



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

(11) PI 1011495-5 B1



(22) Data do Depósito: 30/06/2010

(45) Data de Concessão: 14/05/2019

(54) Título: MÉTODO PARA PRODUÇÃO DE UMA PLANTA TRANSGÊNICA TENDO TOLERÂNCIA AUMENTADA AO ESTRESSE POR CALOR EM RELAÇÃO A UM CONTROLE DO TIPO SELVAGEM

(51) Int.Cl.: C12N 15/82; A01H 5/00; A01H 5/10; C12N 15/00; C12N 15/29; (...).

(30) Prioridade Unionista: 30/06/2009 US 61/221.813.

(73) Titular(es): PERFORMANCE PLANTS, INC..

(72) Inventor(es): JIANGXIN WAN; YAFAN HUANG.

(86) Pedido PCT: PCT IB2010001766 de 30/06/2010

(87) Publicação PCT: WO 2011/001286 de 06/01/2011

(85) Data do Início da Fase Nacional: 29/12/2011

(57) Resumo: MÉTODO PARA PRODUÇÃO DE UMA PLANTA TRANSGÊNICA TENDO TOLERÂNCIA AUMENTADA AO ESTRESSE POR CALOR EM RELAÇÃO A UM CONTROLE DO TIPO SELVAGEM A invenção se refere a métodos para produção de um fenótipo desejado em uma planta pela manipulação da expressão do gene na planta. O método diz respeito aos meios para aumentar o nível de expressão de uma expressão ou atividade de gene bHLH do subgrupo 1b, em que um fenótipo desejado, tal como tolerância ao calor aumentada em relação a uma planta de controle do tipo selvagem após estresse por calor resulta em aborto de flor reduzido e rendimento aumentado. São incluídas plantas produzidas pelos métodos referidos. A invenção também se refere às sequências de ácidos nucleicos e construtos úteis para tais métodos e métodos para geração e isolamento de plantas tendo expressão aumentada de uma expressão ou atividade de bHLH do subgrupo 1b.

“MÉTODO PARA PRODUÇÃO DE UMA PLANTA TRANSGÊNICA TENDO TOLERÂNCIA AUMENTADA AO ESTRESSE POR CALOR EM RELAÇÃO A UM CONTROLE DO TIPO SELVAGEM”

REFERÊNCIA A PEDIDOS RELACIONADOS

[001]Este Pedido reivindica a prioridade para o Pedido Provisório US No. de série 61/221.813, depositado em 30 de junho de 2009, os conteúdos dos quais sendo incorporados neste documento por referência em sua totalidade.

CAMPO DA INVENÇÃO

[002]A invenção está no campo da biologia molecular de plantas e se refere a plantas transgênicas tendo novos fenótipos, métodos de produção de tais plantas e polinucleotídeos e polipeptídeos úteis em tais métodos. Mais especificamente, a invenção se refere à expressão de um regulador transcricional e plantas transgênicas tendo atividade aumentada do regulador transcricional para produzir uma planta tendo um fenótipo benéfico.

FUNDAMENTOS DA INVENÇÃO

[003]Os estresses ambientais são responsáveis pelo aumento de rendimento significativo em culturas agrícolas. A relação entre a variação e produção climática de milho e soja nos Estados Unidos para o período de 1982-1998 foi estudada (Lobell e Asner, 2003) e foi verificado que até mesmo mudanças graduais de temperatura têm um impacto mensurável no rendimento das culturas. Para o milho e soja foi estimado que o rendimento é reduzido por 17% por grau na medida em que a temperatura se eleva acima da ótima para a estação. Tanto monocotiledôneas quanto dicotiledôneas são sensíveis ao estresse calórico, particularmente, durante o desenvolvimento de semente e floração que se traduz em um impacto significativo no rendimento de sementes (Young et al., 2004; Sato et al., 2002; Angadi et al., 2000; Carlson, 1990; Wahid, A., et al. 2007). No campo, o estresse de calor é muitas vezes acompanhado com outros estresses ambientais como a seca que adiciona ainda

ônus à produtividade da planta. O estresse por calor pode ter uma infinidade de efeitos celulares em plantas, tais como, fluidez e permeabilidade da membrana alterada, agregação de proteínas e desnaturação de proteínas, danos celulares resultantes podem resultar em mudanças prejudiciais no crescimento e desenvolvimento das plantas que impacta a capacidade de sobreviver. Foi sugerido que as plantas possuem uma capacidade inerente para termotolerância basal e adquirida e que um mecanismo de resposta ao calor comum está presente em diversas espécies de plantas (Kapoor et al., 1990; Vierling, 1991; Flahaut et al., 1996; Burke et al., 2000; Hong and Vierling, 2000; Massie et al., 2003; Larkindale et al., 2005). Uma série de estudos foi conduzida para identificar e caracterizar genes e vias que estão envolvidos na termotolerância da planta. Por exemplo, os fatores de transcrição de choque térmico (HSF) e proteínas de choque térmico (HSP) têm recebido muita atenção para elucidar os papéis e efeitos desses genes em resposta ao estresse térmico assim como hormônios de crescimento para plantas, tais como o ácido abscísico e etileno. Não está claro como as plantas sentem o calor, no entanto, é evidente que múltiplas vias de sinalização e componentes celulares estão envolvidos (Larkindale et al. 2005, *Physiol Plant.* 138:882-897) e que matéria cruzada de vias de sinalização existe entre os estresses ambientais e nutricionais tais como, estresse de choque térmico, estresse de água/seca, estresse causado pelo frio, estresse oxidativo e estresse de metais pesados.

[004]Os fatores de transcrição são proteínas de ligação ao DNA que interagem como sequências intensificadoras ou promotoras específicas e alteram a expressão genética do gene associado. Quando a sequência específica que se liga ao fator de transcrição está associada a um conjunto de genes, vias inteiras podem ser reguladas coordenadamente com vários genes componentes sendo simultaneamente suprarreguladas ou infrarreguladas. Um fator de transcrição pode alterar coordenadamente um conjunto de genes em resposta a um estímulo, tal como um estresse

ambiental, estado nutricional ou ataque de patógenos, por exemplo, ou pode ser um componente de uma via de sinalização, tal como uma via de sinalização hormonal, por exemplo. Os fatores de transcrição possuem uma estrutura modular e são classificados principalmente com base no domínio de ligação ao DNA. Alguns reguladores de transcrição são participantes em cascatas de sinais múltiplas. A via e os genes a jusante regulados podem variar dependendo da presença ou ausência de outros componentes reguladores e de via. Os reguladores de transcrição interagem como uma rede em que o resultado depende de uma multiplicidade de fatores que interagem.

[005]A ativação transcricional é mediada principalmente através de fatores de transcrição que interagem com elementos intensificadores e promotores. A ligação de fatores de transcrição de para tais elementos de DNA constitui uma etapa crucial na iniciação da transcrição. Cada fator de transcrição se liga a sua sequência de ligação específica em um promotor e ativa a expressão da região de codificação ligada através de interações com coativadores e/ou proteínas que fazem parte do complexo de transcrição.

[006]O fator de transcrição bHLH39 é um membro de uma das maiores famílias de fatores de transcrição em *Arabidopsis thaliana* composto de tanto quanto 162 proteínas (Heim et al., 2003, Toledo-Ortiz et al. 2003, Bailey et al., 2003). Esta família de proteínas é distinguida de outros fatores de transcrição por seu domínio hélice-alça-hélice básico (bHLH). Estudos têm demonstrado que a região básica é crítica para a ligação ao DNA, embora a região hélice-alça-hélice hidrofóbica seja necessária para a formação de homodímero e heterodímero. As proteínas bHLH pode ter vários parceiros de ligação e, conseqüentemente, modulam o nível de expressão um subconjunto diferente de genes dependendo de seu parceiro atual (Zhang et al., 2003). Estudos funcionais implicaram as proteínas bHLH em uma variedade de processos celulares tal como a determinação do destino de células epidérmicas da raiz

(Tominaga, R., et al, 2007), produção de antocianina (Ramsay et al., 2003), e sinalização de luz (Martinez-Garica et al., 2000) e absorção de ferro (Ling et al., 2002).

[007]A região básica de proteínas bHLH, consistindo em aproximadamente 15-17 resíduos, é responsável pela ligação a elementos cis em promotores de genes alvos. Setenta e cinco por cento de todas as proteínas bHLH *Arabidopsis* estão previstos para se ligar ao motivo central conhecido como E-box (5'CANNTG-3') (Toledo-Ortiz et al., 2003). A especificidade para o E-box pode ser prevista de acordo com a presença de dois resíduos críticos, ácido glutâmico-85 e arginina-88. O bHLH39, e seus três homólogos de *Arabidopsis* mais próximos, bHLH38, bHLH100 e bHLH101 todos contêm a arginina e ácido glutâmico crítico e, portanto, estão previstos para se ligar a esse motivo. O tipo de ligação E-box pode ser ainda dividido de acordo com a preferência de ligação de dois nucleotídeos centrais. O bHLH39 está previsto para se ligar ao motivo G-box CACGTG de acordo com a presença de dois resíduos, arginina-89 e histadina-81. A arginina-89 e histadina-81 são responsáveis por contatar os dois nucleotídeos "CG" centrais no G-box e estabilizar a interação. Esses dois resíduos críticos também são conservados em outros três homólogos de *Arabidopsis* bHLH39. Os quatro resíduos críticos são conservados em todos os 95 homólogos, exceto em dois casos, o homólogo de *Cicer arietinum* tem uma arginina no lugar do histadina-81, e um homólogo de *Vitis vinifera* tem uma valina substituída pelo ácido glutâmico-85 (Fig. 1).

[008]Evidências adicionais suportam a especificidade de ligação de bHLH39 vêm de relatórios do ortólogo de arroz OsIRO2 que se liga ao 5'-CACGTGG-3' (Ogo et al., 2006). Como é o caso com OsIRO2, os resíduos externos do motivo de ligação central mais provavelmente também afetam especificidade de ligação.

[009]As proteínas bHLH se ligam como dímeros a seus alvos de DNA, e a especificidade do parceiro de ligação é codificada no domínio hélice-alça-hélice. Evidências de estrutura cristalina de um complexo de Max-DNA humano intacto

mostrou resíduo de leucina 99 como sendo crítico para a formação de dímero (Brownlie et al, 1997). Este resíduo é conservado em todos os homólogos bHLH39, exceto um homólogo de *Vitis vinifera*. Outros resíduos que mostram mais de 95% de conservação em todos os homólogos de bHLH39 incluíram Arginina-100, Tirosina-124, Isoleucina-125, e Prolina-126. A exigência para uma prolina na posição 126 parece ser específica para bHLH39, bHLH38, bHLH100 e bHLH101, sugerindo que este resíduo é importante para facilitar a especificidade dimerização para este grupo.

[0010]De acordo com semelhanças estruturais externas do domínio de ligação ao DNA, bHLH39 é semelhante a 10 outras proteínas bHLH, que juntas formam o subgrupo 1b (Heim et al., 2003). O bHLH39 e seu homólogo de *Arabidopsis* mais próximo, bHLH38, mostram similaridade de 79% e estão localizados em tandem no genoma, o que sugere uma duplicação evolutiva recente. Outros membros do sub-group1b incluem bHLH100 e bHLH101. O percentual de similaridade entre AtbHLH39 e três outros membros do subgrupo 1b AtbHLH (de acordo com o alinhamento Clustal W) são mostrados abaixo na Tabela A.

Tabela A: Percentual de similaridade entre AtbHLH39, AtbHLH38, AtbHLH100 e AtbHLH101

	At bHLH39	At bHLH38	At- bHLH100	At- bHLH101
At- bHLH39	100	79	60	31
At- bHLH38		100	57	32
At- bHLH100			100	29
At- bHLH101				100

[0011] Tanto bHLH39 quanto bHLH38 foram identificados em uma busca de alvos a jusante de uma ligação de DNA com um fator de transcrição de dedo (Dof), fator de ligação ao elemento-ocs (OBP3), que é induzível pelo ácido salicílico. Neste caso bHLH39 foi nomeado ORG3 e bHLH38 foi nomeado ORG2. O bHLH39 e bHLH38 mostraram ter expressão co-regulada e intensificada em linhagens de superexpressão de OBP3 e ser infrarreguladas em OBF3 linhagens de função perda (Kang et al., 2003). Posteriormente, bHLH39 e bHLH38 mostraram-se responsivas ao estresse mediada por deficiência de ferro, juntamente com seus dois homólogos mais próximos no subgrupo 1b de bHLH, bHLH100, e bHLH101. O papel de expressão de bHLH39 e bHLH38 em relação à deficiência de ferro foi estudado. Um regulador transcricional FIT (AtbHLH29) mostrou interagir com bHLH28 ou bHLH39 para regular transcionalmente os genes de absorção do ferro FRO2 e IRT1. A superexpressão de qualquer dentre AtbHLH38 ou AtbHLH39 com FIT foi verificada como alterando o padrão de expressão de FRO2 e IRT1 para ativação constitutiva e resultam na tolerância de deficiência de ferro (Yuan, et al., 2008). A FRO2, uma redutase de quelato férrico, é responsável pelo aumento de ferro no solo, um processo necessário para aumentar a solubilidade e biodisponibilidade do ferro. O ferro é posteriormente transportado através da membrana pelo transportador de ferro, IRT1. A expressão ectópica de qualquer dentre a proteína AtbHLH38 ou AtbHLH39 sob o controle do promotor 35S em plantas de tabaco leva à síntese e excreção de riboflavina, um mecanismo de defesa conhecido para condições deficientes de ferro (Vorwieger et al., 2007).

[0012] Em uma via reguladora de transcrição é comum um primeiro regulador de transcrição regular um segundo regulador de transcrição, dependendo da rede de fatores de interação, um único regulador transcricional pode desempenhar um papel em uma variedade de vias, resultando em uma variedade de resultados fisiológicos

ou bioquímicos. Tal relação tem sido demonstrada entre alguns MYB e algumas proteínas bHLH (Ramsay e Glover, 2005).

[0013]A família MYB de fatores de transcrição é composta de pelo menos 198 genes (Yanhui et al. 2006) e foi proposta como tendo funções reguladoras em um amplo arranjo de processos que variando desde o crescimento e desenvolvimento para respostas de defesa. As proteínas MYB de plantas são classificadas com base na presença e no número de repetições de MYB imperfeitas cada uma composta por cerca de 52 aminoácidos. O domínio MYB forma uma conformação de hélice-volta-hélice e representa o domínio de ligação ao DNA. Os três grandes grupos de proteínas MYB foram classificados como proteínas relacionadas a R1R2R3-MYB, R2R3-MYB e MYB.

[0014]A família de R2R3-MYB de proteínas em *Arabidopsis* consiste em 125 proteínas e é caracterizada por ter um domínio de ligação ao DNA R2R3 em seus N-terminais (Kranz et al. 1998, e Stracke et al., 2001). Estes genes estão envolvidos em uma série de processos biológicos, incluindo ações hormonais de mediação, metabolismo secundário (Paz-Ares et al., 1987), controle de desenvolvimento de morfogênese celular (Oppenheimer et al., 1991), meristema, floral e da semente (Kirik et al, 1998, Schmitz et al, 2002) e resposta a vários fatores ambientais (Kranz et al, 1998; Jin e Martin, 1999; Meissner et al, 1999).

RESUMO DA INVENÇÃO

[0015]Esta invenção é baseada na verificação de que a superexpressão de um regulador transcricional de gene do bHLH de subgrupo 1b resulta em uma planta com um fenótipo alterado tal como, por exemplo, tolerância ao estresse por calor aumentada, aborto de flores reduzido durante o estresse de calor e rendimento aumentado em relação a uma planta do tipo selvagem.

[0016]Mais especificamente, a invenção refere-se à identificação de um bHLH39 ou bHLH101 como reguladores transcricionais que quando superexpressos

irá produzir plantas tendo um fenótipo de tolerância ao estresse por calor.

[0017]Em um aspecto a invenção fornece um método para produção de uma planta transgênica, pela transformação de uma planta, uma cultura de tecido de planta, ou uma célula de planta com um vetor contendo um construto de ácido nucleico que aumenta a expressão ou atividade de um gene do bHLH de subgrupo 1b para obter uma planta, cultura de tecido ou uma célula de planta com atividade ou expressão do bHLH de subgrupo 1b aumentada e crescimento de plantas ou regenerar uma planta a partir da cultura de tecidos de plantas ou células de plantas, em que uma planta tendo tolerância ao estresse por calor aumentada em relação a uma planta do tipo selvagem é produzida.

[0018]Assim, a presente invenção fornece um método para produção de uma planta tendo uma propriedade melhorada, em que o método inclui o aumento da expressão ou atividade de um gene do bHLH de subgrupo 1b endógeno, em que uma planta é produzida tendo um fenótipo vantajoso ou propriedade melhorada. Em uma modalidade particular, a presente invenção fornece um método para produzir plantas tendo tolerância aumentada ao estresse por calor em relação a uma planta do tipo selvagem, em que o método inclui a geração de plantas transgênicas e modificação do genoma de plantas usando os métodos descritos neste documento.

[0019]A tolerância ao estresse por calor se refere à capacidade de uma planta para suportar os efeitos debilitantes do calor que reduz o rendimento de uma planta tipo selvagem e tem melhor desempenho que uma planta do tipo selvagem. Como usado neste documento, o termo "tolerância ao estresse por calor aumentada" se refere a uma planta com tolerância ao estresse por calor maior em comparação com a tolerância ao estresse por calor de uma planta do tipo selvagem correspondente. Por exemplo, uma planta tendo tolerância ao estresse por calor aumentada, em comparação com uma planta do tipo selvagem pode ter 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60% a 70%, 75 % ou maior tolerância ao estresse por calor do que a

planta do tipo selvagem correspondente.

[0020]Os métodos da invenção envolvem o aumento da atividade de um gene bHLH do subgrupo 1b heterólogo ou endógeno por superexpressão ou modificação do promotor, em que uma planta é produzida tendo um fenótipo vantajoso ou propriedade melhorada, tal como a tolerância ao estresse por calor aumentada em relação a uma planta do tipo selvagem. Em um aspecto, a invenção fornece um método para produção de uma planta com tolerância ao estresse por calor aumentada em relação a uma planta do tipo selvagem, pela introdução em uma célula de planta de um construto de ácidos nucleicos que aumenta a expressão ou atividade do gene ou proteína bHLH de subgrupo 1b. Por exemplo, uma planta tendo tolerância ao estresse por calor aumentada em relação a uma planta do tipo selvagem é produzida por a) fornecer um construto de ácidos nucleicos contendo um promotor operavelmente ligado a um construto de ácido nucleico que expressa a atividade de bHLH de subgrupo 1b; b) inserir um construto de ácido nucleico em um vetor; c) transformar uma planta, cultura de tecido, ou uma célula de planta com o vetor para obter uma planta, cultura de tecido, ou uma célula da planta com a atividade de bHLH de subgrupo 1b aumentada; d) crescer a planta ou regenerar uma planta a partir da cultura de tecido ou célula de planta, em que uma planta com tolerância ao estresse por calor aumentada em relação a uma planta do tipo selvagem é produzida. O construto inclui um promotor tal como um promotor constitutivo, um promotor específico de tecido ou um promotor induzível. Preferencialmente, o promotor específico de tecido é um promotor da raiz. Um promotor induzível preferível é um promotor induzível por calor ou secura.

[0021]Uma planta transgênica é fornecida pela invenção tendo um fenótipo vantajoso ou propriedade melhorada, tal como tolerância aumentada ao estresse por calor, produzida pelos métodos descritos neste documento.

[0022]Em outro aspecto a invenção, uma planta é fornecida tendo uma mu-

tação de ocorrência não natural em um gene bHLH, em que a planta tem expressão ou atividade de bHLH de subgrupo 1b aumentada e a planta tem tolerância ao estresse por calor aumentada em relação a uma planta do tipo selvagem. Expressão ou atividade de bHLH aumentada refere-se a um aumento de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60 ou 75 vezes ou maior, no nível de DNA, RNA ou proteína de um gene bHLH em relação a bHLH do tipo selvagem, ou um aumento de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60 ou 75 vezes da atividade de bHLH em relação à atividade de bHLH do tipo selvagem.

[0023]Um aspecto adicional é uma planta tendo um gene bHLH de subgrupo 1b endógeno que tem uma sequência promotora alterada operavelmente associada com ele. A inserção de elementos intensificadores ou mutações promotoras que resultam na expressão do gene aumentada são previstas pela presente invenção.

[0024]A invenção fornece adicionalmente uma semente transgênica produzida pela planta(s) transgênica da invenção, em que a semente produz planta tendo um fenótipo vantajoso ou propriedade melhorada, tal como, por exemplo, tolerância ao estresse por calor aumentada em relação a uma planta do tipo selvagem.

[0025]Em outra modalidade, a invenção fornece ácidos nucleicos para a expressão de ácidos nucleicos em uma célula de planta para produzir uma planta transgênica tendo um fenótipo vantajoso ou propriedade melhorada tais tal como tolerância aumentada ao estresse por calor em relação a uma planta do tipo selvagem.

[0026]As sequências exemplares que codificam um gene bHLH do tipo selvagem ou porção do mesmo que encontram uso nos aspectos da presente invenção são descritas nas SEQ ID NOs: 1-17, 29-52, 79-83 e 89-97. As sequências exemplares que codificam um gene bHLH39 estão descritas na SEQ ID NO é :1-17. As sequências exemplares que codificam um gene bHLH38 estão descritas na SEQ ID NO é :29-52. As sequências exemplares que codificam um gene bHLH101 estão

descritas na SEQ ID NO é :79-83. As sequências exemplares que codificam um gene bHLH100 estão descritas na SEQ ID NO é :89-97. A invenção fornece adicionalmente composições que contêm os ácidos nucleicos da invenção para a expressão em uma célula de planta para produzir as plantas transgênicas descritos neste documento.

[0027]Salvo indicação em contrário, todos os termos técnicos e científicos usados neste documento têm o mesmo significado como comumente entendido por uma pessoa versada na técnica a que pertence esta invenção. Embora os métodos e materiais similares ou equivalentes aos descritos aqui possam ser usados na prática ou teste da presente invenção, métodos e materiais adequados estão descritos abaixo. Todas as publicações, pedidos de patentes, patentes e outras referências mencionadas neste documento são incorporados por referência em sua totalidade. No caso de conflito, o presente relatório descritivo, incluindo definições, prevalecerá. Além disso, os materiais, métodos e exemplos são meramente ilustrativos e não pretendem ser limitantes.

[0028]Outras características e vantagens da invenção serão aparentes e são abrangidas pela seguinte descrição detalhada e reivindicações.

BREVE DESCRIÇÃO DOS DESENHOS

[0029]A Figura 1 é um alinhamento de ClustalW de várias proteínas AtbLHL, incluindo AtbHLH39, AtbHLH38, AtbHLH100 e AtbHLH101.

DESCRIÇÃO DETALHADA

[0030]A invenção é baseada em parte na verificação de que as plantas que têm uma propriedade agrônômica melhorada, tal como, por exemplo, tolerância aumentada ao estresse por calor em relação a um controle do tipo selvagem. Mais especificamente, a invenção é baseada na verificação de que a superexpressão de um regulador transcricional bHLH do subgrupo 1b confere à planta uma propriedade agrônômica melhorada, como por exemplo, tolerância ao estresse por calor aumen-

tada em relação a um controle do tipo selvagem que pode incluir o aborto de flores reduzido e rendimento aumentado

[0031]O resultado surpreendente de que a superexpressão de um gene bHLH39 do subgrupo 1b é suficiente para conferir tolerância ao calor foi demonstrado. De acordo com a análise de EMSA e micromatriz, o bHLH39 é um alvo a jusante de Myb68, um outro fator de transcrição que confere tolerância ao calor, quando superexpresso. A análise de micromatriz mostra que a expressão gênica de IRT1 não é significativamente afetada no 35S-bHLH39, sugerindo que a tolerância ao calor conferida pelo bHLH39 ocorre através de uma via separada a partir da resposta a privação de ferro.

[0032]Desta forma, a invenção fornece métodos para intensificar (por exemplo, aumentando) a tolerância ao estresse por calor de plantas pelo aumento da expressão ou atividade de um gene ou polipeptídeo bHLH do subgrupo 1b. Os métodos para aumentar a expressão ou atividade de um gene bHLH do subgrupo 1b são conhecidos na técnica. Por exemplo, uma planta tendo tolerância ao estresse por calor aumentada em comparação com uma planta do tipo selvagem (por exemplo, controle) é produzida pela introdução para uma célula de planta de um construto de ácido nucléico que aumenta a expressão ou atividade de um gene ou polipeptídeo bHLH do subgrupo 1b. A invenção também inclui as plantas transgênicas produzidas pelos métodos da invenção e as sementes produzidas pelas plantas transgênicas que produzem uma planta tendo tolerância aumentada ao estresse por calor.

[0033]Por conveniência, antes da descrição adicional da presente invenção, certos termos empregados na especificação, exemplos e reivindicações apenas são definidos neste documento. Estas definições devem ser lidas à luz do restante da divulgação e como é entendido por uma pessoa versada na técnica.

[0034]Salvo indicação em contrário, todos os termos técnicos e científicos usados neste documento têm o mesmo significado que comumente entendido por

uma pessoa versada na técnica a que pertence esta invenção. Embora métodos e materiais similares ou equivalentes aos descritos neste documento possam ser usados na prática ou teste da presente invenção, os métodos e materiais adequados estão descritos abaixo. Todas as publicações, pedidos de patentes, patentes e outras referências mencionadas neste são incorporadas por referência em sua totalidade. No caso de conflito, o presente relatório descritivo, incluindo as definições, prevalecerá. Além disso, os materiais, métodos e exemplos são apenas ilustrativos e não pretendem ser limitantes.

[0035]A "sequência de promotora", ou "promotor", significa uma sequência de ácido nucléico capaz de induzir a transcrição de uma sequência de gene operavelmente ligado em uma célula de planta. Os promotores incluem, por exemplo, (mas não estão limitados a) promotores constitutivos, promotores específicos de tecidos, tal como um promotor de raiz, um promotor induzível, tal como um promotor induzível de seca, um promotor induzível de calor ou um promotor endógeno, tal como um promotor, normalmente associado a um gene de interesse, ou seja, um gene bHLH do subgrupo 1b ou sequências de promotores críticos ou sintéticos que são capazes de dirigir a expressão de um gene em uma célula de planta, mas não são normalmente associados com um gene expressível.

[0036]Um "cassete de expressão" significa um construto de vetor em que um gene ou uma sequência de ácidos nucléicos é transcrita. Além disso, o mRNA expresso pode ser traduzido em um polipeptídeo.

[0037]Os termos "expressão" ou "superexpressão" são usados intercambiavelmente e significam a expressão de um gene de tal forma que o transgene ou gene operavelmente ligado é expresso. O nível total de expressão em uma célula pode ser elevado em relação a uma célula do tipo selvagem correspondente.

[0038]A "mutação de ocorrência não natural" se refere a qualquer método que introduz mutações ou mudanças genéticas em uma população de plantas ou

plantas. Por exemplo, a mutagênese química, tal como etano metil sulfonato ou etil éster de ácido metanossulfônico, mutagênese de nêutrons rápidos, meios insercionais de DNA, tais como um método de mutagênese sítio-dirigida ou inserção de T-DNA. Também estão incluídos métodos para induzir mudanças genéticas, tais como métodos de meganuclease que são uma classe particular de "tesouras de DNA", eles são capazes de cortar um cromossomo em um local específico em uma célula viva.

[0039]O termo "estresse térmico" se refere a uma condição onde o crescimento ou produtividade da planta é inibido em relação a uma planta onde o calor não é um fator limitante.

[0040]O termo "tolerância ao estresse por calor" se refere à capacidade de uma planta para ter melhor desempenho que uma planta tipo selvagem em condições de estresse por calor.

[0041]O termo "estresse por secura" se refere a uma condição onde o crescimento ou produtividade da planta é inibido em relação a uma planta onde a água não é limitante. O termo "estresse por água" é usado como sinônimo e intercambiavelmente com o estresse por água e secura.

[0042]A "tolerância à seca" se refere à capacidade de uma planta para ter melhor desempenho que uma planta tipo selvagem em condições de estresse por secura ou condições limitadas por água ou para uso de menos água durante o crescimento e desenvolvimento em relação a uma planta do tipo selvagem.

[0043]O "peso seco" significa tecido vegetal que tenha sido seco para remover a maioria da água celular e é usado como sinônimo e intercambiavelmente com o termo biomassa.

