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

(21) PI 0716954-0 A2

* B R P I 0 7 1 6 9 5 4 A 2 *

(22) Data de Depósito: 25/09/2007
(43) Data da Publicação: 29/10/2013
(RPI 2234)

(51) Int.Cl.:
C12N 1/00

(54) Título: PRODUÇÃO DE ISOPRENÓIDES E PRECURSORES DE ISOPRENÓIDES

(57) Resumo:

(30) Prioridade Unionista: 26/09/2006 US 60/826,970

(73) Titular(es): Amyris Biotechnologies, Inc., The Regents Of The University Of California

(72) Inventor(es): Douglas J. Pitera, Fernaz Nowroozi, Jack D. Newman, Jay D. Keasling, Jennifer Anthony, Larry Anthony

(74) Procurador(es): Dannemann ,Siemsen, Bigler & Ipanema Moreira

(86) Pedido Internacional: PCT US2007020790 de 25/09/2007

(87) Publicação Internacional: WO 2008/039499de 03/04/2008

Relatório Descritivo da Patente de Invenção para **"CÉLULAS HOSPEDEIRAS CAPAZES DE PRODUZIR ISOPRENÓIDES".**

Referência Cruzada

Este pedido de patente reivindica benefício de prioridade do Pedido de Patente Provisório norte-americano N° 60/826,970, depositado em 26 de setembro de 2006, cujo pedido é aqui incorporado por referência em sua totalidade.

Antecedentes

Isoprenóides são abundantes na natureza. Eles compreendem uma família diversa de mais de 40.000 produtos individuais, muitos dos quais são vitais para organismos vivos. Isoprenóides servem para manter a fluidez da célula, transportar elétrons e outras funções metabólicas. Um largo número de isoprenóides naturais e sintéticos é útil como farmacêuticos, cosméticos, perfumes, pigmentos e corantes, fungicidas, antissépticos, nutracêuticos e intermediários químicos finos.

Um produto isoprenóide é tipicamente composto de unidades de difosfato de isopentenila (IPP) de cinco carbonos de repetição, embora isoprenóides e politerpenos irregulares tenham sido relatados. Na natureza, isoprenóides são sintetizados por condensações consecutivas de seu IPP precursor e seu isômero pirofosfato de dimetilalila (DMAPP). Dois caminhos para estes precursores são conhecidos. Eucariotos, com a exceção de plantas, geralmente usam o caminho dependente de mevalonato (MEV) para converter a acetil coenzima A (acetyl-CoA) em IPP, o qual é subsequentemente isomerizado a DMAPP. Procariotos, com algumas exceções, empregam tipicamente apenas o caminho independente de mevalonato ou desoxixilulose-5-fosfato (DXP) para produzir IPP e DMAPP. As plantas usam tanto o caminho MEV como o caminho DXP.

Tradicionalmente, isoprenóides têm sido fabricados através de extração de fontes naturais, tais como plantas, micróbios e animais. Entretanto, o rendimento por meio de extração é usualmente muito lento devido a certas limitações. Em primeiro lugar, a maioria dos isoprenóides se acumulam na natureza apenas em pequenas quantidades. Em segundo lugar, os

organismos de fonte em geral não são propícios para cultivo em larga escala, que é necessário para produzir quantidades comercialmente viáveis de um isoprenóide desejado. Em terceiro lugar, o requisito de certos solventes tóxicos para extração de isoprenóide de necessidade de manuseio especial e procedimentos de eliminação, desse modo complicando a produção comercial de isoprenóides.

O esclarecimento dos caminhos metabólicos MEV e DXP tem tornado possível a produção biossintética de isoprenóides. Por exemplo, micróbios têm sido engenheirados para expressar em excesso uma parte ou todo o caminho do mevalonato para produção do isoprenóide amorfa-4,11-dieno. Outros esforços têm focado no equilíbrio do pool de gliceraldeído-3-fosfato e piruvato, ou aumento na expressão de 1-desóxi-D-xilulose-5-fosfato sintase (dsx) e IPP isomerase (idi).

Apesar disso, dadas as quantidades muito altas de produtos isoprenóides necessários para muitas aplicações industriais, permanece uma necessidade por sistemas de expressão e procedimentos de fermentação que produzam ainda mais isoprenóides do que disponível com as tecnologias atuais.

Literatura

Patentes US números 7.172.886 e 7.192.751; Farmer et al. (2001) *Biotechnol. Prog.* 17:57-61; Kajiwara et al. (1997) *Biochem. J.* 324:421-426; e Kim et al. (2001) *Biotechnol. Bioeng.* 72:408-415; Martin et al. (2003) *Nat. Biotech.* 21(7):796-802; Patente US número 7.183.089;

Sumário da Invenção

A presente invenção proporciona células hospedeiras geneticamente modificadas e uso das mesmas para a produção de compostos isoprenóides.

Breve Descrição dos Desenhos

A figura 1 é uma representação esquemática do caminho do mevalonato (MEV) para a produção de difosfato de isopentenila (IPP).

A figura 2 é uma representação esquemática de caminhos metabólicos de isoprenóide que resultam na produção de poliprenil difosfatos

geranil difosfato (GPP), difosfato de farnesila (FPP) e difosfato de geranilgeranila (GGPPP) como intermediários do caminho biossintético de isoprenóide a partir de IPP e DMAPP.

A figura 3 é um mapa de plasmídeo de expressão pAM39.

A figura 4 é um mapa de plasmídeo de expressão pAM40.

A figura 5 é um mapa de plasmídeo de expressão pAM45.

A figura 6 é um mapa de plasmídeo de expressão pTrc99A-ADS.

A figura 7 é um mapa de plasmídeo de expressão pAM94.

As figuras 8A e 8B mostram produção de amorfa-4,11-dieno em cepas hospedeiras, nas quais o MBIS óperon é expresso a partir de um plasmídeo com número de cópia mais alto e um promotor mais forte.

A figura 9 mostra produção de amorfa-4,11-dieno em cepas hospedeiras, nas quais a dosagem do gene de mevalonato quinase é aumentada.

A figura 10 mostra produção de amorfa-4,11-dieno em cepas hospedeiras, nas quais a dosagem do gene e a expressão de mevalonato quinase são aumentadas

Definições

As expressões "isoprenóide", "composto isoprenóide", "terpeno", "composto de terpeno", "terpenóide" e "composto de terpenóide" são usadas aqui de modo permutável, e referem-se a qualquer composto que seja capaz de ser derivado de IPP. O número de átomos de C presentes nos isoprenóides é tipicamente igualmente divisível por cinco (por exemplo, C5, C10, C15, C20, C25, C30 e C40). Isoprenóides e politerpenos irregulares têm sido relatados e também estão incluídos na definição de "isoprenóide". Compostos isoprenóides incluem, porém não estão limitados a, monoterpenos, diterpenos, triterpenos, sesquiterpenos, e politerpenos.

Como usado aqui, a expressão "difosfato de prenila" é usada de modo permutável com "pirofosfato de prenila", e inclui difosfatos de monoprenila que têm um único grupo prenila (por exemplo, IPP e DMAPP), assim como difosfatos de poliprenila que incluem 2 ou mais grupos prenila. Difosfatos de monoprenila incluem pirofosfato de isopentenila (IPP) e seu isômero

dimetilalil pirofosfato (DMAPP).

Como usado aqui, a expressão "terpeno sintase" (também referida como uma "terpeno ciclase") refere-se a qualquer enzima que enzimaticamente modifique IPP, DMAPP ou um pirofosfato de poliprenila, de modo que um composto precursor de terpenóide seja produzido. A expressão "terpeno sintase" inclui enzimas que catalisam a conversão de um difosfato de prenila em um isoprenóide ou precursor de isoprenóide.

O termo "pirofosfato" é aqui usado de modo permutável com "difosfato". Desse modo, por exemplo, as expressões "difosfato de prenila" e "pirofosfato de prenila" são permutáveis; as expressões "pirofosfato de isopentenila" e "difosfato de isopentenila" são permutáveis; as expressões "difosfato de farnesila" e "pirofosfato de farnesila" são permutáveis; etc.

A expressão "caminho do mevalonato" ou "caminho MEV" é usada aqui para se referir ao caminho biossintético que converte acetil-CoA em IPP. O caminho do mevalonato compreende enzimas que catalisam as seguintes etapas: (a) condensação de duas moléculas de acetil-CoA em acetoacetil-CoA (por exemplo, através da ação da acetoacetil-CoA tiolase); (b) condensação de acetoacetil-CoA com acetil-CoA para formar hidroximetilglutaril-Coenzima A (HMG-CoA) (por exemplo, através da ação da HMG-CoA sintase (HMGS)); (c) conversão da HMG-CoA em mevalonato (por exemplo, através da ação da HMG-CoA redutase (HMGR)); (d) fosforilação do mevalonato em mevalonato 5-fosfato (por exemplo, através da ação da mevalonato quinase (MK)); (e) conversão do mevalonato 5-fosfato em mevalonato 5-pirofosfato (por exemplo, através da ação da fosfomevalonato quinase (PMK)) e (f) conversão do mevalonato 5-pirofosfato em pirofosfato de isopentenila (por exemplo, através da ação da mevalonato pirofosfato descarboxilase (MPD)). O caminho do mevalonato é ilustrado de modo esquemático na figura 1. A "metade superior" do caminho do mevalonato refere-se às enzimas responsáveis pela conversão de acetil-CoA em mevalonato.

A expressão "caminho de 1-desóxi-D-xilulose 5-difosfato" ou "caminho DXP" é usada aqui para se referir ao caminho que converte glice-

raldeído-3-fosfato e piruvato em IPP e DMAPP através de um caminho DXP intermediário.

Como usado aqui, a expressão "prenil transferase" é usada de modo permutável com as expressões "difosfato de isoprenila sintase" e "sintase de poliprenila" (por exemplo, "GPP sintase", "FPP sintase", "GGPP sintase", etc.) para se referir a uma enzima que catalisa a condensação 1'-4 consecutiva de difosfato de isopentenila com substratos iniciador alílicos, resultando na formação de prenil difosfatos de vários comprimentos de cadeia.

As expressões "polinucleotídeo" e "ácido nucléico", usadas aqui de modo permutável, referem-se a uma forma polimérica de nucleotídeos de qualquer comprimento, tanto ribonucleotídeos quanto desoxirribonucleotídeos. Desse modo, esta expressão inclui, porém não está limitada a, DNA ou RNA de filamento único, duplo ou de multifilamentos, DNA genômico, cDNA, híbridos DNA-RNA ou um polímero que compreende bases purina e pirimidina ou outras bases naturais, quimicamente ou bioquimicamente modificadas, não naturais ou derivatizadas de nucleotídeo.

Os termos "peptídeo", "polipeptídeo" e "proteína" são aqui usados de modo permutável e referem-se a uma forma polimérica de aminoácidos de qualquer comprimento, o qual pode incluir aminoácidos codificados e não-codificados, aminoácidos quimicamente ou bioquimicamente modificados ou derivatizados, e polipeptídeos que têm estrutura de peptídeo modificado.

A expressão "de ocorrência natural", como usada aqui, como aplicada a um ácido nucléico, uma célula ou um organismo, refere-se a um ácido nucléico, uma célula ou um organismo que é encontrado na natureza. Por exemplo, uma sequência de polipeptídeo ou polinucleotídeo que está presente em um organismo (incluindo vírus) que pode ser isolado de uma fonte na natureza e o qual não tenha sido intencionalmente modificado por um ser humano em laboratório é de ocorrência natural.

Conforme usado aqui, o termo "isolado" destina-se a descrever um polinucleotídeo ou uma célula que está em um ambiente diferente da-

quele no qual o polinucleotídeo, o polipeptídeo ou a célula ocorre naturalmente. Uma célula hospedeira geneticamente modificada isolada pode estar presente em uma população mista de células hospedeiras geneticamente modificadas.

Conforme usada aqui, a expressão "ácido nucléico exógeno" refere-se a um ácido nucléico que não é normalmente ou naturalmente encontrado em e/ou produzido por uma dada bactéria, organismo ou célula na natureza. Como usada aqui, a expressão "ácido nucléico endógeno" refere-se a um ácido nucléico que é normalmente encontrado em, e/ou produzido por uma dada bactéria, organismo ou célula na natureza. Um "ácido nucléico endógeno" também é referido como "ácido nucléico nativo" ou um ácido nucléico que é "nativo" para uma dada bactéria, organismo ou célula. Por exemplo, os ácidos nucléicos que codificam HMGS, mevalonato quinase e fosfomevalonato quinase na representação de ácidos nucléicos exógenos para *E. coli*.

A expressão "ácido nucléico heterólogo", como usado aqui, refere-se a um ácido nucléico em que pelo menos um dos seguintes é verdade: (a) o ácido nucléico é externo ("exógeno") a (isto é, não naturalmente encontrado em) um dado micro-organismo hospedeiro ou célula hospedeira; (b) o ácido nucléico compreende uma sequência de nucleotídeo que é naturalmente encontrada em (por exemplo, é "endógena a") um dado micro-organismo hospedeiro ou célula hospedeira (por exemplo, o ácido nucléico compreende uma sequência de nucleotídeo que é endógena ao micro-organismo hospedeiro ou célula hospedeira), porém tanto é produzido na célula em uma quantidade não natural (por exemplo, mais do que esperado ou mais do que naturalmente encontrado) quanto difere em sequência da sequência de nucleotídeo endógena, de modo que a mesma proteína codificada (que tem a mesma ou substancialmente a mesma sequência de aminoácido), como encontrada endogenamente é produzida na célula em uma quantidade não natural (por exemplo, mais do que esperada ou mais do que naturalmente encontrada); (c) o ácido nucléico compreende duas ou mais sequências ou segmentos de nucleotídeo que não são encontrados na

mesma relação entre si na natureza, por exemplo, o ácido nucléico é recombinante.

A expressão "polipeptídeo heterólogo", como usada aqui, refere-se a um polipeptídeo que não está naturalmente associado com um dado polipeptídeo. Por exemplo, uma enzima modificadora de precursor de isoprenóide que compreende um "domínio de transmembrana heterólogo" refere-se a uma enzima modificadora de precursor de isoprenóide que compreende um domínio de transmembrana que não está normalmente associado com (por exemplo, não normalmente contíguo a, não normalmente encontrado na mesma cadeia de polipeptídeo com) a enzima modificadora de precursor de isoprenóide na natureza.

"Recombinante", como usado aqui, significa que um ácido nucléico particular (DNA ou RNA) é o produto de várias recombinações de etapas de clonagem, restrição e/ou ligação, resultando em um construto que tem uma sequência estrutural de codificação ou não-codificação distinguível de ácidos nucléicos endógenos encontrados em sistemas naturais. Geralmente, sequências de DNA que codificam a sequência estrutural de codificação podem ser reunidas a partir de fragmentos de cDNA e ligadores curtos de oligonucleotídeos, ou a partir de uma série de oligonucleotídeos sintéticos para proporcionar um ácido nucléico sintético, o qual é capaz de ser expresso a partir de uma unidade transcripcional recombinante contida em uma célula ou em uma transcrição livre de célula e sistema de tradução. Tais sequências podem ser proporcionadas na forma de uma estrutura de leitura aberta ininterrupta por sequências internas não traduzidas ou íntrons, os quais estão tipicamente presentes em genes eucariotos. DNA genômico que compreende as sequências relevantes também podem ser usados na formulação de um gene recombinante ou unidade transcripcional. Sequências de DNA não traduzidas podem estar presentes em 5' ou 3' a partir da estrutura de leitura aberta, onde tais sequências não interferem com a manipulação ou expressão das regiões de codificação e podem, de fato, agir para modular a produção de um produto desejado através de vários mecanismos (vide "sequências regulatórias de DNA", abaixo).

Desse modo, por exemplo, o termo polinucleotídeo "recombinante" ou ácido nucléico "recombinante" refere-se a um que não ocorre naturalmente, por exemplo, é feito através de combinação artificial de outros dois segmentos separados de sequência através de intervenção humana. Esta combinação artificial é frequentemente realizada através de meios de síntese química ou através de manipulação artificial de segmentos isolados de ácidos nucléicos, por exemplo, através de técnicas de engenharia genética. Isto é usualmente feito para substituir um códon com um códom redundante que codifica o mesmo ou um aminoácido conservativo, enquanto tipicamente introduzindo ou removendo um sítio de recognição de sequência. De maneira alternativa, isto é efetuado para unir segmentos de ácido nucléico de funções desejadas para gerar uma combinação desejada de funções. Esta combinação artificial é frequentemente realizada através de meios de síntese química ou através de manipulação artificial de segmentos isolados de ácidos nucléicos, por exemplo, através de técnicas de engenharia genética.

De maneira similar, o termo polipeptídeo "recombinante" refere-se a um polipeptídeo que não é de ocorrência natural, por exemplo, é feito através de combinação artificial de outros dois segmentos separados de sequência de amino através de intervenção humana. Desse modo, um polipeptídeo que compreende uma sequência de aminoácido heterólogo é recombinante.

Por "construto" ou "vetor" deve ser entendido um ácido nucléico recombinante, geralmente DNA recombinante, o qual tem sido gerado com a finalidade da expressão e/ou propagação de uma(s) sequência(s) específica(s) de nucleotídeo, ou pode ser usado na contrução de outras sequências de nucleotídeo recombinante.

Como usadas aqui, as expressões "óperon" e "unidade de transcrição única" são usadas de modo permutável para se referir a duas ou mais regiões contíguas de codificação (sequências de nucleotídeo que codificam um produto de gene tal como um RNA ou uma proteína) que são coordenadamente reguladas através de um ou mais elementos de controle

(por exemplo, um promotor). Como usada aqui, a expressão "produto de gene" refere-se a RNA codificado por DNA (ou vice-versa) ou proteína que é codificada por um RNA ou DNA, onde um gene irá tipicamente compreender uma ou mais sequências de nucleotídeo que codificam uma proteína, e pode também incluir íntrons e outras sequências de nucleotídeo de não-codificação.

A expressão "região de codificação", como usada aqui, refere-se a uma extensão contígua de nucleotídeos (uma sequência de nucleotídeo) que codifica um polipeptídeo. Por exemplo, uma "região de codificação que codifica MK" ou uma "região de codificação de MK" compreende uma sequência de nucleotídeo que codifica mevalonato quinase.

