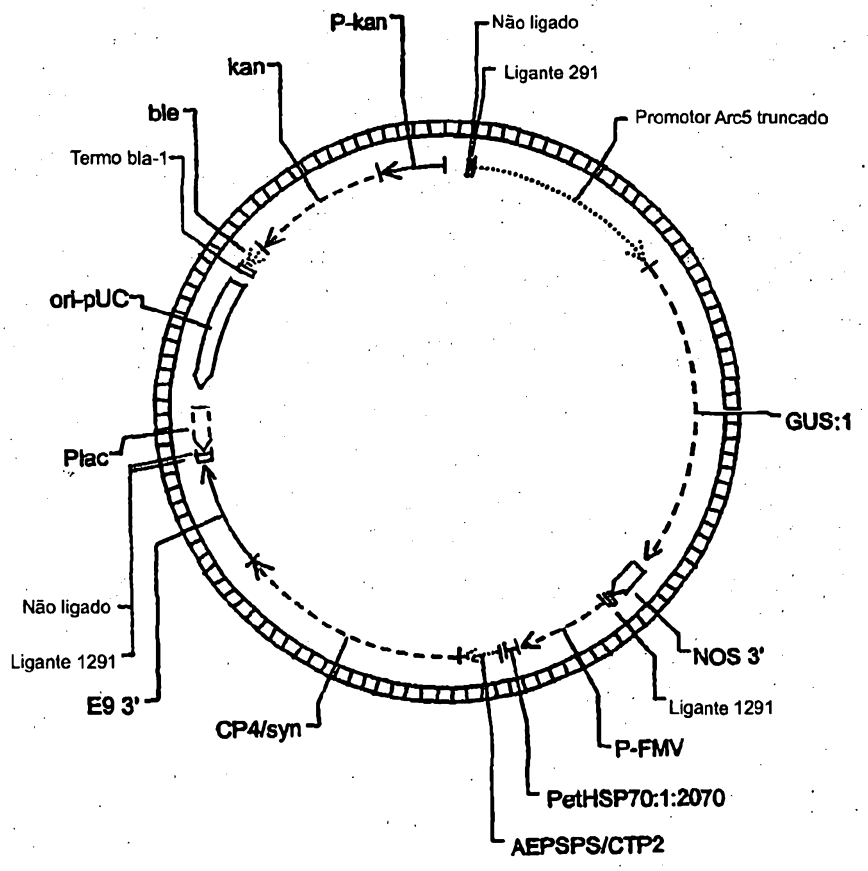


FIG. 6. Um mapa de plasmídeo de pMON55542

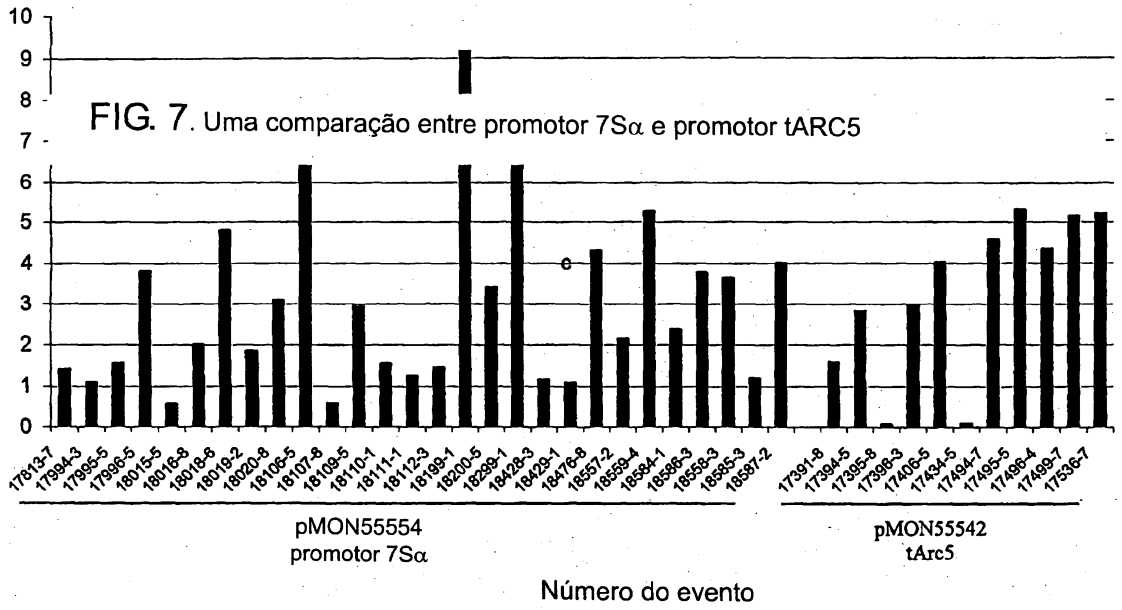


FIG. 7. Uma comparação entre promotor 7Sα e promotor tARC5

FIG. 7. Uma comparação entre promotor 7Sα e promotor tARC5

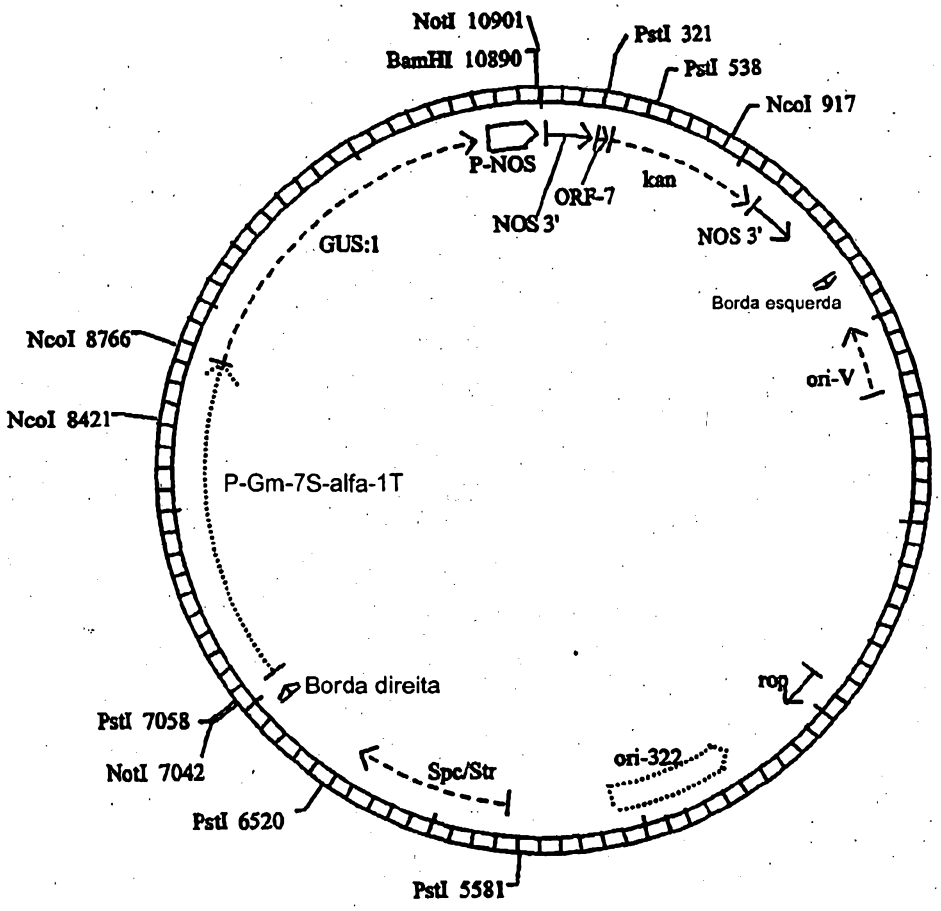


FIG. 8. Um mapa de plasmídeo pMON55553