[0044]O termo "nulo" é definido como um irmão segregado de uma linhagem transgênica que perdeu o transgene inserido e é, por isso, usado como uma linhagem de controle.

[0045]Uma série de várias abreviações padrões tem sido usada em toda a divulgação, tais como g, grama; WT, do tipo selvagem; DW, peso seco; WUE, eficiência do uso da água; d, dia.

[0046]O termo "bHLH do subgrupo 1b" significa um regulador transcricional bHLH39, bHLH38, bHLH100 ou bHLH101. Em alguns casos o termo bHLH é usado para se referir ao bHLH do subgrupo 1b, como apropriado no contexto.

[0047]O termo "ácido nucléico bHLH" refere-se a pelo menos uma porção de um ácido nucléico bHLH. Da mesma forma o termo "proteína bHLH" ou "polipeptídeo bHLH" refere-se a pelo menos uma porção do mesmo. Uma porção é de pelo menos 21 nucleotídeos de comprimento em relação a um ácido nucléico e uma porção de uma proteína ou polipeptídeo é de pelo menos 7 aminoácidos. O termo "AtbHLH" se refere a um gene bHLH de *Arabidopsis thaliana*, o termo "Bn bHLH" se refere a um gene bHLH *Brassica napus*.

[0048]Determinação da homologia entre duas ou mais sequências

[0049]Para determinar o percentual de homologia entre duas sequências de aminoácidos ou entre dois ácidos nucléicos, as sequências são alinhadas para fins de comparação ótima (por exemplo, os gaps (lacunas) podem ser introduzidos em qualquer uma das sequências que estão sendo comparadas para o alinhamento ótimo entre as sequências). Os resíduos de aminoácidos ou nucleotídeos em posições de aminoácidos ou posições de nucleotídeos correspondentes são então comparados. Quando uma posição na primeira sequência é ocupada pelo mesmo resíduo de aminoácido ou nucleotídeo como a posição correspondente na segunda sequência, então, as moléculas são homólogas nessa posição (isto é, como usado neste documento a "homologia" de aminoácido ou ácido nucléico é equivalente à "identidade" de aminoácidos ou ácidos nucleicos).

[0050]A homologia de sequências de ácido nucléico pode ser determinada como o grau de identidade entre duas sequências. A homologia pode ser determina-

da usando programas de computador conhecidos na técnica, tais como software GAP fornecido no pacote do programa GCG. Ver, Needleman e Wunsch (1970). Usando o software GAP GCG com as seguintes configurações para a comparação de sequência de ácidos nucleicos: penalidade de criação de GAP 5,0 e penalidade de extensão de GAP 0,3, a região de codificação das sequências de ácidos nucleicos análogas referidas acima exibe um grau de identidade de pelo, preferencialmente, menos 70%, 75%, 80%, 85%, 90%, 95%, 98%, ou 99%, com a porção de sequência de codificação da sequência de DNA mostrada na SEQ ID NO: 1.

[0051]O termo "identidade de sequência" refere-se ao grau em que duas sequências de polinucleotídeos ou polipeptídeos são idênticas em uma base de resíduo por resíduo de mais de uma região particular de comparação. O termo "percentual de identidade de sequência" é calculado pela comparação de duas sequências perfeitamente alinhadas sobre aquela região de comparação, determinando o número de posições em que a base de ácido nucléico idêntica (por exemplo, A, T, C, G, U ou I, no caso de ácidos nucléicos) ocorre em ambas as sequências para produzir o número de posições correspondentes, dividindo o número de posições correspondentes pelo número total de posições na região de comparação (ou seja, o tamanho da janela), e multiplicando o resultado por 100 para produzir o percentual de identidade de sequência. O termo "identidade substancial" como usado neste documento denota uma característica de uma sequência de polinucleotídeo, em que o polinucleotídeo compreende uma sequência que tem pelo menos 80 por cento de identidade de sequência, preferencialmente, pelo menos 85 por cento de identidade e, frequentemente de 90 a 95 por cento de identidade sequência, mais geralmente, pelo menos 99 por cento de identidade sequência, em comparação com uma sequência de referência sobre uma região de comparação. O termo "percentual de resíduos positivos" é calculado comparando duas sequências perfeitamente alinhadas sobre aquela região de comparação, determinando o número de posições em que as

substituições de aminoácidos idênticas e conservativas, como definido acima, ocorrem em ambas as sequências para produzir o número de posições correspondentes, dividindo o número de posições correspondentes pelo número total de posições na região de comparação (ou seja, o tamanho da janela), e multiplicando o resultado por 100 para produzir o percentual de resíduos positivos.

[0052]Expressão aumentada da expressão e atividade de bHLH do subgrupo 1b

[0053]Um aspecto da invenção refere-se aos meios e métodos para aumentar ou superexpressar a expressão e atividade do gene bHLH do subgrupo 1b, resultando em um aumento da expressão e atividade da proteína bHLH do subgrupo 1b. A "expressão ou atividade bHLH" abrange ambos os níveis de aumento. A expressão ou atividade de bHLH aumentada refere-se a um aumento de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60 ou 75 vezes ou maior, no nível de DNA, RNA ou proteína de um gene bHLH do subgrupo 1b em relação ao bHLH do tipo selvagem, ou um aumento de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60 ou 75 vezes de atividade da proteína bHLH do subgrupo 1b, em comparação com a atividade de bHLH do subgrupo 1b do tipo selvagem.

[0054]As sequências de codificação de um gene bHLH do subgrupo 1b ou porção do mesmo que são úteis na preparação de construtos para a expressão de bHLH do subgrupo 1b incluem, por exemplo, as SEQ ID NOs: 1-17, 29-52, 79-83 e 89-97.

[0055]Os construtos de expressão são para fornecer superexpressão ou constitutivamente em toda a planta, ou em tecidos específicos. Alternativamente, a expressão pode ser projetada para ocorrer em resposta a um estímulo, temporal espacial ou ambientalmente regulado.

[0056]As estratégias da expressão gênica serão evidentes para o técnico no assunto incluindo aquelas que não discutidas neste documento e aquelas desenvol-

vidas no futuro.

[0057]Identificação de homólogos de AtbHLH

[0058]Os homólogos de bHLH do subgrupo 1b de *Arabidopsis thaliana* (AtbHLH) foram identificados usando ferramentas de busca de sequência do banco de dados, tal como a Ferramenta de Busca de Alinhamento Local Básico (BLAST) (Altschul *et al.*, 1990 and Altschul *et al.*, 1997). Os programas de análise de sequência tblastn ou blastn foram empregados usando matriz de escore BLOSUM-62 (Henikoff e Henikoff, 1992). A saída de um relatório BLAST fornece um escore que leva em conta o alinhamento de resíduos similares ou idênticos e quaisquer gaps necessários a fim de alinhar as sequências. A matriz de escore atribui um escore para alinhar qualquer par de sequências possível. Os valores de P refletem quantas vezes espera-se ver um escore ocorrer por acaso. Os escores mais altos são os preferidos e um valor P de baixo limiar é o preferido. Estes são os critérios de identidade de sequência. O programa de análise da sequência tblastn foi usado para consultar uma sequência polipeptídica contra seis modos de traduções de sequências em um banco de dados de nucleotídeos. Os sucessos com um valor P de menos que -25, preferencialmente, menos que -70, e mais preferencialmente, menos que -100, foram identificados como sequências homólogas (critérios de sequência selecionados exemplares). O programa de análise de sequência blastn foi usado para consultar uma sequência de nucleotídeos contra um banco de dados de sequência de nucleotídeos. Também neste caso, os escores mais altos foram preferidos e um valor P limiar preferido foi de menos que -13, preferencialmente, menos que -50, e mais preferencialmente, menos que -100.

[0059]Um gene bHLH do subgrupo 1b pode ser isolados através de técnicas de amplificação de PCR padrão. Uso de iniciadores para regiões conservadas do gene bHLH do subgrupo 1b e amplificação por PCR produz um fragmento ou cópia de comprimento completo do gene desejado. O modelo pode ser o DNA genômico

ou uma biblioteca de cDNA, ou RNA ou mRNA para uso com técnicas de PCR de transcriptase reversa (RTPCR). As regiões conservadas podem ser identificadas usando ferramentas de comparação de sequência, tais como BLAST ou CLUSTALW, por exemplo. Os iniciadores (primers) foram usados e descritos no presente pedido.

[0060]Alternativamente, um fragmento de uma sequência de um gene bHLH do subgrupo 1b é radiomarcada com ^{32}P - por iniciação aleatória (Sambrook et al., 1989) e usada para a triagem de uma biblioteca genômica de plantas (os polinucleotídeos de teste exemplares). Como um exemplo, o DNA total da planta a partir de *Arabidopsis thaliana*, *Nicotiana tabacum*, *Lycopersicon pimpinellifolium*, *Prunus avium*, *Prunus cerasus*, *Cucumis sativus*, ou *Oryza sativa* é isolado de acordo com Stockinger et al. (Stockinger et al., 1996). Aproximadamente de 2 a 10 μg de cada amostra de DNA são digeridos com restrição, transferidos para membrana de nailon (Micron Separations, Westboro, Mass.) e hibridizados. As condições de hibridização são: 42°C. em formamida a 50%, SSC 5X, tampão fosfato 20 mM 1X Denhardt, sulfato de dextrana a10%, e 100 $\mu\text{g/ml}$ de DNA de esperma de arenque. Quatro lavagens de baixa severidade a Temperatura Ambiente (RT) em sarcosil de sódio 0,05% e pirofosfato de sódio 0,02%, 2X SSC, são realizadas antes das lavagens de alta severidade a 55°C em sarcosil de sódio 0,05%, 0,2.vezes de SSC, e pirofosfato de sódio 0,01%. As lavagens de alta severidade são realizadas até que não são detectadas na lavagem de acordo com Walling et al. (Walling et al., 1988). Os isolados positivos são identificados, purificados e sequenciados. Outros métodos estão disponíveis para hibridização, por exemplo, a solução de hibridização de ExpressHybTM disponível a partir de Clontech.

[0061]Células Hospedeiras e Vetores de Expressão Recombinante bHLH

[0062]Outro aspecto da invenção refere-se aos vetores, preferencialmente, vetores de expressão, contendo um ácido nucléico que codifica uma proteína bHLH

do subgrupo 1b, um gene bHLH do subgrupo 1b ou sequência genômica ou porções dos mesmos e análogos ou homólogos dos mesmos. Como usado neste documento o termo vetor de expressão inclui vetores, que são projetados para fornecer a transcrição da sequência de ácidos nucleicos. As sequências transcritas podem ser projetadas para expressar o construto do gene para aumentar a expressão ou atividade total de uma atividade de gene endógeno que correlaciona com a sequência transcrita. A sequência expressa pode ser uma proteína de codificação de bHLH do subgrupo 1b endógena ou a partir de uma espécie heteróloga.

[0063]O ácido nucleico transcrito pode ser traduzido em um produto de polipeptídeo ou proteína. O polipeptídeo pode ser uma variante mutante ou modificada de comprimento não completo da proteína endógena. Como usado neste documento, o termo "vetor" refere-se a uma molécula de ácido nucleico capaz de transportar um outro ácido nucleico ao qual o mesmo foi ligado. Um tipo de vetor é um "plasmídeo", que se refere a uma alça de DNA de dupla fita circular em que segmentos de DNA adicionais podem ser ligados. Um outro tipo de vetor é um vetor viral, em que os segmentos de DNA adicional podem ser ligados no genoma viral. Certos vetores são capazes de replicação autônoma em uma célula hospedeira em que eles são introduzidos (por exemplo, vetores de bactérias tendo uma origem bacteriana da replicação). Outros vetores são integrados no genoma de uma célula hospedeira mediante a introdução na célula hospedeira e, portanto, são replicados junto com o genoma do hospedeiro. Além disso, certos vetores são capazes de dirigir a expressão dos genes aos quais eles estão operativamente ligados. Os vetores são referidos aqui como "vetores de expressão". Em geral, os vetores de expressão de utilidade em técnicas de DNA recombinante estão muitas vezes na forma de plasmídeos. No presente relatório descritivo, "plasmídeo" e "vetor" podem ser usados intercambiavelmente na medida em que o plasmídeo é a forma mais comumente usada de vetor. No entanto, a invenção destina-se a incluir outras formas de vetores de expres-

são, tais como vetores virais ou vetores de transformação de plantas, binários ou não, que têm funções equivalentes.

[0064] Os vetores de expressão recombinante da invenção compreendem um ácido nucléico da invenção em uma forma adequada para a expressão do ácido nucléico em uma célula hospedeira, o que significa que os vetores de expressão recombinantes incluem uma ou mais sequências reguladoras, selecionadas com base nas células hospedeiras para serem usadas para a expressão, que são operativamente ligadas à sequência de ácido nucléico a ser expressa. Dentro de um vetor de expressão recombinante, "operavelmente ligado" destina-se a significar que a sequência de nucleotídeos de interesse está ligada à sequência(s) reguladora de uma maneira que permite a expressão da sequência de nucleotídeos (por exemplo, em um sistema de tradução/transcrição *in vitro* ou em uma célula hospedeira quando o vetor é introduzido na célula hospedeira).

[0065] O termo "sequência reguladora" destina-se a incluir promotores, intensificadores e outros elementos de controle de expressão (por exemplo, sinais de poliadenilação). Tais sequências reguladoras são descritas, por exemplo, em Goeddel (1990). As sequências reguladoras incluem aquelas que dirigem a expressão constitutiva de uma sequência de nucleotídeos em muitos tipos de células hospedeiras e aquelas dirigem a expressão da sequência de nucleotídeos apenas em certas células hospedeiras (por exemplo, sequências reguladoras específicas de tecidos) ou promotores induzíveis (por exemplo, induzidos em resposta a fatores abióticos como condições ambientais, calor, secura, estado nutricional ou estado fisiológico da célula ou bióticos, como responsivo ao patógeno). Exemplos de promotores adequados incluem, por exemplo, promotores constitutivos, promotores induzíveis por ABA, promotores específicos de tecido e promotores induzíveis por estresse abiótico ou biótico. Será apreciado por aqueles versados na técnica que o projeto do vetor de expressão pode depender de fatores como a escolha da célula hospedeira a ser

transformada, o nível de expressão da proteína desejada, bem como horário e local de expressão, etc.. Os vetores de expressão da invenção podem ser introduzidos nas células hospedeiras para assim produzir proteínas ou peptídeos, incluindo proteínas ou peptídeos de fusão, codificados por ácidos nucleicos, como descritos neste documento (por exemplo, as proteínas bHLH do subgrupo 1b, formas mutantes das proteínas bHLH do subgrupo 1b, proteínas de fusão, etc.).

[0066] Os vetores de expressão recombinante da invenção podem ser projetados para a expressão de genes bHLH do subgrupo 1b, proteínas bHLH do subgrupo 1b, ou porção dos mesmos, em células procarióticas ou eucarióticas. Por exemplo, os genes bHLH do subgrupo 1b ou proteínas bHLH do subgrupo 1b podem ser expressos em células bacterianas, como a *Escherichia coli*, células de inseto (usando vetores de expressão de baculovírus)) as células de levedura, células de plantas ou células de mamíferos. As células hospedeiras adequadas são discutidas em Goeddel (1990). Alternativamente, o vetor de expressão recombinante pode ser transcrito e traduzido *in vitro*, por exemplo, usando sequências reguladoras promotoras T7 e polimerase T7.

[0067] Em uma modalidade, um ácido nucleico da invenção é expresso em células de plantas usando um vetor de expressão de plantas. Exemplos de sistemas de vetores de expressão de planta incluem plasmídeo (Ti) de indução de tumor ou porção do mesmo encontrado em *Agrobacterium*, o vírus do mosaico da couve-flor (CaMV) DNA e vetores, tais como pBI121.

[0068] Para expressão em plantas, o cassete de expressão recombinante conterá, além dos ácidos nucleicos bHLH do subgrupo 1b, uma região promotora que funciona em uma célula de planta, um sítio de iniciação de transcrição (se a sequência de codificação para transcrita não tem um) e, opcionalmente, uma sequência de terminação/poliadenilação de transcrição. A região de terminação/poliadenilação pode ser obtida a partir do mesmo gene como a sequência pro-

motora ou podem ser obtidos a partir de genes diferentes. Os sítios de enzima de restrição única nas extremidades 5' e 3' do cassete são tipicamente incluídos para permitir a fácil inserção em um vetor pré-existente.

[0069]Exemplos de promotores adequados incluem promotores de vírus de plantas, tais como o promotor 35S do vírus do mosaico da couve-flor (CaMV) (Odell et al., 1985), os promotores dos genes, tais como actina do arroz (McElroy et al., 1990), ubiquitina (Christensen et al, 1992); pEMU (Last et al., 1991), MAS (Velten et al, 1984.), histona H3 do milho (Lepetit et al, 1992.); e Atanassova et al, 1992), o promotor 5'- ou 3'- derivado de T-DNA de *Agrobacterium tumefaciens*, o promotor de Smas, o promotor de cinamil álcool desidrogenase (Patente US 5.683.439), o promotor Nos, o promotor rubisco, o promotor GRP1-8, promotor ALS, (WO 96/30530), um promotor sintético, tal como Rsyn7, SCP e promotores UCP, ribulose 1,3-difosfato carboxilase, promotores específicos de fruta, promotores de choque térmico (HSP 81,1 ou HSP18.2), promotores específicos de semente, promotores específicos de raiz, ou seja, uclacyanin2 (UCC2, At2g44790) e outras regiões de iniciação de transcrição de vários genes de plantas, por exemplo, incluindo várias regiões de iniciação opina, como por exemplo, octopina, mannopina e nopalina. Em alguns casos, um promotor associado com o gene de interesse (por exemplo, bHLH) pode ser usado para expressar um construto que tem como alvo o gene de interesse, por exemplo, o promotor *AtbHLH* nativo. Elementos reguladores adicionais que podem ser conectados a uma sequência de ácidos nucleicos de codificação bHLH do subgrupo 1b para expressão em células de plantas incluem sequências de poliadenilação, terminadoras, e sequências de ácidos nucleicos que codificam do peptídeos sinais que permitem a localização dentro de uma célula de planta ou secreção da proteína da célula. Esses elementos reguladores e métodos para adicionar ou trocar estes elementos com os elementos reguladores do gene bHLH do subgrupo 1b são conhecidos e incluem, mas não estão limitados a, regiões de poliadenilação e/ou

terminação 3', tais como os do gene (nos) sintase nopalina *Agrobacterium tumefaciens* (Bevan et al, 1983); gene inibidor de proteinase II de batata (PINII) (Keil et al, 1986) e incorporado neste documento por referência); e An *et al.* (1989); e o gene CaMV 19S (Mogen et al, 1990).

[0070]As sequências sinais de planta, incluindo, mas não limitadas a, sequências de DNA/RNA que codificam peptídeo sinal que têm como alvo proteínas para a matriz extracelular da célula de planta (Dratewka-Kos *et al.*, 1989) e o gene extensão *Nicotiana plumbaginifolia* (De Loose *et al.*, 1991), ou peptídeos sinais que têm como alvo as proteínas para o vacúolo tipo o gene sporamin da batata doce (Matsuoka et al., 1991) e o gene de lectina da cevada (Wilkins *et al.*, 1990), ou sinais que fazem com que as proteínas sejam secretadas como as de PR1b (Lund *et al.*, 1992), ou aquelas as proteínas alvos para os plastídios, tais como as de enoil-ACP redutase de colza (Verwoert *et al.*, 1994) são úteis na invenção.

[0071]Em outra modalidade, o vetor de expressão recombinante é capaz de dirigir a expressão do ácido nucleico, preferencialmente, em um tipo de célula particular (por exemplo, elementos reguladores específicos de tecidos são usados para expressar o ácido nucléico). Os elementos reguladores específicos de tecidos são conhecidos na técnica. Por exemplo, o promotor associado a uma sequência de codificação identificada na base de dados TAIR como At2g44790 (P₄₇₉₀) é um promotor específico de raiz. Especialmente úteis em conexão com os ácidos nucléicos da presente invenção são os sistemas de expressão que são operáveis em plantas. Estes incluem os sistemas que estão sob controle de um promotor específico de tecido, bem como aqueles que envolvem os promotores que são operáveis em todos os tecidos da planta.

[0072]Os promotores específicos para órgãos também são bem conhecidos. Por exemplo, o gene de chalcona sintase-A (van der Meer *et al.*, 1990) ou o promotor di-hidroflavonol-4-redutase (dfr) (Elomaa *et al.*, 1998) dirige a expressão em tecidos

dos florais específicos. Também está disponível o promotor patatin classe I que é transcricionalmente ativado somente do tubérculo da batata e pode ser usado para a expressão do gene alvo no tubérculo (Bevan, 1986). Outro promotor específico de batata é o promotor de amido sintase ligada grânulo (GBSS) (Visser et al., 1991).

[0073]Outros promotores específicos de órgãos apropriados para um órgão alvo desejado podem ser isolados através de procedimentos conhecidos. Essas sequências de controle são geralmente associadas com genes expressos exclusivamente no órgão desejado. Em uma planta superior típica, cada órgão tem milhares de mRNAs que estão ausentes de outros sistemas de órgãos (revisto em Goldberg, 1986).

[0074]O sistema ou cassete de expressão resultante é ligado em, ou construído para ser incluído em um vetor recombinante que é apropriado para transformação de plantas. O vetor também pode conter um gene marcador selecionável pelo qual as células de plantas transformadas podem ser identificadas na cultura. O gene marcador pode codificar a resistência a antibióticos. Esses marcadores incluem resistência ao G418, higromicina, bleomicina, canamicina e gentamicina. Alternativamente, o gene marcador pode codificar um gene de tolerância a herbicidas que fornece tolerância aos herbicidas do tipo glufosinato ou glifosato. Depois de transformar as células de planta, as células tendo o vetor serão identificadas por sua capacidade de crescer em meio contendo o antibiótico ou herbicida particular. As sequências de replicação, de origem bacteriana ou viral, geralmente são também incluídas para permitir que o vetor seja clonado em um hospedeiro de bactérias ou fagos, preferencialmente, uma ampla faixa de hospedeiro de origem procariótica de replicação é incluída. Um marcador selecionável para bactérias também deve ser incluído para permitir a seleção de células bacterianas portando o construto desejado. Os marcadores procarióticos selecionáveis adequados também incluem resistência a antibióticos como canamicina ou tetraciclina.

[0075]Outras sequências de DNA que codificam funções adicionais também podem estar presentes no vetor, como é conhecido na técnica. Por exemplo, no caso de transformações de *Agrobacterium*, as sequências de T-DNA também serão incluídas para posterior transferência aos cromossomos de plantas.

[0076]Outro aspecto da invenção refere-se às células hospedeiras em que um vetor de expressão recombinante da invenção foi introduzido. Os termos "célula hospedeira" e "célula hospedeira recombinante" são usados intercambiavelmente neste documento. Entende-se que tais termos não se referem apenas à célula ao sujeito particular, mas também à prole ou progênie potencial dessa célula. Devido a certas modificações poderem ocorrer em sucessivas gerações, quer devido à mutação ou a influências ambientais, tais progênies não podem, de fato, ser idênticas à célula-mãe, mas ainda estão incluídas no âmbito do termo como usados neste documento.

[0077]O vetor de DNA pode ser introduzido em células procarióticas ou eucarióticas através de técnicas de transformação ou transfecção convencionais. Como usado neste documento, os termos "transformação" e "transfecção" destinam-se a se referir a uma variedade de técnicas de reconhecidas na técnica para a introdução de ácido nucléico estranho (por exemplo, DNA) em uma célula hospedeira.

[0078]Uma célula hospedeira da invenção, tal como uma célula hospedeira procariótica ou eucariótica em cultura, pode ser usada para produzir (ou seja, expressar) um polipeptídeo da invenção codificado em uma estrutura de leitura aberta de um polinucleotídeo da invenção. Assim, a invenção ainda fornece métodos para produzir um polipeptídeo com as células hospedeiras da invenção. Em uma modalidade, o método compreende cultivar a célula hospedeira da invenção (em que um vetor de expressão recombinante que codifica um polipeptídeo da invenção foi introduzido) em um meio adequado tal que o polipeptídeo seja produzido. Em outra modalidade, o método ainda compreende isolar o polipeptídeo do meio ou da célula

hospedeira.

[0079]Uma série de tipos celulares pode agir como célula hospedeira adequada para a expressão de um polipeptídeo codificado por uma estrutura de leitura aberta em um polinucleotídeo da invenção. As células hospedeiras da planta incluem, por exemplo, células de plantas que poderiam funcionar como hospedeiros adequados para a expressão de um polinucleotídeo da invenção incluem as células epidérmicas, mesófilo e outros tecidos da terra, e os tecidos vasculares em folhas, caules, órgãos florais e raízes de uma variedade de espécies de plantas, tais como *Arabidopsis thaliana*, *Nicotiana tabacum*, *Brassica napus*, *Zea mays*, *Oryza sativa*, *Gossypium hirsutum* e *Glycine max*.

[0080]Plantas transgênicas e células de plantas transformadas

[0081]A invenção inclui um protoplasto, célula de planta, tecido de plantas e planta (por exemplo, monocotiledôneas ou dicotiledôneas) transformados com um ácido nucléico bHLH do subgrupo 1b, um vetor contendo um ácido nucléico bHLH do subgrupo 1b ou um vetor de expressão contendo um ácido nucléico bHLH do subgrupo 1b. Como usado neste documento, "planta" se destina a incluir não apenas planta no todo, mas também uma porção da mesma (isto é, células e tecidos, incluindo, por exemplo, folhas, caules, brotos, raízes, flores, frutos e sementes).

[0082]A planta pode ser qualquer tipo de planta, incluindo, por exemplo, espécies dos gêneros *Arabidopsis*, *Brassica*, *Oryza*, *Zea*, *Sorgo*, *Brachypodium*, *Miscanthus*, *Gossypium*, *Triticum*, *Glycine*, *Pisum*, *Phaseolus*, *Lycopersicon*, *Trifolium*, *Cannabis*, *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Gerânio*, *Manihot*, *Daucus*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Nicotiana*, *Solanum*, *Petúnia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*, *Lactuca*, *Bromus*, *Espargos*, *Antirrhinum*, *Heterocalis*, *Nemesis*, *Pelargonium*, *Panieum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Lolium*, *Avena*, *Hordeum*, *Secale*, *Picea*, *Caco*, e *Populus*.

[0083]A invenção também inclui células, tecidos, incluindo, por exemplo, folhas, caules, brotos, raízes, flores, frutos e sementes e a progênie derivada da planta transformada.

[0084]Vários métodos para a introdução de genes estranhos em plantas são conhecidos e podem ser usados para inserir um gene em uma planta hospedeira, incluindo protocolos de transformação de plantas biológicas e físicas (Ver, por exemplo, Miki *et al.*, (1993) "Procedure for Introducing Foreign DNA into Plants", In: Methods in Plant Molecular Biology and Biotechnology, Glick e Thompson, eds., CRC Press, Inc., Boca Raton, páginas 67-88; e Andrew Bent em, Clough SJ and Bent AF, (1998) "Floral dipping: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*"). Os métodos escolhidos variam de acordo com a planta hospedeira, e incluem métodos de transfecção química, tais como fosfato de cálcio, transformação com polietileno glicol (PEG), transferência de genes mediada por micro-organismo, tal como *Agrobacterium* (Horsch *et al.*, 1985), eletroporação, transformação por protoplastos, microinjeção, imersão de flor, e bombardeio biobalístico.

[0085]Transformação Mediada por *Agrobacterium*

[0086]O método mais amplamente usado para a introdução de um vetor de expressão em plantas é baseado no sistema de transformação natural de *Agrobacterium tumefaciens* e *A. rhizogenes* que são bactérias fitopatogênicas que transformam geneticamente células de plantas. Os plasmídeos Ti e Ri de *A. tumefaciens* e *A. rhizogenes*, respectivamente, carregam genes responsáveis pela transformação genética de plantas (ver, por exemplo, Kado, 1991). Descrições dos sistemas de vetor de *Agrobacterium* e métodos para transferência de genes mediada por *Agrobacterium* são fornecidos em Gruber *et al.* (1993) e Moloney *et al.*, (1989).