As expressões "sequências regulatórias de DNA", "elementos de controle" e "elementos regulatórios", aqui usadas de modo permutável, referem-se a sequências de controle transcripcional e translacional, tais como promotores, intensificadores, sinais de poliadenilação, terminadores, sinais de degradação de proteína e similares, que providenciam e/ou regulam expresão de uma sequência de codificação e/ou produção de um polipeptídeo codificado em uma célula hospedeira.

O termo "transformação" é usado aqui de modo permutável com "modificação genética" e refere-se à mudança genética permanente ou transiente induzida em uma célula seguinte à intrudução de novo ácido nucléico (isto é, DNA exógeno à célula). Mudança genética ("modificação") pode ser realizada tanto por incorporação do novo DNA ao genoma da célula hospedeira, quanto por manutenção transiente ou estável do novo DNA como um elemento episssomal. Onde a célula é uma célula eucariótica, uma mudança genética permanente é geralmente alcançada por introdução do DNA no genoma da célula. Em células procarióticas, mudanças permanentes podem ser introduzidas nos cromossomos ou através de elementos extracromossomais, tais como plasmídeos e vetores de expressão, os quais podem conter um ou mais marcadores selecionáveis para auxiliar sua manutenção na célula hospedeira recombinante. Métodos adequados de modificação genética incluem infecção viral, transfecção,

conjugação, fusão de protoplasma, eletroporação, tecnologia "particle gun", precipitação de fosfato de cálcio, microinjeção direta e similares. A escolha de método é geralmente dependente do tipo de célula sendo transformada e das circunstâncias sob as quais a transformação ocorre (isto é, *in vitro, ex vivo,* ou *in vivo*). Uma discussão geral destes métodos pode ser encontrada em Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

"Operacionalmente ligado" refere-se a uma justaposição em que os componentes assim descritos estão em uma relação que os permite funcionar em sua maneira pretendida. Por exemplo, um promotor está operacionalmente ligado a uma sequência de codificação se o promotor afeta sua transcrição ou expressão. Como usado aqui, as expressões "promotor heterólogo" e "regiões de controle heterólogas" referem-se a promotores a outras regiões de controle que não estão normalmente associadas com um ácido nucléico particular na natureza. Por exemplo, uma "região de controle transcricional heteróloga a uma região de codificação" é uma região de controle transcricional que não está normalmente associada com a região de controle na natureza.

Uma "célula hospedeira", como usada aqui, denota uma célula eucariótica *in vivo* ou *in vitro*, uma célula procariótica ou uma célula de um organismo multicelular (por exemplo, uma linhagem de célula) cultivado como uma entidade unicelular, cujas células eucarióticas e procarióticas podem ser, ou terem sido, usadas como recipientes para um ácido nucléico (por exemplo, um vetor de expresão que compreende uma sequência de nucleotídeo que codifica um ou mais produtos de gene de caminho biossintético, tais como produtos de gene do caminho do mevalonato), e incluem a progênie da célula original, que foi geneticamente modificada pelo ácido nucléico. Deve ser entendido que a progênie de uma única célula pode não ser necessariamente completamente idêntica em morfologia ou em complemento de DNA genômico ou DNA total como o progenitor devido à mutação natural, acidental ou deliberada. Uma "célula hospedeira recombinante" (também referida como uma "célula hospedeira

geneticamente modificada") é uma célula hospedeira na qual um ácido nucléico heterólogo foi introduzido, por exemplo, um vetor de expressão. Por exemplo, uma célula hospedeira procariótica geneticamente modificada em questão (por exemplo, uma bactéria) é uma célula hospedeira procariótica que, em virtude de introdução em uma célula hospedeira procariótica adequada de um ácido nucléico heterólogo, por exemplo, um ácido nucléico exógeno que é externo (normalmente não encontrado na natureza) à célula hospedeira procariótica, ou um ácido nucléico recombinante que não é normalmente encontrado na célula hospedeira procariótica; e uma célula hospedeira eucariótica geneticamente modificada em questão é uma célula hospedeira eucariótica que, em virtude de entrodução em uma célula hospedeira eucariótica adequada de um ácido nucléico heterólogo, por exemplo, um ácido nucléico exógeno que é externo à célula hospedeira eucariótica, ou um ácido nucléico recombinante que não é normalmente encontrado na célula hospedeira eucariótica.

A expressão "substituição de aminoácido conservativo" refere-se à permutabilidade em proteínas de resíduos de aminoácido que têm cadeias laterais similares. Por exemplo, um grupo de aminoácidos que têm cadeias laterais alifáticas consiste em glicina, alanina, valina, leucina e isoleucina; um grupo de aminoácidos que têm cadeias laterais hidroxil alifáticas consiste em serina e treonina, um grupo de aminoácidos que têm cadeias laterais contendo amida consiste em asparagina e glutamina; um grupo de aminoácidos que têm cadeis laterais aromáticas consiste em fenilalanina, tirosina e triptofano; um grupo de aminoácidos que têm cadeis laterais básicas consiste em lisina, arginina e histidina e um grupo de aminoácidos que têm cadeis laterais contendo enxofre consiste em cisteína e metionina. Grupos de substituição de aminoácidos conservativos exemplares são: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina e asparagina-glutamina.

"Ácidos nucléicos sintéticos" podem ser reunidos a partir de blocos de construção de oligonucleotídeos que são quimicamente sintetizados usando procedimentos conhecidos para aqueles versados na

técnica. Estes blocos de contrução são ligados e enrijecidos para formar segmentos de gene que são então enzimaticamente reunidos para construir o gene inteiro. "Quimicamente sintetizado", como relacionado a uma sequência de DNA, significa que os nucleotídeos componentes foram reunidos *in vitro*. Síntese química manual de DNA pode ser realizada usando procedimentos bem estabelecidos, ou síntese química automatizada pode ser realizada usando um dentre um número de máquinas comercialmente disponíveis. A sequência de nucleotídeo dos ácidos nucléicos pode ser modificada por expressão ótima à base de otimização de sequência de nucleotídeo para refletir o viés de códon da célula hospedeira. O versado na técnica aprecia a possibilidade de expressão bem-sucedida se o uso do códon é enviesado em direção àqueles códons favorecidos pelo hospedeiro. A determinação de códons preferidos pode ser baseada um uma pesquisa de genes derivados da célula hospedeira, onde a informação da sequência ~~está disponível.~~

Um polinucleotídeo ou polipeptídeo tem uma certa "identidade de sequência" percentual a outro polinucleotídeo ou polipeptídeo, significando que, quando alinhados, aquelas porcentagens de bases ou aminoácidos são as mesmas, e na mesma posição relativa, quando comparando as duas sequências. A similaridade de sequência pode ser determinada em um número de diferentes maneiras. Para determinar a identidade de sequência, sequências podem ser alinhadas usando os métodos e programas de computador, incluindo BLAST, disponível na rede mundial em ncbi.nlm.nih.gov/BLAST. Vide, por exemplo, Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10. Outro algoritmo de alinhamento é FASTA, disponível em Genetics Computing Group (GCG) package, de Madison, Wisconsin, USA, uma subsidiária integral de Oxford Molecular Group, Inc. Outras técnicas para alinhamento são descritas em Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., uma divisão de Harcourt Brace & Co., San Diego, California, USA. De particular interesse são programas de alinhamento que permitem lacunas na sequência. O Smith-Waterman é um

tipo de algoritmo que permite lacunas em alinhamentos de sequência. Vide *Meth. Mol. Biol.* 70: 173-187 (1997). Ainda, o programa GAP que usa o método Needleman and Wunsch alignment pode ser utilizado para alinhar sequências. Vide *J. Mol. Biol.* 48: 443-453 (1970).

Um ácido nucléico é "hibridizável" a outro ácido nucléico, tal como um cDNA, DNA ou RNA genômico, quando uma forma de filamento único de ácido nucléico pode enrijecer ao outro ácido nucléico sob condições apropriadas de temperatura e força iônica de solução. Condições de bibridização e lavagem são bem conhecidas e exemplificadas em Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularmente capítulo 11 e Tabela 11.1 na mesma; e Sambrook, J. and Russell, W., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001). As condições de temperatura e força iônica determinam a "estringência" da hibridização. Condições de estringência podem ser filtradas para fragmentos moderadamente similares, tais como sequências homólogas de organismos distancialmente relacionados, a fragmentos altamente similares, tais como genes que duplicam enzimas funcionais de organismo intimamente relacionados. Condições de hibridização e lavagens pós-hibridização são úteis para obter as condições de estringência determinadas desejadas da hibridização. Um conjunto de lavagens ilustrativas pós-hibridização é uma série de lavagens começando com SSX 6 x (onde SSC é NaCl a 0,15 M e tampão citrato a 15 mM), SDS a 0,5% à temperatura ambiente por 15 minutos, então repetida com SSC 2x, SDS a 0,5% a 45°C por 30 minutos, e então reptida duas vezes com SSC 0,2 X, SDS a 0,5% a 50°C por 30 minutos. Outras condições estringentes são obtidas através de uso de altas temperaturas, em que as lavagens são idênticas àquelas acima, exceto pela temperatura das duas lavagens finais de 30 minutos em SSC 2 X, SDS a 0,5%, que é aumentada para 60°C. Outro conjunto de condições altamente estringentes utiliza duas lavagens finais em SSC 0,1 X, SDS a 0,1% a 65°C. Um outro exemplo de condições estringentes de hibridização é hibridização

a 50°C ou mais e SSC 0,1 X (cloreto de sódio a 15 mM/citrato de sódio a 1,5 mM). Um outro exemplo de condições estringentes de hibridização é incubação de um dia para o outro a 42°C em uma solução: formamida a 50%, SSC 5 X (NaCl a 150 mM, citrato de trissódio a 15 mM), fosfato de sódio a 50 mM (pH 7,6), solução de Denhardt 5 X, sulfato de dextrano a 10% e 20 µg/ml de DNA de esperma de salmão "sheared", desnaturado, seguida por lavagem dos filtros em SSC 0,1 X a cerca de 65°C. Condições estringentes de hibridização e condições de lavagem pós-hibridização são condições de hibridização e condições de lavagem pós-hibridização que são pelo menos tão estringentes quanto as condições representativas acima.

Hibridização requer que os dois ácidos nucléicos contenham sequências complementares, embora, dependendo da estringência da hibridização, são possíveis não pareamentos entre bases. A estringência apropriada para hibridizar ácidos nucléicos depende do comprimento dos ácidos nucléicos e do grau de complementação, variáveis bem conhecidas na técnica. Quanto maior o grau de similaridade ou homologia entre duas sequências de nucleotídeos, tanto maior o valor da temperatura de fusão (Tm) para híbridos de ácidos nucléicos que contêm aquelas sequências. A estabilidade relativa (corespondente à Tm mais alta) de hibridizações de ácido nucléicos diminui na seguinte ordem: RNA:RNA, DNA:RNA, DNA:DNA. Para híbridos de mais do que 100 nucleotídeos de comprimento, equações para calcular a Tm têm sido derivadas (vide Sambrook et al., supra, 9,50-9,51). Para hibridizações com ácidos nucléicos mais curtos, isto é, oligonucleotídeos, a posição de não pareamentos se torna mais importante e o comprimento do oligonucleotídeo determina sua especificidade (Vide Sambrook et al., supra, 11,7-11,8). Tipicamente, o comprimento para um ácido nucléico hibridizável é pelo menos cerca de 10 nucleotídeos. Comprimentos ilustrativos mínimos para um ácido nucléico hibridizável são: pelo menos cerca de 15 nucleotídeos, pelo menos cerca de 20 nucleotídeos e pelo menos cerca de 30 nucleotídeos. Além disso, o versado na técnica irá reconhecer que a temperatura e a concentração de sal da solução de lavagem podem ser ajustadas conforme necessário, de

acordo com fatores tal como o comprimento da sonda.

Antes que a presente invenção seja adicionalmente descrita, deve ser entendido que esta invenção não está limitada a modalidades particulares descritas, como tal pode, certamente, variar. Também deve ser entendido que a terminologia usada aqui tem o propósito de apenas descrevar modalidades particulares e não se destina a ser limitativa, visto que o escopo da presente invenção será limitado apenas através des reivindicações anexas.

Onde uma faixa de valores é proporcionada, deve ser entendido que cada valor interveniente, até a décima unidade do limite inferior, a menos que o contexto indique claramente de outra forma, entre o limite superior e o inferior daquela faixa e qualquer outro valor estabelecido ou interveniente, está abrangido pela invenção. Os limites superiores e inferiores destas faixas menores podem, de maneira independente, estar incluídos nas faixas menores, e também estão abrangidos pela invenção, submetidos a qualquer limite especificamente excluído na faixa estabelecida. Onde a faixa estabelecida inclui um ou ambos os limites, as faixas que excluem um ou ambos os limites incluídos também estão incluídas na invenção.

A menos que de outra forma definidos, todos os termos técnicos e científicos têm o mesmo significado conforme comumente entendido por alguém versado na técnica à qual esta invenção pertence. Embora quaisquer métodos e materiais similares ou equivalentes àqueles descritos aqui também possam ser utilizados na prática ou teste da presente invenção, os métodos e materias preferidos são agora descritos. Todas as publicações ora mencionadas são aqui incorporadas por referência para divulgar e descrever os métodos e/ou materias, em conjunto com os quais as publicações são citadas.

Deve ser notado que, como usado aqui e nas reivindicações anexas, as formas singulares "um, uma", "e" e "o, a" incluem referências plurais, a menos que o contexto indique claramento o contrário. Desse modo, por exemplo, referência a "uma célula hospedeira geneticamente

modificada" inclui uma pluralidade de tais células hospedeiras e referência a "o composto isoprenóide" inclui referência a um ou mais compostos isoprenóides e equivalentes dos mesmos conhecidos na técnica, e assim por diante. Também deve ser adicionalmente notado que as reivindicações podem ser formuladas para excluir qualquer elemento opcional. Como tal, esta declaração destina-se a servir como base antecedente para uso de tal terminologia exclusiva como "unicamente", "apenas" e similares em conjunto com a recitação de elementos de reivindicação, ou uso de uma limitação "negativa".

As publicações discutidas aqui são proporcionadas unicamente por sua descrição anterior à data de depóstio do presente pedido de patente. Nada aqui deve ser interpretado como uma admissão de que a presente invenção não tem o direito de anteceder tal publicação por força de invenção prévia. Adicionalmente, as datas de publicação proporcionadas podem ser diferentes das reais datas de publicação, o que pode precisar ser independentemente confirmado.

Descrição Detalhada

A presente invenção proporciona células hospedeiras geneticamente modificadas e o uso das mesmas para produzir isoprenóide e compostos precursores de isoprenóide.

Um método de fazer um isoprenóide ou um precursor de isoprenóide é cultivar uma célula hospedeira, onde a célula hospedeira é capaz de fazer o isoprenóide ou precursor de isoprenóide. Porque o caminho biossintético para fazer um isoprenóide ou um precursor de isoprenóide envolve múltiplas enzimas, o fluxo através de caminho pode não ser ótimo ou devidamente equilibrada. Um método para corrigir tal desequilíbrio é modular os níveis de atividade das enzimas do caminho em relação umas com as outras. Como descrito em maiores detalhes a seguir, aumentar o nível de mevalonato quinase, em relação ao nível de uma ou mais outras enzimas no caminho do mevalonato, proporciona um aumento em níveis de produção de compostos isoprenóides ou compostos precursores de isoprenóide.

Compostos isoprenóides são sintetizados a partir um precursor universal com cinco carbonos, pirofosfato de isopentenila (IPP). Existem dois caminhos principais para converter um substrato em IPP: 1) o "caminho do mevalonato", o qual converte acetil-CoA em IPP e o "caminho do 1-desóxi-D-xilulose 5-difosfato" (também referido como o "caminho DXP"), o qual converte D-gliceraldeído-3-fosfato e piruvato em IPP e DMAPP.

Enzimas do caminho do mevalonato estão representadas na figura 1. O caminho do mevalonato compreende as seguintes reações enzimáticas: (a) condensação de duas moléculas de acetil-CoA em acetoacetil-CoA; (b) condensação de acetoacetil-CoA com acetil-CoA para formar HMG-CoA; (c) conversão de HMG-CoA em mevalonato; (d) fosforilação de mevalonato em mevalonato 5-fosfato; (e) conversão de mevalonato 5-fosfato em mevalonato 5-pirofosfato e (f) conversão de mevalonato 5-pirofosfato em pirofosfato de isopentenila. Enzimas que realizam estas reações incluem acetoacetil-CoA tiolase, hidroximetilglutaril-CoA sintase (HMGS), hidroximetilglutaril-CoA redutase (HMGR), mevalonato quinase (MK), fosfomevalonato quinase (PMK) e mevalonato pirofosfato descarboxilase (MPD).

No caminho DXP, piruvato de D-gliceraldeído-3-fosfato são convertidos através de uma série de reações em IPP e DMAPP. O caminho envolve a ação das seguintes enzimas: 1-desóxi-D-xilulose-5-fosfato sintase (Dxs), 1-desóxi-D-xilulose-5-fosfato reductoisomerase (IspC), 4-difosfocitidil-2-C-metil-D-eritritol sintase (IspD), 4-difosfocitidil-2-C-metil-D-eritritol quinase (IspE), 2C-metil-D-eritritol 2,4-ciclodifosfato sintase (IspF), 1-hidróxi-2-metil-2-(E)-butenil 4-difosfato sintase (IspG) e difosfato de isopentenila isomerase (IspH).

Células eucarióticas exceto células de planta usam o caminho do mevalonato exclusivamente para converter acetil-CoA em IPP, o qual é subsequentemente isomerizado em DMAPP. Plantas usam ambos os caminhos do mevalonato e DXP para síntese de isoprenóide. Procariotos, com algumas exceções, usam o caminho DXP para produzir IPP e DMAPP separadamente través de um ponto de ramificação.