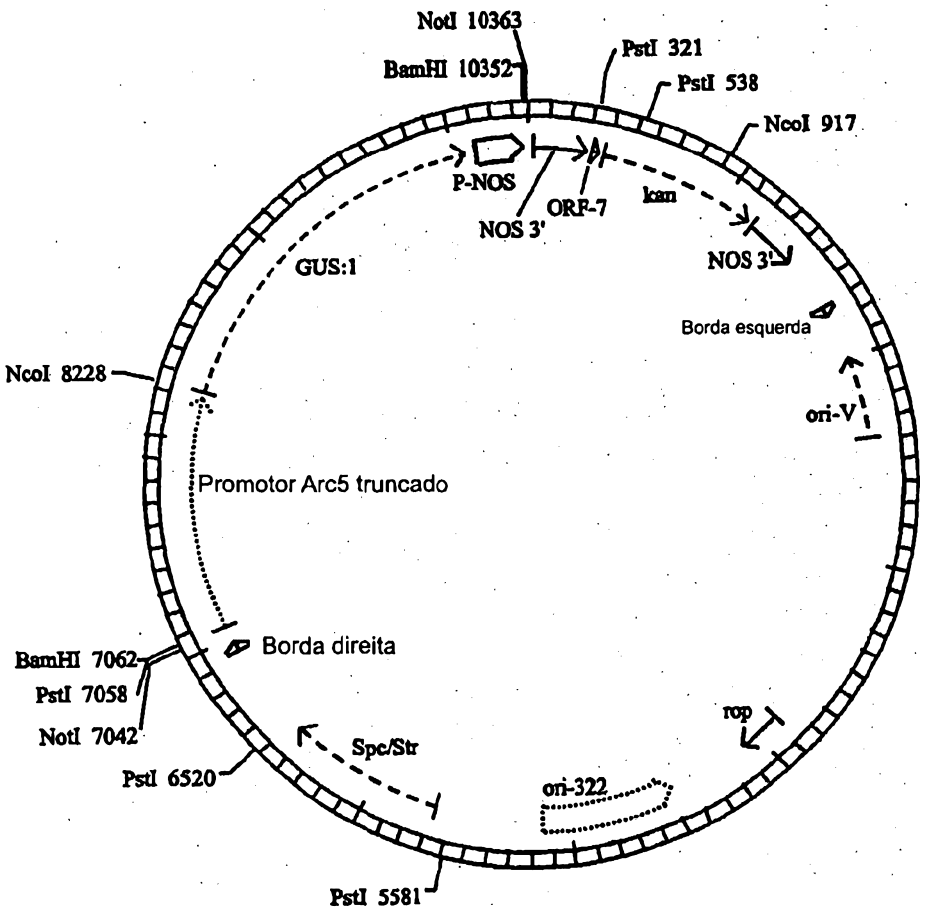


FIG. 9. Um mapa de plasmídeo de pMON55541

RESUMO

Patente de Invenção: **"MÉTODO DE TRANSFORMAÇÃO DE UMA PLANTA, MOLÉCULA DE ÁCIDO NUCLEICO RECOMBINANTE, VETOR E CÉLULA".**

Esta invenção refere-se promotores capazes de transcrever sequencias de ácido nucleico heterólogas em sementes, e métodos de modificar e usar os mesmos.