[0087]As plantas transgênicas de *Arabidopsis* podem ser produzidas facilmente pelo método de imersão de plantas com flores em uma cultura de *Agrobacte-*

rium, com base no método de Andrew Bent in, Clough SJ and Bent AF, 1998. Floral dipping: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. As plantas do tipo selvagem são crescidas até que a planta possua tanto flores em desenvolvimento quanto flores abertas. As plantas são invertidas por 1 minuto em uma solução de cultura de *Agrobacterium* carregando o construto de gene apropriado. As plantas são então deixadas na horizontal em uma bandeja e mantidas cobertas por dois dias para manter a humidade e depois endireitadas e ensacadas para continuar o desenvolvimento das sementes e crescimento. A semente madura é colhida em massa

[0088]Transferência de Gene direta

[0089]Um método aplicável em geral de transformação de plantas é a transformação mediada por microprojétil, onde o DNA é carregado na superfície dos microprojéteis medindo cerca de 1 a 4 μm . O vetor de expressão é introduzido em tecidos vegetais com um dispositivo biobalístico que acelera os microprojéteis para velocidades de 300 a 600 m/s que é suficiente para penetrar a parede celular da planta e das membranas (Sanford et al, 1993;.. Klein et al, 1992).

[0090]A transformação das plantas também pode ser alcançada pelo método de Injector de Feixe de Aerosol (ABI) descrito na Patente US 5.240.842 e Patente US 6.809.232. A tecnologia de feixe de aerosol é usada para acelerar as partículas molhadas ou secas para velocidades que permitem que as partículas penetrem as células vivas. A tecnologia de feixes de aerossol emprega a expansão de jato de um gás inerte na medida em que o mesmo passa de uma região de maior pressão de gás para uma região de baixa pressão do gás através de um pequeno orifício. O gás em expansão acelera as gotículas do aerossol, contendo moléculas de ácido nucléico para serem introduzidas em uma célula ou tecido. As partículas aceleradas são posicionadas para o impacto de um alvo preferido, por exemplo, uma célula de planta. As partículas são construídas como gotículas de um tamanho suficientemente

pequeno para que a célula sobreviva à penetração. A célula ou tecido transformado é cultivado para produzir uma planta por meio de técnicas padrões conhecidas por aqueles versados na técnica aplicável.

[0091]Regeneração de Transformantes

[0092]O desenvolvimento ou regeneração de plantas a partir de qualquer protoplastos de plantas únicas ou vários explantes é bem conhecido na técnica (Weissbach e Weissbach, 1988). Este processo de regeneração e crescimento geralmente inclui as etapas de seleção de células transformadas, cultivo dessas células individualizadas através dos estágios usuais do desenvolvimento embrionário durante o estágio de plântula enraizada. Os embriões e sementes transgênicas são igualmente regenerados. Os brotos enraizados transgênicos resultantes são posteriormente plantados em um meio de crescimento da planta apropriado, tal como o solo.

[0093]O desenvolvimento ou regeneração de plantas contendo o gene exógeno, estranho que codifica um polipeptídeo de interesse introduzido por *Agrobacterium* partir de explantes de folhas podem ser alcançado por métodos conhecidos na técnica, como descrito em Horsch et al., 1985. Neste procedimento, os transformantes são cultivados na presença de um agente de seleção e em um meio que induz a regeneração de brotos na cepa da planta que está sendo transformada como descrito em Fraley et al., 1983. Em particular, Patente US. 5.349.124 (especificação incorporada neste documento por referência) detalha a criação de células de alfaca geneticamente transformadas e as plantas resultantes que expressam proteínas de cristal híbrido que conferem atividade inseticida contra as larvas de Lepidoptera para essas plantas.

[0094]Este procedimento normalmente produz brotos dentro de dois a quatro meses e os brotos são então transferidos para um meio de indução de raiz apropriado contendo o agente seletivo e um antibiótico para impedir o crescimento bacteriana-

no. Os brotos que enraizaram na presença do agente seletivo para formar mudas são então transplantados para o solo ou outros meios para permitir a produção de raízes. Estes procedimentos variam dependendo da cepa de planta determinada empregada, tais variações sendo bem conhecidas na técnica.

[0095] Preferencialmente, as plantas regeneradas são autopolinizadas para fornecer plantas transgênicas homozigóticas, ou os pólen obtidos das plantas regeneradas é cruzado com plantas de sementes crescidas de interesse agrônômico, preferencialmente, de linhagens puras. Por outro lado, o pólen de plantas dessas linhagens importantes é usado para polinizar plantas regeneradas. A planta transgênica da presente invenção que contém um polipeptídeo desejado é cultivada usando métodos conhecidos por um técnico no assunto.

[0096] A planta transgênica preferida é um segregado independente e pode transmitir o construto do gene bHLH do subgrupo 1b para sua progênie. Uma planta transgênica mais preferida é homozigótica para o construto de gene, e transmite esse construto de gene para todos os descendentes mediante o acasalamento sexual. A semente de uma planta transgênica pode ser crescida no campo ou estufa, e as plantas transgênicas sexualmente maduras resultantes são autopolinizadas para gerar plantas de reprodução verdadeira. Os progênies destas plantas se tornam linhagens de reprodução verdadeira que são avaliadas para a expressão aumentada do gene bHLH do subgrupo 1b.

[0097] Método de produção de plantas transgênicas

[0098] Também incluídos na invenção estão os métodos de produção de uma planta transgênica tendo tolerância aumentada ao estresse por calor, aborto de flores reduzido e rendimento aumentado em relação a uma planta do tipo selvagem após um estresse por calor. O método inclui a introdução em uma ou mais células da planta de um composto que aumenta a expressão ou atividade de bHLH do subgrupo 1b na planta para gerar uma célula de planta transgênica e regenerar uma

planta transgênica a partir da célula transgênica. O composto pode ser, por exemplo, (i) um polipeptídeo bHLH do subgrupo 1b, (ii) um ácido nucléico, análogo, homólogo, ortólogo, porção, variante de bHLH do subgrupo 1b ou complemento dos mesmos; (iii) um ácido nucléico que aumenta a expressão de um ácido nucléico bHLH do subgrupo 1b. Um ácido nucléico que aumenta a expressão de um ácido nucléico bHLH do subgrupo 1b pode incluir elementos promotores ou intensificadores. Os ácidos nucleicos bHLH do subgrupo de 1b podem ser endógenos ou exógenos, por exemplo, um ácido nucléico bHLH do subgrupo 1b de *Arabidopsis* pode ser introduzido em uma espécie de milho ou *Brassica*. Preferencialmente, o composto é uma sequência endógena de ácido nucléico bHLH do subgrupo 1b para a espécie que está sendo transformada. Alternativamente, o composto é uma sequência exógena de ácido nucléico bHLH do subgrupo 1b para a espécie que está sendo transformada, tendo pelo menos 70%, 75%, 80%, 85%, 90% ou mais de homologia com a sequência alvo endógena.

[0099]Em vários aspectos a planta transgênica tem um fenótipo alterado em comparação com uma planta do tipo selvagem (isto é, não transformado). Pelo fenótipo alterado quer-se dizer que a planta tem uma ou mais características que é diferente da planta do tipo selvagem. Por exemplo, quando a planta transgênica foi contactada com um composto que aumenta a expressão ou atividade de um ácido nucléico bHLH do subgrupo 1b, a planta tem um fenótipo, tal como tolerância aumentada ao estresse por calor, aborto de flores reduzido e rendimento aumentado em relação a uma planta do tipo selvagem após o estresse por calor.

[00100]A planta pode ser qualquer tipo de planta, incluindo, por exemplo, espécies dos gêneros

[00101]*Arabidopsis*, *Brassica*, *Oryza*, *Zea*, *Sorghum*, *Brachypodium*, *Miscanthus*, *Gossypium*, *Triticum*, *Glycine*, *Pisum*, *Phaseolus*, *Lycopersicon*, *Trifolium*, *Cannabis*, *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*,

Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Cichorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocalis, Nemesis, Pelargonium, Panieum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Lolium, Avena, Hordeum, Secale, Picea, Caco, e Populus.

EXEMPLOS

Exemplo 1: Identificação de Genes homólogos

[00102]Pesquisas de blasto de bHLH39 e seus homólogos *Arabidopsis* bHLH38, bHLH100, e bHLH101 foram realizadas contra a bancos de dados de proteína NCBI's, nucleotídeo, e EST e bancos de dados unigene TIGR(1e-01). O banco de dados de sequência genômica foi usado para espécies cuja sequência completa é conhecida, como Arroz e Sorgo. Para confirmar uma sequência foi de fato um ortólogo das sequências de interesse, todos os homólogos putativos foram então *blasted* contra a o banco de dados completo de proteína *Arabidopsis*, e qualquer sequência cujo *Arabidopsis* superior alcançado não foi uma das quatro sequências de interesse foi filtrada de qualquer análise. A redundância mínima entre vários *hits*, todas as sequências EST *hits* foram montadas usando o programa de montagem cap3 garantindo que tinham um mínimo de 90% de identidade entre pelo menos 40 nucleotídeos com um máximo de 5 gaps. Fases de leitura aberta foram determinadas usando o programa EMBOSS *getorf*.

[00103]As pesquisas de blasto de AtbHLH39, AtbHLH38, AtbHLH100, AtbHLH101 geraram 96 sequências homólogas de 40 espécies diferentes. A conservação de sequência foi limitada a domínios de ligação de DNA e dimerização, com pouca ou nenhuma conservação foram desta área. Homólogos foram descobertos em ambas monocotiledôneas e dicotiledôneas sugerindo um requisito funcional importante nas plantas. Homólogos de AtbHLH39 em espécies de importância agrônômica como arroz e milho foram recuperados bem como espécies usadas em

biocombustíveis como *Brachypodium* e *Switchgrass*.

[00104]De acordo com a pesquisa de blasto contra o banco de dados genômico de *Brachypodium*, há três homólogos para este grupo de genes, e todos os três são mais semelhantes a AtbHLH38. Há uma evidência EST para dois dos três homólogos. Se o terceiro homólogo é um pseudogene ou códigos para uma proteína funcional ainda deve ser determinado.

[00105]Em outras duas espécies de monocotiledôneas, arroz e Sorgo, há somente um homólogo para este grupo, e em ambas as espécies o homólogo é mais próximo de AtbHLH38. Em ambos os arroz e Sorgo, o gene codifica pelo menos duas variantes unidas.

[00106]Há pelo menos dois homólogos para este grupo em *Brassica napus*, uma proteína como homologia próxima a AtbHLH39, e outra com homologia próxima a AtbHLH38. No entanto, o homólogo BnbHLH39 contém uma mutação não-senso que resulta na codificação de uma proteína parcial sem ligação ao domínio de DNA ou de dimerização. Este achado é confirmado com vários EST, e ainda ocorre em *Brassica rapa*.

Exemplo 2: Construção de Vetor

[00107]O vetor binário pBI121 foi otimizado para a transformação de *Arabidopsis* e diferentes culturas. O gene GUS em pBI121 foi deletado por digestões e religamentos de Small e EcoCR1, resultando no vetor pBI121ΔGUS. Este vetor foi usado para clonar gene para superexpressão. A porção C-terminal de 1,1-Kb do gene GUS foi isolado de pBI121 como um fragmento EcoR V-Sac I (posições 6613-7715 em pBI121) e clonado nos sítios Sma I e Sac I em pBI121ΔGUS, resultando em pBI121tGUS (com porção N-terminal do gene GUS deletado). O vetor foi usado para criar genes Construtos *Hairpin* de regulação para baixo com a sequência parcial GUS como a alça ou espaçador.

[00108]O gene NPTII no vetor pBI121 e seus derivados contém uma muta-

ção pontual (G para T na posição 3383 em pBI121, substituição de aminoácido E182D). A enzima mutante mostrou atividade enzimática várias vezes inferior do que seu tipo selvagem (PNAS, 87:3435-3439, 1990). Para melhorar a eficiência da transformação de diferentes culturas, os vetores pBIΔGUS e pBIItGUS foram restabelecidos com o gene WT-NPTII: um fragmento Nhe I-BstB I (0,9 kb, posições 2715-3648) foi substituído com um fragmento Nhe I-BstB I de exatamente a mesma sequência exceto pela única diferença de nucleotídeo. O fragmento foi isolado por digestão de restrição do plasmídeo pRD400 que continha genes WT-NPTII (PNAS, 87:3435-3439, 1990; Gene, 122:383-384, 1992). Os vetores modificados foram nomeados pBI300ΔGUS e pBI300tGUS, respectivamente. O gene WT-NPTII foi ainda isolado de pBI300ΔGUS como um fragmento Nhe I-Hind III (2,2 kb) e clonado nos sítios correspondentes em pBI121. Isto gerou o vetor pBI300GUS. Para distinguir estes vetores dos outros, vetores binários baseados em pBI121 continha o gene WT-NPTII foram desenhados série pBI300.

[00109]Para usar Basta como um agente de seleção, um marcador de resistência Basta foi subclonado em pBI121. Um fragmento de 1,3-kb Ase I englobando o marcador de seleção Basta (35S-Bar-nosT) foi amplificado por PCR do vetor pEGAD usando um iniciador direto contendo um sítio Pme I e um iniciador reverso contendo um sítio Hind III. O fragmento foi clonado, entre os sítios de Pme I (posição 2492) e Hind III (posição 4950) em pBI121, pBIΔGUS e pBIItGUS. Como resultado, o marcador de seleção canamicina (Pnos-NPTII-nosT) foi substituído com o marcador de seleção Basta nestes vetores. Para distinguir estes vetores dos outros, vetores binários a base de pBI121 contendo o gene marcador de seleção Basta foram designados série pBI800.

[00110]Os vetores pBI300 contém o gene WT-NPTII direcionando pelo promotor nopalina sintase (nos). Uma vez que o promotor não é ativo em plantas monocotiledôneas, foi necessário substituí-lo com um promotor forte para transforma-

ção de plantas monocotiledôneas. Para este fim, o gene promotor *Brachyposium* TIF1 (BdTIF1, também chamado BdGOS2) foi clonado em vetores da série pBI300 para direcionar o gene WT-NPTII. A sequência de promotor BdTIF1 (-1 até -2548 com relação ao códon inicial ATG, incluindo o primeiro éxon, primeiro intron e parte do segundo éxon) foi amplificado por PCR como um fragmento PmeI-NheI e clonado nos sítios correspondentes em HSP81.1-AtMyb68-pBI300 (um vetor a base de pBI300 contendo a proteína 81.1 promotora de choque térmico de *Arabidopsis* e a sequência de codificação Myb68 de *Arabidopsis*, vide abaixo para detalhes). Como resultado o promotor TIF1p foi colocado 5' a montante ao gene WT-NPTII com um pequeno fragmento de cerca de 120 bp (incluindo 65 bp do promotor nos) entre estes. Para delimitar o fragmento de 120-bp, a sequência de codificação WT-NPTII e vetor de sequência de flanqueamento foi amplificado como o fragmento Nhe I-Sal I (2,1 Kb) e ligado ao plasmídeo digerido com Nhe I e Sal I. A clonagem resultou no vetor HSP81.1-AtMyb68-pBI500. Vetores a base de pBI121 contendo o gene WT-NPTII direcionados pelo promotor BdTIF1 foram designados série pBI500. A sequência HSP81.1-AtMyb68 foi ainda substituída com a sequência de promotor 35S como um fragmento Hind III-BamH I isolado por digestão de restrição de pBI300ΔGUS. Isto gerou o vetor pBI500ΔGUS.

Exemplo 3: bHLH subgrupo b1 Construtos de superexpressão

[00111]Construto 35S-AtbHLH39:

[00112]A sequência de codificação AtbHLH39 (At3g56980) foi amplificada por RT-PCR de *Arabidopsis* usando iniciador direto BHLH039FW-XbaI e iniciador reverso BHLH039RV-BamH I. O produto de PCR (0,8 Kb) foi clonado no sítio Xba I e BamH I no vetor binário pBI300ΔGUS e pBI800ΔGUS, gerando o Construto 35S-AtbHLH39-pBI300 e 35S-AtbHLH39-pBI800, respectivamente.

[00113]Construto 35S-AtbHLH101:

[00114]A sequência de codificação AtbHLH101 (At5g04150) foi amplificada

por RT-PCR de *Arabidopsis* usando iniciador direto BHLH101FW-XbaI e iniciador reverso BHLH101RV-BamH I. O produto de PCR (0,7 Kb) foi clonado nos sítios Xba I e BamH I no vetor binário pBI300ΔGUS e pBI800ΔGUS, gerando Construto 35S-AtbHLH101-pBI300 e 35S-AtbHLH101-pBI800, respectivamente.

[00115]Construto 35S-AtbHLH38:

[00116]Usando a mesma estratégia conforme descrito acima, a sequência de codificação AtbHLH38 (at3g56970) é amplificada por RT-PCR de *Arabidopsis* usando um iniciador direto contendo um sítio Xba I e iniciador reverso contendo um sítio Bam HI, e clonado no sítio Xba I e BamH I no vetor binário pBI300ΔGUS, gerando o Construto 35S-AtbHLH38-pBI300.

[00117]Construto 35S-AtbHLH100:

[00118]Usando a mesma estratégia conforme descrito acima, a sequência de codificação AtbHLH100 (At2g41240) é amplificada por RT-PCR de *Arabidopsis* usando um iniciador direto contendo um sítio Xba I e iniciador reverso contendo um sítio Bam HI, e clonado nos sítios Xba I e BamH I no vetor binário pBI300ΔGUS, gerando o Construto 35S-AtbHLH100-pBI300.

[00119]Construtos HSP81.1-AtbHLH39:

[00120]Várias etapas foram envolvidas no desenvolvimento do Construto. Primeiro, a sequência promotora (-401 a -1 com relação ao códon de início ATG) do gene de proteína HSP81.1 de choque térmico de *Arabidopsis* (At5g52640) foi amplificado por PCR com iniciadores contendo extremidades Sal I e Xba I de DNA genômico de *Arabidopsis*, e clonados nos mesmos sítios em pBI101, assim substituindo o promotor 35S. Este vetor foi chamado de pHSP81.1-GUS. O sequenciamento revelou mutações pontuais de T para C na posição -266 e C para T em -121 no promotor. A coloração GUS de mudas de *Arabidopsis* transformadas com o Construto mostrou os mesmos perfis de indução de calor como relatado na literatura. Por isso estas mutações não afetam aparentemente a funcionalidade do promotor. Segundo,

MCS2-oligo de New England Biolabs foi anelado e ligado a pHSP81.1-GUS que foi digerido com Xba I e Sma I. Isto resultou no vetor pHSP81.1MCS-GUS. Terceiro, a sequência GUS foi deletada por digestão SmaI e EcoICRI, e auto-ligação do vetor. Isto levou ao vetor pHSP81.1MCSΔGUS. Quarto, a sequência de codificação AtMyb68 (At5g65790) foi clonada nos sítios XbaI e BamH I para superexpressão de Myb68. Quinto, a sequência mutante NPTII foi substituída com esta sequência WT conforme descrito acima, produzindo o vetor HSP81.1-AtMyb68-pBI300. Por fim, a sequência de codificação AtbHLH39 como um fragmento Xba I-BamH I foi clonada nos sítios Xba I e BamH I assim, substituindo a sequência AtMyb68, resultando no Construto HSP81.1-AtbHLH39-pBI300. Igualmente, o fragmento Xba I-BamH I da sequência de codificação AtbHLH39 substituiu a sequência AtMyb68 em HSP81.1-AtMyb68-pBI500, resultando no t vetor HSP81.1-AtbHLH39-pBI500.

[00121]Construto UCC2-AtbHLH39:

[00122]Gene uclacyanin2 de *Arabidopsis* (UCC2, At2g44790) é expresso em níveis muito altos nas raízes. Sua expressão é detectável mas em níveis muito baixos em outras partes da planta. Seu perfil de expressão específico em célula na raiz é semelhante ao de AtbHLH39, ou seja, predominantemente expresso em endoderme e córtex e estela. A sequência promotora do gene UCC2 (-1 até -1475 com relação ao códon de início ATG) foi amplificada por PCR usando iniciador direto contendo um sítio Sal I (P790-Sal-F) e iniciador reverso contendo um sítio XbaI (P790-Xb-R). O produto de PCR do promotor UCC2 foi clonado nos sítios Sal I e Xba I no vetor HSP81.1-AtMyb68-pBI300 (vide acima), substituindo o promotor HSP81.1. A sequência de codificação AtbHLH39 foi então clonada neste vetor como um fragmento Xba I-BamH I, substituindo a sequência de codificação AtMyb68. O vetor resultante é chamado de UCC2-AtbHLH39-pBI300.

[00123]Construto BdBS-AtbHLH39-pBI500:

[00124]O gene promotor de biotina sintase de *Brachypodium* (BdBS) (-1 até

-553 com relação ao códon de início ATG) foi amplificado por PCR com um fragmento Sal I-Xba I e foi clonado para substituir o promotor HSP81.1 em HSP81.1-AtbHLH39-pBI500. O vetor resultante foi chamado de BdBS-AtbHLH39-pBI500.

[00125]Construto BdUCC-AtbHLH39-pBI500:

[00126]O homólogo mais perto de *Brachypodium* do gene uclacyanin2 de *Arabidopsis* foi encontrado na sequência genômica super_67. A sequência foi 34% idêntica às regiões alinhadas junto a uclacyanin2 de *Arabidopsis*. As fases de leitura aberta foram determinadas. Um blasto reverso da sequência de proteína traduzida para as proteínas TAIR8 de *Arabidopsis* demonstraram ser mais semelhantes à Uclacyanin1 de *Arabidopsis*, um homólogo próximo de uclacyanin2. Em *Arabidopsis*, uclacyanin1 compartilha o padrão de expressão semelhante a uclacyanin2 mas com uma expressão geral mais fraca. A sequência promotora de promotor homólogo de uclacyanin de *Brachypodium* (BdUCC) (-22 até -1405 em relação ao códon de início ATG) foi amplificada como um fragmento Sal I-Xba I conforme foi clonado para substituir o promotor HSP81.1 em HSP81.1-AtbHLH39-pBI500. O vetor resultante foi chamado de BdUCC-AtbHLH39-pBI500.

Exemplo 4: Clonagem de bHLH39 de *Brachypodium*

[00127]Há três genes bHLH de *Brachypodium* como alta homologia ao bHLH subgrupo 1b de *Arabidopsis* compreendendo bHLH39, bHLH 38, bHLH 100 e bHLH 101. Todos os três homólogos de *Brachypodium* são mais proximamente relacionados a AtbHLH38. Há uma forte evidência EST para o homólogo #1 (super_13.506): a porção 5' é idêntica à sequência em EST DV488230 enquanto que a porção 3' é idêntica à DV488393 que contém polyAs. No entanto, DV488393 ainda contém sequência aparentemente que pertence a um intron. Portanto, a fase de leitura aberta exata permanece ainda a ser determinada. A sequência de codificação putativa (BdbHLH39H1) é clonada por RT-PCR usando iniciador direto BdH1-Xb-F (contendo um sítio Xba I) e iniciador reverso BdH1-Bm-R (contendo um sítio BamH I), e clona-

do nos sítios correspondentes em um vetor para criar Construtos HSP81.1-BdbHLH39H1-pBI500, BdBS-BdbHLH39-BI500 e BdUCC-BdbHLH39H1-pBI500.

[00128]Iniciadores

Tabela 1: Iniciadores de Clonagem

SEQ ID	Nome	Sequência (5' a 3')	Produto de PCR
106	BHLH039FW-XbaI	aaaTCTAGAATGTGTGCATTAGTAC-CTCCATTGTTTC	At-bHLH39
107	BHLH039RV-BamHI	aaaGGATCCTCATATATATGAG-TTTCCACATTCTCATAC	CDS, 0,8 Kb
108	BHLH101FW-XbaI	aaatCTAGAATGGAGTATCCATGGCTGCAG-TCTC	At-bHLH101
109	BHLH101RV-BamHI	aaaGGATCCTTATGATTGGCGTAATCCCAA-GAGC	CDS, 0,7 Kb
110	P790-Sal-F	acgtGTCGAC CTT AGC CAA TGG ATG AGG ATG	AtUCC2 promotor,
111	P790-Xb-R	acgtTCTAGA TTT TTG TTT ACT GTA GAA GAG	1,5 Kb
112	BdGOS-Pm-F10	acgtGTTTAAAC GCA TAG ACT CTC AGC GGA GAG	BdTIF1 promotor,
113	BdGOS-Nh-R	acgtGCTAGC gaaaactcctggtgagagtgg	2,5 Kb
114	NPTII-Nh-F	acgtGCTAGC atgattgaacaagatggattgcac	WT-
115	NPTII-Sal-R	acgtGTCGAC CTG CAG GCA TGC AAG CTT	NPTII e

SEQ ID	Nome	Sequência (5' a 3')	Produto de PCR
		GG	sequência de flaqueamento, 2,1 Kb
116	BdBSp-Sal-F	acgtGTCGAC ctctggatgcctaaacaaacgac	BdBS promotor, 0,5 Kb
117	BdBSp-Xb-R	acgtTCTAGA ggcttttgcggtcggcctg	
118	BdUCCp-Sa-F4	acgtGTCGAC GGA GGT GCA GTT TGC AGC AG	BdUCC promotor, 1,4 Kb
119	BdUCCp-Xb-R4	acgtTCTAGA TAT AGA GAG AGG GTG ATC AAC GA	
120	BdH1-Xb-F	acgtTCTAGA ATG GGG CAC AAG CAG CTG TTC	BdbHLH 39 homólogo, super_13,506 (0,7 Kb)
121	BdH1-Bm-R	acgtGGATCC TCA CTG ATG CAT ATG CAG TCC	

Exemplo 5: Transformações de Planta

[00129]Os Construtos descritos acima foram ou são transformados em *Ara-bidopsis* e *Brachypodium* conforme apropriado. Outras espécies são transformadas

com um vetor apropriado e as plantas transformadas são produzidas.

Tabela 2: Transformação de Construtos

CONSTRUTO	ESPÉCIES ALVOS	Transformado
35S-AtbHLH39-pBI300	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	At, Bn
35S-AtbHLH39-pBI800	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	
35S-AtbHLH101-pBI300	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	At
35S-AtbHLH101-pBI800	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	
35S-AtbHLH38-pBI300	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	
35S-AtbHLH100-pBI300	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	
HSP81.1-AtbHLH39-pBI300	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	At, Bn
HSP81.1-AtbHLH39-pBI500	Monocotiledôneas, i.e. Brachypodium	
UCC2-AtbHLH39-pBI300	Dicotiledôneas i.e.Arabidopsis(At), Brassica(Bn)	
HSP81.1-AtbHLH39-pBI500	Monocotiledôneas, i.e. Brachypodium	
BdBS-AtbHLH39-pBI500	Monocotiledôneas, i.e. Brachypodium	
HSP81.1-AtbHLH39-pBI500	Monocotiledôneas, i.e. Brachypodium	
BdUCC-AtbHLH39-pBI500	Monocotiledôneas, i.e. Brachypodium	
HSP81.1-BdbHLH39H1-pBI500	Monocotiledôneas, i.e. Brachypodium	

CONSTRUTO	ESPÉCIES ALVOS	Transforma- do
BdBS-BdbHLH39-BI500	Monocotiledôneas, i.e. Brachypodium	
BdUCC-BdbHLH39H1- pBI500	Monocotiledôneas, i.e. Brachypodium	

Exemplo 6: Métodos de transformação

[00130]Plantas transgênicas *Arabidopsis* foram feitas pelo método de mergulhar as plantas com flores em uma cultura de *Agrobacterium*, baseado no método de Andrew Bent in, Clough SJ and Bent AF, 1998. Floral dipping: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. As plantas tipo selvagem cresceram sob condições padrões como 16 horas, ciclo de dia de 8 horas de luz a escuro, até que a planta tivesse flores em desenvolvimento e flores abertas. A planta foi invertida por 2 minutos em uma solução de cultura de *Agrobacterium* contendo o Construto de gene apropriado. As plantas foram então deixadas na horizontal em uma bandeja e mantidas cobertas por dois dias para manter a umidade e em seguida ajeitadas e ensacadas para continuar o crescimento e desenvolvimento da semente. A semente madura foi colhida.

[00131]As plantas transformadas T1 foram selecionadas por germinação e crescimento em placas MS contendo 50 µg/ml de canamicina ou um meio apropriado de seleção. Mudas verdes, resistentes a canamicina (Kan^R) foram identificadas após 2 semanas de crescimento e transplantadas para solo. As plantas foram ensacadas para garantir auto-fertilização e a semente T2 de cada planta colhida separadamente. Durante o crescimento das plantas T1, amostras de folhas foram colhidas, o DNA foi extraído e análises de *Southern blot* e PCR foram realizadas.

[00132]As sementes T2 foram analisadas para segregação de Kan^R. A partir destas linhagens que apresentaram um fenótipo resistente 3:1, as plantas T2 sobreviventes cresceram, foram ensacadas durante o conjunto de sementes, e sementes

T3 foram colhidas de cada linhagem. A semente T3 foi novamente usada para análise de segregação Kan^R e aquelas linhagens que apresentaram fenótipo 100% Kan^R foram selecionadas como linhagens homozigotas. Outra análise molecular e fisiológica foi realizada usando mudas T3.