O IPP produzido através de caminho do mevalonato pode ser isomerizado para produzir DMAPP. O IPP e/ou o DMAPP pode ser atuado através de preniltransferase para produzir polipirofosfato de prenila. Por exemplo, conforme mostrado na figura 2, IPP ou DMAPP pode ser modificado através de prenila transferase para gerar o poliprenil difosfatos geranil difosfato (GPP), difosfato de farnesila (FPP) e geranil geranil difosfato (GGPP). GPP e FPP são adicionalmente modificados através de terpeno sintases para gerar monoterpenos e sesquiterpenos, respectivamente; e GPP é adicionalmente modificado através de terpeno sistases para gerar diterpenos e carotenóides. IPP e DMAPP são gerados através de um ou dois caminhos: o caminho do mevalonato (MEV) e o caminho do 1-desóxi-D-xilulose-5-fosfato (DXP).

Células Hospedeiras Geneticamente Modificadas

A presente invenção proporciona células hospedeiras geneticamente modificadas e métodos de uso das mesmas para produzir compostos isoprenóides. Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão é modificada geneticamente de tal modo ela que produz mevalonato quinase em um nível que é maior do que o nível de pelo menos uma dentre acetoacetil-CoA tiolase, hidroximetilglutaril-CoA sintase (HMGS), hidroximetilglutaril-CoA redutase (HMGR), fosfomevalonato quinase (PMK), e mevalonato pirofosfato descarboxilase (MPD). Um método em questão, abaixo descrito em maiores detalhes, geralmente envolve cultura *in vitro* de uma célula hospedeira geneticamente modificada em questão em um meio adequado, de tal modo que a célula hospedeira geneticamente modificada converte um substrato em IPP e produz um composto isoprenóide.

Em algumas modalidades, o nível de MK produzido em uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos

cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 3 vezes, pelo menos cerca de 4 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 7 vezes ou pelo menos cerca de 10 vezes, ou mais, mais alto do que o nível de uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK, e MPD.

Por exemplo, em algumas modalidades, o nível de MK produzido em uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 3 vezes, pelo menos cerca de 4 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 7 vezes ou pelo menos cerca de 10 vezes, ou mais, mais alto do que o nível de HMGS na célula.

Em outras modalidades, o nível de MK produzido em uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 3 vezes, pelo menos cerca de 4 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 7 vezes ou pelo menos cerca de 10 vezes, ou mais, mais alto do que o nível de HMGR na célula.

Em outras modalidades, o nível de MK produzido em uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo

menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 3 vezes, pelo menos cerca de 4 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 7 vezes ou pelo menos cerca de 10 vezes, ou mais, mais alto do que o nível de ambos HGMS e HMGR na célula.

Em outras modalidades, o nível de MK produzido através de uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 3 vezes, pelo menos cerca de 4 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 25 vezes, pelo menos cerca de 50 vezes ou pelo menos cerca de 100 vezes, ou mais, mais alto do que o nível de MK produzido através de uma célula hospedeira geneticamente modificada com pMBI ou com pMBIS. pMBI e pMBIS são descritos na patente US Nº 7,192,751.

Desse modo, por exemplo, em algumas modalidades, onde uma célula hospedeira geneticamente modificada em questão é uma célula procariótica, o nível de MK produzido através de uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 3 vezes, pelo menos cerca de 4 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos

cerca de 25 vezes, pelo menos cerca de 50 vezes ou pelo menos cerca de 100 vezes, ou mais, mais alto do que o nível de MK produzido através de mesma célula geneticamente modificada com pMBI ou com pMBIS.

Em algumas modalidades, a razão molar de polipeptídeo MK em uma célula hospedeira geneticamente modificada em questão em relação a uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e polipeptídeo MPD é a partir de cerca de 1,25 : 1 a cerca de 100 : 1 ou maior do que 100 : 1. Por exemplo, em algumas modalidades, a razão molar de MK em relação a uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD é a partir de cerca de 1,25 : 1 a cerca de 1,5 : 1, a partir de cerca de 1,5 : 1 a cerca de 2 : 1, a partir de cerca de 2 : 1 a cerca de 2,5 : 1, a partir de cerca de 2,5 : 1 a cerca de 3 : 1, a partir de cerca de 3 : 1 a cerca de 5 : 1, a partir de cerca de 5 : 1 a cerca de 10 : 1, a partir de cerca de 10 : 1 a cerca de 25 : 1, a partir de cerca de 25 : 1 a cerca de 50 : 1 ou a partir de cerca de 50 : 1 a cerca de 100 : 1.

O nível de MK produzido em uma célula hospedeira geneticamente modificada em questão pode ser controlado de várias maneiras. Em algumas modalidades, o número de cópia de regiões de codificação que compreendem sequências de nucleotídeo que codificam MK é mais alto do que o número de cópia de regiões de codificação que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD. Em outras modalidades, uma sequência de nucleotídeo que codifica MK está sob controle transcripcional de (por exemplo, é operacionalmente ligada a) um promotor mais forte do que o promotor ao qual uma ou mais sequências de nucleotídeo, que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD, está operacionalmete ligada. Em outras modalidades, o nível de MK é aumentado em relação ao nível de uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD através de aumento tanto do número de cópia de regiões de codificação que codificam MK em relação ao número de cópia de regiões de codificação que compreendem sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD, quanto através de

aumento da força do promotor de um promotor ao qual uma sequência que codifica MK está operacionalmente ligada, em relação à força do promotor de um promotor ao qual uma sequência de nucleotídeo que codifica uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD está operacionalmente ligada.

Em outras modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que compreende uma sequência de nucleotídeo que codifica MK, onde a sequência de nucleotídeo que codifica MK está operacionalmente ligada a um promotor mais forte do que o promotor ao qual um óperon, que compreende sequências de nucleotídeo que codificam acetoacetil-CoA tiolase, HMGS e HMGR, está operacionalemente ligado. Em outras modalidades, uma célula hospedeira geneticamente modificada em questão compreende um primeiro ácido nucléico que compreende uma sequência de nucleotídeo que codifica MK, onde o primeiro ácido nucléico é um vetor de expressão com alto número de cópia; e um segundo ácido nucléico que compreende sequências de nucleotídeo que codificam acetoacetil-CoA tiolase, HMGS e HMGR, onde o segundo ácido nucléico é um vetor de expressão com baixo número de cópia.

Controle dos níveis de MK através de número de cópia de regiões de codificação que codificam MK

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma pluralidade de regiões de codificação que compreendem uma sequência de nucleotídeo que codifica MK, um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato exceto mevalonato quinase; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase. A pluralidade de (por exemplo, duas ou mais) regiões de codificação que codificam MK pode estar no mesmo ácido nucléico ou em ácidos nucléicos fisicamente separados. Por exemplo, as duas ou mais regiões de codificação que codificam MK estão, em algumas modalidades, todas

presentes no DNA genômico da célula hospedeira. Como um outro exemplo, em algumas modalidades, uma primeira região de codificação que codifica MK está presente no DNA genômico da célula hospedeira; e uma segunda região de codificação que codifica MK está presente no vetor recombinate extracromossomal. Ainda como um outro exemplo, em algumas modalidades, as duas ou mais regiões de codificação que codificam MK estão incluídas em um ou mais vetor (es) recombinate (s) extracromossomal (ais).

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma única cópia de uma região de codificação que compreende sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD; duas ou mais regiões de codificação que codificam MK, cada uma das quais compreende sequências de nucleotídeo que codificam MK; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase. Por exemplo, em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma única cópia de uma região de codificação de ácido nucléico que compreende sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD, onde a região de codificação está integrada ao genoma da célula hospedeira; duas ou mais regiões de codificação que codificam MK, cada uma das quais compreende sequências de nucleotídeo que codificam MK, onde as duas ou mais regiões de codificação que codificam MK são extracromossomais, tais como vetores de expressão extracromossomais que não estão integrados ao genoma hospedeiro; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Em outras modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que compreende duas ou mais regiões de codificação que codificam MK, uma região de codificação de ácido nucléico que compreende sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR,

PMK e MPD; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase, onde a razão do número de cópia das regiões de codificação de MK para o número de cópia da região de codificação de ácido nucléico que compreende sequências de nucleotídeo que codificam uma ou dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD é a partir de cerca de 1,5 : 1 a cerca de 100 : 1, por exemplo, a razão do número de cópia das regiões de codificação que codificam MK para o número de cópia das regiões de codificação de ácido nucléico que compreendem sequências de nucleotídeo que codificam uma ou dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD é a partir de cerca de 1,5 : 1 a cerca de 2 : 1, a partir de cerca de 2 : 1 a cerca de 2,5 : 1, a partir de cerca de 2,5 : 1 a cerca de 3 : 1, a partir de cerca de 3 : 1 a cerca de 5 : 1, a partir de cerca de 5 : 1 a cerca de 10 : 1, a partir de cerca de 10 : 1 a cerca de 25 : 1, a partir de cerca de 25 : 1 a cerca de 50 : 1 ou a partir de cerca de 50 : 1 a cerca de 100 : 1.

Por exemplo, em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende pelo menos duas regiões de codificação que codificam MK e uma única cópia de uma região de codificação que codifica uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD. Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende duas regiões de codificação que codificam MK e uma única cópia de uma região de codificação que codifica acetoacetil-CoA tiolase, HMGS, HMGR.

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma pluralidade de regiões de codificação que codificam MK, onde a uma região de codificação que codifica MK está presente em um vetor de expressão com alto número de cópia, tal como um plasmídeo com alto número de cópia; uma região de codificação de ácido nucléico que compreende sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD, onde região de codificação que codifica uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD está presente em um

vetor de expressão com baixo número de cópia, tal como um plasmídeo com baixo número de cópia; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Em algumas modalidades, o número de cópia de uma região de codificação que codifica MK em uma célula hospedeira geneticamente modificada em questão é mais alto do que o número de cópia de regiões de codificação que codificam MK em uma célula hospedeira de controle geneticamente modificada com pMBI. Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma pluralidade de regiões de codificação que compreende uma sequência de nucleotídeo que codifica MK, onde as regiões de codificação que codificam MK estão presentes em um ácido nucléico que compreende uma origem p15A de replicação (para uma sequência de nucleotídeo p15A ori, vide, por exemplo, nucleotídeos 4525-5321 de pAM39 (SEQ ID Nº: 3); e Selzer et al. (1983) Cell 32:119-129).

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma pluralidade de regiões de codificação que codificam MK, onde as regiões de codificação que codificam MK estão presentes em um plasmídeo com alto número de cópia; uma região de codificação de ácido nucléico que compreende sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD, onde região de codificação que compreende sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD estão presente em um plasmídeo com médio número de cópia; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma pluralidade de regiões de codificação que codificam MK, onde as regiões de codificação que codificam MK estão presentes em um plasmídeo com médio número de cópia; uma região de codificação de ácido nucléico que compreende

sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD, onde região de codificação que compreende sequências de nucleotídeo que codificam uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD está presente em uma plasmídeo com baixo número de cópia; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende um primeiro ácido nucléico que compreende uma região de codificação que codifica MK, onde o primeiro ácido nucléico é um vetor de expressão com alto número de cópia; um segundo ácido nucléico que compreende sequências de nucleotídeo que codificam acetoacetil-CoA tiolase, HMGS e HMGR, onde o segundo ácido nucléico é um vetor de expressão com baixo número de cópia e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Plasmídeos com baixo número de cópia geralmente proporcionam menos do que cerca de 20 cópias de plasmídeo por célula (por exemplo, a partir de cerca de 1 cópia de plasmídeo por célula a cerca de 5 cópias de plasmídeo por célula, a partir de cerca de 5 cópias de plasmídeo por célula a cerca de 10 cópias de plasmídeo por célula, a partir de cerca de 10 cópias de plasmídeo por célula a cerca de 15 cópias de plasmídeo por célula ou a partir de cerca de 15 cópias de plasmídeo por célula a cerca de 20 cópias de plasmídeo por célula); plasmídeos com médio número de cópia geralmente proporcionam a partir de cerca de 20 cópias de plasmídeo por célula a cerca de 50 cópias de plasmídeo por célula, ou a partir de cerca de 20 cópias de plasmídeo por célula a cerca de 80 cópias de plasmídeo por célula; e plasmídeos com alto número de cópia geralmente proporcionam a partir de cerca de 80 cópias de plasmídeo por célula a cerca de 200 cópias de plasmídeo por célula, ou mais do que 200 cópias de plasmídeo por célula.

Vetores de expressão com baixo número de cópia adequados

para células procarióticas (por exemplo, *Escherichia coli*) incluem, mas não estão limitados a, pACYC184, pBeloBac11, pBR332, pBAD33, pBBR1MCS e seus derivados, pSC101, SuperCos (cosmídeo) e pWE15 (cosmídeo). Vetores de expressão com médio número de cópia adequados para células procarióticas (por exemplo, *E. coli*) incluem, mas não estão limitados a, pTrc99A, pBAD24, e vetores de expressão que contêm uma origem ColE1 de replicação e seus derivados. Vetores de expressão com alto número de cópia adequados para células procarióticas (por exemplo, *E. coli*) incluem, mas não estão limitados a, vetores pUC, pBluescript, pGEM e pTZ.

Vetores de expressão com baixo número de cópia adequados (centroméricos) para células eucarióticas (por exemplo, células de levedura) incluem, mas não estão limitados a, pRS415 e pRS416 (Sikorski & Hieter (1989) Genetics 122:19-27). Vetores de expressão de 2 mícrons com alto número de cópia adequados para células eucarióticas (por exemplo, células de levedura) incluem, mas não estão limitados a, pRS425 e pRS426 (Christainson et al. (1992) Gene 110:119-122). Vetores de expressão de 2 mícrons alternativos incluem variantes não selecionáveis do vetor de 2 mícrons (Bruschi & Ludwig (1988) Curr. Genet. 15:83-90) ou plasmídeos de 2 mícrons intactos que sustentam um cassete de expressão (como exemplificado no Pedido de Patente US N° 20050084972).

Controle dos níveis de MK através de força do promotor

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma pluralidade de regiões de codificação de ácido nucléico que compreendem sequências de nucleotídeo que codificam MK, onde as regiões de codificação que codificam MK são, cada uma, operacionalmente ligadas a (por exemplo, sob controle transcricional de) um primeiro elemento de controle transcricional (por exemplo, um primeiro promotor); um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto mevalonato quinase (por exemplo, que codifica uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD), onde a (s) sequência (s) de nucleotídeo que codifica (m) uma ou mais dentre

acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD estão operacionalmente ligados a (por exemplo, sob controle transcricional de) um segundo elemento de controle transcricional (por exemplo, um segundo promotor); e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase, onde o primeiro elemento de controle transcricional é mais forte do que o segundo elemento de controle transcricional.

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende uma pluralidade de regiões de codificação de ácido nucléico que codifica sequências de ácido nucléico que codificam MK, onde as regiões de codificação que codificam MK são, cada uma, operacionalmente ligadas a (por exemplo, sob controle transcricional de) um primeiro promotor, um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas da caminho do mevalonato, exceto mevalonato quinase (por exemplo, que codifica uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD), onde a (s) sequência (s) de nucleotídeo que codifica (m) uma ou mais dentre acetoacetil-CoA tiolase, HMGS, HMGR, PMK e MPD estão operacionalmente ligados a (por exemplo, sob controle transcricional de) um segundo promotor, onde o primeiro promotor é mais forte do que o segundo promotor; e

um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Em algumas modalidades, uma região de codificação está contida em um óperon. Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que compreende um primeiro óperon que compreende sequências de nucleotídeo que codificam MK, PMK e MPD, onde o primeiro óperon está operacionalmente ligado a um primeiro promotor, um ácido nucléico que compreende um segundo óperon que compreende sequências de nucleotídeo que codificam acetoacetil-CoA tiolase, HMGS e HMGR, onde o segundo óperon está operacionalmente ligado a um segundo promotor, onde

o primeiro promotor é mais forte do que o segundo promotor; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Elementos regulatórios incluem, por exemplo, promotores e operadores. Um promotor é uma sequência de nucleotídeos que inicia e controla a transcrição de uma sequência de ácido nucléico através de uma enzima RNA polimerase. Um operador é uma sequência de nucleotídeos adjacente ao promotor, que funciona para controlar a transcrição da sequência de ácido nucléico desejada. O operador contém um domínio ligado a proteína, onde uma proteína repressora específica pode se ligar. Na ausência de uma proteína repressora adequada, a transcrição inicia através de promotor. Na presença de uma proteína repressora adequada, a proteína repressora se liga ao operador e, desse modo, inibe a transcrição a partir do promotor.

Em algumas modalidades da presente invenção, promotores usados em vetores de expressão são induzíveis. Em outras modalidades, os promotores usados em vetores de expressão são constitutivos. Em algumas modalidades, uma ou mais sequências de ácido nucléico estão operacionalmente ligadas a um promotor induzível, e uma ou mais sequências de ácido nucléico estão operacionalmente ligadas a um promotor constitutivo.

Exemplos não-limitativos de promotores adequados para uso em células hospedeiras procarióticas incluem um promotor da RNA polimerase do bacteriófago T7; um promotor de trp; um promotor do óperon lac; um promotor híbrido, por exemplo, um promotor híbrido de lac/tac, um promotor híbrido de tac/trc, um promotor de trp/lac, um promotor de T7/lac; um promotor de trc, um promotor de tac e similares; um promotor araBAD; promotores regulados *in vivo*, tal como um promotor ssaG ou um promotor relacionado (vide, por exemplo, a publicação da Patente US N° 20040131637), um promotor pagC (Pulkkinen e Miller, J. Bacteriol. (1991) 173(1):86-93; Alpuche-Aranda et al. (1992) Proc. Natl. Acad. Sci. U S A. 89 (21) : 10079-83), um promotor *nirB* (Harborne et al. (1992) *Mol. Micro.* 6 :

2805 - 2813), e similares (vide, por exemplo, Dunstan et al. (1999) *Infect. Immun.* 67 : 5133 - 5141; McKelvie et al. (2004) *Vaccine* 22 : 3243 - 3255; e Chatfield et al. (1992) *Biotechnol.* 10 : 888 - 892); um promotor sigma70, por examplo, um promotor sigma70 de consenso (vide, por exemplo, números de acesso no GenBank AX798980, Ax798961 e AX798183); um promotor de fase estacionária, por exemplo, um promotor *dps*, um promotor *spv* e similares; um promotor derivado da ilha de patogenicidade SPI-2 (vide, por exemplo, WO 96/17951); um promotor actA (vide, por exemplo, Shetron-Rama et al. (2002) *Infect. Immun.* 70 : 1087 - 1096); um promotor rpsM (vide, por exemplo, Valdivia e Falkow (1996) *Mol. Microbiol.* 22 : 367 378); um promotor tet (vide, por exemplo, Hillen et al. (1989) In Saenger W. and Heinemann U. (eds) *Topics in Molecular and Structural Biology, Protein–Nucleic Acid Interaction*. Macmillan, London, UK, Vol. 10, pp. 143 – 162); um promotor SP6 (vide, por exemplo, Melton et al. (1984) *Nucl. Acids Res.* 12 : 7035 - 7056); e similares.