[00133]Plantas transgênicas *Brassica napus*, *Glycine max* e *Zea maize* são produzidas usando transformação mediada por *Agrobacterium* de tecido pecíolo de cotiledônea. As sementes são esterilizadas como a seguir. As sementes são umedecidas com etanol 95% por um curto período de tempo como 15 segundos. Aproximadamente 30 ml de solução esterilizante I são adicionados (70% Javex, 100µl Tween20) e deixados por aproximadamente 15 minutos. A solução I é removida e substituída com 30 ml de solução II (0,25% cloreto mercúrico, 100µl Tween20) e incubada por cerca de 10 minutos. As sementes são rinsadas com pelo menos 500 ml de água estéril duplamente destilada e armazenadas em um prato estéril. As sementes são germinadas em placas de Meio 1/2 MS, pH 5,8, complementada com sacarose 1% e agar 0,7%. Cotiledôneas totalmente expandidas são colhidas e colocadas em Meio I (Murashige orgânico mínimo (MMO), sacarose 3%, 4,5 mg/L benzil adenina (BA), 0,7% fitoagar, pH5,8). Uma cultura de *Agrobacterium* contendo o Construto de ácido nucleico de interesse é crescido por 2 dias em meio AB Mínimo. Os explantes de cotiledônea são mergulhados de modo que somente a parte cortada do pecíolo entre em contato com a solução de *Agrobacterium*. Os explantes são então embebidos em Meio I e mantidos por 5 dias a 24°C, com ciclos de 16,8 hr de luz/escuro. Os explantes são transferidos para o Meio II (Meio I, 300 mg/L timentina) por outros 7 dias e em seguida para o Meio III (Meio II, 20 mg/L canamicina). Qualquer tecido de raiz ou brotos que se desenvolveram neste tempo é dessecado. Transferir os explantes para placas novas de Meio III após 14 -21 dias. Quando regenerado, o tecido de broto se desenvolve o tecido regenerado é transferido para o Meio IV (MMO, sacarose 3%, 10% fitoagar, 300 mg/L timentina, 20 mg/L 20 mg/L

canamicina). Uma vez que o tecido de broto tenha se desenvolvido, o tecido de broto de qualquer calo é mergulhado em 10X IBA e transferido para o Meio V (Murashige e Skooge (MS), sacarose 3%, 0,2 mg/L ácido indol butírico (IBA), 0,7% agar, 300 mg/L timentina, 20 mg/L 20 mg/L canamicina) para enraizar. As plântulas saudáveis são transferidas para o solo. O que foi acima mencionado, com ou sem modificação, é apropriado para a transformação de várias espécies de plantas incluindo *Glycine max*, *Zea maize* e algodão.

[00134]*Glycine max*, *Zea maize* e algodão transgênicos são produzidos usando métodos a base de meio de *Agrobacterium* que são conhecidos aos especialistas na técnica. De modo alternativo, pode ser usado um método de transformação por bombardeamento biolístico de partícula ou não partícula. Um exemplo de transformação biolística não partícula é mostrado no Pedido de Patente U.S. 20010026941. Este método foi usado para produzir plantas transgênicas de *Glycine max* e *Zea maize*. Plantas viáveis são propagadas e linhagens homozigotas são geradas. As plantas são testadas para a presença de fenótipos de tolerância, fisiológicos e bioquímicos conforme descrito em outro ponto.

[00135]A transformação de tecido de planta como *Zea maize*, por exemplo, é conseguida por sonicação de cultura de tecido de calo. O tecido de calo foi produzido como a seguir. *Ears* de milho foram colhidas 18 dias após espigamento e esterilizada na superfície em alvejante 50% v/v por 20 minutos seguido por três lavagens com água estéril destilada. Embriões imaturos variando em tamanho de 2 a 4 mm foram colhidos dos *kernel*s. Os embriões foram colocados em meio MSD_{1.5} (2% sacarose, 1X MS sais de macronutrientes e micronutrientes, vitaminas 1X MS, 1,5 mg / L 2,4-D, 0,8% agar, pH 5,8) com o escutelo voltado para cima. Os embriões foram incubados a 26-28 °C no escuro. As células friáveis de culturas de 2 semanas de vida foram transferidas para meio novo MSD_{1.5} e em seguida incubados a 26-28 °C no escuro. O calo friável foi subcultivado para meio novo MSD_{1.5} a cada 21 dias.

[00136]A transformação do tecido de calo foi realizado conforme prescrito abaixo. O Construto foi introduzido em GV3101 *Agrobacterium* por inoculação de uma única colônia de GV3101 *Agrobacterium* contendo o plasmídeo HPR-GUS em 10 mL de LB unido com 150 µg/mL rifampicina, 100 µg/mL sulfato de gentamicina, e 50 µg/mL canamicina. A cultura foi deixada crescer durante a noite a 28°C com agitação a 200 rpm. O calo de milho foi cortado em pedaços de aproximadamente 3-5 mm em tamanho. A cultura de *Agrobacterium* foi centrifugada a 1500 x g por 10 minutos e lavada duas vezes com 10 mL de MSD_{1.5} líquido (2% sacarose, 1X MS sais de macronutriente e micronutriente, vitaminas 1X MS, 1.5 mg / L 2,4-D, pH 5,8). A bactéria foi ressuspensa em MSD_{1.5} líquido para um OD_{600nm} de 0,25 e 1 mL de *Agrobacterium* diluído ou MSD_{1.5} líquido, para controles negativos, foi colocada em tubos de microcentrífuga de 1,5 mL contendo quatro pedaços de calo adicionado a cada tubo. O calo e a cultura *Agrobacterium* foram sonicados em um Branson 200 Ultrasonic Cleaner por 0, 3, 10, 30, 100, ou 300 segundos com a bactéria ou 0 ou 300 segundos sem a bactéria (em MSD_{1.5} líquido isolado). Após sonicação, o calo foi rasgado sobre papel de filtro estéril e colocado em meio MSD_{1.5}A (meio sólido MSD_{1.5} modificado com 100 µM acetosiringona). O período de co-cultivo foi de 4 dias no escuro a 28°C. O calo foi rinsado em MSD_{1.5} líquido, rasgado sobre papel de filtro estéril, e colocado em meio MSD_{1.5}T (meio sólido MSD_{1.5} modificado com 400 µg/mL Timentina) por 3 dias no escuro a 28°C. Sete dias após a sonicação, o calo foi adicionado a 1 mL de solução de coloração GUS (50 mM NaPO₄, pH 7,0, 0,1% Triton X-100, 1 mM EDTA, 2 mM DTT, 0,5 mg/mL X-GlcA) e deixado para incubar durante a noite a 37 °C. A solução de coloração foi substituída com 1 mL de tampão de fixação (10% de formaldeído, 50% etanol) e incubado por 30 minutos em temperatura ambiente. O tampão de fixação foi substituído com 80% etanol e incubado por 1 hora em temperatura ambiente. O etanol 80% foi substituído com 100% etanol e incubado por 1 hora em temperatura ambiente. O calo foi avaliado para coloração

azul, indicando atividade GUS.

[00137]Outros métodos de transformação de planta são usados quando apropriados e são comumente descritos e conhecidos na técnica.

Exemplo 7: Análise de expressão

[00138]O RNA total foi isolado de 22 linhagens transgênicas 35S-AtbHLH39 e tipo selvagem de linhagens *Arabidopsis*. Aproximadamente 10 µg de RNA total foram carregados em cada linha. O *Northern* foi avaliado com HPR cDNA radiomarcado em solução de hibridização ExpressHyb (Clontech) e exposto usando uma tela de fosfo imagem. Para quantificação os blots foram reavaliados com tubulina, um gene constitutivamente expresso, para um padrão de comparação.

[00139]O nível de expressão de AtbHLH39 foi elevado 4 vezes a 126 vezes superior ao do controle WT Columbia. As linhagens de melhor desempenho (linhagem 34 e linhagem 97) baseadas em tolerância ao calor e secura, tinham aumentos de 50 a 40 vezes na expressão de AtbHLH39, respectivamente.

Exemplo 8: Microarranjo e análise de expressão

[00140]O subgrupo 1b bHLH inclui AtbHLH38, AtbHLH39, AtbHLH100 e AtbHLH101. Os dados de microarranjo de linhagens transformadas como um Construto de superexpressão para aumentar a expressão de MYB68 ou atividade (35S-AtMYB68) mostraram que a expressão de RNA de AtbHLH39 e AtbHLH101 aumentou 32 vezes 15 vezes, respectivamente quando AtMYB68 foi super expresso 37 vezes. Os genes AtbHLH38 e AtbHLH100 não estavam presentes no chip de microarranjo, assim a expressão destes genes intimamente relacionados ainda está por ser determinada. Os quatro membros AtbHLH38, AtbHLH39, AtbHLH100, AtbHLH101 têm alta sequência de homologia no grupo 1b, e podem ser pelo menos parcialmente redundantes.

[00141]A expressão de AtbHLH39 é principalmente co-localizada com AtMYB68 no periciclo da raiz como é o de Myb68 (Birnbaum et al. 2003; Schmid et

al. 2005). A análise do promotor endógeno MYB68 e do promotor AtbHLH39 demonstra que estes promotores se expressam predominantemente no tecido da raiz.

[00142]Outras análises de microarranjo, a expressão do gene de IRT1 não é significativamente afetado nas plantas transgênicas que super expressão AtbHLH39 (Construto 35S-AtbHLH39) sugerindo que o fenótipo de tolerância ao calor conferido pela superexpressão de AtbHLH39 não regula a expressão do gene IRT1 e que o envolvimento relatado de bHLH39 nas respostas à privação de ferro ocorre por uma via separada.

[00143]As características de transcrição dos membros de bHLH subgrupo 1b família de gene de trator de transcrição AtbHLH39, AtbHLH38, AtbHLH100, e AtbHLH101 foram analisadas. AtbHLH39 expressa predominantemente na raiz, enquanto que AtbHLH38, AtbHLH100, e AtbHLH101 têm baixa expressão em tecidos de folha, broto, flor e raiz. A transcrição de AtbHLH38 é pouco detectável em raiz e folha, AtbHLH100 se expressa em nível muito baixo no broto e na flor e AtbHLH101 ainda se expressa em nível muito baixo na raiz e na folha.

Exemplo 9: MYB68 se liga a promotor bHLH39

[00144]Ensaio de troca de mobilidade eletroforética (EMSA) foram realizados para determinar se a proteína MYB68 poderia se ligar à sequência promotora de AtbHLH39. Quatro diferentes segmentos de promotor foram marcados e testados: P1 (-1 a -224, com relação ao códon de início ATG), P2 (-204 a -419), P3 (-401 a -623) e P4 (-601 a -788). A sonda de DNA marcada foi incubada com proteína de fusão purificada MYB68. Um complexo de proteína-DNA foi detectado com a sonda P1 enquanto outras regiões do promotor não apresentaram ligação com MYB68. A formação do complexo P1/Myb68 foi eliminada pela adição de um competidor frio de DNA com a mesma sequência que à sonda, mas não afeta pelos competidores frios de P3. Além disso, nenhum complexo proteína-DNA foi formado quando a sonda P1 foi incubada com uma proteína truncada MYB68 (MYB68CD) na qual os domínios de

ligação a R2R3 DNA foram deletados. Os dados demonstraram que a proteína MYB68 se liga especificamente ao promotor AtbHLH39 na região -1 a -224.

[00145]Para ainda localizar o sítio de ligação MYB68, quatro oligonucleotídeos de dupla fita cobrindo a região promotora -1 a -232 foram testados como competidores nos ensaios de ligação: P11 (-1 a -61), P12 (-52 a -118), P13 (-109 a -178) e P14 (-169 a -232). A ligação de MYB68 à sonda P1 foi abolida pelo competidor P12, mas não foi afetado por outros fragmentos, indicando que o sítio de ligação MYB68 está dentro da sequência -52 a -118.

Exemplo 10: Expressão constitutiva de At-bHLH39 em *Arabidopsis* resulta em aborto de flor reduzido após estresse por calor

[00146]Um experimento foi definido com 14 linhagens transgênicas 35S-AtbHLH39 e um controle tipo selvagem (WT). As plantas cresceram em potes de 2,25 polegadas sob condições ótimas (22°C, 18hr de luz de 200uE, 60% RH) em uma câmara de crescimento até três dias após o aparecimento da primeira flor. Um tratamento de estresse por calor foi aplicado colocando as plantas a 42°C por 2 horas. Uma semana após o período de estresse, as plantas foram avaliadas para um número de flores abortadas. Os resultados são mostrados na Tabela 3. Dez das linhagens transgênicas tiveram o aborto de flor reduzido em relação aos controles WT. Uma das linhagens (34-1) teve flores menos abortadas de maneira estatisticamente significativa em relação ao WT.

Tabela 3: Aborto médio de flor após 2 hr a 42°C (n=6 a 11) $\pm$ SE

Entrada	# de flores abortadas	Aborto de flor como % de WT
97-7	2,5 $\pm$ 0,5	44%
34-1	2,9 $\pm$ 0,3	51%
46-10	3,2 $\pm$ 0,3	56%
57-2	3,6 $\pm$ 0,4	64%
1-7	3,8 $\pm$ 0,2	67%

Entrada	# de flores abortadas	Aborto de flor como % de WT
41-2	4,1±0,3	72%
38-10	4,5±0,7	79%
96-4	4,5±0,5	79%
12-4	4,7±0,5	82%
30-2	4,8±0,5	85%
94-10	5,1±0,9	90%
10-1	5,7±0,5	100%
52-4	6,2±0,7	108%
103-1	6,2±0,6	109%
WT	5,7±0,6	100%

Exemplo 11: Expressão constitutiva de AtbHLH39 em *Arabidopsis* resulta em tolerância à secura das plantas

[00147]A tolerância à secura foi avaliada em seis linhagens transgênicas At-bHLH39 que ainda apresentaram aborto de flor reduzida após estresse por calor. As plantas foram crescidas (pote 5 por 3 polegadas, n=8) sob condições ótimas em uma câmara de crescimento (22°C, 18hr luz de 200uE, 60% RH) até que a primeira flor se abra. O tratamento para secura foi aplicado molhando todas as plantas no mesmo nível de saturação. Mais água foi retida. As plantas foram pesadas diariamente para determinar a perda de peso diária e todas as plantas foram colhidas no dia quatro de tratamento, em cujo tempo todas as plantas estavam visivelmente murchas. A perda de água em relação à biomassa do broto final foi calculada e é um indicador representativo de tolerância à secura. Os dados foram normalizados para WT que foi ajustado como 100% (Tabela 4). Todas as seis linhagens mostraram perda de água reduzida em relação à biomassa de broto, 3 das quais foram estatisticamente significativas. Isto é indicativo de fenótipo de tolerância à secura. Duas das linhagens que melhor desempenharam tolerância à secura foram ainda as

que melhor se desempenharam tolerância ao calor indicando uma relação próxima entre os dois traços: tolerância à seca e tolerância ao calor como resultado de uma expressão constitutiva de *AtbHLH39* em *Arabidopsis*.

Tabela 4: Perda de água em relação ao peso seco do broto e tolerância à seca em linhagens transgênicas 35S-bHLH39

Entrada	Perda da água em 3d/broto DW d4	Tolerância à seca (% de WT)	Broto DW d4 (g)
34-1	125±3	128%	0,56±0,01
97-7	140±5	119%	0,51±0,02
94-10	150±5	113%	0,46±0,02
65-1	153±6	112%	0,47±0,02
30-2	162±4	106%	0,43±0,01
1-7	166±3	104%	0,43±0,01
WT	173±5	100%	0,40±0,01

Exemplo 12: Expressão constitutiva de *At-bHLH101* em *Arabidopsis* resulta em aborto de flor reduzido após estresse por calor

[00148]Um experimento foi estabelecido com 17 linhagens transgênicas 35S-*AtbHLH101* e um controle WT. As plantas foram crescidas em potes de 2,25 polegadas sob condições ótimas (22C, 18hr de luz de 200uE, 60% RH) em uma câmara de crescimento até três dias após o aparecimento da primeira flor. Um tratamento de estresse por calor foi aplicado colocando as plantas a 42°C por 1,75 hours. Uma semana após o período de estresse as plantas foram avaliadas para o número de flores abortadas. Os resultados são mostrados na Tabela 5 e mostram que dez das linhagens transgênicas tinham pelo menos 10% de aborto de flor reduzidos em relação aos controles WT e duas linhagens tiveram uma redução de 50% em aborto de flor em relação aos controles WT.

Tabela 5: Aborto de flor após estresse por calor (n=12)

Entrada	# de flores abortadas	Aborto de flor (% de WT)
61-1	1,3±0,3	41%
88-3	1,3±0,4	43%
18-1	2,0±0,5	65%
90-1	2,3±0,5	73%
15-2	2,5±0,5	81%
49-5	2,5±0,5	81%
19-4	2,7±0,4	86%
70-8	2,7±0,3	86%
75-2	2,7±0,5	86%
16-2	2,8±0,4	89%
39-11	2,8±0,4	92%
11-1	2,9±0,4	95%
81-3	3,0±0,4	97%
79-6	3,1±0,4	100%
97-2	3,1±0,4	100%
45-2	3,2±0,5	103%
9-1	3,2±0,4	103%
WT	3,1±0,3	100%

Exemplo 13: Expressão constitutiva de AtbHLH101 em *Arabidopsis* resulta em tolerância à secura das plantas

[00149]A tolerância à secura foi avaliada em dez linhagens transgênicas At-bHLH101 que ainda apresentaram aborto reduzido de flor após estresse ao calor. As plantas cresceram (ponte 5 por 3 polegadas) sob condições ótimas em uma câmara de crescimento (22°C, 18hr de luz de 200uE, 60% RH) até que a primeira flor se abrisse. O tratamento para secura foi aplicado em seguida molhando todas as plan-

tas para o mesmo nível de saturação. Mais água foi retida. As plantas foram pesadas diariamente para determinar a perda de água e todas as plantas foram colhidas no dia quatro de tratamento, em cujo tempo todas as plantas estavam visivelmente murchas. A perda de água em relação à biomassa de broto final foi calculada como um indicador representativo de tolerância à secura. Os dados foram normalizados para WT que foi definido como 100% (Tabela 6). Oito das dez linhagens avaliadas mostraram algum grau de tolerância à secura e uma linhagem teve uma tolerância à secura estatisticamente significativo de 27% superior em relação a WT. Esta linhagem foi ainda uma das quatro que apresentou melhor desempenho de tolerância ao calor como observado pelo reduzido aborto de flor.

Tabela 6: Perda de água em relação ao peso seco do broto e tolerância à secura em linhagens transgênicas 35S-bHLH101

Entrada	Perda de água 3d/brotoDW d4	Tolerância à secura (% WT)	Broto DW d4 (g)
90-1	147±5	127%	0,48±0,02
16-2	174±6	114%	0,41±0,02
61-1	176±5	113%	0,40±0,01
70-8	177±7	112%	0,41±0,02
6-5	181±3	110%	0,39±0,01
11-1	187±7	107%	0,38±0,02
45-2	189±8	106%	0,37±0,02
19-4	191±5	105%	0,37±0,01
79-6	203±7	100%	0,35±0,01
82-4	203±3	99%	0,34±0,01
WT	202±6	100%	0,34±0,01

Exemplo 14: Expressão constitutiva de At-bHLH39 em *Arabidopsis* resulta em rendimento aumentado de semente em relação a um controle tipo selvagem

após o estresse ao calor

[00150]As plantas cresceram (pote 3 por 3") sob condições ótimas em uma câmara de crescimento (22°C, 18hr luz de 200uE, 60% RH) até florescerem. Nas plantas em floração são divididas em dois grupos onde o primeiro grupo é exposto ao estresse por calor (todas as plantas floresceram dentro de dois dias) e o segundo grupo é mantido sob condições ótimas até a maturidade. O tratamento ao estresse por calor consiste de uma exposição diária a 45°C. As temperaturas são elevadas de 22 a 45°C em um período de uma hora e mantidas a 45°C por um período de tempo de 2hr a 3 hr. Os tratamentos de estresse por calor diários são aplicados por um período de 10 dias. Após o estresse de aquecimento, as plantas são devolvidas às condições ótimas e crescem até amadurecerem. Todas as plantas são colhidas na maturidade e o rendimento final das sementes é determinado. Comparando o rendimento (como % do ótimo) de plantas transgênicas aos WT o grau de proteção de rendimento é calculado.

Tabela 7: QUADRO DE REFERÊNCIA ID SEQUÊNCIA ID

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
ARABIDOPSIS THALIANA	1	bHLH39	AT3G56980	nucle- otídeo	777
AVENA SATIVA	2	bHLH39	CN817002	nucle- otídeo	222
BRACHYPODIUM DISTACHYON	3	bHLH39	su- per_13.506_gen	nucle- otídeo	2091
BRACHYPODIUM DISTACHYON	4	bHLH39	su- per_13.506_cds	nucle- otídeo	729
BRASSICA NAPUS	5	bHLH39	EE515575	nucle- otídeo	587

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
BRASSICA NAPUS	6	bHLH39	TC84782	nucle- otídeo	595
BRASSICA NAPUS	7	bHLH39	TC88840	nucle- otídeo	631
BRASSICA RAPA	8	bHLH39	Contig2	nucle- otídeo	693
GLYCINE MAX	9	bHLH39	TC269627	nucle- otídeo	723
HORDEUM VULGARE	10	bHLH39	AK251746	nucle- otídeo	738
PANICUM VIRGATUM	11	bHLH39	Contig2	nucle- otídeo	723
SOLANUM LYCOPERSICUM	12	bHLH39	DV105842	nucle- otídeo	351
TRITICUM AESTIVUM	13	bHLH39	TC358765	nucle- otídeo	714
TRITICUM AESTIVUM	14	bHLH39	TC343683	nucle- otídeo	617
TRITICUM AESTIVUM	15	bHLH39	TC300244	nucle- otídeo	434
TRITICUM AESTIVUM	16	bHLH39	CA618726	nucle- otídeo	261
ZEA MAYS	17	bHLH39	TC429418	nucle- otídeo	228

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
ARABIDOPSIS THALIANA	18	bHLH39	AT3G56980	proteína	258
AVENA SATIVA	19	bHLH39	CN817002	proteína	158
BRACHYPODIUM DISTACHYON	20	bHLH39	su- per_13.506_ORF	proteína	242
GLYCINE MAX	21	bHLH39	TC269627	proteína	241
HORDEUM VULGARE	22	bHLH39	AK251746	proteína	246
PANICUM VIRGATUM	23	bHLH39	Contig2	proteína	241
SOLANUM LYCOPERSICUM	24	bHLH39	DV105842	proteína	117
TRITICUM AESTIVUM	25	bHLH39	TC358765	proteína	238
TRITICUM AESTIVUM	26	bHLH39	TC343683	proteína	206
TRITICUM AESTIVUM	27	bHLH39	TC300244	proteína	144
TRITICUM AESTIVUM	28	bHLH39	CA618726	proteína	87
ARABIDOPSIS THALIANA	29	bHLH38	AT3G56970	nucle- otídeo	762
BRASSICA NAPUS	30	bHLH38	TC95626	nucle- otídeo	753
BRASSICA	31	bHLH38	AM061155	nucle-	765

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
OLERACEA				otídeo	
BRASSICA RAPA	32	bHLH38	EX134222	nucle- otídeo	435
CICER ARIETINUM	33	bHLH38	FE670123	nucle- otídeo	300
HORDEUM VULGARE	34	bHLH38	TC164142	nucle- otídeo	501
HORDEUM VULGARE	35	bHLH38	BAF30424.1	nucle- otídeo	759
MEDICAGO TRUNCATULA	36	bHLH38	TC127269	nucle- otídeo	767
MEDICAGO TRUNCATULA	37	bHLH38	TC115041	nucle- otídeo	780
MEDICAGO TRUNCATULA	38	bHLH38	AJ496888	nucle- otídeo	300
PANICUM VIRGATUM	39	bHLH38	Contig1	nucle- otídeo	690
POPULUS	40	bHLH38	TC89850	nucle- otídeo	549
POPULUS	41	bHLH38	EEF05011.1	nucle- otídeo	795
POPULUS	42	bHLH38	EEE91492.1	nucle- otídeo	474
RICINUS COMMUNIS	43	bHLH38	EEF30834	nucle- otídeo	774

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
RICINUS COMMUNIS	44	bHLH38	EEF30835	nucle- otídeo	555
SOLANUM LYCOPERSICUM	45	bHLH38	TC194645	nucle- otídeo	717
SORGO BICOLOR	46	bHLH38	TC117663	nucle- otídeo	642
TRITICUM AESTIVUM	47	bHLH38	TC337566	nucle- otídeo	672
TRITICUM AESTIVUM	48	bHLH38	CA650144	nucle- otídeo	390
TRITICUM AESTIVUM	49	bHLH38	CA502657	nucle- otídeo	459
VIGNA UNGUICULATA	50	bHLH38	FF388259	nucle- otídeo	732
VITIS VINIFERA	51	bHLH38	CAO17950.1	nucle- otídeo	1563
VITIS VINIFERA	52	bHLH38	CAN79614	nucle- otídeo	735
ARABIDOPSIS THALIANA	53	bHLH38	AT3G56970	proteína	253
BRASSICA NAPUS	54	bHLH38	TC95626	proteína	251
BRASSICA OLERACEA	55	bHLH38	AM061155	proteína	255
BRASSICA RAPA	56	bHLH38	EX134222	proteína	145
CICER	57	bHLH38	FE670123	proteína	100

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
ARIETINUM					
HORDEUM VULGARE	58	bHLH38	TC164142	proteína	167
HORDEUM VULGARE	59	bHLH38	BAF30424.1	proteína	252
MEDICAGO TRUNCATULA	60	bHLH38	TC127269	proteína	246
MEDICAGO TRUNCATULA	61	bHLH38	TC115041	proteína	260
MEDICAGO TRUNCATULA	62	bHLH38	AJ496888	proteína	100
ORYZA SATIVA	63	bHLH38	NP_001045424.1	proteína	247
PANICUM VIRGATUM	64	bHLH38	Contig1	proteína	230
POPULUS	65	bHLH38	TC89850	proteína	183
POPULUS	66	bHLH38	EEF05011.1	proteína	264
POPULUS	67	bHLH38	EEE91492.1	proteína	158
RICINUS COMMUNIS	68	bHLH38	EEF30834.1	proteína	257
RICINUS COMMUNIS	69	bHLH38	EEF30835.1	proteína	184
SOLANUM LYCOPERSICUM	70	bHLH38	TC194645	proteína	239
SORGO BICOLOR	71	bHLH38	TC117663	proteína	214
TRITICUM	72	bHLH38	TC337566	proteína	224

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
AESTIVUM					
TRITICUM AESTIVUM	73	bHLH38	CA650144	proteína	130
TRITICUM AESTIVUM	74	bHLH38	CA502657	proteína	153
VIGNA UNGUICULATA	75	bHLH38	FF388259	proteína	244
VITIS VINIFERA	76	bHLH38	CAN64266.1	proteína	245
VITIS VINIFERA	77	bHLH38	CAO17950.1	proteína	520
VITIS VINIFERA	78	bHLH38	CAN79614.1	proteína	244
ARABIDOPSIS THALIANA	79	bHLH101	AT5G04150	nucle- otídeo	723
BRASSICA OLERACEA	80	bHLH101	AM060621	nucle- otídeo	546
BRASSICA RAPA	81	bHLH101	Contig1	nucle- otídeo	678
ORYZA SATIVA	82	bHLH101	CI296230	nucle- otídeo	261
ORYZA SATIVA	83	bHLH101	TC345105	nucle- otídeo	450
ARABIDOPSIS THALIANA	84	bHLH101	AT5G04150	proteína	240
BRASSICA OLERACEA	85	bHLH101	AM060621	proteína	182
BRASSICA RAPA	86	bHLH101	Contig1	proteína	226