Em outra modalidade da presente invenção, a atividade total de uma mevalonato quinase heteróloga em relação a outras enzimas do caminho do mevalonato em um micro-organismo hospedeiro é aumentada através de expressão da enzima a partir de um promotor forte. Em algumas modalidades, a sequência de nucleotídeo que codifica mevalonato quinase está operacionalmente ligada a um promotor que é mais forte do que o promotor p_{LAC} ((SEQ ID Nº: 21), por exemplo, a sequência de nucleotídeo que codifica MK está operacionalmente ligada a um promotor que é pelo menos cerca de 10%, pelo menos cerca de 25%, pelo menos cerca de 50%, pelo menos cerca de duas vezes, pelo menos cerca de 5 vezes, ou pelo menos cerca de 10 vezes, ou mais, mais forte do que um promotor p_{LAC} que tem uma sequência de nucleotídeo conforme mostrada em SEQ ID Nº: 21.

Promotores fortes adequados para uso em células procarióticas (por exemplo, *Escherichia coli*) incluem, mas não estão limitados a, um promotor *lac*UV5 (vide, por exemplo, SEQ ID Nº: 22), Trc, Tac, T5, T7 e P_{Lambda}. Em outra modalidade da presente invenção, a atividade total da mevalonato quinase em um organismo hospedeiro é aumentada através de

expressão da enzima a partir de um promotor forte em um palsmídeo com alto número de cópia. Exemplos adequados para uso em células procarióticas (por exemplo, *Escherichia coli*) incluem, mas estão limitados ao uso de promotores Trc, Tac, T5, T7 e P_{Lambda} com vetores pBAD24, pBAD18, pGEM, pBluescript, pUC e pTZ.

Exemplos não-limitativos de promotores adequados para uso em células hospedeiras eucarióticas incluem, mas estão limitados a, um promotor precoce imediato de CMV, um promotor de timidina quinase HSV, um promotor precoce ou tardio de SV40, LTRs de retrovírus e um promotor da metalotioneína I de camundongo.

Exemplos não-limitativos de promotores constitutivos adequados para uso em células procarióticas incluem um promotor sigma70 (por exemplo, um promotor sigma70 de consenso). Exemplos não-limitativos de promotores induzíveis adequados para uso em células hospedeiras bacterianas incluem o pL do bacteriófago λ; Plac; Ptrp; Ptac (promotor híbrido Ptrp-lac); um promotor induzível por isopropil-beta-D44 tiogalactopiranosídeo (IPTG), por exemplo, um promotor *lacZ*; um promotor induzível por tetraciclina; um promotor induzível por arabinose, por exemplo, PBAD (vide, por exemplo, Guzman et al. (1995) *J. Bacteriol.* 177 : 4121 - 4130); um promotor induzível por xilose, por exemplo, Pxyl (vide, por exemplo, Kim et al. (1996) *Gene* 181 : 71 - 76); um promotor GAL1; um promotor de triptofano; um promotor lac; um promotor induzível por álcool, por exemplo, um promotor induzível por metanol, um promotor induzível por etanol; um promotor induzível por rafinose; um promotor induzível por calor, por exemplo, promotor PL lambda induzível por calor; um promotor controlado por um repressor sensível ao calor (por exemplo, vetores de expressão à base de lamba reprimidos por CI857; vide, por exemplo, Hoffmann et al. (1999) *FEMS Microbiol Lett.* 177 (2) : 327 - 34); e similares.

Exemplos não-limitativos de promotores constitutivos adequados para uso em células hospedeiras de levedura incluem um promotor ADH1, um promotor ADH2, um promotor PGK ou um promotor LEU2. exemplos não-limitativos de promotores induzíveis adequados para uso em células

hospedeiras de levedura incluem, mas não estão limitados a, um promotor divergente induzível por galactose, tal como um promotor GAL 1 ou um promotor GAL 10 (West at al. (1984) Mol. Cell. Biol. 4(11):2467-2478) ou um promotor CUP1. Onde desejado, o promotor é mais forte do que o promotor Lac nativo de *E. coli*.

Exemplos não-limitativos de operadores para uso em células hospedeiras bacterianas incluem um operador promotor de lactose (proteína represora de LacI muda a conformação quando colocada em contato com lactose, desse modo impedindo a proteína repressora de LacI de se ligar ao operador), um operador promotor de triptofano (quando complexado com triptofano, a proteína repressora de TrpR tem uma conformação que se liga ao operador; na ausência de triptofano, a proteína repressora de TrpR tem uma conformação que não se liga ao operador, e um operador promotor de tac (vide, por exemplo, deBoer et al. (1983) Proc. Natl. Acad. Sci. U.S.A. 80:21-25.).

Enzimas do caminho do mevalonato

O caminho do mevalonato compreende: (a) condensação de duas moléculas de acetil-CoA em acetoacetil-CoA; (b) condensação de acetoacetil-CoA com acetil-CoA para formar HMG-CoA; (c) conversão de HMG-CoA em mevalonato; (d) fosforilação de mevalonato em mevalonato 5-fosfato; (e) conversão de mevalonato 5-fosfato em mevalonato 5-pirofosfato; e (f) conversão de mevalonato 5-pirofosfato em pirofosfato de isopentenila. As enzimas do caminho do mevalonato necessárias para produção de IPP variam dependendo das condições de cultura.

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que compreende sequências de nucleotídeo que codificam uma mevalonato quinase, como descrito acima; em ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto mevalonato quinase, e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Sequências de nucleotídeo que codificam enzimas do caminho do mevalonato

Sequências de nucleotídeo que codificam produtos de gene do caminho do mevalonato (MEV) são conhecidas na técnica e qualquer sequência de nucleotídeo que codifica produto de gene do caminho MEV pode ser usada para gerar uma célula hospedeira geneticamente modificada em questão. Por exemplo, sequências de nucleotídeo que codificam acetoacetil-CoA tiolase, HMGS, HMGR, MK, PMK, MPD e IDI são conhecidas na técnica. O seguinte são exemplos de sequências de nucleotídeo conhecidas que codificam produto de gene do caminho MEV, com números de acesso no GenBank e organismo seguinte a cada enzima do caminho MEV, em parêntesis: acetoacetil-CoA tiolase: (REGIÃO NC_000913: 2324131..2325315; *E. coli*), (D49362; *Paracoccus denitrificans*), e (L20428; *Saccharomyces cerevisiae*); HMGS: (NC_001145. complement 19061..20536; *Saccharomyces cerevisiae*), (X96617; *Saccharomyces cerevisiae*), (X83882; *Arabidopsis thaliana*), (AB037907; Kitasatospora griseola), e (BT007302; *Homo sapiens*); HMGR: (NM_206548; *Drosophila melanogaster*), (NM_204485; Gallus gallus), (AB015627; *Streptomyces sp.* KO-3988), (AF542543; *Nicotiana attenuata*), (AB037907; Kitasatospora griseola), (AX128213, que proporciona a sequência que codifica uma HMGR truncada; *Saccharomyces cerevisiae*), e (NC_001145: complement (115734..118898; *Saccharomyces cerevisiae*)); MK: (L77688; *Arabidopsis thaliana*), e (X55875; *Saccharomyces cerevisiae*); PMK: (AF429385; *Hevea brasiliensis*), (NM_006556; *Homo sapiens*), (NC_001145. complement 712315..713670; *Saccharomyces cerevisiae*); MPD: (X97557; *Saccharomyces cerevisiae*), (AF290095; *Enterococcus faecium*), e (U49260; *Homo sapiens*); e IDI: (NC_000913, 3031087..3031635; *E. coli*), e (AF082326; *Haematococcus pluvialis).*

Um exemplo não-limitativo de sequências de nucleotídeo que codificam acetoacetil-CoA tiolase, HMGS e HMGR está ilustrado nas figuras 13A – C (SEQ ID Nº: 1) da Patente US N° 7.183.089. Um exemplo não limitante de sequências de nucleotídeo codificando MK, PMK, MPD, e

isomerase difosfato de isopentenila (IDI) é apresentado nas figuras 16A-D da patente americana US 7.183.089. Um exemplo não-limitante de sequências de nucleotídio codificando HK, PMK, MPD e isomerase isopentenil difosfato (IDI) é apresentado nas figuras 16 A-D da patente americana US 7.183.089.

Em algumas modalidades, a região de codificação da HMGR está ilustrada em SEQ ID Nº: 13 da Patente US Nº 7.183.089 (vide também as figuras 20A-C da Patente US N° 7.183.089), a qual codifica uma forma truncada de HMGR ("tHMGR), que carece do domínio de transmembrana de HMGR do tipo selvagem. O domínio de transmembrana de HMGR contém as porções regulatórias da enzima e não tem atividade catalítica.

A sequência de codificação de qualquer enzima do caminho MEV conhecida pode ser alterada de várias formas conhecidas na técnica para gerar mudanças direcionadas na sequência de aminoácido da enzima codificada. O aminoácido de uma enzima variante do caminho MEV será substancialmente similar à sequência de aminoácido de qualquer enzima do caminho MEV conhecida, isto é, irá diferir apenas por pelo menos um aminoácido, e pode diferir por pelo menos dois, pelo menos 5, pelo menos 10 ou pelo menos 20 aminoácidos, porém tipicamente não mais do que cerca de 50 aminoácidos. As mudanças da sequência podem ser substituições, inserções ou anulações. Por exemplo, como descrito abaixo, a sequência de nucleotídeo pode ser alterada para o viés de códon de uma célula hospedeira particular. Em adição, uma ou mais diferenças da sequência de nucleotídeo pode ser introduzida, que resulta em mudanças conservativas de aminoácido na proteína codificada.

Enzimas do Caminho DXP

O caminho DXP compreende: 1-desóxi-D-xilulose-5-fosfato sintase (Dxs), 1-desóxi-D-xilulose-5-fosfato reductoisomerase (IspC), 4-difosfocitidil-2-C-metil-D-eritritol sintase (IspD), 4-difosfocitidil-2-C-metil-D-eritritol quinase (IspE), 2C-metil-D-eritritol 2,4-ciclodifosfato sintase (IspF) e 1-hidróxi-2-metil-2-(E)-butenil 4-difosfato sintase (IspG).

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que

compreende sequências de nucleotídeo que codificam uma mevalonato quinase, conforme descrito acima; um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto mevalonato quinase; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase, onde um caminho DXP endógeno na célula hospedeira é funcionalmente deficiente.

Preniltransferases

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que compreende sequências de nucleotídeo que codificam uma mevalonato quinase, conforme descrito acima; um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto mevalonato quinase; um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma preniltransferase; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

Preniltransferases constituem um grupo externo de enzimas que catalizam a condensação consecutiva de IPP, resultado na formação de difosfatos de prenila de vários comprimentos de cadeia. Preniltransferases adequadas incluem enzimas que catalizam a condensação de IPP com substratos iniciador alílicos para formar compostos isoprenóides com a partir de cerca de duas unidades de isopreno a cerca de 6000 unidades de isopreno ou mais, por exemplo, duas unidades de isopreno (pirofosfato de geranila sintase), 3 unidades de isopreno (pirofosfato de farnesila sintase), 4 unidades de isopreno (pirofosfato de geranilgeranila sintase), 5 unidades de isopreno, 6 unidades de isopreno (pirofosfato de hexadecila sintase), 7 unidades de isopreno, 8 unidades de isopreno (fitoeno sintase, pirofosfato de octaprenila de sintase), 9 unidades de isopreno (pirofosfato de nonaprenila sintase, 10 unidades de isopreno (pirofosfato de decaprenila sintase), a partir de cerca de 10 unidades de isopreno a cerca de 15 unidades de isopreno, a partir de cerca de 15 unidades de isopreno a cerca de 20 unidades de iso-

preno, a partir de cerca de 20 unidades de isopreno a cerca de 25 unidades de isopreno, a partir de cerca de 25 unidades de isopreno a cerca de 30 unidades de isopreno, a partir de cerca de 30 unidades de isopreno a cerca de 40 unidades de isopreno, a partir de cerca de 40 unidades de isopreno a cerca de 50 unidades de isopreno, a partir de cerca de 50 unidades de isopreno a cerca de 100 unidades de isopreno, a partir de cerca de 100 unidades de isopreno a cerca de 250 unidades de isopreno, a partir de cerca de 250 unidades de isopreno a cerca de 500 unidades de isopreno, a partir de cerca de 500 unidades de isopreno a cerca de 1000 unidades de isopreno, a partir de cerca de 1000 unidades de isopreno a cerca de 2000 unidades de isopreno, a partir de cerca de 2000 unidades de isopreno a cerca de 3000 unidades de isopreno, a partir de cerca de 3000 unidades de isopreno a cerca de 4000 unidades de isopreno, a partir de cerca de 4000 unidades de isopreno a cerca de 5000 unidades de isopreno ou a partir de cerca de 5000 unidades de isopreno a cerca de 6000 unidades de isopreno ou mais.

Preniltransferases adequadas incluem, mas não estão limitadas a, uma difosfato *E*-isoprenil sintase, incluindo, mas não limitada a, difosfato de geranila (GPP) sintase, difosfato de farnesila (FPP) sintase, difosfato de geranilgeranila (GGPP) sintase, difosfato de hexaprenila (HexPP) sintase, difosfato de heptaprenila (HepPP) sintase, octaprenil (OPP) difosfato sintase, difosfato de solanesila (SPP) sintase, difosfato de decaprenila (DPP) sintase, "chicle" sintase e guta-percha sintase; e uma *Z*-isoprenil difosfato sintase, incluindo, mas não limitada a, difosfato de nonaprenila (NPP) sintase, difosfato de undecaprenila (UPP) sintase, difosfato de deidrodolichila sintase, difosfato de eicosaprenila sintase, borracha natural sintase e outras *Z*-isoprenil difosfato sintases.

As sequências de nucleotídeo de numerosas prenil transferases de uma variedade de espécies são conhecidas e podem ser usadas ou modificadas para uso na geração de uma célula hospedeira geneticamente modificada em questão. Sequências de nucleotídeo que codificam prenil transferases são conhecidas na técnica. Vide, por exemplo, pirofosfato de farnesila sintetase mRNA humana (N° de acesso no GenBank J05262;

Homo sapiens); difosfato de farnesila sintetase (FPP) gene (N° de acesso no GenBank J05091; *Saccharomyces cerevisiae*); difosfato de isopentenila:dimetilalil difosfato isomerase gene (J05090; *Saccharomyces cerevisiae*); Wang e Ohnuma (2000) *Biochim. Biophys.* Acta 1529:33-48; Patente U.S. Patent N° 6.645.747; *Arabidopsis thaliana* pirofosfato de farnesila sintetase 2 (FPS2)/FPP sintetase 2/difosfato de farnesila sintase 2 (At4g17190) mRNA (N° de acesso no GenBank NM_202836); Ginkgo biloba difosfato de geranilgeranila sintase (ggpps) mRNA (N° de acesso no GenBank AY371321); *Arabidopsis thaliana* pirofosfato de geranilgeranila sintase (GGPS1)/GGPP sintetase/farnesiltranstransferase (At4g36810) mRNA (N° de acesso no GenBank NM_119845); Synechococcus elongatus gene para farnesila, geranilgeranila, geranilfarnesila, hexaprenila, difosfato de heptaprenila sintase (SelF-HepPS) (N° de acesso no GenBank AB016095); etc.

Terpeno Sintases

Um ácido nucléico que compreende uma sequência de nucleotídeo que codifica qualquer terpeno sintase conhecida pode ser usado. Terpeno sintases adequadas incluem, mas não estão limitadas a, amorfa-4,11-dieno sintase (ADS), beta-cariofileno sintase, germacreno A sintase, 8-epicedrol sintase, valenceno sintase, (+)-delta-cadineno sintase, germacreno C sintase, (E)-beta- farneseno sintase, Casbeno sintase, vetispiradieno sintase, 5-epi-aristoloqueno sintase, Aristolqueno sintase, beta-cariofileno, alfa-humuleno, (E,E)-alfa-farneseno sintase, (-)-beta-pineno sintase, Gama-terpineno sintase, limoneno ciclase, Linalool sintase, 1,8-cineolo sintase, (+)-sabineno sintase, E-alfa-bisaboleno sintase, (+)-difosfato de bornila sintase, levopimaradieno sintase, Abietadieno sintase, isopimaradieno sintase, (E)-gama-bisaboleno sintase, taxadieno sintase, pirofosfato de copalila sintase, kaureno sintase, longifoleno sintase, gama-humuleno sintase, Delta-selineno sintase, beta-felandreno sintase, limoneno sintase, mirceno sintase, terpinoleno sintase, (-)-canfeno sintase, (+)-3-careno sintase, difosfato de sin-copalila sintase, alfa-terpineol sintase, sin-pimara-7,15-dieno sintase, ent-sanda-aracopimaradieno sintase, estemer-

13-eno sintase, E-beta-ocimeno, S-linalool sintase, geraniol sintase, gama-terpineno sintase, linalool sintase, E-beta-ocimeno sintase, epi-cedrol sintase, alfa-zingibereno sintase, guaiadieno sintase, cascariladieno sintase, cis-muroladieno sintase, afidicolina-16b-ol sintase, elizabetatrieno sintase, sandalol sintase, patchoulol sintase, Zinzanol sintase, cedrol sintase, escareol sintase, copalol sintase, manool sintase e similares .