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
ORYZA SATIVA	87	bHLH101	CI296230	proteína	87
ORYZA SATIVA	88	bHLH101	TC345105	proteína	150
ARABIDOPSIS THALIANA	89	bHLH100_ 2	AT2G41240.2	nucle- otídeo	726
ARABIDOPSIS THALIANA	90	bHLH100_ 1	AT2G41240.1	nucle- otídeo	729
ORYZA SATIVA	91	bHLH100_ 1	TC340917	nucle- otídeo	621
PANICUM VIRGATUM	92	bHLH100_ 1	FL920216	nucle- otídeo	711
SORGO BICOLOR	93	bHLH100_ 1	TC113263	nucle- otídeo	732
TRITICUM AESTIVUM	94	bHLH100_ 2	TC317240	nucle- otídeo	417
TRITICUM AESTIVUM	95	bHLH100_ 1	TC303529	nucle- otídeo	705
TRITICUM AESTIVUM	96	bHLH100_ 1	CD865039	nucle- otídeo	691
ZEA MAYS	97	bHLH100_ 1	TC409749	nucle- otídeo	465
ARABIDOPSIS THALIANA	98	bHLH100_ 2	AT2G41240.2	proteína	241
ARABIDOPSIS THALIANA	99	bHLH100_ 1	AT2G41240.1	proteína	242
ORYZA SATIVA	100	bHLH100_ 1	TC340917	proteína	207

<u>ESPÉCIE</u>	<u>SEQ</u> <u>ID NO:</u>	<u>bHLH</u>	<u>Referência</u>	<u>Tipo de</u> <u>Seq.</u>	<u>Compri-</u> <u>mento</u>
		1			
PANICUM VIRGATUM	101	bHLH100_ 1	FL920216	proteína	237
SORGO BICOLOR	102	bHLH100_ 1	TC113263	proteína	244
TRITICUM AESTIVUM	103	bHLH100_ 2	TC317240	proteína	139
TRITICUM AESTIVUM	104	bHLH100_ 1	TC303529	proteína	235
ZEA MAYS	105	bHLH100_ 1	TC409749	proteína	155

Sequências

>SEQIDNO:1

ATGTGTGCATTAGTACCTCCATTGTTTCCAACTTTGGGTGGCCATCAACGGGA
 GAGTACGACAGCTACTACCTCGCCGGAGATATCCTCAACAACGGCGGGTTTCTT
 GATTTTCCGGTACCGGAGGAGACTTATGGAGCTGTTACAGCGGTGACTCAACAT
 CAGAATAGCTTTGGTGTTTCTGTTTCGTCTCGGAGGGAAATGAAATAGACAACAATC
 CGGTGGTCGTCAAGAAGCTTAATCACAATGCTAGTGAGCGTGACCGTCGCAGG
 AAAATTA ACTCTTTGTTCTCATCTCTCCGTT CATGTCTTCCTGCCTCTGGCCAATC
 GAAGAAGCTAAGCATTCTGCGACGGTTTCTCGAAGCTTGAAGTACATACCAGA
 GCTGCAAGAGCAAGTGAAGAAGCTAATAAAAAAGAAGGAAGAGCTCTTGGTGCA
 AATTT CAGGTCAAAGAAACACTGAATGTTACGTTAAGCAGCCACCAAAGGCCGT
 CGCGAATTATATCTCGACCGTTTCTGCGACTAGGCTTGGTGACAACGAAGTGAT
 GGTCCAAATCTCATCGTCCAAGATTCATAACTTTTCGATATCTAATGTTTTAAGTG
 GGTTAGAAGAAGATAGGTTTGTCTTGTGGACATGTCATCTTCAAGGTCTCAAG

GAGAAAGGCTTTTCTACACTTTGCATTTACAAGTGGAGAAGATTGAAAATTACAA
GCTGAATTGCGAAGAGTTAAGTCAGAGGATGTTGTACTTGTATGAGGAATGTGG
AAACTCATATATATGA

>SEQIDNO:2

GCACGAGGCCTCCCTCCGCTCCCTCCTCCCCGACACCGATCACAGCAAGAAGC
TGAGCATCCCCATCACTGTGACGCGGGTGCTCAAGTACATCCCGGAGCTGCAG
AAGCAGGTGGACACGCTGGAGAAGAAGAAGGAAGAGCTGACCCAGGCGAACT
GCAAACCAGGAGTTGTGGCCATGAAGGAGAACACGGCTCCGATCGTGTCCGCC
ACCTGCCTCGA

>SEQIDNO:3

ATGGGGCACAAGCAGCTGTTCGTGGACGACCCGTTGCGGAGCAGCATCTCGTC
TCTGGAGGCGGAGGCCATCTTCTCCGGCGCCGGCGGGCAGTGGCGCGCCGG
CGGCGGCCTCGACGACCGTGACCTCTCCGCCATGCCGGCGGCGGCCAACACC
TCGTGGGCGGCTCCGGCTCTCCCGGCGGCGGCGGCAGGAAGATGAGCCACA
ACGCGTACGAGCGCGACCGCCGCAAGCAGCTCAACGAGCTCTACTCCTCCCTC
CGCTCCCTCCTCCCCGACGCCGACCACACCGTATGCAAATCAAATTGAAGCCAT
AGATCATAATTTGATCCTGAATCCTGATGGATCTGGTGATGATTTGACTAATTGC
AGAAGAAGCTGAGCATCCCGATCACAGTGTGCGGGTGCTAAAGTACATCCCG
GAGCTGCAGAAGGAAGTGGACGGGCTGGAGAGGAAGAAGGAGGAGCTGACGC
GCGCCAACTGCAAGCCCGGGGTGATCGCCATGAAGGACCAGAACGTTGCTCCC
GTCGTCTCCGCGACCTGCCTCGACGACAAGGACATCATGGTTCAGGTCAGCTT
GCTCAGCGGCATGGCGGCGGCGGCTCTGCCGATGTCCACGTGCATAAAGATTG
TGGAGAACGAAGGTCTTCGCCTCGTCAGCTCGTCCACTTCTGCCTTTGGGAACA
GGACGTTCTATAACCTCCATCTTCAGGTAATTGGTACATCTGTCTGCATGAAGCC
TTAATTTCTATTGGTAATTATCAATGTCATCGATCCATGCTTGCTCGATTCATTT

TGGCAAATTTGCCTCATACTTACTCGACATCGTAGTAGAAGAGAAGGAAAAAAAAA
AAGCCCAGCATTTTTCTTTGAGAAACAAAGCACAAACATTTTCTGCTCCCAAGATC
CATTCCGAAAGCCGGGTGCACAACACTACGAGTAAAAATCGTTTCCTCTTAGTTA
TATAACGTGTGGTACCTTCAACTTAAATGAGCATCAGTTGATGCAGAGTGGTACC
CCTGTTTCGAAAAGACTTCAACATGGCACCATCTTCACTAGAGGTCCCCCAGCC
CCTAATTCAGACGAAGGCATCATTTTAGTCCATTAACCCCGGTAGGTTGAGATT
CAAAGGCTAGATTTATATAACGGTTTTGCTATTTCTAAGGCGACTGAGAAATAAG
TATTTCTAACAACCATTTGATCCAATCACTGAGATTCTGTGCAAGATCCATACATAT
ACAAAGTATTAGAAAACCTTAATCGGATGGATATGATCGTTGACTGAGAAATAGA
CACTTCTTTTATGTAAAGGGTGTTTACTATATATATGTGGGTTTTGATATACTCAT
ATTGAATAGATTTGAAGATCATGTCAGGCCCTACTTAGGGTAATCTGTTGTGAAA
TTAAACTGTGTTTCATAGGGCAAAAACATTGTCTTATAAATCAGCACAAAATCAAC
GAATTGGGAGTTTTTACGTAACAAATATAAATTGTAGCAACACAAATTAATTGGCT
ACAATACAATCCAAGAACAACAAGACGAGTATACACGGCAAACGATCATGCATA
TGGTGAGTTGGTGACCAGATCACCCGCTCTACTAGATGCTCCTAGATGCATATG
GTGCCATAATTTTAGGAAAACCGAACTGGAAGAAGCACCTAAAAACGAGTATG
AAATCATCAATCGACCAAGATTAGAGGTGGTCTCATCATCGACAACATGAACAA
TTAAGATGCATGCTGGAGAACAGATAATCTAACACAGCCACAGGTTTATTACAAA
AAGCTTAACAGAACTTCGCTAAAGCAACCAAAGAATGAGAACAAAAAATATAT
CTTCTAATAACATGTGTGTGCTGTTGTATAGCATCTGAAGGCGTAATGCGAACT
CTAATTTATCTGAAGTATGTAGTGCTTATATGCTTATATAACATGTAAATAAGCAA
TATATATTCATTAATTTCAATTAATTTGTATACTGAAACAGAGAAACCAGCGAACG
ATGAGCAAGGAGTGCCCAGCGTTCTGTGACGAGCTGGAGAAAGCCATCAAGAA
AAAGGCAGGACTGCATATGCATCAGTGA

>SEQIDNO:4

ATGGGGCACAAGCAGCTGTTCTGTGGACGACCCGTTGCGGAGCAGCATCTCGTC

TCTGGAGGCGGAGGCCATCTTCTCCGGCGCCGGCGGGCAGTGGCGCGCCGG
CGGCGGCCTCGACGACCGTGACCTCTCCGCCATGCCGGCGGCGGCCAACACC
TCGTGGGCGGCTCCGGCTCTCCCGGCGGCGGCGGCAGGAAGATGAGCCACA
ACGCGTACGAGCGCGACCGCCGCAAGCAGCTCAACGAGCTCTACTCCTCCCTC
CGCTCCCTCCTCCCCGACGCCGACCACACCAAGAAGCTGAGCATCCCGATCAC
AGTGTCGCGGGTGCTAAAGTACATCCCGGAGCTGCAGAAGGAAGTGGACGGG
CTGGAGAGGAAGAAGGAGGAGCTGACGCGCGCCAACTGCAAGCCCGGGGTGA
TCGCCATGAAGGACCAGAACGTTGCTCCCGTCGTCTCCGCGACCTGCCTCGAC
GACAAGGACATCATGGTTCAGGTCAGCTTGCTCAGCGGCATGGCGGCGGCGG
CTCTGCCGATGTCCACGTGCATAAAGATTCTGGAGAACGAAGGTCTTCGCCTCG
TCAGCTCGTCCACTTCTGCCTTTGGGAACAGGACGTTCTATAACCTCCATCTTCA
GAGAAACCAGCGAACGATGAGCAAGGAGTGCCCAGCGTTCTGTGACGAGCTGG
AGAAAGCCATCAAGAAAAAGGCAGGACTGCATATGCATCAGTGA

>SEQIDNO:5

AAAAAAAAATTAGAAAAGAAGAAAAGAGTTTATCGGGTCTCTCTCACGAGTCACG
GCGTCGACGAACTGGAGGTGAAGCCATCTCCGATCGAGAGTCGCCGTCATCC
CTATCCCACTCCTGCAATATCTACTCCGACGCTGTGTCGTGTTAATTTCCGGCACCT
CCTCGCCAAGTCGCCATCGTCCCTGTGCTGTGCGAGCTGTCACTGTCCATGCA
ACGAATAAATACTCTGTGTCTGCATCGCTTTGACAGTGCTTGGTGCTTCTATGAT
ATGTCCTCCGGTTTATGCTTTGCCAGATTCCGAAGAAGATGGGTTTGTTCCTTGT
TGATGTTTCATCTTCTAGGTCTCATGGAGAAAGGCTCGTCTACAGTTTGCATCTT
CAAATGGGAACATAAATAATCACGAGCTGACGTGCGAAGAGCTAAGCCAGAGA
ATGTATACTTGTATGAGGAATGCGGAACTCGTTTAGATGATAATCTGTTCTTGT
TTCTTTTAGTTATGTCATCTGTTTCTCAACATGTAACATTCATGTAGCCGAGTTGT
TTCGTTTATTTTTCTTGTGAAATCAATGATCGATTATCGC

>SEQIDNO:6

GACGGCGTCGACGAACTGGAGGTGAAGCCATCTCCGATCGAGAGTCGCCGTC
ATCCCTATCCCACTCCTGCAATATCTACTCCGACGCTGTCGTCGTTAATTTCGGC
ACCTCCTCGCCAAGTCGCCATCGTCCCTGTCGTGTAAGCCCTAGGTCTGATTTG
TCTTTCCTTTTAAATCGACGAAATTGAAATTGTGCGAGCTGTCACTGTCCATGCA
ACGAATAAATACTCTGTGTCTGCATCGCTTTGACAGTGCTTGGTGCTTCTATGAT
ATGTCCTCCGGTTTATGCTTTGCCAGATTCCGGAAGAAGATGGGTTTGTTCCTTGT
TGATGTTTCATCTTCTAGGTCTCATGGAGAAAGGCTCGTCTACAGTTTGCATCTT
CAAATGGGAAACATAAATAATCACGAGCTGACGTGCGAAGAGCTAAGCCAGAGA
ATGTATACTTGTATGAGGAATGCGGAAACTCGTTTAGATGATAATCTGTTCTTGT
TTCTTTTAGGTTATGTCATCTGTTTCTCAACATGTAACATTCATGTAGCCGAGTTG
TTTCGTTTATTTTCTTGTGCGAAATTCAATGATCGATTATCACATTTAC

>SEQIDNO:7

GACGACAAAAAAAAAATTAGAAAAGAAGAAAAGAGTTTATCGGGTCTCTCTCAC
GAGTCACGGCGTCGACGAACTGGAGGTGAAGCCATCTCCGATCGAGAGTCGC
CGTCATCCCTATCCCACTCCTGCAATATCTACTCCGACGCTGTCGTCGTTAATTT
CGGCACCTCCTCGCCAAGTCGCCATCGTCCCTGTCGTGTAAGCCCTAGGTGCG
AGCTGTCACTGTCCATGCAACGAATAAATACTCTGTGTCTGCATCGCTTTGACAG
TGCTTGGTGCTTCTATGATATGTCTCCGGTTTATGCTTTGCCAGATTCCGGAAG
AAGATGGGTTTGTTCCTTGTGATGTTTTCATCTTCTAGGTCTCATGGAGAAAGGCT
CGTCTACAGTTTGCATCTTCAAATGGGAAACATAAATAATCACGAGCTGACGTGC
GAAGAGCTAAGCCAGAGAATGTATACTTGTATGAGGAATGCGGAAACTCGTTTA
GATGATAATCTGTTCTTGTTCCTTTAGTTATGTCATCTGTTTCTCAACATGTAAC
ATTCATGTAGCCGAGTTGTTTCGTTTATTTTCTTGTGCGAAATCAATGATCGATTAT
CACATTTACATGCTACTGAATATTGACTG

>SEQIDNO:8

ACGACAAAAAAATTAGAAAAGAAGAAAAGAGTTTATCGGGTCTCTCTCACGAG
TCACGGCGTCGACGAACTGGAGGTGAAGCCATCTCCGATCGAGAGTCGCCGT
CATCCCTATCCCCTCCTGCAATATCTACTCTGACGCTGTCGTGTTAATTTCCG
CACCTCCTCGCCAAGTCGCCATCGTCCCTGTCGTGTAAGCCCTAGGTCTGATTT
GTCTTTCTTTTAAATCGACGAAATTGAAATTGTGCGAGCTGTCACTGTCCATGC
AACGAATAAATACTCTGTGTCTGCATCGCTTTGACAGTGCGTGGTGCTTCTATGA
TATGTCCTCTGGTTTATGCTTTGCCAGATTCGGAAGAAGATGGGTTTGTTCTTT
GTTGATGTTTCATCTTCTAGGTCTCATGGAGAAAGGCTCGTCTACAGTTTGCATC
TTCAAATGGGAAACATAAATAATCACGAGCTGACGTGCGAAGAGCTAAGCCAGA
GAATGTATACCTGTATGAGGAATGCGGAACTCGTTTAGATGATAATCTGTTCTT
GTTTCTTTTTGTTATGTCGTCTGTTTCTCAACATGTAACATTCATGTAGCCGAGTT
GTTTCGTTTATTTTCTTGTCGAAATCAATGATCGATTATCACATTTACATGCTACT
GAATATTGACTGATTACTATGAAATTCACTAATAT

>SEQIDNO:9

ATGGTTGCTTTGTTTTCCCCTCCGGTGTTCTCAACCAAGGGATGGTTCTTAGAAG
AAGAGCCATTAAGCTATGATGTGTCTTCAGATTACTCATTTCCCTATCAATTTTT
GCACCACAGACACAGATTGAACTTGAAATAGAAAGGTCCTCTGCACCATCCCCT
GAAGACCCTGCCATGGTCAAAAAGCTTAGCCACAACGCTAGTGAACGTGATCGC
CGCAAGAAGGTTAATGACTTGTTTTCTTCACTTCGTTCACTTCTTCTGCGCCAG
ATCAAACGAAAAAATGAGCATTCCAGCTACAGTTTCGCGAGTTTTAAATACAT
ACCTGAGTTACAACATCAAGTGCAAGCACTAACTAAGAAAAAAGAGGAGCTTCT
GTGCAGAATTTCAAAAAATCTCAAAGGAGATTCGGTGAACAAAGAATCTCAAAG
GAGAATTTCCCATCACAATTCTGATTTTGCTGTTTCACTAGTAGGCTCAACGAT
TGTGAAGCTGTTGTTACATTTCTTATGAGGCTCACAAGGCTCCACTATCCG
ACATCTTGCAATGTTTAGAAAATAATGGCCTTTATTTGCTAAATGCTTCTTCTCT

GAAACTTTTGGAGGAAGGGTCTTCTACAACCTTGCAATTTCCAGGTGGAAAAAACT
CATAGATTAGAGTCCGAAATTCTAACTGAGAAGCTTTTGTCAATATATGAGAAGC
AAAGGATTTTC

>SEQIDNO:10

ATGGGGCACCAGACCCAGATGTTGACGACCCGTTGCGGAGCAGTATGTCGTC
CCTGGACGCAGACATCTTCTCCGTCGCCGGCGGCCTCCACCCATCGCAGTGGC
CGGGACTCGACCACGACGTCTCGCTGGCGCCGGCTGCCAACAACGGCACCTC
CTCCGGCGGGCTACGGCTCCCCCGGGGGCGGCGATGGCTCGGGCTCCCACCG
CAAGATCAGCCACAACGCGTACGAGCGCGACCGCCGCAAGCAGCTCAACGAG
CTCTACTCCGACCTCCGCTCCCTCCTCCCCGACTCCGATCACACCAAGAAGCTG
AGCATTCCGATCACGGTGTGCGGGGTGCTCAAGTACATCCCGGAGCTGCAGAA
GCAGGTGGACGGACTGGAGAAGAAGAAGGAGGAGCTTACGAGGGCCAGCTGC
AAGCCAGGCGTATTGACCATGAAGGAGAACACGGTCCCGATCGTGTCCGCCAC
CTGCCTCGACGAAAGGGAGATCATGGTCCAGGTTAGCTTGGTGAGCACCATGG
CCGGAGCTCTGCCCATGTCCAAGTGCATCAAAGTGCTGGAGAACGAAGGCCTC
CGCCTCATCAGCTCGTCCACTTCTGCTTTCCAGAACAGGACGTTCTATAGCCTC
CATCTTCAGAGAACCCAACGGACGATGAGCAAAGAGTGTCCGGCATTTTGTGAA
GAACTGGAGAATGCCCTGACGCAGAAGGCGGGACTACGTCTACATCACCAG

>SEQIDNO:11

ATGACCCCTTCTCGAGCAGCATCTCGTCGCTGGAGGCGGACATCTTCTCCGCC
GGCGGCCAGCGGCGCGTCGCCGCCGTGGCCGGACCTCGAACTCGACCTCGAC
CTCGACGACGACGACATCCACGACCTCTCCGCGCCGGCGGCCAACGCCACCT
CCTCAGGAGGCTATGGCTCGGGCGGAGGCTCCGGCGGGCTCCCACAGGAAGCT
CAGCCACAACGCGTACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTCTACT
CCTCGCTCCGCTCCCTCCTCCCCGACGCTGACCACACTAAGAAGCTGAGCATC

CCCACCACGGTCTCCCGAGTTCTCAAGTACATCCCCGAGCTGCAGAAGCAGGT
GGACAACCTGGAGAGGAGGAAGAAGGAGCTGACGAACGCCAACTGCAAACCA
GGAGTTCTGAAGACGAGCCAGATTGTAACCTCCATTGTTTCTGCTACCTGCCTC
AACGATACGGAGATCATGGTTCAGGTCAGCCTGCAGAGCAATGTGGCTGCCAC
AAGTCTTCCTCTGTCCAAGTGCATAAAAGTGCTGGAGAACGAAGGCCTTCACCT
GATTAGTTCATCAACTTACTCCACCTTCGACAACAGGACATTCTATAGCCTCCAT
CTTCAGAGAAGTCAAAGAACGATGAAGGAGGAGTGCCCAGCATTCTGCGATGA
ACTGGAGAGGATTATCAAGAAGAAAGCAGGGGCG

>SEQIDNO:12

ATGTTAGCCATTTCTTCTTCTCCTCCTTTATTTTCTACTACTACTAATAATTTT
GGTTGGCTTTTGAAGATCTTATAAGCCATGAATTAACAAATAGTGGAGAACTT
CAAATTCATCTCAAAAAAGCCTTCAACATTGTGATTCAAATAAATTTGATCAAATT
ATTATCAACAGTGGTGATCAGTATCAACCTGATCAGACGGTTAAGAAGCTTAATC
ATAACGCAAGTGAACGTGACCGTAGAAAGAAAATCAACAGCTTATATTCTTCTCT
TCGTTCTTTACTACCTCCTTCTGATCATACGAAAAAGCTAAGCATTCCATCAACA
GTATCAAGAATTCTAAAG

>SEQIDNO:13

ATGGGGCACCAGCACCAGATGTTCAACGACCCCTTCGCGAGCAGCATGTCGTC
ACTGGAGGAAGACATGTTCTCCGGTGCCGGAGGCTACCACCACCTCACGCCGT
CCATGCAGTGGCCGGGCTTGGATAACGACATACCGTCGGCGCCGGCTGCCAAC
AACGCCACCTCCTCCGGTGGCTCTGGATCACACCGCAAGATGAGTCACAACGC
GTACGAACGTGACCGCCGCAAGCAGCTCAACGAGCAATATTCCTCCCTCCGCT
CCCTCCTCCCCGACGATGATCACACCAAGAAGATGAGCATTCCGACCACGGTG
TCGCGGGTGATCAACTACATCCCGGAGCTGCAGAAGGAGGTAGACCGCCTGGA
GAAGAAGAAGGAGGAGCTGAGGCGGGGCAGCTGCGAGCAAGGCGCCATGAGG

CAGAACACGGCCCCGATCGTGTCCGCCACCTGCCTCGACGACAGGGAGATCAT
GGTCCAGGTCAGCCTGGTGAGCACCATGGCGGGAGCTCTGCCCATGTCCAAGT
GCATCAAGGTGCTGGAGAACCAAGGCCTTCGCCTCATAAATTCCTCGACTTCCG
CGTTTCAGAACAGGACGTTCTACAGCCTCCATCTTCAGAGAACCCAACGGACAA
TGAGCAAGGAGGGCCAAACATTTTGTAAACGAATTGGAGAACGCTGTGAAGCAAA
AGGCGGGACTACATCTACATCAT

>SEQIDNO:14

ATGGGGCACCAGCACCAGATGTTCTGAAGACCCGTTTCGCGAGCAGCATATCGTC
GCTGGAGGCCGAGATATTCTCCGTGCGCCGGCGGCCACCACCATACGCAGTGG
CCGGGCCTCGACCACGACATCCCGCTGGCCCCGGCTGCCAATAACGGCACCT
CCTCCGGCGGCTACGGCTCCCCCGGGGGCGGCGATGGCTCGGGCTCCCATCG
CAAGATCAGCCACAACGCCTACGAGCGCGACCGCCGCAAGCAGCTCAACGAGC
TCTACTCCGACCTCCGCTCCCTCCTCCCCGACACCGATCACACGAAGAAGCTTA
GCATTCCGATCACGGTGTGCGGGGTGCTCAAGTACATCCCGGAGCTGCAGAAG
CAGGTGGACGGCCTGGAGAAGAAGAAGGAGGAGCTGACGCGCGCCAGCTGCA
AGCCCGGCGTGCTGACCATGAAGGGGGACACGGCTCCGATCGTGTCCNGCCA
CTGCCTCGACGACAGGGAGATCATGGTCCANNGTCAGCTGGTGAGCACCATGG
GCGGAGTCTGCCATGTCAAGTGCTCAGGTGCTGAGACGAAGGCTCGGCTCATA
GTCGTCACTCCGGTTCAGACAGACTCTATATTCATCTC

>SEQIDNO:15

ATGGGGCACCAGCACCAGATGTTCTGAAGACCCGTTTCGCGAGCAGCATATCGTC
GCTGGAGGCCGACATCTTCTCCGTGCGCCGGCGGCCACCACCATCCGCAGTGG
CCGGGCCTCGACCACGACGTCCCGTTGGCGCCGGCTGCCAACAACGGCACAT
CCTCCGGCGGCTACGGCTCCCCCGGTGGCGGGCGACGGCTCGGGCTCCCACCG
CAAGATCAGCCACAACGCGTACGAGCGCGACCGCCGCAAGCAGCTCAACGAG

CTCTACTCCGACCTCCGCTCCCTCCTCCCCGACACCGATCACACGAAGAACTG
AGCATTCCGATCACGGTGTGCGGGGTGCTCAAGTACATCCCGGAGCTGCAGAA
GCAGGTGGACGGAAGTGGAGAAGAAGAAGGAGGAGCTGACGCGCGCCAACTGC
AGCCCGGCGTGCTG

>SEQIDNO:16

AACAACGCCACCTCCTCCGGCGGGCTCTGGATCACACCGAAAGATGAGTCACAA
CGCGTACGAGCGTGACCGCCGCAAGCAGCTCAACGAGCAATATTCCTCCCTCC
GCTCCCTCCTCCCCGATGACGACCACAATAAGAAGATGAGCATTCCGACCACG
GTGTCGCGGGTGATCAAGTACATCCCGGAGCTGCAGAAGGAGGTAGACGGCCT
GGAGAAGAAGAAGGAGGAGCTGAGGCGGGGCCAGCTACGAGCAAGCGCCA

>SEQIDNO:17

CGTGTCCCGAGTTCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGGACAACC
TGGAGAGGAGGAAGAAGGAGCTGACCAACGCCAACTGCAAGCCGGGAGTTCT
GAAAACCACCAAGGCCGTAAGTCCCATTTGTTTCTGCTACCTGCCTCAACGACAC
GGAGATCATGGTTCAGGTCAGCCTGCACAGCGATGTGGCCGCCACAGCTCTCC
CTCTCTCCAAGTGCAT

>SEQIDNO:18

MCALVPPLFPNFGWPSTGEYDSYYLAGDILNNGGFLDFPVPEETYGAVTAVTQHQN
SFGVSVSSEGNEIDNNPVVVKLNHNASERDRRRKINSLFSSLRSCLPASGQSKKL
SIPATVSRSLKYIPELQEQVKKLIKKEELLVQISGQRNTECYVKQPPKAVANYISTVS
ATRLGDNEVMVQISSSKIHNFSISNVLSGLEEDRFVLVDMSSSRSQGERLFYTLHLQ
VEKIENYKLNCEELSQRMLYLYEECGNSYI

>SEQIDNO:19

RHEASLRSLLPDTHSKKLSIPITVTRVLKYIPELQKQVDTLEKKKEELTQANCKPGV
VAMKENTAPIVSATCLDDRDIMVQVSLLSNMAGALPVSKCIKVLENEGLRLVSSSTS
AFQNKTFYSLHVQRTQRTISKVCPAFCDELENAIKRAGMRLQQ

>SEQIDNO:20

MGHKQLFVDDPFASSISSLEAEAIFSGAGGQWRAGGGLDDRDLSAMPAAANTSSG
GSGSPGGGGRKMSHNAYERDRRKQLNELYSSLRSLLPDADHTKKLSIPITVSRVLK
YIPELQKEVDGLERKKEELTRANCKPGVIAMKDQNVAPVVSATCLDDKDIMVQVSL
SGMAAAALPMSTCIKILENEGLRLVSSSTSFAFQNRFTFYNLHLQRNQRTMSKECPAF
CDELEKAIKKKAGLHMHQ

>SEQIDNO:21

MVALFSPPVFSTKGWFLEEEPLSYDVSSDYSFPYQFFAPQTQIELEIERSSAPSPED
PAMVKKLSHNASERDRRKKVNDLVSSLRSLLPDQTKKMSIPATVSRVLKYIPELQ
HQVQALTKKKEELLCRISKNLKGDSVNKESQRRISHHNSDFAVSTSRLNDCEAVVHI
SSYEAHKAPLSDILQCLENNGLYLLNASSSETFGGRVFYNLHFQVEKTHRLESEILTE
KLLSIYEKQRIF

>SEQIDNO:22

MGHQTQMFDDPFASSMSSLDADIFSVAGGLHPSQWPGLDHDVSLAPAANNGTSS
GGYGSPGGGDGSGSHRKISHNAYERDRRKQLNELYSDLRSLLPDSDHTKKLSIPIT
VSRVLKYIPELQKQVDGLEKKKEELTRASCKPGVLTMKENTVPIVSATCLDEREIMV
QVSLVSTMAGALPMSKCIKVLENEGLRLISSSTSFAFQNRFTFYNLHLQRTQRTMSKEC
PAFCEELENALTQKAGLRLHHQ

>SEQIDNO:23

MTPSRAASRRWRRRTSSPPAASGASPPWPDLELDLDDDDIHDLSAPAANATSSG

GYGSGGGSGGSHRKL SHNAYERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTVSR
VLKYIPELQKQVDNLERRKKELTNANCKPGVLKTSQIVTPIVSATCLNDTEIMVQVSL
QSNVAATSLPLSKCIKVLENEGLHLISSSTYSTFDNRTFYSLHLQRSQRTMKEECPA
FCDELERIIKKKAGA

>SEQIDNO:24

MLAISSSSPPLFSTTTNNFGWLLEDLISHELTNSGETSNSSQKSLQHCDSENKFDQIII
NSGDQYQPDQTVKKLNHNASERDRRKINSLYSSLRSLLPDHTKKLSIPSTVSRI
LK

>SEQIDNO:25

MGHQHQMFNDPFASSMSSLEEDMFSGAGGYHHLTPSMQWPGLDNDIPSAPAANN
ATSSGGSGSHRKMSHNAYERDRRKQLNEQYSSLRSLLPDDHTKKMSIPTTVSRVI
NYIPELQKEVDRLEKKKEELRRGSCEQGAMRQNTAPIVSATCLDDREIMVQVSLVST
MAGALPMSKCIKVLENQGLRLINSSTSAFQNRTFYSLHLQRTQRTMSKEGQTFCNE
LENAVKKQAGLHLHH

>SEQIDNO:26

MGHQHQMFEDPFASSISSLEAEIFSVAGGHHHTQWPGLDHDIP LAPAANNGTSSG
GYGSPGGGDGSGSHRKISHNAYERDRRKQLNELYSDLRSLLPDTHHTKKLSIPITVS
RVLKYIPELQKQVDGLEKKKEELTRASCCKPGVLT MKGDTAPIV SXHCLDDREIMVXX
QLVSTMGGVCHVKCSGAETKARLIVVTPVQTD SIFIS

>SEQIDNO:27

MGHQHQMFEDPFASSISSLEADIFSV AAGHHHPQWPGLDHDVPLAPAANNGTSSG
GYGSPGGGDGSGSHRKISHNAYERDRRKQLNELYSDLRSLLPDTHHTKKLSIPITVS
RVLKYIPELQKQVDGLEKKKEELTRAN CSPAC

>SEQIDNO:28

NNATSSGGSGSHRKMSHNAYERDRRKQLNEQYSSLRSLLPDDDHNNKKMSIPTTVS
RVIKYIPELQKEVDGLEKKKEELRRASYEQAP

>SEQIDNO:29

ATGTGTGCATTAGTCCCTTCATTTTTCACAACTTCGGTTGGCCGTCAACGAATC
AATACGAAAGCTATTACGGTGCCGGAGATAACCTAAATAACGGCACATTTCTTGA
ATTGACGGTACCACAGACTTATGAAGTGAATCATCATCAGAATAGCTTGGGAGT
ATCTGTTTCGTCAGAAGGAAATGAGATAGACAACAATCCGGTTGTGGTCAAGAA
GCTTAATCACAATGCTAGTGAACGTGACCGACGCAAGAAGATCAACACTTTGTT
CTCATCTCTCCGTTTCATGTCTTCCAGCTTCTGATCAATCGAAGAAGCTAAGTATT
CCTGAAACGGTTTCAAAGAGCTTAAAGTACATACCAGAGCTGCAACAGCAAGTG
AAGAGGCTAATACAAAAGAAGGAAGAAATTTTGGTACGAGTATCGGGTCAAAGA
GACTTTGAGCTTTACGATAAGCAGCAACCAAAGGCGGTGCGAGTTATCTCTCA
ACGGTTTCTGCCACTAGGCTTGGTGACAACGAAGTGATGGTCCAAGTCTCATCG
TCCAAGATTCATAACTTTTCGATATCAAATGTGTTGGGTGGGATAGAAGAAGATG
GGTTTGTTCTTGTGGATGTTTCATCATCAAGATCTCAAGGAGAGAGGCTCTTCTA
CACTTTGCATCTTCAAGTGGAGAATATGGATGATTACAAGATTAATTGCGAAGAA
TTAAGTGAAAGGATGTTGTACTTGTACGAGAAATGTGAAAACTCGTTTAACTAG

>SEQIDNO:30

ATGTGTGCATTAGTCCCTCCATTGTTCCCAAACCTTTGGGTGGCCGTGACAGGA
GAGTACGAGAGTAACTACCTGGCCGGAGTGAACCTCGAGGACTTTACGTTTCTT
GATTTTCCGGCACCAGAGACATATGGAGTGGAACATCATCAGGAGATTCAGGAA
ATGTTGGGGGTCTCTGTTCCGTCCGAGGGGAATGGAGTTGTAACCAAGAAGCTT
AATCACAATGCTAGTGAGCGTGACCGTCGCAAGAAGATCAACTCTTTGTTCTCG

TCTCTCCGTTTCATGTCTCCCAGCTTCTGATCAAACGAAGAAGCTAAGTATTCCTC
AGACGGTTTCTCGGAGCTTGAAGTACATTCCAGAGCTGCAAGAGCAAGTGAAGA
AGCTAATACAAAAGAAGGAAGAACTCTTGGTGCGAGTATCAGGTCAAAGAGCCA
TTGAACATTATGTTGAGCCGCAGCCAAAGGCCGTTGCACGTTACGTCTCGACCA
TTTCTGCGACTAAGCTTGGAGACAACGAAGTGCTGGTCCAAATCTCATCGTCCA
AGAATCATAACTTTTCGATATCTAATGTGTTGAGTGGGTAGAAGAAGATGGGTT
TGTTCTTGTTGATGTTTCATCTTCCAGGTATCATGGAAAATGGCTCTTCTACTCTT
TGCATCTTCAAATGGGAAATAAAGATAATCACAACTGAAGTGCGAAGAGCTAA
GCCAGAGAATTTTGTACTTGTATGAGGAATGTGAAAACCTCATTTAGA

>SEQIDNO:31

ATGTGTGCATTAGTCCCTCCACTGTTCCCCGACTTTGGGTGGCCGTGACGGCA
GGTTACGAGAGCTACTACCTCGGCGGAGAAAACCTCAACAACGACATGTTTCTT
GATTTTCCGGTTGTGGAACTTATGGAGTATTGGCTCATCATCAGAACAGCTTAG
GAGTTTCTGTTTCGTGCGAGGGAAATGGAATAGACAACAACCCGGTTGTTAAAA
AGAAGCTTAATCACAATGCTAGTGAGCGTGACCGTCGCAAGAAGATCAACTCTT
TGTTTGCATCTCTCCGCTCATGTCTTCCAACCTCAGATCAATCGAAAAAGCTAAG
CATTTCAGCCACCGTTTCACGAAGCTTGAAGTACATACCAGAGTTGCAAGAGCA
AGTGAAGAAGTTATTACAAAAGAAGGAAGAACTCTTGGTTTCGAGTATCAGGTCA
ACGAGACATTGAACTTTACGTTAAGCCACAACCAAAGGCAATTGCAAGTTATGTC
TCCACTGTTTCCGCGACTAGGCTTGGAGACAACGAAGTGATGGTCCAAATCTCA
TCATCCAAGATTCATAACTTCTCGATATCTAAAGTGTTAACTGGGTAGAAGAAG
ATGGTTTTGTTCTTGTGGATGTTTCATCTTCAAGGTTTCAAGGGGAAAGGCTTTT
CTACACTTTGCATCTTCAAGTAGAAAATATGGATGATCATTACAAAATGAATTGC
GAAGAGTTAAGTGAAAGGATGTTGTACTTGTACGAGGAATGTGAAAATTNNNTTA
GG

>SEQIDNO:32

ATTCCTCAGACGGTTTCTCGGAGCTTGAAGTACATACCAGAGCTACAAGAGCAA
GTGAAGAAGCTAATACAAAAGAAGGAAGAACTCTTGGTGCGAGTATCAGGTCAA
AGAGACATTGAACATTACGTTGAGCCGCACCCAAAGGCCGTTGCACGTTACGTC
TCGACCATTTCTGCGACTAAGCTTGGAGACAACGAAGTGATGGTCCAAATCTCA
TCGTCCAAGAATCATAACTTTTCGATATCTAATGTGTTGAGTGGGTTAGAAGAAG
ATGGGTTTGTTCCTTGTGATGTTTCATCTTCAAGGTCTCATGGAGAAAGGCTCTT
CTACACTTTGCATCTTCAAATGGGAAATAAAGATGATTACAACTGACATGCGAA
GAGCTACGCCAGAGAATGTTATACTTGTATGAGGAATGTGGAAACTCGTTTAGA

>SEQIDNO:33

ATTCCAACCTCCCTCACAAGCCACAAGCAGCGACCTTAGCATGGTCAAGAACTT
ATCCGCAATGCTAGTGAACGAGATCGCCGCAAGAAAATCAATACTTTGTATTCTT
CACTTCGTTCACTTCTTCCTGTGGCAGAACAGATGAAGAAGTTGAGCAATCCGG
CAACAATTTACGAGTCCTAAAGTACATACGTGAGTTACAGAAGCAGGTAGAAG
GACTACTTACGAGAAAGGAGGCGATTTTATTGAACTATCTCCAGAAGTAGATGA
GGTGAAGAGTAAAGAATCTGAGAGGAAG

>SEQIDNO:34

AGCGACCGCCGCAAGCAGCTCAACGAGCAATATTCCTCCCTCCGCTCCCTCCT
CCCCGATGACGATCACAATAAGAAGATGAGCATTCCGACCACGGTGTGCGGGG
TGATCAAGTACATCCCGGAGCTACAGAAGGAGGTGACGGTCTGGAGAAGAAG
AAGGAGGAGCTCAGGCGAGCTAGCAGCGAGCAAGGCGTGCTGACTATGAGGC
AGAACACGGCTCCTGTGCTCTCCGCCACCTGCCTCGACGACAGGGAAATCATG
GTCCAGGTCAGTCTGGTGAGCACCATGGCCGCAGCTCTGCCCATGTCCAAGTG
CATCAAGGTGCTGGAGAACGAAGGCCTTCGCCTCATAAATTCCTCGACTTCCGC
GTTCCAGAACAGGACCTTCTATAGCCTCCATCTTCAGAGAACCCAACGAACAAT

GAGCAAGGAGGGCCAAACATTTTGTAACTGGAGAACGCCGTGAAGCAA
AGGCAGGACTGCATCTGCATCAT

>SEQIDNO:35

ATGGGGCACCAGACCCAGATGTTTCGACGACCCGTTTCGCGAGCAGTATGTCGTC
CCTGGACGCAGACATCTTCTCCGTCGCCGGCGGCCTCCACCCATCGCAGTGGC
CGGGACTCGACCACGACGTCTCGCTGGCGCCGGCTGCCAACAACGGCACCTC
CTCCGGCGGGCTACGGCTCCCCCGGGGGCGGCGATGGCTCGGGCTCCCACCG
CAAGATCAGCCACAACGCGTACGAGCGCGACCGCCGCAAGCAGCTCAACGAG
CTCTACTCCGACCTCCGCTCCCTCCTCCCCGACTCCGATCACACCAAGAAGCTG
AGCATTCCGATCACGGTGTGCGCGGTGCTCAAGTACATCCCGGAGCTGCAGAA
GCAGGTGGACGGACTGGAGAAGAAGAAGGAGGAGCTTACGCGGGCCAGCTGC
AAGCCAGGCGTATTGACCATGAAGGAGAACACGGTCCCGATCGTGTCCGCCAC
CTGCCTCGACGAAAGGGAGATCATGGTCCAGGTTAGCTTGGTGAGCACCATGG
CCGGAGCTCTGCCCATGTCCAAGCGCATCAAAGTGCTGGAGAACGAAGGCCTC
CGCCTCATCAGCTCGTCCACTTCTGCTTTCCAGAACAGGACGTTCTATAGCCTC
CATCTTCAGAGAACCCAACGGACGATGAGCAAAGAGTGTCCGGCATTTCGTGAA
GAACTGGAGAATGCCCTGACGCAGAAGGCGGGGTACGTCTACATCACCAGTAG
ATTATATGTAGCAGAATAA

>SEQIDNO:36

ATGTTAGCGATATCTCCTCCTATGTTTTCAACAATTGGATGGCCCTTTGAGGAGC
CTTTAAGCCATAACCAGCATCAGAATTCATTCTACAAAGACACTGTTGATCAATT
ATTTAATTTTCATGATCAAGTTGAGGCAGAAATTAATTCAACAGATCCCTCACAAT
CCACAAGCAGTGACCTTAGCATGGTCAAGAAGCTTGTTTCATAATGCCAGTGAAC
GCGATCGCCGCAAGAAGATCAATAATTTGTATTCATCACTTCGATCACTCCTTCC
TGTTTCTGATCAAATGAAATTAAGCATTCCGGGAACAATTTCTAGAGTCCTGAAA

TACATACCTGAATTACAGAATCAAGTAGAGGGACTAATTAAGAGAAAGGATGAG
ATCTTATTGGGACTTTCTCCACAAGTAGAAGAGTTTATTCTAAGCAAAGAATCTC
AAAGGAAGAAGCATAGTTACAACCTCTGGTTTTGTAGTTTCAAGTAGTAGGCTCAA
TGATAGTGAAATTACCATTGAGATTTTCATGTTACACTGTCCAAAAGATTCCACTTT
CTGAGATCTTGATTTGTTTGAAAATGATGGCCTTTTGCTGCTTAATGTTTCTTCA
TCAAAGACCTTTGGAGGGAGGGTCTTCTATAATTTGCATTTCCAGGTGGATAAAA
CACAGATATTAGAATCTCATATTCTAAATGAGAAGCTCTTATCAATAATGGAGAA
GGAAGGAGAGTTTTTAAACAATAATTAAAGTTTAGGATTTGGCTCTTTAA

>SEQIDNO:37

ATGGTTGCATTCTGCCACCTCAGTTCTCATACTCAAACATGGGATGGCTCTTAG
AGGAGTTAGAGCCAGAGTCCTTAATTAGTCATAAAGAGAAGAACTATGCATCTTT
AGAGTACTCGTTACCGTATCATCAATTCTCTTCACCAAAGGAACATGTTGAAATT
GAAAGGCCACCATCCCCTAACTTATGGCCAAGAACTTAACCACAATGCTAGT
GAACGTGATCGCCGCAAGAAGATTAATAGCTTGATTTCTTCACTTCGTTCACTTC
TTCCCGGTGAAGATCAAACGAAAAAATGAGCATTCCGGTAACAATTTACGTGT
CTTAAAATACATCCCTGATTTACAAAAGCAGGTGCAAGGACTTACCAAGAAAAAA
GAAGAGCTTCTATCAAGAATTTCTCATCGACAAGAATATGCAGTTAACAAAGAAT
CACAAAGGAAGAAAATTCCAAATTACAATTCTGCTTTTGTAGTTTCAACAAGTAG
GCTTAATGATACTGAGCTTGTTATTCATATTTGCTCTTATGAGGCCAACAAGATT
CCTCTATCTGAGATCTTGATGTGTTTAGAAAATAATGGTCTTCTTCTACTTAACTC
TTCTTCTTCTAAAACCTTTGGAGGGAGGCTCTTCTATAACTTGCATTTTCAGGTG
GATAAACTCAAAGATATGAGTGTGATGATCTGATTCAAAGCTTTCTTCAATATA
TGAGAAGCAGCAAATAATCATTTGGGCACTATGGATCAAACGATCAATAGTGG
TCTGATATAT

>SEQIDNO:38

ATGTTAGCGATATCTCCTCCTATGTTTTCAACAATTGGATGGCCCTTTGAGGAGC

CTTTAAGCCATAACCAGCATCAGAATTCATTCTACAAAGACACTGTTGATCAATT
ATTTAATTTTCATGATCAAGTTGAGGCAGAAATTAATTCAACAGATCCCTCACAAT
CCACAAGCAGTGACCTTAGCATGGTCAAGAAGCTTGTTTATTATGCCAGTGAAC
GCGATCGCCGCAAGAAGATCAATAATTTGTATTCATCACTTCGATCACTCCTTCC
TGTTTCTGATCAAATGGTACTTAAT

>SEQIDNO:39

ATGGAGCACCAGCTGTTGATGACCCCTTCTCTAGCAGCATCTCGTCGCTGGAG
GCGGACATCTTCTCCGCCGGCGGCCAGCTGCCGTGCCGCCGTGGCCGGACC
TCGACCTCGACCTCGACGACGACGACATCCACGACCTCTCCGCGCCGACCGGC
AACCCACCTCCTCAGGAGGCTATGGCTCGGGCGGAGGCTCCGGAGGCTCCC
ACAGGAAGCACAGCCACAACGCGTACGAGCGCGACCGCCGGAAGCAGCTCAA
CGAGCTCTACTCCTCGCTCCGCTCCCTCCTCCCCGACGCTGACCACACTAAGAA
GCTGAGCATCCCCACCACGGTCTCCCGAGTTCTCAAGTACATCCCCGAGCTGC
AGAAGCAGGTGGACAACCTGGAGAGGAGGAAGAAAGAGCTGACGAACGCCAA
CTGCAAACCAGGAGTTCTGAACACGAGCCAGATTGTAACCTCCATTGTTTCTGC
TACTTGCCTCAACGATACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAACGT
GGCTGCCACAAGTCTTCCTCTGTCCAAGTGCATAAAAGTGATGGAGAATGAAGG
CCTTCACCTAATTAGTTCATCACTTACTCCACCTTCGACAACAGGACATTCTAT
AGCCTCCATGTTTCAGAGAAGTCAAAGAACGATGAAAGAGGAGTGCCCAGCA

>SEQIDNO:40

ATGGTCAAGAACTTAGCCACAACGCTAATGAACGTGACCGTCGCAAGAAGATT
AAAAGTTTGTATTCTTCACTTCGTTCACTTCTCCCAGCAGCAGATCAAATGAAGA
AATTAAGCGTGCCGGCCACTGTTTCACGTGCGCTTAAGTACCTACCAGAGCTTC
AACAGCAAGTGGAGAGACTGGTTCAAAGAAAGGAGGAGCTTTTATCAAAGTTAT
CAAAGCAAGGTGGTATAATTCATCAAGAAAATCAAAGAAATGACACCGTGTATAG

CTCTTTATCATCGGTATCGGCAAGCCAGCTTAGTGATAGAGAAGTTGTCGTTTCAT
ATTTCCACTTACAAGAACCATAAAAAGTCCATTATCAGAAATCTTGCTCACCTTAGA
GGAAGATGGACTTGTTCTAAAAAACTCTTCTTCCTTTGAGTCATTTGGGGACAGG
GTCTTCTATAATTTACATCTTCAGGTCATGGAAGGAACTTACACATTGGATAGTG
AGGCCATGAGGGCGAAGCTTGTGTCTTTATCTGTAAAGAGGGAATCATCGTCTC
TA

>SEQIDNO:41

ATGTTAGAAGAATTATCTCCCATCAGTTTGTCTCAACATTTGGATGGCCCTTGG
AGGAAGCCATAAGCCATGAACAGCACTACAGCTTTAGAGATGGTGAACTCCAG
AGTCATTTACTCACTTCCCTCCATCTCAGCCAGATGTAAGACAGCTTGATCGCTC
CACATCATTCACGGCCACAGTGGAAGCGGTGACCCTAGCATGGCTAAGAAGC
TTAACCACAACGCTAGCGAACGTGACCGTCGCAAAAAGATCAACAGTTTGTATT
CTTCACTCCGTTCACTACTTCCTGCAGCCGATCAAAGGAAGAAATTAAGCATACC
GTATACAGTTTCACGTGTGCTTGTATACATACCAAACTTCAACAACAAGTGGAG
AGACTGATTCAAAGGAAGGAGGAGCTTCTATCGAAGTTATCTAGGCAAGCTGAC
GATTTAACTCATCAAGAAAATCAAAGAAAAGGCACCATGTATAGCTCTTTATCAT
CGGTATCGGCGAGCCGGCTCAGTGACAGGGAAGTTGTCATTCATATCTCAACTA
ACAAGCTCCATAGAAGTTCATTATCAGAAATCTTGGTTAATTTAGAGGAGGCTGG
ACTTCTTCTACTAAATTCTTCTTCCTTCGAGTCCTTTGGAGGCAGAGTCTTCTATA
ATTTACACCTTCAGGCCATGGAAGGAACTTACACAGTAGAGTGCGAGGCCTTGA
ATGAGAGGCTTGTGTCCTTGTGCGAGAAGAGGGAGTCATTGTTTCCATTAAATT
CAAGTTCTCCATATTCTAACTGTGTATTCTAG

>SEQIDNO:42

GATCCTAACATGGTTAAGAAGCTTAACCACAACGCTAGCGAACGTGATCGTCGC
AAGAAGATCAACAGTTTGTATTCTTCACTCCGTTCACTTCTTCCAGCTTCCGATG

GAATGAAGAAATTAAGCATACCGTCCACAATTTACGTGTGCTTAAGTACATACC
AGAACTTCAACAGCAAGTGGAGAGACAGATCCAAAGGAAGGAGGAGCTTCTATC
AAATCTATCTCGGCAAGATGATTTAATTCATCAAGAAAATCAAAGAAAAGACACC
ATGTATAGCTCTTTATCATCGGTATCGGCAAGCCGGCTTGGTGATAGAGAAGTT
GTCGTTCAAATTTCCACTTGCAAGGTCCTTAAAAGCCCAATATCAGAAATCTTGC
TTAATTTAGAGGAAAATGGACTTGTTCTAATAAATTCTTCTTCCTTTGAGTCCTTT
GGAGGCAACGTCTTCTACCATTACATCTTCAGGTA

>SEQIDNO:43

ATGTTAGCATTATCTCCTCCTGTATTTCCAACACCTGAATGGCCCTTAGAGGACC
CCTTAGGCATTGACCAAATCTCCTACTTCTGTAGAGAACTCAGCCTGCTACTGC
TGCTTTTCTTCCATCTTATCAGCAAGAGTTATTATTATTAGAGCTTGATCATCAAC
AATCCACATCTTTCACAGCCTATAATAGCAGTGGTGGTGACGCTAACGATATGG
TGAAGAAGCTTAATCATAATGCAAGCGAACGTGATCGTCGCAAGAAGATGAACA
CCCTCTATTCTTCCCTCCGATCACTATTTCCGGCCGCGATGAAATGAAGAAGC
TGAGTATACCTGCCACAATTTGAGGGTGTTGAAGTACATACCAGAACTACAAG
AACAGTTAGAGAGATTGGTCCAAAGGAAGGAAGAGATTTTGCTAAGAATATCTAA
GCAAAATCATATTGTTAATCCCCAAATAAACCAAGAAAAGGCACTTCTCACAGC
AGTTTATCAGTAGTATCAGCTAATCAAATTAGTGACAAAGAAGCCATTATTCAAAT
TTCTACGTACAGTAATACTATCCATACAAGTCCACTATCAGAAATCTTGCTTCTTT
TGGAGGAGGAAGGCCTTCTTTTGATTAATTCTTCTTCCGCTGAATCCTTTGGTGG
CAGGGTCTTCAACAATTTACATGTTGAGGTTGATGATACTTATACATTGGAATGT
GATGCTTTAAGTGAGAAGCTTGCACTCTGTATGCCAAGAGGGACGGGCTGTTC
CCATGA

>SEQIDNO:44

ATGGTCAAGAAGCTTAATCATAATGCAAGCGAAAGGGATCGCCGCAAGAAGATG

AACACTCTCTATTCTTCCCTCCGATCACTTCTTCCGGCCTCCGATCAAATGAAGA
AGCTGAGCATACCTGCCACAATTTCCAGGGTGTTGAAGTACATACCAGAACTAC
AACACAATTGGAGAGATTTCGTCCAAAGGAAAGAAGAACTATTACTGAGAATATC
TAAGCAGAATCATATTATTAATCCCCAAATAAACCAAAGAAAAGGCACTACTCAC
AGCACCTTATCAGTAGTATCAGCTAATCAAATTAGTGACAAAGAAGTTGTTATTC
AAGTTTCTACTTACAATAATACTATCCATACAAGTCCATTATCAGAAATCTTGCTT
CTTCTGGAGGAGGAAGGCCTTCTTCTGATTAATTCTTCCTCCTTTGAGTCCTTTG
GAGGCAGGGTCTTCTACAATTTACATCTTCAGGTTGATGGAAGTTATATATTGGA
GTGTGATGCTTTAAGCGAGAAGCTTGCAGCTTTATATGAGAGAGACGGGTTATT
TCCATGA

>SEQIDNO:45

ACAACGATAATAACTACGCCTCAATTTCAAAGTATCAGAATAACAAGTTGTTTG
AAGGTTTACGTGCCGATAATACTATTGATTTACCTTCATCTCATCATTATCAACAA
CAATGTTTGAAAGGAAGTGAGTTTGATGTTGATGAGTTAGGGGTAGAAAGGTCA
TTAATGGAGAAGAAGCTAAATCATAATGCAAGTGAACGTAATAGAAGGAAGAAG
ATGAATTTTCTTTATTCAACTCTTCGTTCTTTGCTTCCTCCTCCTACTAATAAACAT
CAAAAGAAAAAATTAAGCTTTCCAGCAACAGTATCATATGTACAAGAATACATCC
CAGAGTTGAAGAAAGAAATAGAGAGGCTAAGCAAAACAAAAGATTTGCTTTTATC
AAAGAAATCAAATTATTCATTACTCAAATTGATGATAATAATAAGAGAAAATTAA
TTATTGGTGGAAGTTCTTGTAATTCTTCAACAACATCAATTTGTGCAAGTCAACTA
AGTAATTCACAAGTTTTGGTACAAATTTCAACAAGTCAAGAAAATAATTTTCCAAT
TTCACAAGTATTTGCAAGTGTAGAGGAAGATGGATTAATTTTGCTAAATGCATCA
TCCTTTAAATCTTTTGGAGACAAGATTTTTCACAGCTTGCAATTTTCAGATGCAAG
GACCAATTGAAATGGACATTCAGGTTTTGAAGACTAAGCTTTTAGTAATGTGT

>SEQIDNO:46

ATGGACCATCAGCTGTTTCGACGACCCCTTCGGGAGCAGCATCTCGTCGCTGGA
GGCGGACATCTTCTCCGCCGGCGGCGGCGGACAGCTGCCGTGCGCCGCGTGG
CCGGACCTCGACCTCGACGACGACTACGACATACACGACCTCTCCGCGCCGGC
CGCCAACGCCGCCACCTCCTCGGGAGGAGGCTATGGCTCCGGCGGCTCCGGC
AGGAAGCTCAGCCACAACGCATACGAGCGCGACCGCCGGAAGCAGCTCAACG
AGCTCTACTCCTCGCTCCGATCCCTCCTCCCGGACGCTGATCACACTAAGAAGC
TGAGCATCCCCACCACCGTGTCCCGAGTTCTCAACACCAAAGAGATCGTAACTC
CCATTGTTTCTGCTACTTGCCTTAACGACACGGAGATCATGGTTCAGGTCAGCC
TGCACAGCAATGTGGCCGCCACAGCTCTCCCTCTCTCCAAGTGCATAAAGGTGC
TAGAAAACGAAGGCCTTCTCCTCGTCAGCTCATCAACCTACTCCACCTTCGAGA
ACAAGACATTCTATAGCCTCCATCTTCAGAGAAGTCAAAGAACGATGAAGGAGC
AGTGCCCAGGATTCTGCGACGAACTGGAGAAGATCGTCAGGAAGAAAGCAGGG
GCG