Sequências de nucleotídeo que codificam terpeno sintases são conhecidas na técnica, e qualquer sequência de nucleotídeo que codifica terpeno sintase pode ser usada para modificar geneticamente uma célula hospedeira. Por exemplo, as seguintes sequências de nucleotídeo que codificam terpeno sintases, seguidas por seus números de acesso no *Genbank* e os organismos nos quais elas foram identificadas, são conhecidas e podem ser usadas: (-)-germacreno D sintase mRNA (AY438099; *Populus balsamifera subsp. trichocarpa* x *Populus deltoids*); E,E-alfa-farneseno sintase mRNA (AY640154; *Cucumis sativus*); 1,8-cineol sintase mRNA (AY691947; *Arabidopsis thaliana*); terpeno sintase 5 (TPS5) mRNA (AY518314; *Zea mays*); terpeno sintase 4 (TPS4) mRNA (AY518312; *Zea mays*); mirceno/ocimeno sintase (TPS10) (At2g24210) mRNA (NM_127982; *Arabidopsis thaliana*); geraniol sintase (GES) mRNA (AY362553; *Ocimum basilicum*); pineno sintase mRNA (AY237645; *Picea sitchensis*); mirceno sintase 1e20 mRNA (AY195609; *Antirrhinum majus*); (E)-β-ocimeno sintase (0e23) mRNA (AY195607; *Antirrhinum majus*); E-β-ocimeno sintase mRNA (AY151086; *Antirrhinum majus*); terpeno sintase mRNA (AF497492; *Arabidopsis thaliana*); (-)-canfeno sintase (AG6.5) mRNA (U87910; *Abies grandis*); (-)-4S-limoneno sintase gene (por exemplo, sequência genômica) (AF326518; *Abies grandis*); delta-selineno sintase gene (AF326513; *Abies grandis*); amorfa-4,11-dieno sintase mRNA (AJ251751; *Artemisia annua*); E-α-bisaboleno sintase mRNA (AF006195; *Abies grandis*); gama-humuleno sintase mRNA (U92267; *Abies grandis*); δ-selineno sintase mRNA (U92266; *Abies grandis*); pineno sintase (AG3.18) mRNA (U87909; *Abies grandis*); mirceno sintase (AG2.2) mRNA (U87908; *Abies grandis*); etc.

Sequências de aminoácido das seguintes terpeno sintases são encontradas com os números de acesso no GenBank mostrados em parênteses, junto com o organismo no qual cada uma foi identificada, seguinte a cada terpeno sintase: (-)-germacreno D sintase (AAR99061; *Populus balsamifera subsp. trichocarpa* x *Populus deltoids*); D-cadineno sintase (P93665; *Gossypium hirsutum*); 5-epi-aristoloqueno sintase (Q40577; *Nicotiana tabacum*); E,E-alfa-farneseno sintase (AAU05951; *Cucumis sativus*); 1,8-cineol sintase (AAU01970; *Arabidopsis thaliana*); (R)-limoneno sintase 1 (Q8L5K3; *Citrus limon*); difosfato de sin-copalila sintase (AAS98158; *Oryza sativa*); uma taxadieno sintase (Q9FT37; *Taxus chinensis*; Q93YA3; *Taxus bacca*; Q41594; *Taxus brevifolia*); uma D-cadineno sintase (Q43714; *Gossypium arboretum*); terpeno sintase 5 (AAS88575; *Zea mays*); terpeno sintase 4 (AAS88573; *Zea mays*); terpenóide sintase (AAS79352; *Vitis vinifera*); geraniol sintase (AAR11765; *Ocimum basilicum*); mirceno sintase 1e20 (AAO41727; *Antirrhinum majus*); 5-epi-aristoloqueno sintase 37 (AAP05762; *Nicotiana attenuata*); (+)-3-careno sintase (AAO73863; *Picea abies*); (-)-canfeno sintase (AAB70707; *Abies grandis*); abietadieno sintase (AAK83563; *Abies grandis*); amorfa-4,11-diene sintase (CAB94691; *Artemisia annua*); tricodieno sintase (AAC49957; *Myrothecium roridum*); gama-humuleno sintase (AAC05728; *Abies grandis*); δ-selineno sintase (AAC05727; *Abies grandis*); etc.

Otimização de Códon

Em algumas modalidades, uma sequência de nucleotídeo que codifica uma enzima (por exemplo, MK; uma enzima do caminho do mevalonato, exceto MK, uma preniltransferase; uma terpeno sintase) é modificada para refletir a preferência do códon para a célula hospedeira particular. Por exemplo, a sequência de nucleotídeo será, em algumas modalidades, modificada para preferência de códon de levedura. Vide, por exemplo, Bennetzen e Hall (1982) *J. Biol. Chem.* 257(6): 3026-3031. Como outro exemplo não-limitativo, a sequência de nucleotídeo será, em algumas modalidades, modificada para preferência de códon de *E. coli*. Vide, por exemplo, Gouy e Gautier (1982) *Nucleic Acids Res.* 10(22):7055-7074; Eyre-

Walker (1996) *Mol. Biol. Evol.* 13(6):864-872. Vide também Nakamura et al. (2000) *Nucleic Acids Res.* 28(1):292. Tal modificação de códon também é referida como "otimização de códon".

Construtos de Expressão

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que compreende sequências de nucleotídeo que codificam MK (como descrito acima); um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto MK; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase; onde um ou mais dos ácidos nucléicos está presente em um vetor de expressão. Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão que compreende um ácido nucléico que compreende sequências de nucleotídeo que codificam MK (como descrito acima); um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto MK; um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase; e um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma preniltransferase; onde um ou mais dos ácidos nucléicos está presente em um vetor de expressão.

Vetores de expressão dequados incluem, mas não estão limitados a, vetores de baculovírus, vetores bacteriófagos, vetores de plasmídeos, fagemídeos, cosmídeos, fosmídeos, cromossomos artificiais bacterianos, virais (por exemplo, vetores virais à base de vírus vaccinia, poliovírus, adenovírus, vírus associados a adeno, SV40, vírus herpes simplex e similares), cromossomos artificiais à base de P1, plasmídeos de levedura, cromossomos artificiais de levedura e quaisquer outros vetores específicos para específicos hospedeiros de interesse (tal como *E. coli* e levedura). Vetores adequados incluem sequências cromossomais, não-cromossomais e sintéticas de DNA.

Numerosos vetores de expressão adequados são conhecidos

para aqueles versados na técnica e muitos estão comercialmente disponíveis. Os vetores a seguir são proporcionados a título de exemplo; para células hospedeiras bacterianas: vetores pQE (Qiagen), plasmídeos pBluescript, vetores pNH, vetores lambda-ZAP (*Stratagene*); pTrc99a, pKK223-3, pDR540, e pRIT2T (*Pharmacia*); para células hospedeiras eucarióticas: pXT1, pSG5 (*Stratagene*), pSVK3, pBPV, pMSG e pSVLSV40 (*Pharmacia*). Entretanto, qualquer outro plasmídeo ou outro vetor pode ser usado desde que ele seja compatível com a célula hospedeira.

Dependendo do sistema hospedeiro/vetor utilizado, qualquer dentre um número de elementos de controle de transcrição e tradução, incluindo promotores constitutivos e induzíveis, elementos intensificadores de transcrição, terminadores de transcrição, etc, pode ser usado no vetor de expressão (vide, por exemplo, Bitter et al. (1987) *Methods in Enzymology*, 153 : 516 - 544).

Promotores adequados para uso em células hospedeiras procarióticas incluem, mas não estão limitados a, um promotor da RNA polimerase do bacteriófago T7; um promotor de trp; um promotor do óperon lac; um promotor híbrido, por exemplo, um promotor híbrido de lac/tac, um promotor híbrido de tac/trc, um promotor de trp/lac, um promotor de T7/lac; um promotor de trc, um promotor de tac e similares; um promotor araBAD; promotores regulados *in vivo*, tal como um promotor ssaG ou um promotor relacionado (vide, por exemplo, a publicação da Patente US N° 20040131637), um promotor pagC (Pulkkinen e Miller, J. Bacteriol. (1991) 173 (1) : 86 - 93; Alpuche-Aranda et al., PNAS 1992; 89 (21) : 10079 - 83), um promotor *nirB* (Harborne et al. (1992) Mol. Micro. 6 : 2805 - 2813), e similares (vide, por exemplo, Dunstan et al. (1999) *Infect. Immun.* 67 : 5133 - 5141; McKelvie et al. (2004) *Vaccine* 22 : 3243 - 3255; e Chatfield et al. (1992) *Biotechnol.* 10 : 888 - 892); um promotor sigma70, por examplo, um promotor sigma70 de consenso (vide, por exemplo, números de acesso no GenBank AX798980, Ax798961 e AX798183); um promotor de fase estacionária, por exemplo, um promotor *dps*, um promotor *spv* e similares; um promotor derivado da ilha de patogenicidade SPI-2 (vide, por exemplo,

WO 96/17951); um promotor actA (vide, por exemplo, Shetron-Rama et al. (2002) *Infect. Immun.* 70 : 1087 - 1096); um promotor rpsM (vide, por exemplo, Valdivia e Falkow (1996) *Mol. Microbiol.* 22 : 367 378); um promotor tet (vide, por exemplo, Hillen, W. e Wissmann, A. (1989) In Saenger W. and Heinemann U. (eds) *Topics in Molecular and Structural Biology, Protein–Nucleic Acid Interaction*. Macmillan, London, UK, Vol. 10, pp. 143 – 162); um promotor SP6 (vide, por exemplo, Melton et al. (1984) *Nucl. Acids Res.* 12 : 7035 - 7056); e similares.

Exemplos não-limitativos de promotores eucarióticos adequados incluem um promotor precoce imedaito de CMV, promotor de timidina quinase HSV, promotor precoce ou tardio de SV40, LTRs de retrovírus e metalotioneína I de camundongo. Promotores adequados para expressão em levedura incluem, mas não estão limitados a, CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO e TP1; e, por exemplo, AOX1 (por exemplo, para uso em *Pichia*). A seleção do vetor e promotor adequados está dentro do nível de conhecimento do versado na técnica. O vetor de expressão também pode conter um sítio de ligação ao ribossomo para início de tradução e um terminador de transcrição. O vetor de expressão também pode incluir sequências apropriadas para expressão por amplificação.

Em adição, os vetores de expressão incluem um ou mais genes marcadores selecionáveis para proporcionar uma característica fenotípica para seleção de células hospedeiras transformadas, tal como resistência a dihidrofolato redutase ou neomicina para cultura de célula eucariótica, ou tal como resistência à tetraciclina ou ampicilina em células hospedeiras procarióticas, tal como *E. coli*.

Geralmente, um vetor de expressão irá incluir origens de replicação e marcadores selecionáveis que permitem a transformação da célula hospedeira, por exemplo, o gene de resistência à ampicilina de *E. coli*, o gene TRP1 de S. cerevisiae, etc; e um promotor derivado de um gene altamente expresso para direcionar a transcrição da sequência de codificação. Tais promotores podem ser derivados de operons que codificam

enzimas glicolíticas, tal como 3-fosfoglicerato quinase (PGK), fator-α, fosfatase ácida ou proteínas de choque térmico, entre outras.

Em algumas modalidades, uma sequência de nucleotídeo que codifica uma enzima (por exemplo, uma enzima do caminho do mevalonato; uma terpeno sintase; uma preniltransferase) está operacionalmente ligada a um promotor induzível, Promotores induzíveis são bem conhecidos na técnica. Promotores induzíveis adequados incluem, mas não estão limitados a, o pL do bacteriófago λ; Plac; Ptrp; Ptac (promotor híbrido Ptrp-lac); um promotor induzível por isopropil-beta-D-tiogalactopiranosídeo (IPTG), por exemplo, um promotor *lacZ*; um promotor induzível por tetraciclina; um promotor induzível por arabinose, por exemplo, P_{BAD} (vide, por exemplo, Guzman et al. (1995) *J. Bacteriol.* 177 : 4121 - 4130); um promotor induzível por xilose, por exemplo, Pxyl (vide, por exemplo, Kim et al. (1996) *Gene* 181 : 71 - 76); um promotor GAL1; um promotor de triptofano; um promotor lac; um promotor induzível por álcool, por exemplo, um promotor induzível por metanol, um promotor induzível por etanol; um promotor induzível por rafinose; um promotor induzível por calor, por exemplo, promotor P_L lambda induzível por calor; um promotor controlado por um repressor sensível ao calor (por exemplo, vetores de expressão à base de lambda reprimidos por CI857; vide, por exemplo, Hoffmann et al. (1999) *FEMS Microbiol Lett.* 177 (2) : 327 - 34); e similares.

Em algumas modalidades, uma sequência de nucleotídeo que codifica uma enzima (por exemplo, uma enzima do caminho do mevalonat; uma preniltransferase) está operacionalmente ligada um promotor constitutivo. Promotores constitutivos adequados pra uso em células procarióticas são conhecidos na técnica e incluem, mas não estão limitados a, um promotor sigma70, por exemplo, um promotor sigma70 de consenso.

Em levedura, um número de vetores contendo promotores constitutivos e induzíveis pode ser usado. Para uma revisão, vide *Current Protocols in Molecular Biology*, Vol. 2, 1988, Ed. Ausubel, et al., Greene Publish. Assoc. & Wiley Interscience, Ch. 13; Grant, et al., 1987, *Expression and Secretion Vectors for Yeast*, in Methods in Enzymology, Eds. Wu &

Grossman, 31987, Acad. Press, N.Y., Vol. 153, pp. 516 - 544; Glover, 1986, *DNA Cloning*, Vol. II, IRL Press, Wash., D.C., Ch. 3; e Bitter, 1987, *Heterologous Gene Expression in Yeast*, Methods in Enzymology, Eds. Berger & Kimmel, Acad. Press, N.Y., Vol. 152, pp. 673 - 684; e *The Molecular Biology of the Yeast Saccharomyces*, 1982, Eds. Strathern et al., Cold Spring Harbor Press, Vols. I e II. Um promotor de levedura constitutivo, tal como ADH ou LEU2 ou um promotor induzível, tal como GAL, pode ser usado (*Cloning in Yeast*, Ch. 3, R. Rothstein In: *DNA Cloning* Vol. 11, *A Practical Approach*, Ed. DM Glover, 1986, IRL Press, Wash., D.C.). De modo alternativo, vetores podem ser usados, os quais promovem a integração de sequências de DNA externas com cromossomos de levedura.

Em algumas modalidades, uma célula hospedeira geneticamente modificada em questão compreende um ácido nucléico que compreende sequências de nucleotídeo que codificam enzimas (por exemplo, uma enzima do caminho do mevalonato; uma terpeno sintase; uma preniltransferase), como descrito acima, onde cada um dos ácidos nucléicos está contido nos vetores de expressão em separado. Em outras modalidades, dois ou mais ácidos nucléicos estão contidos em um único vetor de expressão. Onde dois ou mais ácidos nucléicos estão contidos em um único vetor de expressão, em algumas modalidades, as sequências de nucleotídeo serão operacionalmente ligadas a um elemento de controle comum (por exemplo, um promotor). Onde dois ou mais ácidos nucléicos estão contidos em um único vetor de expressão, em algumas modalidades, as sequências de nucleotídeo serão operacionalmente ligadas a elementos de controle diferentes (por exemplo, promotores), por exemplo, elementos de controle diferentes são operacionalmente ligados a sequências de nucleotídeo que codificam enzima separadamente em um único vetor de expressão. Por exemplo, como notado acima, em algumas modalidades, uma sequência de nucleotídeo que codifica MK está operacionalmente ligada a um primeiro promotor; e uma sequência de nucleotídeo que codifica acetoacetil-CoA tiolase, HMGS e HMGR está operacionalmente ligada a um segundo promotor.

Células Hospedeiras

Em algumas modalidades, a célula hospedeira geneticamente modificada é uma célula eucariótica. Células hospedeiras eucarióticas adequadas incluem, mas não estão limitadas a, células de fungo, de alga, de inseto e de planta. Células hospedeiras de fungo adequadas incluem, mas não estão limitadas a, *Ascomycota*, *Basidiomycota*, *Deuteromycetes*, *Fungi imperfecti*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia kodamae*, Pichia membranaefaciens, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia quercuum*, *Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia sp.*, *Pichia angusta*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, Trichoderma reesei, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium graminearum*, *Fusarium venenatum* e *Neurospora crassa*. Células hospedeiras de alga adequadas incluem, mas não estão limitadas a, *Chlamydomonas reinhardtii* e *Phormidium sp.* ATCC29409.

Em outras modalidades, a célula hospedeira é uma célula procariótica. Células procarióticas adequadas incluem, mas não estão limitadas a, qualquer de uma variedade de cepas laboratoriais de *Escherichia coli*, *Lactobacillus sp.*, *Salmonella sp.*, *Shigella sp.*, e similares. Vide, por exemplo, Carrier et al. (1992) *J. Immunol.* 148 : 1176 - 1181; Patente US N° 6.447.784; e Sizemore et al. (1995) *Science* 270 : 299 - 302. Exemplos de cepas de salmonela, que podem ser empregadas na presente invenção, incluem, mas não estão limitadas à, *Salmonella typhi* e *S. typhimurium*. Cepas de *Shigella* adequadas incluem, mas não estão limitadas a, *Shigella flexneri*, *Shigella sonnei* e *Shigella disenteriae*. Tipicamente, a cepa de laboratório é uma não patogênica. Exemplos não-limitativos de outras bactérias adequadas incluem, mas não estão limitadas a, *Bacillus subtilis*, *Pseudomonas pudita*, *Pseudomonas aeruginosa*, *Pseudomonas mevalonii*, *Rhodobacter sphaeroides*, *Rhodobacter capsulatus*, *Rhodospirillum rubrum*, *Rhodococcus sp.* e similares. Em

algumas modalidades, a célula hospedeira é *Escherichia coli*.

Em outras modalidades, a célula hospedeira empregada no método de produção é uma célula bacteriana. Hospedeiros bacterianos adequados incluem, mas não estão limitados a, qualquer de uma variedade de bactérias gram-positivas, gram-negativas ou gram-variáveis, tais como micro-organismos que pertencem aos gêneros *Escherichia, Corynebacterium, Brevibacterium, Bacillus, Microbacterium, Serratia, Pseudomonas, Agrobacterium, Alicyclobacillus, Anabaena, Anacystis, Arthrobacter, Azobacter, Chromatium, Erwinia, Methylobacterium, Rhodobacter, Rhodopseudomonas, Rhodospirillum, Scenedesmun, Strepromyces, Synnecoccus e Zymomonas*. Exemplos de micro-organismos hospedeiros adequados usados aqui incluem em *Escherichia coli, Lactobacillus sp., Lactococcus lactis, Salmonella sp., Salmonella enterica, Salmonella typhi, Salmonella typhimurium, Shigella sp., Shigella flexneri, Shigella sonnei, Shigella dysenteriae, Enterobacter sakazakii, Pseudomonas* sp. D-0110, *Pseudomonas pudica, Pseudomonas aeruginosa, Pseudomonas mevalonii, Rhodobacter sphaeroides, Rhodobacter capsulatus, Rhodospirillum rubrum, Rhodospirillum salexigens, Rhodospirillum salinarum, Rhodococcus sp., Mesorhizobium loti, Clostridium acetobutylicum, Clostridium tetani E88, Clostridium lituseburense, Clostridium saccharobutylicum, Clostridium perfringens, Clostridium beijerinckii, Fusobacterium nucleatum, Thermoanaerobacterium thermosaccharolyticum, Butyrivibrio fibrisolvens, Bacillus thuringiensis, Bacillus anthracis, Bacillus megaterium, Bacillus subtilis, Bacillus amyloliquefacines, LactoBacillus johnsonii, Acinetobacter, Roseburia* sp., *Faecalibacterium prausnitzii*, and *Coprococcus* sp., *Staphylococcus epidermidis, Staphylococcus haemolyticus, Staphylococcus aureus, Brevibacterium ammoniagenes, Brevibacterium immariophilum* ATCC14068, *Brevibacterium saccharolyticum* ATCC14066, *Brevibacterium flavum* ATCC14067, *Brevibacterium lactofermentum* ATCC13869, *Corynebacterium glutamicum* ATCC13032, *Corynebacterium glutamicum* ATCC14297, *Corynebacterium acetoacidophilum* ATCC13870, *Microbacterium ammoniaphilum* ATCC15354, *Serratia ficaria, Serratia fonticola, Serratia liquefaciens, Serratia marcescens, Agrobacterium radiobacter, Agrobacteri-*

um rhizogenes, Agrobacterium rubi, Anabaena cylindrica, Anabaena doliolum, Anbaena flos-aquae, Arthrobacter aurescens, Arthrobacter citreus, Arthrobacter globformis, Arthrobacter hydrocarboglutamicus, Arthrobacter mysorens, Arthrobacter nicotianae, Arthrobacter paraffineus, Arthrobacter protophonniae, Arthrobacter roseoparaffinus, Arthrobacter sulfureus, Arthrobacter ureafaciens, Chromatium buderi, Chromatium tepidum, Chromatium vinosum, Chromatium warmingii, Chromatium fluviatile, Erwinia uredovora, Erwinia carotovora, Erwnia ananas, Erwinia herbicola, Erwinia punctata, Erwinia terreus, Methylobacterium rhodesianum, Methylobacterium extorquens, Rhodopseudomonas blastica, Rhodopseudomonas marina, Rhodopseudomonas palustris, Streptomyces ambofaciens, Streptomyces aureofaciens, Streptomyces aureus, Streptomyces fungicidicus, Streptomyces griseochromogenes, Streptomyces griseus, Streptomyces lividans, Streptomyces olivogriseus, Streptomyces rameus, Streptomyces tanashiensis, Streptomyces vinaceus, Zymomonas mobilis e similares (vide, por exemplo, Carrier et al. (1992) *J. Immunol.* 148 : 1176 - 1181; Patente US N° 6.447.784; e Sizemore et al. (1995) *Science* 270 : 299 - 302).

Em ainda outras modalidades, a cepa hospedeira bacteriana é não patogênica para seres humanos. Exemplos de cepas de *Escherichia coli* que podem ser empregadas incluem cepas de clonagem comuns, tais como DH1, B, MG1655, W3110, BL21, DH10B, JM109, DH5alfa, XL1-Blue, XL2-Blue, MC1000, KY3276, W1485, HB101, N°49, NY49, MP347, NM522 e derivadas das mesmas. Em uma modalidade, uma cepa RecA de *Escherichia coli* é empregada nos métodos da invenção. Em algumas modalidades, a cepa de *Escherichia coli* também produz IPP através de caminho DXP. Em outras modalidades, a cepa de *Escherichia coli* compreende um caminho DXP que é funcionalmente deficiente.

Em outras modalidades da presente invenção, uma célula hospedeira geneticamente modificada em questão é geneticamente modificada, de modo que uma enzima do caminho endógeno é funcionalmente deficiente. Enzimas que podem ser inativadas em um microorganismo hospedeiro para aumentar a produção de IPP e compostos

derivados do mesmo incluem, mas não estão limitadas a piruvato sintase, que, quando ativa, usa acetil-CoA (e dióxido de carbono e ferrodoxina reduzida) para produzir piruvato e, desse modo, reduz o fornecimento de acetil-CoA disponível para a produção de IPP e compostos derivados do mesmo; acetil-CoA sintetase, que, quando ativa, usa coenzima A (e propionato e ATP) para produzir propionil-CoA e, desse modo, reduz o fornecimento de coenzima A disponível para a produção de acetil-CoA; e piruvato formato liase, que, quando ativa, usa coenzima A (e 2-oxobutanoato) para produzir propionil-CoA e, desse modo, reduz o fornecimento de coenzima A disponível para a produção de acetil-CoA.

Métodos para desativar genes que codificam essas enzimas são bem-conhecidos na técnica e incluem, mas não estão limitados à, inserção de um elemento genético móvel (por exemplo, em transposon), supressão de todo ou parte do gene, de modo que o produto de gene não é feito, ou é truncado e é não funcional (isto é, enzimaticamente inativo); mutação do gene de modo que o produto de gene não é feito, ou é truncado e é não funcional; deleção ou mutação de um ou mais elementos de controle que controlam a expressão do gene, de modo que o produto do gene não é feito, e similar.

Métodos de Produção de um Composto Isoprenóide

A presente invenção proporciona métodos de produção de um composto isoprenóide, os métodos geralmente envolvem cultivo de uma célula hospedeira geneticamente modificada em questão em um meio adequado, sob condições que providenciam a produção das enzimas do caminho do mevalonato e da terpeno sintase (e opcionalmente também uma preniltransferase heteróloga, como descrito acima), de modo que um composto isoprenóide é produzido pela célula em uma quantidade recuperável. Em algumas modalidades, um método em questão adicionalmente compreende recuperação do composto isoprenóide, por exemplo, da célula, do meio de cultura da célula ou tanto da célula quanto do meio de cultura da célula.

A produção de um isoprenóide ou um precursor de isoprenóide é

aumentada em uma célula hospedeira geneticamente modificada em questão, em comparação à célula-mãe de controle, que não é assim geneticamente modificada. Desse modo, por exemplo, a produção de um isoprenóide ou um precursor de isoprenóide é aumentada em pelo menos cerca de 10%, pelo menos cerca de 20%, pelo menos cerca de 50%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 20 vezes, pelo menos cerca de 30 vezes, pelo menos cerca de 40 vezes, pelo menos cerca de 50 vezes, pelo menos cerca de 75 vezes, pelo menos cerca de 100 vezes, pelo menos cerca de 200 vezes, pelo menos cerca de 300 vezes, pelo menos cerca de 400 vezes, ou pelo menos cerca de 500 vezes, ou mais, na célula hospedeira geneticamente modificada, em comparação com a célula hospedeira de controle.

Por exemplo, a produção de um composto isoprenóide em uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 10%, pelo menos cerca de 20%, pelo menos cerca de 50%, pelo menos cerca de duas vezes, pelo menos cerca de 2,5 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, pelo menos cerca de 20 vezes, pelo menos cerca de 30 vezes, pelo menos cerca de 40 vezes, pelo menos cerca de 50 vezes, pelo menos cerca de 75 vezes, pelo menos cerca de 100 vezes, pelo menos cerca de 200 vezes, pelo menos cerca de 300 vezes, pelo menos cerca de 400 vezes, ou pelo menos cerca de 500 vezes, ou mais, mais alta do que o nível de produção do composto isoprenóide em uma célula hospedeira geneticamente modificada com pMevT, pMBIS, e um ácido nucléico heterólogo que codifica uma terpeno sintase. As sequências de nucleotídeo de pMevT e pMBIS podem ser encontradas na Patente US N^{os} 7.192.751 e 7.183.089; vide, por exemplo, SEQ ID Nº: 3 da Patente US N° 7.183.089; óperon MecT, SEQ ID Nº: 8 da Patente US N° 7.192.751; pMBIS, SEQ ID Nº: 4 da Patente US N° 7.183.089; e óperon MBIS, SEQ ID Nº: 13 da Patente US N° 7.192.751).

Como outro exemplo, em algumas modalidades, o nível de um composto isoprenóide produzido em uma célula hospedeira geneticamente

modificada em questão é pelo menos cerca de 500 mg/L, pelo menos cerca de 600 mg/L, pelo menos cerca de 700 mg/L, pelo menos cerca de 800 mg/L, pelo menos cerca de 900 mg/L, pelo menos cerca de 1000 mg/L, pelo menos cerca de 1200 mg/L, pelo menos cerca de 1400 mg/L, pelo menos cerca de 1600 mg/L, pelo menos cerca de 1800 mg/L ou pelo menos cerca de 2000 mg/L, após cerca de 10 horas a cerca de 20 horas, cerca de 20 horas a cerca de 30 horas, cerca de 30 horas a cerca de 40 horas, ou cerca de 40 horas a cerca de 50 horas, em cultura. Por exemplo, em algumas modalidades, o nível de um composto isoprenóide produzido em uma célula hospedeira geneticamente modificada em questão é a partir de cerca de 500 mg/L a cerca de 600 mg/L, a partir de cerca de 600 mg/L a cerca de 700 mg/L, a partir de cerca de 700 mg/L a cerca de 800 mg/L, a partir de cerca de 800 mg/L a cerca de 1000 mg/L, a partir de cerca de 1000 mg/L a cerca de 1200 mg/L, a partir de cerca de 1200 mg/L a cerca de 1400 mg/L, a partir de cerca de 1400 mg/L a cerca de 1600 mg/L, a partir de cerca de 1600 mg/L a cerca de 1800 mg/L, ou a partir de cerca de 1800 mg/L a cerca de 2000 mg/L, ou maior do que 2000 mg/L, após cerca de 10 horas a cerca de 20 horas, cerca de 20 horas a cerca de 30 horas, cerca de 30 horas a cerca de 40 horas, ou cerca de 40 horas a cerca de 50 horas, em cultura. Em algumas modalidades, o nível de um composto isoprenóide produzido em uma célula hospedeira geneticamente modificada em questão é pelo menos cerca de 500 mg/L, pelo menos cerca de 600 mg/L, pelo menos cerca de 700 mg/L, pelo menos cerca de 800 mg/L, pelo menos cerca de 900 mg/L, pelo menos cerca de 1000 mg/L, pelo menos cerca de 1200 mg/L, pelo menos cerca de 1400 mg/L, pelo menos cerca de 1600 mg/L, pelo menos cerca de 1800 mg/L, ou pelo menos cerca de 2000 mg/L, onde o nível é após cerca de 10 horas a cerca de 20 horas, cerca de 20 horas a cerca de 30 horas, cerca de 30 horas a cerca de 40 horas, ou cerca de 40 horas a cerca de 50 horas, após indução (por exemplo, onde uma ou mais das sequências de nucleotídeo que codificam uma enzima (por exemplo, enzima de mevalonato, exceto MK, preniltransferase, terpeno sintase) está operacionalmente ligada a um promotor induzível).

Em algumas modalidades, a taxa de crescimento de uma célula hospedeira geneticamente modificada em questão é maior do que a taxa de crescimento de uma célula de controle. Por exemplo, em algumas modalidades, uma célula hospedeira geneticamente modificada em questão cresce a uma taxa que é pelo menos cerca de 10%, pelo menos cerca de 15%, pelo menos cerca de 20%, pelo menos cerca de 25%, pelo menos cerca de 30%, pelo menos cerca de 35%, pelo menos cerca de 40%, pelo menos cerca de 45%, pelo menos cerca de 50%, pelo menos cerca de 60%, pelo menos cerca de 70%, pelo menos cerca de 80%, pelo menos cerca de 90%, pelo menos cerca de 100% ou 2 vezes, pelo menos cerca de 5 vezes, pelo menos cerca de 10 vezes, ou mais, mais alta do que a txa de crescimento de uma célula de controle.O crescimento da célula o é prontamente determinado usando métodos bem conhecidos, por exemplo, medição de densidade ótica (OD) a cerca de 600 nm (OD_{600}) de culturas líquidas de bactérias; tamanho de colônia; taxa de crescimento e similares.

Isoprenóides que podem ser produzidos usando o método da invenção incluem, mas não estão limitados a, monoterpenos, incluindo, mas não limitados a, limoneno, citranelol, geraniol, mentol, perilil álcool, linalol, thujone, sesquiterpenos, incluindo, mas não limitados a, periplanona B, ginkgolide B, amorfadieno, artemisinina, ácido artemisínico, valenceno, nootkatona, epi-cedrol, epi-aristoloqueno, farnesol, gossipol, santonina, periplanona e froskolina; diterpenos, incluindo, mas não limitados a, casbeno, eleuterobina, paclitaxel, prostatina e pseudopterosina; triterpenos, incluindo, mas não limitados a, arbrusideE, bruceantina, testosterona, progesterona, cortisona, digitoxina. Isoprenóides também incluem, mas não estão limitados a, carotenóides, tal como licopeno, caroteno α e β, criptoxantina α e β, bixina, zeaxantina, astaxantina e luteína. Isoprenóides também incluem, mas não estão limitados a, triterpenos, compostos esteróides e compostos que são compostos de isoprenóides modificados por outros grupos químicos, tais como alcalóides misturados com terpeno, menaquinonas (por exemplo, vitamina K-2) e coenzima Q-10.

Em algumas modalidades, uma célula hospedeira

geneticamente modificada em questão é cultivada em um meio adequado (por exemplo, caldo Luria-Bertoni, opcionalmente suplementado com um ou mais agentes adicionais, tal como um indutor (por exemplo, onde uma ou mais sequências de nucleotídeo que codificam enzimas (por exemplo, uma sequência de nucleotídeo que codifica MK, uma (s) enzima (s) do caminho do mevalonato, exceto MK, uma preniltransferase, uma terpeno sintase) estão sob controle d eum promotor induzível), etc.); e o meio de cultura é sobreposto com um solvente orgânico, por exemplo, dodecano, formando uma camada orgânica. O composto isoprenóide produzido pela célula hospedeira geneticamente modificada em questão é se divide na camada orgânica, a partir da qual ele pode ser purificado. Em algumas modalidades, onde a sequência de nucleotídeo que codifica enzima que modifica isoprenóide está operacionalmente ligada a um promotor induzível, um indutor é adicionado ao meio de cultura; e, após um tempo adequado, o composto isoprenóide é isolado da camada orgânica sobreposta no meio de cultura.

Em algumas modalidades, o composto isoprenóide produzido por uma célula hospedeira geneticamente modificada será separado de outros produtos que podem estar presentes na camada orgânica. A separação do composto isoprenóide de outros produtos que podem estar presentes na camada orgânica é prontamente alcançada usando, por exemplo, técnicas cromatográficas padrão.

Em algumas modalidades, um composto isoprenóide sintetizado por um método em questão é adicionalmente quimicamente modificado em uma reação livre de célula. Por exemplo, em algumas modalidades, ácido artemisínico é isolado do meio de cultura e/ou um lisato de célula, e o ácido artemisínico é adicionalmente quimicamente modificado em uma reação livre de célula para gerar artemisinina.

Em algumas modalidades, o composto isoprenóide é puro, por exemplo, pelo menos cerca de 40% puro, pelo menos cerca de 50% puro, pelo menos cerca de 60% puro, pelo menos cerca de 70% puro, pelo menos cerca de 80% puro, pelo menos cerca de 90% puro, pelo menos cerca de

95% puro, pelo menos cerca de 98% ou mais do que 98% puro, onde "puro", no contexto de um composto isoprenóide, refere-se a um composto isoprenóide que é livre de outros compostos isoprenóides, macromoléculas, contaminantes, etc.

Exemplos

Os exemplos a seguir são utilizados para proporcionar àqueles versados na técnica uma divulgação e descrição completas de como realizar e usar a presente invenção, e não se destinam a limitar o escopo do que os inventores consideram como sua invenção, nem se destinam a simbolizar que os experimentos abaixo são todos ou os únicos experimentos realizados. Foram feitos esforços para garantir a precisão com relação aos números usados (por exemplo, quantidades, temperatura, etc.), porém alguns erros e desvios experimentais devem ser levados em conta. A menos que indicado de outra forma, partes são partes por peso, peso molecular é peso molecular médio, a temperatura é em graus Celsius e a pressão é a pressão atmosférica ou próxima da mesma. Abreviações-padrão podem ser usadas, por exemplo, bp, par (es) de base; kb, kilobase (s); pl, picolitro (s); s ou sec, segundo (s); min, minuto (s); h hora (s); aa, aminoácido (s); kb, kilobase (s), bp par (es) de base; nt, nucleotídeo (s); i.m., intramuscular (mente); i.p., intraperitoneal (mente); s.c., subcutanea (mente); e similares.