>SEQIDNO:47

ATGGAGCATCAACTATTCGACGACGCCGTCCCGAGCAGCATGATCTGGCCGTT
AGAGGCAGAAAACGGTTTCACCGACGAGCTGCCGTCTTTGCAGTTACCGGACG
TGGACCTTGACTTCGACATCCACGAGTTCTCCGCACCGGCAACGGCACCGGCG
AAAGCGGCCTCCTCGGGTGGCTCCGGATTGGTTGGTTCCGGTTCAGGATCGCA
TAAGAAGCTCAACCACAACGCGTACGAGCGCGACCGGCGGACGCAGCTCAATC
AGCTCTACTCGACTCTCCGTTCTCTCATCCCCAACGCAGATCACACAAAGAAGC
TGAGCATTCCGACGACGGTGTGTCAGGTCCTCGACTACATACCCAAGCTGCAG
AAGCAGGTCGAGGATCTCAAGAAGAAGAAACAGGAGCTCAGTACAGCCAAATG
CAGAGAAAGACTGCAGCGCGTCAAGGACAACACATGCCGTATTGTTTCTGCCAC
TCCTCTCGATGGCAACGAAATCATGGTCCAGGTTAGCCTGCTGAGCAACATGGC
TGCAAGTCTTCCTCTATCCAAGTGCATAAACGTATTTGAGAACAAAGGGCTTCAC
CTCATCAGTTCATCGACTTTCTCCACCGAGGTCAATAGAACATTTTACAGCTTCC

ACTTTGAGGTACGTTTTTACATGCGCCCT

>SEQIDNO:48

GGCGATGGCTCGGGCTCCCATCGCAAGATCAGCCACAACGCCTACGAGCGCG
ACCGCCGCAAGCAGCTCAACGAGCTCTACTCCGACCTCCGCTCCCTCCTCCCC
GACACCGATCACACGAAGAAGCTTAGCATTCCGATCACGGTGTGCGGGGTGCT
CAAGTACATCCCGGAGCTGCAGAAGCAGTGGGCTTGNAGAAANAAAGAAGGAG
GAGNTGACGCGCGCCAACTGCAACCCGGNGTGNTGACCATGAAGGGGAACAC
GGTCCGATTGTTCCGCCACCTGCCTCGACGACAGGGANATTATGGTCCAAGTC
AACCTGGTGAGCACATGGCCGGANTNTGCCATTTCAAATGCATCAAAGTGCTG
GAAAACAANNGCTCCGGTCA

>SEQIDNO:49

ATGGAGCAGCTGTTGTCGACGACCCAGCCTTCGCGAGCAGCATGTCGTCGCT
TGAGGCGGACATCTTCTCCGGCGCCGGCCAGCTGCCGTCTCGCCGTGGCTG
GACCTAGACCTCGACGACGATGTCCAAGACCTCTCCATGGCGCCGACGACGGC
GAACGCGGTGTCCTCCGGCTACGGCTCCGGCGGATCCGGCTCCCACAGGAAG
CTCAGCCACAACGCCTACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTCTA
CTCCTCCCTCCGCGCTCTCCTCCCCGACGCCGATCACACTAAGAAGCTGAGCA
TCCCGACGACGGTGTCTCGCGTGCTCAAATACATACCCGAGCTGCAGAAGCAG
GTGGAGAATCTGGAGAGGAAGAAGAAGGAGCTGACGACGACGAGCACCACCAA
CTGCCAACCAAGAGTGTTGGGGAGCCAGCTGATGAGC

>SEQIDNO:50

ATGGTCGCATTGTTCTCTCCTCCTCTCTTCTCAACCAAAGGGTGGCTCTTAGAG
GAGGAGCCATTCCGGCTATAATAATACCCATAATCTCTCCTACAAAGATGATGCGT
CTTCTCAGTACTCCTTTCCCTATCAATTTTATTCACCACAGACACAGATTGAGGTT

GAAATTGAAAGGTCCACTGCACCATCCTCTGACCCTGCCATGGTCAAGAACTT
AGCCACAATGCTAGCGAACGTGATCGCCGCAAGAAGGTCAACAACTTGGTTTCT
TCACTTCGCTCACTTCTTCCAATGGCAGATCAAACGAAAAAATGAGCATTCTG
CAACAGTTTCCAGAGTGTTGAAATACATACCTGAACTACAACAGCAAGTGCAAG
CACTAACAAAGAGAAAAGAGGAGCTTCTGTGCAGAATTTCTCGGCAATTGCAAG
GAGAAGCAGTGAACAAAGAATCTCAGAGAAAAATTTCCCATCACAACTCTTCTTT
TGTTGTCTCAACGACTAGGCTTAACGATTGTGAAGCTGTAGTTCACATTTTCATCT
CATGAGACACACAAGGCTCCACTATCAGAGATTCTGCAGTGCTTAGAAAATGAT
GGCCTTTTTCTGCTACATGCTTCTTCCTCAGAAACCTTTGGAGGAAGGTTCTTCT
ACAATTTGCATTTTTCACGTGGAGAAAACTGATAGATTAGAGACCGAGATTTTAAC
TGAGAAGCTTTTACCAATATAT

>SEQIDNO:51

ATGTTAGCATTCTCCCCTCCATTGTTTCCAACCCTTGGATGGCCCTTGGAGGAT
CCCATAAGCCATGCACAGAACTACATATATGGAGAAACAGAACTTCAGAATCG
TTTCTTCACTTGCCCTCATCTCAGCCACAAGTGGAAGTCAATTGCTCCACCCCAT
ATGCAGCAGTTAGTGGTAAATCCCACGATGGTTAAGAACTTAACCACAACGTCA
GTGTGCGGGATCGTCGGAAGAAGATCAACAGCTTGTAATCCTCTCTGCGTTCAC
TACTTCCATCAGCTGATCAAGTGAAGAAATTAAGCATTCTTCGACAGTTTCATG
TGTGCTAAAATACATACCAGAGCTGCAACGGCAAGTGGAGAGACTGATCCAAAA
GAAAGAAGAGTTTTTATCAAAGATTTCTAGGGAAGGAGATCTAATTCACCTAGAA
AATCAAAGAAATGGCACACTTGGAAGCTCTTTATCTGCTGTTTCAGCAAGAAGG
CTTAGTGACAGGGAAATTGTGGTTCAGATATCCACATTTAAGGTCCATGAGAGT
CCACTTTCTGAGGTTTTGTAAATTTGGAGGAGGATGGGCTTCTTGTAATCAATG
CATCATCTTTTGAGTCCTTTGGAGGGAGGGTCTTCTACAACTTACATCTTCAGGT
TGAAGGAACTCAAGGAATGGAGTCGACCGCCCACTTGGAGATGACAAGAAGTCT
TAAAAACAAACACATGAATATATTGGTGATTCATATGGACTTTCCTCCTTTTTTT

CTTAAGATGTTCTGATCTTTACAAGAGTATTTACCAATCATATATCAACTTCCTA
CCAATGCTATGTTGGTAAATTAATTATCATTCTACTGCTACATGTAATTAAGAAAG
AAAATTTTGAACTTCTAAACATCAATTGGCCAGCGCCGAACCAATAATTGTAGC
AGGTCAAGCTGCCAAAAAATTGGATGATGAGCTTCTCTTGGAACCAAGATGAA
AACTGAAGGGATGGGAGTACTGGAAACACCAATATTGATTAAAGCAAAGAATGG
TACCAAAGAAATGGTGGAGAGAATCCTTGATCTTTACCCCATGGCAATTCATGAC
ATAGACTCCAACAAGAAGAATATAGTGCTATTGGCGGTGGAGAATAGGCACCCC
CATGTGTATGAGCTCTTCCTGAAGAGAAATATTGTGAAAGATAGTGTATTTGGTG
CAGTTGATAATAAAGGCAACAGTGCATTGCATCTGGCTGCCATGTTTGCAGATTA
TCGGCCTTGGGTCACTCCTGGTGTTCATTGCAAATGCAATGGGAAGTCAAATG
GTATGAGTATGTGAAGAAGTCCATGCCACCAAATTTCTTCCGTTTCCACAACAAT
GAAAACAAGTCTACAAAGCAGATTTTCACCCGTGAACACAGAGATCTGGTGCAA
AAGGGTGGGCAATGGCTAAATAACACAGCCACCTCATGCTCGTTGGTAGTAACA
CTCATTGCAACAGTTGCCTTCGCCACATCAACTGCTGTACCGGGCGGCACCAAG
GAGGGGACTGATTCATGTCCTCTCAATGGTCCCTAA

>SEQIDNO:52

ATGTTAGCATTCTCCCCTCCATTGTTTCCAACCCTTGGATGGCCCTTGGAGGAT
CCCATAAGCCATGCACAGAACTACATATATGGAGAAACAGAACTTCAGAATCG
TTTCTTCACTTGTCCTCATCTCAGCCACAAGTGGAACCTCAATTGCTCCACCCCAT
CTGCAGCAGTTAGTGGTAATCCCACGATGGTTAAGAACTTAACCACAACGCCA
GTGAGCGGGATCGTCGGAAGAAGATCAACAGCTTGTAATCCTCTATGCGTTCAC
TACTTCCATCAGCTGATCAAGNGAAGAAATTAAGCATTCTTCGACAGTTTCACG
TGTGCTAAAATACATACCAGAACTGCAACGACAAGTGGAGAGATTGATTCAAAA
GAAAGAAGAGTTTTTATCAAAGATTTGTAGGGAAGGAGATCCAATTCACCTAGAA
AATCAAAGAAATGGCACACTTGGAAGCTCTTTATCTGCTGTTTCAGCAAGAAGG
CTTAGTGACAGGGAAATTGTGGTTCAGATATCCACATTTAATGTCCATGAGAGTC

CTCTTTCTGAGGTTTTGTTAAATTTGGAGGAGGATGGGCTTCTTGTAATCAATGC
ATCATCTTTTGAGTCCTTTGGAGGGAGGGTCTTCTACAACTTACATCTTCAGGTT
GAAGGAACTCAAGGAATGGAGTGTGAGTTGTTGAGCGAGAAGCTACTTTCATTG
TGTGAAAGGAGAGAGAGGCTTTTCCATGA

>SEQIDNO:53

MCALVPSFFTNGFWPSTNQYESYYGAGDNLNNGTFLELTVPQTYEVTHHQNSLGV
SVSSEGNEIDNNPVVVKLNHNASERDRRKKINTLFSSLRSCLPASDQSKKLSIPET
VSKSLKYIPELQQQVKRLIQKKEELVRVSGQRDFELYDKQQPKAVASYLSTVSATRL
GDNEVMVQVSSSKIHNFSISNVLGIEEDGFVLVDVSSSRSQGERLFYTLHLQVEN
MDDYKINCEELSERMLLYEKCENSFN

>SEQIDNO:54

MCALVPPLFPNFGWPSTGEYESNYLAGVNLEDFTFLDFPAPETYGVEHHQEIQEML
GVSVPSEGNVVTKKLNHNASERDRRKKINSLFSSLRSCLPASDQTKKLSIPQTVSR
SLKYIPELQEQVKKLIQKKEELLVRVSGQRAIEHYVEPQPKAVARYVSTISATKLGDN
EVLVQISSSKNHNFSISNVLSGLEEDGFVLVDVSSSRYPYHGWLFYSLHLQMGNKDN
HKLKCEELSQRILLYEEECENSFR

>SEQIDNO:55

MCALVPPLFPDFGWPSTAGYESYYLGGENLNNDMFLDFPVVETYGVLAAHHQNSLG
VSVSSEGNIDNNPVVKKLNHNASERDRRKKINSLFASLRSCLPSTSDQSKKLSISA
TVSRSLKYIPELQEQVKKLLQKKEELLVRVSGQRDIELYVKPQPKAIASYVSTVSATR
LGDNEVMVQISSSKIHNFSISKVLTGLEEDGFVLVDVSSSRFQGERLFYTLHLQVEN
MDDHYKMNCEELSERMLLYEEECENXXR

>SEQIDNO:56

IPQTVSRSLKYIPELQEQVKKLIQKKEELLVRVSGQRDIEHYVEPHPKAVARYVSTISA
TKLGDNEVMVQISSSKNHNFSISNVLSGLEEDGFVLVDVSSSRSHGERLFYTLHLQ
MGNKDDYKLTCEELRQRMLYLYEECGNSFR

>SEQIDNO:57

IPTPSQATSSDLSMVKKLIRNASERDRRKKINTLYSSLRSLLPVAEQMKKLSNPATIS
RVLKYIRELQKQVEGLLTRKEAILLKLSPEVDEVKSKESEK

>SEQIDNO:58

SDRRKQLNEQYSSLRSLLPDDDHNNKMSIPTTVSRVIKYIPELQKEVDGLEKKKEEL
RRASSEQGVLTMRQNTAPVVSATCLDDREIMVQVSLVSTMAAALPMSKCIKVLNE
GLRLINSSTSAFQNRTFYSLHLQRTQRTMSKEGQTFCNELENAVKQKAGLHLHH

>SEQIDNO:59

MGHQTQMFDDPFASSMSSLDADIFSVAGGLHPSQWPGLDHDVSLAPAANNGTSS
GGYGSPGGGDGSGSHRKISHNAYERDRRKQLNELYSDLRSLLPDSHTKKLSIPIT
VSRVLKYIPELQKQVDGLEKKKEELTRASCKPGVLTMKENTVPIVSATCLDEREIMV
QVSLVSTMAGALPMSKRIKVLNEGLRLISSSTSAFQNRTFYSLHLQRTQRTMSKEC
PAFCEELENALTQKAGYVYITSRLYVAE

>SEQIDNO:60

MLAISPPMFSTIGWPFEEPLSHNQHQNSFYKDTVDQLFNFHDQVEAEINSTDPSQS
TSSDLSMVKKLVHNASERDRRKKINNLYSSLRSLLPVSDQMKLSIPGTISRVLKYIPE
LQNQVEGLIKRKDEILLGLSPQVEEFILSKESQRKKHSYNSGFVVSSSRLNDSEITI
SCYTVQKIPLSEILICLENDGLLLLNVSSSKTFGGRVFYNLHFQVDKTQILESHILNEKL
LSIMEKEGEFLKQ

>SEQIDNO:61

MVAFCPPQFSYSNMGWLLEEELEPESLISHKEKNYASLEYSLPYHQFSSPKEHVEIER
PPSPKLMAKKLNHNASERDRRKKINSLISSLRSLLPGEDQTKKMSIPVTISRVLKYIPD
LQKQVQGLTKKKEELLSRISHRQEYAVNKESQRKKIPNYNSAFVVSTSRLNDTELVI
HISSYEANKIPLSEILMCLENNGLLLLNSSSSSKTFGGRLFYNLHFQVDKTQRYECDDL
IQKLSSSIYEKQQNNHLGTMDQTINSGLIY

>SEQIDNO:62

MLAISPPMFSTIGWPFEEPLSHNQHQNSFYKDTVDQLFNFHDQVEAEINSTDPSQS
TSSDLSMVKKLVHYASERDRRKKINNLYSSLRSLLPVSDQMVLN

>SEQIDNO:63

MEQLFVDDPAFASSMSSLEADIFSGAGQLPSSPWLDLDLDDDDVQDLSPAPTTANAV
SSGYGSGGSGSHRKLSHNAYERDRRKQLNELYSSLRALLPDADHTKLSIPTTVSRV
LKYIPELQKQVENLERKKKELTTTSTTNCKPGVLGSQLMSEGMPIVSATCINDMEI
MVQVSLLSNVAGSVLPLSKCIKVLNEGLHFISSTSSGFGNRTFYSIHLQRSEGTIN
EECPAFCEERLEKVVRNKAKL

>SEQIDNO:64

MEHQLFDDPFSSSISSLEADIFSAGGQLPSPWPDLDLDDDDIHDLSAPTGNPTS
SGGYGSGGGSGGSHRKHSHNAYERDRRKQLNELYSSLRSLLPADADHTKKLSIPTT
VSRVLKYIPELQKQVDNLERRKKELTNANCKPGVLNTSQIVTPIVSATCLNDTEIMVQ
VSLHSNVAATSLPLSKCIKVMENEGLHLISSSTYSTFDNRTFYSLHVQRSQRTMKEE
CPA

>SEQIDNO:65

MVKKLSHNANERDRRKKIKSLYSSLRSLLPADQMKKLSVPATVSRALKYLPQLQ

QVERLVQRKEELLSKLSKQGGIIHQENQRNDTVYSSLSSVSASQLSDREVVVHISTY
KNHKSPLEILLTLEEDGLVLKNSSSFESFGDRVFYNLHLQVMEGTYTLDSEAMRAK
LVSLSVKRESSSL

>SEQIDNO:66

MLEELSPISLFSTFGWPLEEAISHEQHYSFRDGETPESFTHFPPSQPDVRQLDRSTS
FTHSGSGDPSMAKKLNHNASERDRRKKINSLYSSLRSLPAADQRKKLSIPYTVSR
VLVYIPKLQQQVERLIQRKEELLSKLSRQADDLTHQENQRKGTMYSSLSSVSASRLS
DREVVIIHISTNKLHRSSLSEILVNLEEAGLLLLNSSSFESFGGRVFYNLHLQAMEGTY
TVECEALNERLVSLCEKRESLFPLNSSSPYSNCVF

>SEQIDNO:67

DPNMVKKLNHNASERDRRKKINSLYSSLRSLPASDGMKKLSIPSTISRVLKYIPELQ
QQVERQIQRKEELLSNLSRQDDLIHQENQRKDTMYSSLSSVSASRLGDREVVVQIS
TCKVLKSPISEILLNLEENGLVLINSSSFESFGGNVFYHLHLQV

>SEQIDNO:68

MLALSPPVFPTPEWPLEDPLGIDQISYFCRETQPATAAFLPSYQQELLLLLELDHQQS
TSFTAYNSSGGDANDMVKKLNHNASERDRRKKMNTLYSSLRSLFPAADEMKKLSIP
ATISRVLKYIPELQEQLERLVQRKEEILLRISKQNHIVNPQINQRKGTSHSSLSVVSAN
QISDKEAIIQISTYSNTIHTSPLSEILLLEEEGLLLINSSSAESFGGRVFNNLHVQVDD
TYTLECDALSEKLASLYAKRDGLFP

>SEQIDNO:69

MVKKLNHNASERDRRKKMNTLYSSLRSLPASDQMKKLSIPATISRVLKYIPELQQQ
LERFVQRKEELLLRISKQNHIIINPQINQRKGTTHSTLSVVSANQISDKEVVIQVSTYNN
TIHTSPLSEILLLEEEGLLLINSSSFESFGGRVFYNLHLQVDGTYILECDALSEKLAAL

YERDGLFP

>SEQIDNO:70

TTIITTPQFQTDQNNKLFEGLRADNTIDLPSHHYQQQCLKGSEFDVDELGVERSLM
EKKLHNHNASERNRRKKMNFLYSTLRSLPPPTNKHQKKLSFPATVSYVQEYIPELK
KEIERLSKTKDLLLSKKSNYSLKIDDNNKRKLIIGGTSCNSSTTSICASQLSNSQVLV
QISTTQENNFPISQVFASVEEDGLILLNASSFKSFGDKIFHSLHFQMGGPIEMDIQVLK
TKLLVMC

>SEQIDNO:71

MDHQLFDDPFGSSISSLEADIFSAGGGGQLPSPWPDLDDLDDYDIHDLSPAANA
ATSSGGGYGSGGSGRKLSHNAYERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTV
SRVLNTKEIVTPIVSATCLNDTEIMVQVSLHSNVAATALPLSKCIKVLENEGLLLVSSS
TYSTFENKTFYSLHLQRSQRTMKEQCPGFCDELEKIVRKKAGA

>SEQIDNO:72

MEHQLFDDAVPSSMIWPLEAENGFTDELPSLQLPDVDLDFDIHEFSAPATAPAKAAS
SGGSGLVGSGSGSHKKLNHNAYERDRRTQLNQLYSTLRSLIPNADHTKKLSIPTTV
CQVLDYIPKLQKQVEDLKKKKQELSTAKCRERLQRVKDNTCRIVSATPLDGNEIMVQ
VSLLSNMAASLPLSKCINVFENKGLHLISSSTFSTEVNRTFYSFHFEVRFYMRP

>SEQIDNO:73

GDGSGSHRKISHNAYERDRRKQLNELYSDLRSLLPDTDHTKKLSIPITVSRVLKYIPE
LQKQWAXRXKEGGXDARQLQPGVXTMKGNTVRLFRHLPRRQGXYGPSQPGEHM
AGXCPFQMHQSAGKQXLRS

>SEQIDNO:74

MEQLFVDDPAFASSMSSLEADIFSGAGQLPSSPWLDLDLDDDVQDLSPAPTTANAV
SSGYGSGGSGSHRKL SHNAYERDRRKQLNELYSSLRALLPDADHTKKLSIPTTVSR
VLKYIPELQKQVENLERKKKELTTTSTTNCQPRVLGSQLMS

>SEQIDNO:75

MVALFSPPLFSTKGWLLEEEFPGYNNTHNLSYKDDASSQYSFPYQFYSPQTQIEVEI
ERSTAPSSDPAMVKKLSHNASERDRRKVNVLVSSLRSLLPADQTKKMSIPATVS
RVLKYIPELQQQVQALTKRKEELLCRISRQLQGEAVNKESQRKISHHNSSFVSTTR
LNDCEAVVHISSHETHKAPLSEILQCLENDGLFLLHASSSETFGGRFFYNLHFHVEKT
DRLETEILTEKLLPIY

>SEQIDNO:76

MLAFSPPLFSTFGWPWEDPXSHEQNYIYQETEASESFLHLPSSSEPQAEINYSTPSA
AVSGNPTMVKKLNHNASERDRRKINSLYSSLRSLLPADQAKKLSIPSTVSRVLKYI
PELQKQVERLIQKKEELLSKISRQGDIIHQEKQRKATLASSLSAVSANRLSDREIVVQI
STFKVHESPLSEVLLNLEEDGLLVINASSFESFGGRVFYNLHLQVEGTHRMECEVLS
EKLLSLCEKRRDAFP

>SEQIDNO:77

MLAFSPPLFPTLGWPLEDPISHAQNYIYGETETSESFLHLPSSQPQVELNCSTPYAA
VSGNPTMVKKLNHNVSVRDRRKINSLYSSLRSLLPADQVKKLSIPSTVSCVLKYIP
ELQRQVERLIQKKEEFLSKISREGDLIHLENQRNGTLGSSLSAVSARRLSREIVVQI
STFKVHESPLSEVLLNLEEDGLLVINASSFESFGGRVFYNLHLQVEGTQGMESTAHL
EMTRTLKNKHMNILVIHMDFPFFLKMFLIFTRVFTNHISTSYQCYYGKLIILLHVIKK
ENFETSKHQLASAEPPIVAGQA AKKLDELLETMKMTEGMGVLETPILIKAKNGTKE
MVERILDLYPMAIHIDSNKKNIVLLAVENRHPHVYELFLKRNIVKDSVFGAVDNKGN
SALHLAAMFADYRPWVTPGVALQMQWEVKWYEYVKKSMPPNFFRFHNNENKSTK
QIFTREHRDLVQKGGQWLNNTATSCSLVVTIATVAFATSTAVPGGTKEGTDSCPLN

GP

>SEQIDNO:78

MLAFSPPLFPTLGWPLEDPISHAQNYIYGETETSESFLHLSSSQPQVELNCSTPSAA
VSGNPTMVKKLNHNASERDRRKINSLYSSMRSLLPSADQXKKLSIPSTVSRVLKYI
PELQRQVERLIQKKEEFLSKICREGDPIHLENQRNGTLGSSLSAVSARRLSREIVV
QISTFNVHESPLSEVLLNLEEDGLLVINASSFESFGGRVFYNLHLQVEGTQGMECEL
LSEKLLSLCERREAFP

>SEQIDNO:79

ATGGAGTATCCATGGCTGCAGTCTCAAGTTCATTCTTTTACCTACTCTCCATT
TTCCTTCCTTCCTTCATCCTTTAGATGATTCCAAGAGCCATAACATCAATCTTCAT
CATATGAGTCTTAGTCACAGCAATAATACTAACAGTAACAATAACAATTATCAAGA
AGAAGATCGAGGAGCGGTGGTTTTGGAGAAGAACTGAATCACAACGCAAGCG
AACGAGACCGCCGTAGAAAACCTTAACGCCTTGACTCTTCACTTCGTGCTCTCTT
GCCTCTTTCTGATCAAAGAGGAAGCTGAGCATTCTATGACGGTAGCGAGAGT
AGTGAAATACATACCAGAGCAGAAGCAAGAACTTCAACGTTTGTCTCGGAGAAA
AGAAGAGCTCTTGAAGAGGATCTCGAGAAAACTCACCAAGAGCAGCTGAGAAA
CAAAGCAATGATGGACTCAATAGATTCTTCTTCCTCTCAACGGATCGCAGCAAAT
TGGCTCACTGACACAGAGATTGCTGTCCAGATTGCTACGTCGAAATGGACATCT
GTTTCAGACATGTTGCTTAGGTTAGAAGAAAACGGGCTTAATGTCATAAGCGTCT
CTTCTTCCGTTTCTTCCACCGCAAGGATCTTCTACACTCTACATCTTCAGATGAG
AGGAGATTGCAAAGTGAGACTGGAGGAACTCATCAATGGTATGCTCTTGGGATT
ACGCCAATCATAA

>SEQIDNO:80

ATGTGTGCCTTAACACCAATGTTTCCAAGTAACCAACAAGAATGGTACTCTACTT

CAACAATGGAGTATCCATGGCTTGATTCTTCTCTCCTACTCTCCCTTCTTCTCT
TTATCCTTCTTTTCGACCAACTAGATGAATTCAAGAGCTATAACATCAATCTTCTTC
CTCATCATATGAATCTTGCTGACATAAATGGTACTAACAATGATCAAGAAGAACA
TCAAGGATCGGTTTTGGAAAAGAACTGAATCACAACGCAAGTGAACGCGACCG
CCGTAGAAAGCTAAACGCCTTATACGCTTCACTTCGTGCTCTCTTGCCTCCTTCT
GATCAAAAGAGAAAGTTGAGCATTCCAAAGACCATAGCGGGAGTGGTGAAGTAT
ATACCAGAGCAGAAGCAAGAACTTCAACGTTTGTCTAGGAGGAAAGAAGAGCTT
ATGAAGAGAATCTCCAATAAGACAGAGACTTTGAATCATCAACAAGAACAGCTGA
GAAATAGAGCATTAAATGATGGAGTCAATAGATTCTTCTTCACAAAAGATCGCT

>SEQIDNO:81

ATGGAGTATCCATGGCTTGATTCTTCTCTCCTACTCTCCCTTCTTCTCTTTATCC
TTCTTTTCGACCAACTAGATGAATTCAAGAACTATAACATCAATCTTCTTCCTCATC
ATATGAATCTTGCCGACATAAATGGTACTAACAATACCAGTAACAATGATCAAGA
AGAACATCAAGGATCGGTTTTGGAAAAGAACTGAATCACAACGCAAGTGAACG
CGACCGCCGTAGAAAGCTAAACGCCTTATACGCTTCACTTCGTGCTCTCTTGCC
TCCTTCTGATCAAAAGTCGGCGAATCAGAGAAAGTTGAGCATTCCAAAGACCGT
AGCGGGAGTGGTGAAGTATATACCAGAGCAGAAGCAAGAACTTCAACGTTTGT
TAGGAGGAAAGAAGAGCTTATGAAGAGGATCTCCAATAAGATAGAGACTTTGAA
TCATCAACAAGAACAGCTGAGAAATAGAGCATTAAATGATGGAGTCAATAGATTCT
TCTTCACAAAAGATCGCTGCAAATTGGATCACCAACACAGAAATAGCTGTCCAG
ATTGCTACATGGAAATGGACATCTATCTCAGACATGTTGCTTAGGTTAGAAGAAA
ACGGGCTTAATGTCATAAGCGTCTCTTCTTCGGTTTCTTCCACCGCAAGGATCTT
CTACACACTGCATCTTCAGATG

>SEQIDNO:82

GCTTTCTCTTTCAGCTCGATCGATCCACAGCTCAACGAGCTCTACTCCTCCCTC

CGCGCTCTCCTCCCCGACGCCGATCACACTAAGAAGCTGAGCATCCCGACGAC
GGTGTCTCGCGTGCTCAAGTACATACCCGAGCTGCAGAAGCAGGTGGAGAATC
TGGAGAGGAAGAAGAAGGAGCTGACGACGACGAGCACCACCAACTGCAAACCA
GGAGTGTTGGGGAGCCAGCTGATGAGCGAGGGCATGGCTCCCATCGTT