Exemplo 1: Gerando plasmídeos de expresão que codificam enzimas do caminho do mevalonato

Plasmídeo de expressão pMevT foi gerado por inserção do óperon MevT no vetor pBAD33. O óperon MevT codifica o conjunto de enzimas do caminho MEV que juntas transformam a acetil-CoA precursora ubíqua em (R)-mevalonato, a saber, acetoacetil-CoA tiolase, HMG-CoA sintase e HMG-CoA redutase. O óperon MevT foi gerado através de amplificação por PCR a partir da sequência de codificação do DNA genômico de *Escherichia coli* do gene atoB (número de acesso no GenBank NC_000913 REGIÃO: 2324131..2325315) (codifica uma acetocetil-CoA tiolase), a partir da sequência de codificação do DNA genômico de *Saccharomyces cerevisiae* do gene ERG13 (número de acesso no GenBank

X96617, REGIÃO: 220..1695) (codifica uma HMG-CoA sintase) e a partir de um segmento do DNA genômico de *Saccharomyces cerevisiae* da região de codificação do gene HMG1 (número de acesso no GenBank M22002, REGIÃO: 1660..3165) (codifica uma HMG-CoA redutase truncada (tHMGR)). O iniciador de PCR a montante usado para a amplificação do fragmento do gene HMG1 incluiu um gene de iniciação artificial. Os fragmentos amplificados foram unidos usando extensões de sobreposição (SOEing), sendo que, durante este processo, sítios de ligação ao ribossomo foram introduzidos após as sequências de codificação do atoB e do ERG13. Após a adição de protuberâncias 3' A, o óperon MevT foi ligado ao vetor de clonagem pCR4 TA (Invitrogen, Carlsbad, CA). O óperon MevT foi subsequentemente ligado ao sítio de restrição para *XmaI PstI* do vetor pBAD33 (Guzman *et al.* (1995) *J. Bacteriol.* 177(14): 4121-4130). Para colocar o óperon sob o controle do promotor P_{Lac}, o fragmento *araC-*P_{BAD} *NsiI-XmaI* do pBAD33 foi substituído pelo fragmento NsiI-XmaI do pBBR1MCS, produzindo o plasmídeo de expressão pMevT (vide Patente US número 7.192.751).

O plasmídeo de expressão pAM36-MevT66 foi gerado através de inserção do óperon MevT66 no vetor pAM36. O vetor pAM36 foi gerado através de inserção de um cassete de oligonucleotídeo contendo sítios de restrição para *AscI-SfiI-AsiSI-XhoI-PacI-FseI-PmeI* no vetor pACYC184 (número de aceso no GenBank XO6403) e através de remoção do gene que confere resistência à tetramicina em pACYC184. O óperon MevT66 foi sinteticamente gerado usando SEQ ID Nº: 1 como um modelo, o qual compreende o gene atoB de *Escherichia coli* (número de acesso no GenBank NC_000913 REGIÃO: 2324131..2325315), o gene ERG13 de *Saccharomyces cerevisiae* (número de aceso no GenBank X96617, REGIÃO: 220..1695), e uma versão truncada do gene HMG1 de *Saccharomyces cerevisiae* (número de aceso no GenBank M22002, REGIÃO: 1777..3285), todas as três sequências sendo códon-otimizadas para expressão em *Escherichia coli*. O óperon MevT66 sinteticamente gerado foi flanqueado através de um sítio de restrição para EcoRI em 5' e

um sítio de restrição para Hind III em 3', e poderia, assim, ser clonado em sítios de restrição compatíveis de um vetor de clonagem, tal como um vetor de origem pUC ou pACYA padrão. A partir deste construto, o óperon MevT66 foi amplificado por PCR com sítios de restrição para *SfiI* e *AsiSI* de flanqueamento, o fragmento de DNA amplificado foi totalmente digerido usando enzimas de restrição *SfiI* e *AsiSI*, a mistura da reação foi resolvida através de eletroforese em gel, o fragmento de DNA de aproximadamente 4,2 kb foi extraído por gel usando um kit de purificação de gel (Qiagen, Valencia, CA), e o fragmento de DNA isolado foi ligado ao sítio de restrição para *SfiI AsiSI* do vetor pAM36, produzindo o plasmídeo de expressão pMevT66.

O plasmídeo de expressão pAM25 foi gerado através de inserção do óperon MevT66 no vetor pAM29. O vetor pAM29 foi criado através de montagem da origem p15A de replicação e do gene que confere resistência à kanamicina do pZS24-MCS1 (Lutz e Bujard (1997) *Nucl Acids Res.* 25:1203-1210) com um promotor *lacUV5* gerado por oligonucleotídeo. A sequência de nucleotídeo de pAM29 é dada como SEQ ID Nº: 2. O construto de síntese de DNA que compreende o óperon MevT66 (vide descrição para pAM36-MevT66 acima) foi totalmente digerigo usando enzimas de restrição *EcoRI* and *Hind III*, a mistura da reação foi resolvida através de eletroforese em gel, o fragmento de DNA de aproximadamente 4,2 kb foi extraído por gel, e o fragmento de DNA isolado foi ligado ao sítio de restrição para *EcoRI HindIII* do pAM29, produzindo o plasmídeo de expressão pAM25.

O plasmídeo de expressão pMevB-Cm foi gerado através de inserção do óperon MevB no vetor pBBR1MCS-1. O óperon MevB codifica o conjunto de enzimas que juntas convertem (R)-mevalonato em IPP, a saber, mevalonato quinase, fosfomevalonato quinase e mevalonato pirofosfato descarboxilase. O óperon MevB foi gerado através de amplificação por PCR a partir das sequências de codificação do DNA genômico de Sacchromyces cerevisiae do gene ERG12 (número de aceso no GenBank X55875, REGIÃO: 580..1911) (codifica uma mevalonato quinase), o gene ERG8

(número de aceso no GenBank Z49939, REGIÃO: 3363..4718) (codifica uma fosfomevalonato quinase), e o gene MVD1 (número de aceso no GenBank X97557, REGIÃO: 544..1734) (codifica uma mevalonato pirofosfato descarboxilase), e através de união de fragmentos de PCR usando estensões de sobreposição (SOEing). Através de escolha apropriada de iniciador, os códons de parada de *ERG12* e *ERG8* foram mudados de TAA para TAG durante amplificação para introduzir sítios de ligação ao ribossomo. Após a adição de protuberâncias 3' A, o óperon MevB foi ligado ao vetor de clonagem pCR4 TA (Invitrogen, Carlsbad, CA). O óperon MevB foi extirpado por digestão total do construto de clonagem usando enzima de restrição *PstI*, resolvendo a mistura da reação através de eletroforese em gel, extraindo por gel o fragmento de DNA de aproximadamente 4,2 kb e ligando o fragmento de DNA isolado ao sítio de restrição para *PstI* do vetor pBBR1MCS-1 (Kovach *et al.*, *Gene* 166(1): 175-176 (1995)), produzindo o plasmídeo de expressão pMevB-Cm.

O plasmídeo de expressão pMBI foi gerado através de inserção do óperon MBI no vetor pBBR1MCS-3. Além das enzimas do óperon MevB, o óperon MBI também codifica uma isopentenil pirofosfatase isomerase, a qual cataliza o conversão de IPP em DMAPP. O óperon MBI foi gerado através de amplificação por PCR a partir da sequência de codificação do DNA genômico de *Escherichia coli* do gene *idi* (número de aceso no GenBank AF119715) usando iniciadores que continham um sítio de restrição para *XmaI* nas extremidades 5', digerindo o fragmento de DNA amplificado por completo usando enzima de restrição *XmaI*, resolvendo a mistura da reação através de eletroforese em gel, extraindo por gel o fragmento de DNA de aproximadamente 0,5 kb e ligando o fragmento de DNA isolado ao sítio de restrição para *XmaI* do plasmídeo de expressão pMevB-Cm, desse modo colocando *idi* na extremidade 3' do óperon MevB. O óperon MBI foi subclonado no sítio de restrição para *SalI SacI* do vetor pBBR1MCS-3 (Kovach *et al.*, *Gene* 166(1): 175-176 (1995)), produzindo o plasmídeo pMBI (vide patente US número 7.192.751).

O plasmídeo de expressão pMBIS foi gerado através de

inserção do gene *ispA* no pMBI. O gene *ispA* codifica uma pirofosfato de farnesila sintase, que cataliza a condensação de duas moléculas de IPP com uma molécula impermeável de DMAPP para fazer pirofosfato de farnesila (FPP). A sequência de codificação do gene *ispA* (número de aceso no GenBank D00694, REGIÃO: 484..1383) foi amplificada por PCR a partir do DNA genômico de *Escherichia coli* usando um iniciador para diante com um sítio de restrição para *SacII* e um iniciador reverso com um sítio de restrição para *SacI*. O produto amplificado por PCR foi digerido por completo usando enzimas de restrição *SacII e SacI*, a mistura da reação foi resolvido através de eletroforese em gel e o fragmento de DNA de aproximadamente 0,9 kb foi extraído por gel, e o fragmento de DNA isolado foi ligado ao sítio de restrição para *SacII SacI* do pMBI, desse modo colocando o gene *ispA* em 3' de *idi* e do óperon MevB, produzindo o plasmídeo de expressão pMBIS (vide patente US número 7.192.751).

O plasmídeo de expressão pAM47 foi gerado através de inserção do óperon MBIS no vetor pAM37. O vetor pAM37 foi gerado através de montagem da origem p15A de replicação e do gene que confere resistência ao cloranfenicol do vetor do pZA31luc (Lutz e Bujard (1997) *Nucl Acids Res.* 25 : 1203-1210) com um promotor constitutivo forte sintaticamente gerado através de ligação de oligonucleotídeos. O óperon MBIS foi amplificado por PCR a partir de pMBIS usando iniciadores 9-38A (SEQ ID Nº: 7) e 9-38B (SEQ ID Nº: 8), o produto PCR de aproximadamente 5,5 kb foi purificado e digerido por completo usando enzimas de restrição *ApaI* e *MluI*, e o fragmento de DNA foi ligado ao sítio de enzima de restrição para *ApaI* e *MluI* do pAM37, produzindo o plasmídeo de expressão pAM47.

Plasmídeos de expressão pAM39 e pAM40 foram gerados através de combinação dos plasmídeos de expressão pAM47 e pAM25. Os plasmídeos de expressão pAM47 e pAM25 foram digeridos usando enzima de restrição *SacI*, e ligados entre si em duas orientações, produzindo os plasmídeos de expressão pAM39 e pAM40. As sequências de nucleotídeo dos plasmídeos de expressão pAM39 e pAM40 são dadas como SEQ ID Nº: 3 e SEQ ID Nº: 4, respectivamente, e mapas de plasmídeos são mostrados

nas figuras 3 e 4, respectivamente.

O plasmídeo de expressão pAM45 foi gerado através de inserção do óperon MBIS em pAM36-MevT66 e adição de promotores *lacUV5* na frente dos operons MBIS e MevT66. O óperon MBIS foi amplificado por PCR a partir de pMBIS usando iniciadores que compreendem um sítio de restrição para *XhoI* em 5' e um sítio de restrição para *PacI* em 3', o produto amplificado por PCR foi digerido por completo usando enzimas de restrição *XhoI* e *PacI*, a mistura da reação foi resolvida através de eletroforese em gel, o fragmento de DNA de aproximadamente 5,4 kb foi extraído por gel e o fragmento de DNA foi ligado ao sítio de restrição para *XhoI PacI* do pAM36-MevT66, produzindo o plasmídeo de expressão pAM43. Um fragmento de DNA que compreende uma sequência de nucleotídeo que codifica o promotor *lacUV5* foi sintetizado a partir de oligonucleotídeos e ligado aos sítios de restrição para *AscI SfiI* e *AsiSI XhoI* do pAM43, produzindo o plasmídeo de expressão pAM45. A sequência de nucleotídeo do pAM45 é dada como SEQ ID Nº: 5 e um mapa de plasmídeo é mostrado na figura 5.

O plasmídeo de expressão pAM29-MK foi gerado através de enserção de uma sequência de nucleotídeo que codifica uma mevalonato quinase ("MK") no vetor pAM29. Uma sequência de nucleotídeo que codifica uma mevalonato quinase foi amplificada por PCR a partir de pMBIS usando iniciador MK-SR (SEQ ID Nº: 9) e MK-HR (SEQ ID Nº: 10). O produto PCR foi purificado, digerido por completo usando enzimas de restrição *SalI* e *HindIII*, e ligado ao sítio de enzima de restrição para *SalI HindIII* do pAM29, produzindo o plasmídeo de expressão pAM29-MK.

O plasmídeo de expressão pAM29-PMK foi gerado através de inserção de uma sequência de nucleotídeo que codifica uma fosfomevalonato quinase ("PMK") no vetor pAM29. Uma sequência de nucleotídeo que codifica uma fosfomevalonato quinase foi amplificada por PCR a partir de pMBIS usando iniciadores PMK-SR (SEQ ID Nº: 11) ePMK-HR (SEQ ID Nº: 12). O produto PCR foi purificado, digerido por completo usando enzimas de restrição *SalI* e *HindIII* e ligado ao sítio de enzima de

restrição para *SalI HindIII* do pAM29, produzindo o plasmídeo de expressão pAM29-PMK.

O plasmídeo de expressão pAM29-MPD foi gerado através de inserção de uma sequência de nucleotídeo que codifica uma mevalonato pirofosfato descarboxilase ("MPD") no vetor pAM29. Uma sequência de nucleotídeo que codifica uma mevalonato pirofosfato descarboxilase foi amplificada por PCR a partir de pMBIS usando iniciadores MPD-ER (SEQ ID Nº: 13) e MPD-SR (SEQ ID Nº: 14). O produto PCR foi purificado, digerido por completo usando enzimas de restrição *SalI* e *HindIII* e ligado ao sítio de enzima de restrição para *EcoRI SalI* do pAM29, produzindo o plasmídeo de expressão pAM29-MPD.

O plasmídeo de expressão pAM29-idi foi gerado através de inserção de uma sequência de nucleotídeo que codifica uma pirofosfato de isopentenila isomerase ("idi") no vetor pAM29. Uma sequência de nucleotídeo que codifica uma pirofosfato de isopentenila isomerase foi amplificada por PCR a partir de pMBIS usando iniciadores idi-EF (SEQ ID Nº: 15) e idi-SR (SEQ ID Nº: 16). O produto PCR foi purificado, digerido por completo usando enzimas de restrição *SalI* e *HindIII* e ligado ao sítio de enzima de restrição para *EcoRI SalI* do pAM29, produzindo o plasmídeo de expressão pAM29-*idi*.

O plasmídeo de expressão pAM29-ispA foi gerado através da inserção de uma sequência de nucleotídeo que codifica uma pirofosfato de farnesila sintase ("ispA") no vetor pAM29. Uma sequência de nucleotídeo que codifica uma pirofosfato de farnesila sintase foi amplificada por PCR a partir de pMBIS usando iniciadores ispA-EF (SEQ ID Nº: 17) e ispA-SR (SEQ ID Nº: 18). O produto PCR foi purificado, digerido por completo usando enzimas de restrição *SalI* e *HindIII* e ligado ao sítio de enzima de restrição para *EcoRI SalI* do pAM29, produzindo o plasmídeo de expressão pAM29-*ispA*.

Exemplo 2: Geração de plasmídeos de expressão que codificam enzimas que convertem FPP.

O plasmídeo de expressão pTrc99A-ADS foi gerado através de

inserção de uma sequência de nucleotídeo que codifica uma amorfa-4,11-dieno sintase ("ADS") no vetor pTrc99A. A sequência de amorfa-4,11-dieno sintase foi gerada sinteticamente, de modo que, após a tradução, a sequência de aminoácido seria idêntica àquela descrita por Merke et al. (2000) *Ach. Biochem. Biophys.* 381:173-180, de modo que a sequência de nucleotídeo que codifica a amorfa-4,11-dieno sintase foi otimizada para expressão em *Escherichia coli,* de modo que a sequência de nucleotídeo foi flanqueada através do sítio de enzima de restrição para *NcoI* em 5' e para *XmaI* em 3' (vide patente US número 7,192,751). A sequência de nucleotídeo foi digerida por completo usndo enzimas de restrição *NcoI* e *XmaI*, a mistura da reação foi resolvida através de eletroforese em gel, o fragmento de DNA de aproximadamente 1,6 kb foi extraído por gel e o fragmento de DNA isolado foi inserido no sítio de enzima de restrição para *NcoI XmaI* do vetor pTrc99A (Amman et al. (1985) *Gene* 40:183-190), produzindo o plasmídeo de expressão pTrc99A-ADS (vide figura 6 para um mapa de plasmídeo).

O plasmídeo de expressão pAM94 foi gerado através da inserção de uma sequência de nucleotídeo que codifica a *Saccharomyces cerevisiae* melavonato quinase (MK), códon-otimizado para expressão em *Escherichia coli*, no vetor pTrc99A. O sequência de nucleotídeo códon-otimizada foi amplificada por PCR a partir de pMBISopt usando iniciadores 9-153C (SEQ ID Nº: 19) e 9-153D (SEQ ID Nº: 20). O produto PCR foi digerido por completo usando enzimas de restrição *BamHI* e *HindIII* e ligado ao sítio de restrição para *BamHI HindIII* do plasmídeo de expressão pTrc99A-ADS, produzindo o plasmídeo de expressão pAM94. A sequência de nucleotídeo do pAM94 é dada como SEQ ID Nº: 6 e um mapa de plasmídeo é mostrado na figura 7.

Exemplo 3: Geração de cepas hospedeiras de *Escherichia coli*

Tabela 1 - Cepas hospedeiras de *Escherichia coli*

Cepas Hospedeiras	Plasmídeos de Expressão	Seleção de Antibiótico
B32	pMevT pMBIS pTrc-ADS	100 ug/mL de Carbenicilina 35 ug/mL de Cloranfenicol 10 ug/mL de Tetraciclina
B59	pAM39 pTrc-ADS	100 µg/ml de Carbenicilina 35 de µg/ml Cloranfenicol
B60	pAM40 pTrc-ADS	
B125	pAM45 pTrc-ADS	
1	pMBIS pTrc-ADS pAM29	100 µg/mL de Ampicilina 50 µg/mL de Kanamicina 5 µg/mL de Tetraciclina
2	pMBIS pTrc-ADS pAM29-MK	
3	pMBIS pTrc-ADS pAM29-PMK	
4	pMBIS pTrc-ADS pAM29-MPD	
5	pMBIS pTrc-ADS pAM29-idi	
6	pMBIS pTrc-ADS pAM29-ispA	
B177	pMevT pMBIS pAM94	100 µg/mL de carbenicilina 35 µg/mL de cloranfenicol 10 µg/mL de tetraciclina

Transformantes de célula hospedeira foram selecionados em

ágar Luria Bertoni (LB) contendo antibióticos, como detalhado na Tabela 1. Colônias simples de cepas B59 e B60 foram transferidas de ágar LB para tubos de cultura contendo 5 mL de meio líquido LB e antibióticos. Colônias simples de cepas B32, B125 1 até 6, e B177 foram transferidas de ágar LB para tubos de cultura contendo 5 mL de M9-MOPS (Tabela 2) contendo 8 – 10 g/L de D-glicose. As culturas foram incubadas a 30°C em um agitador rotatório a 250 rpm por 30 horas, neste ponto o crescimento celular foi atrasado por arrefecimento das culturas em gelo. As células foram armazenadas a -80°C em frascos do tipo "cryo" em 1 mL de alíquotas estoque feitas de 400 µL de glicerol estéril a 50% e 600 µL de cultura líquida.