>SEQIDNO:83

ATGGAGCAGCTGTTTCGTGACGACCCAGCCTTCGCGAGCAGCATGTCGTGCTG
TGAGGCGGACATCTTCTCCGGCGCCGGCCAGCTGCCGTCTCGCCGTGGCTG
GACCTAGACCTCGACGACGATGTCCAAGACCTCTCCATGGCGCCGACGACGGC
GAACGCGGTGTCCTCCGGCTACGGCTTCGGCGGATCCGGCTCCCACAGGAAG
CTCAGCCACAACGCCTACGAGCGCGACCGCCGGAAGCAGCTCAACGAGCTCTA
CTCCTCCCTCCGCGCTCTCCTCCCCGACGCCGATCACACTAAGAACTGAGCAT
TTCGACGAACGTGTCCTGCGTGTTTCAGTACATAACCGAACCTGCAGAAACAAG
TGGAGAATATGGAGAAGAAAAAAAAGAGCTGACGACGACGAGCACCACCAACT
GTCAACCCCAAGATGTGGGTAGAAGC

>SEQIDNO:84

MEYPWLQSQVHSFSPTLHFPSFLHPLDDSKSHNINLHHMSLSHSNNTNSNNNNYQ
EEDRGAVVLEKKLNHNASERDRRRKLNALYSSLRALLPLSDQKRKLSIPMTVARVVK
YIPEQKQELQRLSRRKEELLKRISRKTHQEQLRNKAMMDSIDSSSSQRIANWLTDT
EIAVQIATSKWTSVSDMLLRLEENGLNVISVSSSVSSTARIFYTLHLQMRGDCKVRLE
ELINGMLLGLRQS

>SEQIDNO:85

MCALTPMFPSNQEWYSTSTMEYPWLDSFSPTLPSSLYPSFDQLDEFKSYNINLLP
HHMNLADINGTNNDQEEHQGSVLEKKLNHNASERDRRRKLNALYASLRALLPPSD
QKRKLSIPKTIAGVVKYIPEQKQELQRLSRRKEELMKRISNKTETLNHQQEQLRNRA

LMMESIDSSSQKIA

>SEQIDNO:86

MEYPWLDSFSPTLPSSLYPSFDQLDEFKNYNINLLPHHMNLADINGTNNTSNNDQE
EHQGSVLEKKLNHNASERDRRRKLNALYASLRALLPPSDQKSANQRKLSIPKTVAG
VVKYIPEQKQELQRLYRRKEELMKRISNKIETLNHQQEQLRNRLMMESIDSSSQKI
AANWITNTEIAVQIATWKWTSISDMLLRLEENGLNVISVSSSVSSTARIFYTLHLQM

>SEQIDNO:87

AFSFSSIDPQLNELYSSLRALLPDADHTKKLSIPTTVSRVLKYIPELQKQVENLERKKK
ELTTTSTTNCKPGVLGSQLMSEGMAPIV

>SEQIDNO:88

MEQLFVDDPAFASSMSSLEADIFSGAGQLPSSPWLDLDDLDDDVQDLSMAPTTANAV
SSGYGFGGSGSHRKLSHNAYERDRRKQLNELYSSLRALLPDADHTKKLSISTNVSC
VVQYITEPAETSGEYGEEKELTTTSTTNCQPQDVGRS

>SEQIDNO:89

ATGTGTGCACTTGTCCCTCCATTATATCCCAATTTCTGGCTGGCCTTGCGGAGAT
CATAGCTTCTATGAAACCGACGACGTATCCAACACGTTTCTTGATTTTCCGTTGC
CGGACTTGACGGTGACTCATGAGAATGTGTCTGTCTGAGAATAACAGAACATTAC
TAGACAATCCCGTGGTGATGAAGAAGCTTAATCACAACGCGAGTGAACGTGAGC
GTCGCAAGAAGATCAACACAATGTTCTCATCTCTTCGTTCTTGTCTTCCTCCAC
CAATCAAACGAAGTTAAGTGTTTCGGCAACAGTTTCACAAGCATTGAAGTACATA
CCAGAGCTGCAAGAGCAAGTTAAAAAGCTCATGAAGAAGAAAGAAGAGCTCTCG
TTTCAAATTTCTGGGTCAAAGAGATCTCGTTTACACCGACCAAACAGTAAGTCAG
AGGAAGGGGTTACAAGCTATGCGTCGACAGTTTCTTCGACTAGGCTCAGTGAGA

CTGAAGTGATGGTCCAAATTTTCATCGTTACAGACTGAAAAATGTTGTTTTGGGAA
TGTCTTGAGTGGTGTAGAAGAAGATGGGTTGGTTCTTGTGGGTGCTTCATCTTC
AAGGTCTCATGGAGAGCGACTCTTTTACTCTATGCATCTTCAGATAAAAAATGGC
CAGGTGAATTCCGAAGAATTAGGTGATAGATTGTTGTACTTGTACGAGAAATGTG
GACACTCGTTTACATGA

>SEQIDNO:90

ATGTGTGCACTTGTCCCTCCATTATATCCCAATTTTCGGCTGGCCTTGCGGAGAT
CATAGCTTCTATGAAACCGACGACGTATCCAACACGTTTCTTGATTTTCCGTTGC
CGGACTTGACGGTGACTCATGAGAATGTGTCGTCTGAGAATAACAGAACATTAC
TAGACAATCCCGTGGTGATGAAGAAGCTTAATCACAACGCGAGTGAACGTGAGC
GTCGCAAGAAGATCAACACAATGTTCTCATCTCTTCGTTCTTGTCTTCCTCCAC
CAATCAAACGAAGAAGTTAAGTGTTTCGGCAACAGTTTCACAAGCATTGAAGTAC
ATACCAGAGCTGCAAGAGCAAGTTAAAAAGCTCATGAAGAAGAAAGAAGAGCTC
TCGTTTCAAATTTTCGGGTCAAAGAGATCTCGTTTACACCGACCAAAACAGTAAGT
CAGAGGAAGGGGTTACAAGCTATGCGTCGACAGTTTCTTCGACTAGGCTCAGTG
AGACTGAAGTGATGGTCCAAATTTTCATCGTTACAGACTGAAAAATGTTGTTTTGG
GAATGTCTTGAGTGGTGTAGAAGAAGATGGGTTGGTTCTTGTGGGTGCTTCATC
TTCAAGGTCTCATGGAGAGCGACTCTTTTACTCTATGCATCTTCAGATAAAAAAT
GGCCAGGTGAATTCCGAAGAATTAGGTGATAGATTGTTGTACTTGTACGAGAAA
TGTGGACACTCGTTTACATGA

>SEQIDNO:91

CCACGTCTGTCGCGGCCGACTCCTTCTACGTGGCCTCCCGATCCCGGTGG
TGTCCGCGGCGAGGAGCTGATGGCGGGGCGAACCTCATCCACAACGCCTAC
GAGCGCGACCGCCGGAAGCAGCGCAACGAGCTCTACTCCTCCCTCCGCGCTCT
CCTCCCCGACGCCGATCACACTAAGAAGCTGAGCATCCCGACGACGGTGTCTC

GCGTGCTCAAGTACATACCCGAGCTGCAGAAGCAGGTGGAGAATCTGGAGAGG
AAGAAGAAGGAGCTGACGACGACGAGCACCACCAACTGCAAACCAGGAGTGTT
GGGGAGCCAGCTGATGAGCGAGGGCATGGCTCCCATCGTTTCGGCTACCTGCA
TCAATGACATGGAGATCATGGTTCAGGTCAGCTTGTTGAGCAATGTGGCGGGTT
CAGTTCTTCCTCTCTCCAAGTGTATCAAAGTACTGGAGAACGAAGGTCTTCACTT
CATCAGTTCATCGACTTCCTCCGGATTTGGGAACAGGACATTCTACAGTATCCAT
CTTCAGAGAAGTGAAGGAACGATCAACGAGGAGTGCCCAGCATTTTGTGAAAG
GTTGGAGAAAGTCGTCAGGAACAAAGCAAAGCTT

>SEQIDNO:92

ATGGAGCACCAGCTGTTTCGATGACCCCTTCTCTAGCAGCATCTCGTCGCTGGAG
GCGGACATCTTCTCCGCCGGCGGCCAGCTGCCGTCGCCGCCGTGGCCGGACC
TCGACCTCGACCTCGACGACGACGACGGCATCCACGACCTCTCCGCGCCGGC
CGGCAACCCACCTCTTCAGGAGGCTATGGCTCGGGCGGAGGCTCCCACAGG
AAGATCAGCCACAACGCGTACGAGCGTGACCGCCGGAAGCAGCTCAACGAGCT
CTACTCCTCGCTCCGCTCCCTCCTCCCCGACGCTGACCACACTAAGAAGCTGA
GCATCCCCACCACGGTCTCCCGAGTTCTCAAGTACATCCCCGAGCTGCAGAAG
CAGGTGGACAACCTGGAGAGGAGGAAGAAGGAGCTGACGAACGCCAACTGCA
AACCAGGAGTTCTGAACACGAGCCAGATTGTAACCTCCATTGTTTCTGCTACTTG
CCTCAACGATACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAACGTGGCTG
CCACAAGTCTTCCTCTGTCCAAGTGCATAAAAGTGATGGAGAACGAAGGCCTTC
ACCTAATTAGTTCATCAACTTACTCCACCTTCGACAACAGGACATTCTATAGCCT
CCATGTTTCAGAGAAGTCAAAGAACGATGAAGGAGGAGTGCCCAGCATTCTGCG
ATGAACTGGAGAGGATTATC

>SEQIDNO:93

ATGGACCATCAGCTGTTTCGACGACCCCTTCGGGAGCAGCATCTCGTCGCTGGA

GGCGGACATCTTCTCCGCCGGCGGGCGGCGGACAGCTGCCGTGCGCCGCCGTGG
CCGGACCTCGACCTCGACGACGACTACGACATACACGACCTCTCCGCGCCGGC
CGCCAACGCCGCCACCTCCTCGGGCGGGCGGCTATGGCTCCGGCGGCTCCGGC
AGGAAGCTCAGCCACAACGCATACGAGCGCGACCGCCGGAAGCAGCTCAACG
AGCTCTACTCCTCGCTCCGATCCCTCCTCCCGGACGCTGATCACACTAAGAAGC
TGAGCATCCCCACCACCGTGTCCCGAGTTCTCAAGTACATCCCGGAGCTGCAG
AAGCAGGTGGATAACCTGGAGAGGAGGAAGAAGGAGCTGACCAACGCCAACTG
CAAGCCGGGAGTTCTCAACACCAAAGAGATCGTAACTCCCATTGTTTCTGCTAC
TTGCCTTAACGACACGGAGATCATGGTTCAGGTCAGCCTGCACAGCAATGTGGC
CGCCACAGCTCTCCCTCTCTCCAAGTGCATAAAGGTGCTAGAAAACGAAGGCCT
TCTCCTCGTCAGCTCATCAACCTACTCCACCTTCGAGAACAAAGACATTCTATAGC
CTCCATCTTCAGAGAAGTCAAAGAACGATGAAGGAGCAGTGCCCAGGATTCTGC
GACGAACTGGAGAAGATCGTCAGGAAGAAAGCAGGGGCG

>SEQIDNO:94

GCCTCGTGCCGGCGGGTGCTCAAGTACATCCCGGAGCTGCAGAAGCAGGTGG
ACGGACTGGAGAAGAAGAAGGAGGAGCTGACGCGCGCCAACTGCAAGCCCGG
CGTGCTGACCATGAAGGAGAACATGGCTCCGATCGTGTCCGCCACCTGCCTCG
ATGACAGAGAAATCATGGTCCAGGTCAGCCTGGTGAGCACCATGGCCGGAGTT
CTGCCCATGTCCAAGTGCATCAAGGTGCTGGAGAACGAAGGCCTACGCCTCAT
CAGCTCGTCCACTTCCGCGTTTCAACAACAGGACGTTCTATAGCCTCCATCTTCA
GAGAACCCAACGGACGATGAGCAAGGAGTGTCCGGCATTGTTGTGAAGAACTGG
AGAACGCCCTGACGCAAAAGGCAGGACTACGTCTACATCACCACCAG

>SEQIDNO:95

ATGGATCACCAGCTGTACGGCGACCCCTCCGCGAGCAGCTTCTCTCCGCTGGA
GGCACAGATCTTCTCCGGCCAGCTGCCGCCGTGTCACGCCATGGCCAAATC
TCGACGTTGACCTCGCCCTGGACCTCGACGTTCTCGAGGATGACATCGTCCGG

GAGCTCTCTGCTGGCACAGTGGCAAACGCGGCATCGTCAGGTTCCGGCTCCGG
CGCCCACAAGAAGCTCAGCCACAACGCGTACGAGCGCGACCGCCGGAAGCAG
CTCAACGAGCTATACCTCTCGCTCCGTTCTCTCCTCCCGGACGCCGACCACACC
AAGAAGCTGAGTATTCCGACGACGGTGTGTGCGAGCGCTCAAGTACATCCCCGA
GCTGCAGAAACAGGTCGAGAATCTGGAGAAGAAGAAAGAGAAACTGGCTAGTG
CCAACTGCAAACCAGGGGTACTGAGCGTGACCGGCAGCATAGCTCCAACTGTG
TCCGCTACTTGCCTCAACCACAAGGAAATCATGGTTCAGATTAGCTTGCTGAGA
GATACAGATGCTTCTACAGCTCTACCTCTTTCCAAGTGTATAAATGTACTGGAGA
ACGAAGGACTTCAGCTCATCAGTTCATCGACTTCCTCCACCTTTGGGAACAAAA
CGTTCTATAACCTCCATCTTCAGAGAAGTCAAGGAGCCACTAACATGGAGTGC
CCATCGTTTTG

>SEQIDNO:96

CGCAGATCTTCTCCAGCCAGCTGCCGCCGTACCGCCGTGGCCGAATCTCGAT
GTTGACGTTGACCTGGACCTCGACGTTCTTGAGGACGACGTCGTCCGCGAACT
CTCAGGGAGGCCGGCAAACGCGGCATCGTCAGGCTCCGGCTCCGGCGGGCCCC
GGCTCCCACAAGAAGCTCAGTCACAACGCGTACGAGCGCGACCGCCGGAAGC
AGCTCAACGAGCTCTACCTCTCACTCCGTTCTCTCCTGCCGGACGCCGACCACA
CTAAGAAGCTGAGTATTCCGACGATGGTGTGTGCGAGCGCTCAAGTACATCCCGA
GCTGCAGAAACAGGTCGAGAATCTGGAGAAGAAGAAAGAGAAACTTGCTAGTTC
CAACTGCAAACCAGAGGTACTGAGCGCAAGCGGCAGCATAGCTCTAACTGTGT
CCGCTACTTGCCTCAACGACAAGGAAATCATGGTTCAGATTAGCTTGCTGAGAC
ATACGGATGCTGCTACAGCTCTACCTCTTTCCAAGTGTATAAATGTACTGGAGAA
CGAAGGACTTGAGCTCGTCAGTTCATCGACTTCCTGCACCTTTGGGAACAAAAT
GTTCTATAACCTCCATCTTCAGAGAAGTCAAGGAGCGCTAACATGGGAGTGTCC
ATCCTTCTGTGACAAATTGGAACAAGCAATCAGGAAAACAGCAGGATTA

>SEQIDNO:97

CGACGGAAGCAGCTCAACGACCTTTACTCCTCGCTCCGCTCCCTCCTCCCGGA
CGCTGACCACACCAAGAAGCTGAGCATCCCCACCACCGTGTCCCGAGTCCTCA
AGTACATCCCGGAGCTGCAGAAGCAGGTGGACAACCTGGAGAGGAGGAAGCG
GGAGCTGACCAACGCCAACTGCAAGCCGGGAGTTCTCAACACCAGCGAGATCG
TAACTACTCCCATTGTTTCTGCTACTTGCCTCAACGACACGGAGATCATGGTTCA
GGTCAGCCTGCACAGCAATGTGGCAGCCACGGCTCTCCCTCTCTCCAAGTGCA
TAAAGGTGCTGGAGGACGCAGGCCTTCACCTCATCAGCTCATCAACCTACTCCA
CCTTTGGGAACAAGACATTCTATAGCCTCCATCTTCAGGTGTGCATGCATGTTCA
TTCAATGGTTCCTGCCGTTTCCTTCAATTTTTTTATC

>SEQIDNO:98

MCALVPPLYPNFGWPCGDHSFYETDDVSNTFLDFPLPDLTVTHENVSSENNRTLLD
NPVVMKKLNHNASERERRKKINTMFSSLRSCLPPTNQTKLSVSATVSQALKYIPELQ
EQVKKLMKKKEELSFQISGQRDLVYTDQNSKSEEGVTSYASTVSSTRLSETEVMVQ
ISSLQTEKCSFGNVLSGVEEDGLVLVGASSSRSHGERLFYSMHLQIKNGQVNSEEL
GDRLLYLYEKC GHSFT

>SEQIDNO:99

MCALVPPLYPNFGWPCGDHSFYETDDVSNTFLDFPLPDLTVTHENVSSENNRTLLD
NPVVMKKLNHNASERERRKKINTMFSSLRSCLPPTNQTKKLSVSATVSQALKYIPEL
QE QVKKLMKKKEELSFQISGQRDLVYTDQNSKSEEGVTSYASTVSSTRLSETEVMV
QISSLQTEKCSFGNVLSGVEEDGLVLVGASSSRSHGERLFYSMHLQIKNGQVNSEE
LGDRLLYLYEKC GHSFT

>SEQIDNO:100

PRRLRGRLLLRRPPDPGGVRRRGADGGANLIHNAYERDRRKQRNELYSSLRALLP

DADHTKKLSIPTTVSRVLKYIPELQKQVENLERKKKELTTTSTTNCKPGVLGSQLMS
EGMAPIVSATCINDMEIMVQVSLLSNVAGSVLPLSKCIKVLENEGLHFISSTSSGFG
NRTFYSIHLQRSEGTINEECPAFCEERLEKVVRNKAKL

>SEQIDNO:101

MEHQLFDDPFSSSISSLEADIFSAGGQLPSPWPDLDDLDDDDGIHDLSAPAGNPT
SSGGYGSGGGSHRKISHNAYERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTVSR
VLKYIPELQKQVDNLERRKKELTNANCKPGVLNTSQIVTPIVSATCLNDTEIMVQVSL
HSNVAATSLPLSKCIKVMENEGLHLISSSTYSTFDNRTFYSLHVQRSQRTMKEECPA
FCDELERII

>SEQIDNO:102

MDHQLFDDPFGSSSISSLEADIFSAGGGGQLPSPWPDLDDLDDDDYDIHDLSAPAANA
ATSSGGGYGSGGSGRKL SHNAYERDRRKQLNELYSSLRSLLPDADHTKKLSIPTTV
SRVLKYIPELQKQVDNLERRKKELTNANCKPGVLNTKEIVTPIVSATCLNDTEIMVQV
SLHSNVAATALPLSKCIKVLENEGLLLVSSSTYSTFENKTFYSLHLQRSQRTMKEQC
PGFCDELEKIVRKKAGA

>SEQIDNO:103

ASCRRVLKYIPELQKQVDGLEKKKEELTRANCKPGVLTMKENMAPIVSATCLDDREI
MVQVSLVSTMAGVLPMSKCIKVLENEGLRLISSSTSFAHNRTFYSLHLQRTQRTMSK
ECPAFCEELENALTQKAGLRLHHHQ

>SEQIDNO:104

MDHQLYGDPSASSFSPLAQIFSGQLPPSSTPWPNLDVDLALDLVDLEDDIVRELSA
GTVANAASSGSGSGAHKKL SHNAYERDRRKQLNELYLSLRSLLPDADHTKKLSIPTT
VCRALKYIPELQKQVENLEKKKEKLANCKPGVLSVTGSIAPT VSATCLNHKEIMV

QISLLRDTASTALPLSKCINVLENEGLQLISSSTSSTFGNKTFYNLHLQRSQGATKH
GVPIVL

>SEQIDNO:105

RRKQLNDLYSSLRSLLPDADHTKKLSIPTTVSRVLKYIPELQKQVDNLERRKRELTNA
NCKPGVLNTSEIVTTPIVSATCLNDTEIMVQVSLHSNVAATALPLSKCIKVLEDAGLHL
ISSSTYSTFGNKTFYSLHLQVCMHVHSMVPAVSFNFFI

Literatura

1. Bailey, P.C., Martin C., Toledo-Ortiz ,G., Quail ,P.H., Huq, E., Heim, M.A., Jakoby, M., Weber, M., Weisshaar, B.(2003). Update on the Basic Helix-Loop-Helix Transcription Factor Gene Family in *Arabidopsis thaliana*. Plant Cell 15(11), 2497-2502.

2. Brownlie, P., Ceska, T.A., Lamers, M., Romier, C., Stier, G., Teo, H., Suck, D. (1997). The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. Structure 5, 509-520.

3. Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P.C. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. Mol. Biol. Evol. 20, 735-747.

4. Kang, H.-G., Foley, R.C., Onate-Sanchez, L., Lin, C., Singh, K.B. (2003). Target genes for OBP3, a Dof transcription factor, include novel basic helix-loop-helix domain proteins inducible by salicylic acid. Plant J 35, 362-372.

5. Ling H.Q., Bauer P., Berczky Z., Keller B., Ganai M. (2002). The tomato fer gene encoding a bHLH protein controls iron-uptake responses in roots. Proc. Natl. Acad. Sci. USA 2002; 99:13938-13943

6. Martinez-Garcia, J.F., Huq, E., Quail, P.H. (2000). Direct Targeting of Light Signals to a Promoter Element-Bound Transcription Factor. Science 288, 859-863.

7.Ogo, Y., Itai, R.N., Nakanishi, H., Inoue, H., Kobayashi, T., Suzuki, M., Takhashi, M., Mori, S., Nishizawa, N.K. Isolation and characterization of IRO2, a novel iron-regulated bHLH transcription factor in graminaceous plants. J. Exp. Bot. 57, 2867-2878.

8.Ramsay, N.A., Walker, A.R., Mooney, M., Gray, J.C. (2003). Two basic-helix-loop-helix genes (*MYC-146* and *GL3*) from *Arabidopsis* can activate anthocyanin biosynthesis in a white-flowered *Matthiola incana* mutant. Plant Mol. Biol. 52, 679-687.

9.Toledo-Ortiz, G., Huq, E., and Quail, P.H. (2003). The *Arabidopsis* basic/helix-loop-helix transcription factor family. Plant Cell 15(8), 1749-1770.

10.Tominaga, R., Iwata, M., Okada, K., Wada, T. (2007). Functional Analysis of the Epidermal-Specific MYB Genes *CAPRICE* and *WEREWOLF* in *Arabidopsis*. Plant Cell 19(7), 264-2277.

11.Vorwieger, A., Gryczka, C., Czihal, A., Douchkov, D., Tiedemann, J., Mock, H.-P., Jakoby, M., Weisshaar, B., Saalback, I., Baumlein, H. (2007). Iron assimilation and transcription factor controlled synthesis of riboflavin in plants. Planta 226, 147-158.

12.Wang H.-Y., Klatte, M., Jakoby, M., Baumlein, H., Weisshaar, B., Bauer, P. (2007) Iron deficiency-mediated stress regulation of four subgroup Ib BHLH genes in *Arabidopsis thaliana*. Planta 226, 897-908.

13.Yuan, Y., Wu, H., Wang, N., Li, J., Zhao, W., Du, J., Wang, D, Ling, H.-Q. (2008). FIT interacts with AtbHLH38 and AtbHLH39 in regulating iron uptake gene expression for iron homeostatis in *Arabidopsis*.

14.Zhang, F., Gonzalez, A., Zhao, M., Payne, T., Lloyd, A. (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. Development 130, 4859-4869.

REIVINDICAÇÕES

1. Método para produção de uma planta transgênica tendo tolerância aumentada ao estresse por calor em relação a um controle do tipo selvagem, o método **CARACTERIZADO** pelas etapas de transformar uma planta, uma cultura de tecido de planta ou uma célula de planta com um vetor compreendendo um construto de ácido nucléico que compreende uma sequência de ácido nucléico que codifica um polipeptídeo hélice-alça-hélice básico (bHLH) do subgrupo 1b selecionado a partir do grupo que consiste em bHLH38, bHLH39, bHLH100 e bHLH101 para obter uma planta transformada, uma cultura de tecido transformada ou uma célula de planta transformada com expressão ou atividade de bHLH aumentada, e crescer a dita planta transformada ou regenerar uma planta a partir da dita cultura de tecido de planta transformada ou célula de planta transformada, em que o dito construto de ácido nucléico compreende uma sequência de ácido nucléico selecionada a partir das SEQ ID NOs: 1-17, 29-52, 79-83 e 89-97, e em que uma planta tendo tolerância aumentada ao estresse por calor em relação a um controle do tipo selvagem é produzida.

2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito construto de ácido nucléico compreende um promotor constitutivo, um promotor induzível ou um promotor específico de tecido.

3. Método, de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de que o dito promotor específico de tecido é um promotor de raiz.

4. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito construto de ácido nucléico compreende uma sequência de ácido nucléico selecionada a partir das SEQ ID NOs: 1-17.

5. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito construto de ácido nucléico compreende uma sequência de ácido nucléico selecionada a partir das SEQ ID NOs: 29-52.

6. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito construto de ácido nucléico compreende uma sequência de ácido nucléico selecionada a partir das SEQ ID NOs: 79-83.

7. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o dito construto de ácido nucléico compreende uma sequência de ácido nucléico selecionada a partir das SEQ ID NOs: 89-97.

FIGURA 1

Alinhamento de ClustalW de proteínas AtbHLH

```

AtbHLH39      -MCALVPPLFENFGWPSTGEYDSYLAGQILNNGGFLEDFVPEETYGAVTAVTQHQNSFG
AtbHLH38      -MCALVPSEFTNFGWPESTNGYESYYGAGDMLNNGTFLELTVR-QTY----EVTTHQNSLG
AtbHLH100     -MCALVPPELYPNFGWPCG--DRSEYETDDVSN--TFLDEPLF-----DLTVTHEN---
AtbHLH101     MEYFWLQSQVHSFSEPTLNEPSELHPLQDSKSNENLHMSLS-----HSNNTNENN
               . . . . . * . . . . . : : . . . . . : *

AtbHLH39      VSVSSEGN-EIDNNEVVVKKLNHNASERDRRRKINSLESSLRSCLPASGQSKKLSIPATV
AtbHLH38      VSVSSEGN-EIDNNEVVVKKLNHNASERDRRRKINTLFSSLSRCLPASDQSKKLSIPETV
AtbHLH100     --VSENNRTLLDNEFVVMKKLNHNASERDRRRKINTMESSLRSCLPPTNQTKKLSVSATV
AtbHLH101     NNYQEEDR----GAVVLEKKLNHNASERDRRRRLNALYSSLRALLELDQKKKLSIPATV
               . . . . . * : ***** : * : : : : : : : : * : : : : *

AtbHLH39      GRSLKYIPELQEQVKKLIKKNDELLVQISQQRNTECYVR--QPPKAVANYISIVSATRLG
AtbHLH38      SKSEKYIPELQQQVKRLIQSKREILVRYSGQEDFELYDK--QQPKAVASYLSTVSATRLG
AtbHLH100     SQALKYIPELQEQVKKLMKKKEELSEFQISGQRDLVYTDQNSKSEEGVTSYASTVSSTRLS
AtbHLH101     ARVVKYIPQQKQELQELSRKEELLKRISPKTHQEQLENKAMMDSIDSSSSQRIANWLT
               : : ***** : : : : : : : * : : : : : : : : : : :

AtbHLH39      DNEVMVQISSSKIHNFESISNVLGGLIEDRFVLVDMSSSRSQGERLFYTLHLQVEKIENYK
AtbHLH38      DNEVMVQVSSSKIHNFESISNVLGGIEDGFEVLVDVSSSSRSQGERLFYTLHLQVENMDQYK
AtbHLH100     ETEVMVQISSLQTEKQSEGNVLSCVEEDGLVLVVGASSSRSHGERLFYSMHLQIK---NGQ
AtbHLH101     DTEIADVQIATSKWT--SVSDMLRLLEENGLNVI SVSSSVSSTARIIFYTLHLQMEG--LCK
               : . : * : : : * . : : : : : : : : : : : : : : : : :

AtbHLH39      LNCEELSQRMVLYVEECGNSYI 252 (SEQ ID NO: 122)
AtbHLH38      INCEELSERMELYVEKCENSPN 253 (SEQ ID NO: 123)
AtbHLH100     VNSEELQDRLLYVEECGHSET 242 (SEQ ID NO: 124)
AtbHLH101     VLEELINCMLLGLRQS----- 240 (SEQ ID NO: 125)
               : : * : : : * : :

```