Exemplo 4: Produção de amorfa-4,11-dieno em cepas hospedeiras de *Escherichia coli*

Culturas de produção de cepas hospedeiras de B32, B59, B60 e B125 foram estabelecidas através da adição de uma alíquota estoque de cada cepa para separar frascos de 250 mL contendo 40 mL de meio (meio TB-1% de glicerol para cepas B32, B59 e B60; meio M9-MOPS contendo 20 g/L de D-glicose para cepas B32 e B125) e antibióticos, conforme detalhado na Tabela 1. As culturas foram incubadas a 30°C em um agitador rotatório a 250 rpm até que eles alcançassem um OD_{600} de aproximadamente 0,2 a 0,3, neste ponto a produção de amorfa-4,11-dieno nas células hospedeiras foi induzida pela adição de 40 µL de IPTG a 1 M ao meio de cultura. No momento da indução, as culturas foram cobertas com 8 mL de dodecano para capturar a amorpa-4,11-dieno. Amostras foram tomadas em vários pontos de tempo através da adição da camada de dodecano a 990 µL de acetato de etila em um frasco GC de vidro claro e formação de vórtice por 30 segundos.

Os extratos de cultura de acetato de etila foram analisados em um cromatógrafo a gás/espectômetro de massa (GC/MS) Hewlett-Packard 6890, conforme descrito em Martin *et al.* ((2001) *Biotechnol. Bioeng.* 75:497-503), através de varredura para o íon molecular (204 m/z) e o íon 189 m/z. Para acelerar os tempos de execução, o programa de temperatura e matriz de coluna foram modificados para alcançar uma resolução de pico ótima e o

menor tempo de execução total. Compostos em uma maostra de 1 µL foram separados usando uma coluna DB-XLB (disponível por Agilent Technologies, Inc., Palo Alto, CA) e gás veículo hélio. O ciclo de forno para cada amostra era ou 80°C mantidos por 2 minutos, aumentando a temperatura a 30°C/minuto até uma temperatura de 160°C, aumentando a temperatura a 3°C/minuto até uma temperatura de 170°C, aumentando a temperatura a 50°C/minuto até 300°C, e mantendo a 300°C por 2 minutos (protocolo GC 1), ou 100°C mantidos por 0,75 minuto, aumentando a temperatura a 60°C/minuto até uma temperatura de 300°C, e mantendo a 300°C por 0,5 minuto (protocolo GC 2). As amostras resolvidas foram analisadas por um detector seletivo de massa Hewlett-Packard, modelo 5973. Espectros de massa anteriores demonstraram que o produto da amorfa-4,11-dieno sintase foi amorfa-4,11-dieno, e que amorfa-4,11-dieno teve um tempo de retenção de 7,9 minutos usando protocolo GC 1. Beta- ou trans-cariofileno (Sigma-Aldrich, St. Louis, MO) foi usado como um padrão interno para quantificação. Titragem de amorfa-4,11-dieno foi calculada com base em uma curva de calibração quantitativa do amorfa-4,11-dieno purificado em acetato de etila "embebido" com cariofileno.

As figuras 8A e 8B mostram que cepas hospedeiras, em que o óperon MBIS é expresso a partir de um plasmídeo com número de cópia mais alto e um promotor mais forte, produzem níveis mais altos de amorfa-4,11-dieno.

Exemplo 5: Produção de amorfa-4,11-dieno em cepas hospedeiras de *Escherichia coli* que abrigam uma pluralidade de cópias de sequências de nucleotídeo que codificam mevalonato quinase.

Custuras de produção de cepas hospedeiras e até 6 foram estabelecidas através da adição de uma alíquota estoque de cada cepa para separar frascos de 250 mL contendo 50 mL de meio M9-MOPS contendo 10 g/L de D-glicose e antibióticos, conforme detalhado na Tabela 1, a um OD_{600} de início de aproximadamente 0,5. As culturas foram incubadas a 30°C em um agitador rotatório a 250 rpm até que elas alcançassem um OD_{600} de aproximadamente 0,2 a 0,3, neste ponto a produção de amorfa-4,11-dieno

nas células hospedeiras foi induzida através da adição de 40 µL de IPTG a 1 M ao meio de cultura. No momento da indução, as culturas foram também suplementadas com 20 mM de D,L-mevalonato e foram cobertas com 8 mL de dodecano para capturar o amorfa-4,11-dieno. Amostras foram tomadas em vários pontos de tempo através da transferência de 10 µL do dodecano coberto para 990 µL de acetato de etila, e analisadas usando protocolo GC 1, conforme descrito no Exemplo 4.

A figura 9 mostra que cepas hospedeiras em que a dosagem de gene de mevalonato quinase é aumentada produz níveis mais altos de amorfa-4,11-dieno.

Exemplo 6: Produção de amorfa-4,11-dieno em cepas hospedeiras de *Escherichia coli* que abrigam uma pluralidade de cópias de sequências de nucleotídeo que codificam mevalonato quinase.

Cepas hospedeiras B32 e B177 foram cultivadas e a produção de amorfa-4,11-dieno analisada, conforme descrito no Exemplo 4.

A figura 10 mostra que cepas hospedeiras em que a dosagem de gene e expressão de mevalonato quinase é aumentada produzem níveis mais altos de amorfa-4,11-dieno.

Embora a presente invenção tenha sido descrita com referência às modalidades específicas da mesma, deve ser entendido pelos versados na técnica que várias mudanças podem ser feitas e equivalentes podem ser substituídos sem se afastar do verdadeira espírito e escopo da invenção. Além disso, muitas modificações podem ser feitas para adaptar uma situação particular, material, composição de matéria, processo, etapa ou etapas de proceso ao objetivo, espírito e escopo da presente invenção. Todas estas modificações devem estar dentro do escopo das reivindicações anexas.

LISTAGEM DE SEQUÊNCIA

<110> Keasling, Jay D.
Nowroozi, Farnaz
Pitera, Douglas J.
Newman, Jack D.
Anthony, Jennifer
Anthony, Larry

<120> Produção de isoprenóides e precursores de isoprenóides

<130> BERK-059WO

<150> 60/826,970
<151> 2006-09-26

<160> 22

<170> FastSEQ para windows versão 4.0

<210> 1
<211> 4247
<212> DNA
<213> Sequência artificial

<220>
<223> vetor recombinante

<400> 1
gaattcaaag gaggaaaata aaatgaagaa ctgtgtgatt gtttctgcgg tccgcacggc 60
gatcggcagc tttaacggct ctttagcgag cacctctgca atcgatctgg gtgcgacggt 120
cattaaggcc gccattgaac gcgccaaaat cgacagccag cacgttgatg aggtgatcat 180
gggcaatgtg ttacaagccg gcctgggtca aaacccagcg cgtcaagcac tgttaaaatc 240
tggtctggcc gagaccgtgt gtggcttcac cgtcaataag gtttgcggct ctggcctgaa 300
gagcgtggcc ctggcagcac aagcgattca agccggtcag gcacaaagca tcgttgcggg 360
tggcatggag aacatgtctc tggcgccgta cttattagat gccaaagccc gcagcggtta 420
tcgcctgggc gatggtcagg tgtacgacgt catcttacgc gatggcttaa tgtgcgcgac 480
ccacggttac cacatgggta ttacggccga aaacgtggcg aaagaatacg gcattacgcg 540
cgagatgcag gatgaattag cactgcactc tcagcgcaaa gcagcagccg cgatcgagtc 600
tggtgcgttt acggcggaaa tcgtgccagt taacgtggtc acgcgcaaga agacgttcgt 660
tttcagccag gacgagttcc cgaaggcaaa cagcaccgcg gaggccttag gtgccttacg 720
cccagccttt gacaaagcgg gcacggtcac cgccggtaat gcgagcggca tcaatgatgg 780
tgcagcggca ctggtcatca tggaagagag cgccgcatta gcagcgggtc tgaccccatt 840
agcgcgcatt aaatcttatg ccagcggcgg cgtcccacca gccctgatgg gcatgggtcc 900
ggtcccagcc acgcaaaaag ccctgcaatt agcgggcctg caactggccg acattgatct 960
gatcgaggcg aacgaggcgt ttgcagcgca gttcctggcg gtgggtaaga atctgggctt 1020
cgacagcgag aaagtcaatg tgaacggtgg cgcgattgcg ttaggccatc cgattggtgc 1080
aagcggcgca cgcatcttag tgacgttact gcacgccatg caggcacgcg acaagacctt 1140
aggcctggcg accttatgta ttggtggcgg tcaaggtatc gccatggtga tcgaacgcct 1200
gaactgaaga tctaggagga aagcaaaatg aaactgagca ccaagctgtg ctggtgtggc 1260
atcaagggtc gcctgcgccc acaaaagcag caacagctgc acaacacgaa cctgcaaatg 1320
accgagctga aaaagcagaa gacggccgag caaaagaccc gcccgcagaa cgttggcatc 1380
aagggcatcc agatttatat cccgacgcag tgtgtcaacc aatctgagct ggagaaattc 1440
gatggcgtca gccagggtaa gtacaccatc ggcctgggcc agaccaacat gagcttcgtg 1500
aacgaccgtg aggacatcta ttctatgagc ctgacggtgc tgtctaagct gatcaagagc 1560
tacaacatcg acacgaataa gatcggtcgt ctggaggtgg gtacggagac gctgattgac 1620
aagagcaaaa gcgtgaagtc tgtcttaatg cagctgttcg gcgagaacac ggatgtcgag 1680
ggtatcgaca ccctgaacgc gtgttacggc ggcaccaacg cactgttcaa tagcctgaac 1740
tggattgaga gcaacgcctg ggatggccgc gatgcgatcg tcgtgtgcgg cgatatcgcc 1800
atctatgaca agggtgcggc acgtccgacc ggcggtgcag gcaccgttgc gatgtggatt 1860
ggcccggacg caccaattgt cttcgattct gtccgcgcgt cttacatgga gcacgcctac 1920

gacttttaca agccggactt cacgagcgaa tacccgtacg tggacggcca cttctctctg 1980
acctgctatg tgaaggcgct ggaccaggtt tataagtctt atagcaaaaa ggcgatttct 2040
aagggcctgg tcagcgaccc ggcaggcagc gacgccctga acgtgctgaa gtatttcgac 2100
tacaacgtgt tccatgtccc gacctgcaaa ttagtgacca aatcttatgg ccgcctgtta 2160
tataatgatt tccgtgccaa cccgcagctg ttcccggagg ttgacgccga gctggcgacg 2220
cgtgattacg acgagagcct gaccgacaag aacatcgaga agaccttcgt caacgtcgcg 2280
aagccgttcc acaaagagcg tgtggcccaa agcctgatcg tcccgaccaa cacgggcaac 2340
atgtataccg cgtctgtcta cgcggcattc gcgagcctgc tgaattacgt cggttctgac 2400
gacctgcagg gcaagcgcgt tggcctgttc agctacggta gcggcttagc ggccagcctg 2460
tatagctgca aaattgtcgg cgacgtccag cacatcatca aggagctgga catcaccaac 2520
aagctggcga agcgcatcac cgagacgccg aaagattacg aggcagcgat cgagttacgc 2580
gagaatgcgc atctgaagaa gaacttcaag ccgcaaggta gcatcgagca cctgcagagc 2640
ggcgtctact acctgacgaa cattgacgac aagttccgcc gttcttatga cgtcaaaaag 2700
taactagtag gaggaaaaca tcatggtgct gacgaacaaa accgtcatta gcggcagcaa 2760
ggtgaagtct ctgagcagcg cccaaagctc tagcagcggc ccgtctagca gcagcgagga 2820
ggacgacagc cgtgacattg agtctctgga caagaagatc cgcccgctgg aggagttaga 2880
ggccctgctg agcagcggca acaccaagca gctgaagaac aaggaagttg cagcgctggt 2940
gatccacggt aagctgccac tgtatgcgct ggaaaagaaa ctgggcgata cgacgcgtgc 3000
ggtcgcggtg cgtcgcaaag ccttaagcat cttagcggag gccccggtgt tagccagcga 3060
ccgcctgccg tacaagaact acgactacga ccgcgtgttt ggcgcgtgct gcgagaatgt 3120
cattggctac atgccgttac cggttggtgt gatcggcccg ctggtcattg atggcacgag 3180
ctatcacatt ccaatggcga ccacggaagg ttgcttagtc gccagcgcca tgcgtggctg 3240
taaggcgatt aacgccggcg gtggcgcgac gaccgtgtta accaaggatg gtatgacgcg 3300
cggtccggtc gtccgcttcc caacgctgaa gcgcagcggc gcgtgtaaga tttggctgga 3360
ttctgaggag ggccaaaacg cgatcaagaa agccttcaac tctacgagcc gtttcgcgcg 3420
tttacagcat atccagacct gcctggccgg cgacctgctg ttcatgcgct tccgcaccac 3480
cacgggcgat gcgatgggca tgaacatgat cagcaagggc gtcgaatata gcctgaaaca 3540
aatggtggaa gaatatggct gggaggacat ggaggttgtc tctgtgagcg gcaactattg 3600
caccgacaag aagccggcag ccattaactg gattgagggt cgcggcaaaa gcgtcgtggc 3660
agaagcgacc atcccaggcg acgtggtccg taaggttctg aagagcgacg tcagcgccct 3720
ggttgagtta aatatcgcga aaaacctggt cggcagcgcg atggcgggca gcgtgggtgg 3780
ctttaacgca catgcagcga atctggttac ggcggttttc ttagccttag gtcaggaccc 3840
agcccaaaat gtcgagagca gcaactgcat taccttaatg aaagaggttg acggtgacct 3900
gcgcatcagc gtttctatgc cgtctatcga ggtcggcacg atcggcggcg gcaccgtttt 3960
agaaccgcaa ggtgcgatgc tggatctgct gggcgtgcgc ggcccacatg caacggcccc 4020
aggcaccaat gcccgccaac tggcccgtat cgtggcctgc gcggttctgg cgggtgagct 4080
gagcctgtgc gccgcattag ccgcgggcca tttagttcaa tctcacatga cccacaaccg 4140
caagccggca gaaccaacca agccaaataa cctggacgca accgacatta accgtctgaa 4200
ggatggcagc gtcacgtgca ttaaaagctg agcatgctac taagctt 4247

<210> 2
<211> 2174
<212> DNA

<213> Sequência artificial

<220>

<223> vetor recombinante

<400> 2
gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg 60
ccctttcgtc ttcacctcga gaggctttac actttatgct tccggctcgt ataatgtgtg 120
gaattgtgag cggataacaa ttgaattcta cggatccgat gtcgactcaa agcttcgatc 180
ccatggtacg cgtgctagag gcatcaaata aaacgaaagg ctcagtcgaa agactgggcc 240
tttcgtttta tctgttgttt gtcggtgaac gctctcctga gtaggacaaa tccgccgccc 300
tagacctagg gatatattcc gcttcctcgc tcactgactc gctacgctcg gtcgttcgac 360
tgcggcgagc ggaaatggct tacgaacggg gcggagattt cctggaagat gccaggaaga 420
tacttaacag ggaagtgaga gggccgcggc aaagccgttt ttccataggc tccgcccccc 480
tgacaagcat cacgaaatct gacgctcaaa tcagtggtgg cgaaacccga caggactata 540
aagataccag gcgtttcccc ctggcggctc cctcgtgcgc tctcctgttc ctgcctttcg 600
gtttaccggt gtcattccgc tgttatggcc gcgtttgtct cattccacgc ctgacactca 660
gttccgggta ggcagttcgc tccaagctgg actgtatgca cgaacccccc gttcagtccg 720
accgctgcgc cttatccggt aactatcgtc ttgagtccaa cccggaaaga catgcaaaag 780
caccactggc agcagccact ggtaattgat ttagaggagt tagtcttgaa gtcatgcgcc 840

ggttaaggct aaactgaaag gacaagtttt ggtgactgcg ctcctccaag ccagttacct 900
cggttcaaag agttggtagc tcagagaacc ttcgaaaaac cgccctgcaa ggcggttttt 960
tcgttttcag agcaagagat tacgcgcaga ccaaaacgat ctcaagaaga tcatcttatt 1020
aatcagataa aatatttcta gatttcagtg caatttatct cttcaaatgt agcacctgaa 1080
gtcagcccca tacgatataa gttgttacta gtgcttggat tctcaccaat aaaaaacgcc 1140
cggcggcaac cgagcgttct gaacaaatcc agatggagtt ctgaggtcat tactggatct 1200
atcaacagga gtccaagcga gctctcgaac cccagagtcc cgctcagaag aactcgtcaa 1260
gaaggcgata gaaggcgatg cgctgcgaat cgggagcggc gataccgtaa agcacgagga 1320
agcggtcagc ccattcgccg ccaagctctt cagcaatatc acgggtagcc aacgctatgt 1380
cctgatagcg gtccgccaca cccagccggc cacagtcgat gaatccagaa aagcggccat 1440
tttccaccat gatattcggc aagcaggcat cgccatgggt cacgacgaga tcctcgccgt 1500
cgggcatgcg cgccttgagc ctggcgaaca gttcggctgg cgcgagcccc tgatgctctt 1560
cgtccagatc atcctgatcg acaagaccgg cttccatccg agtacgtgct cgctcgatgc 1620
gatgtttcgc ttggtggtcg aatgggcagg tagccggatc aagcgtatgc agccgccgca 1680
ttgcatcagc catgatggat actttctcgg caggagcaag gtgagatgac aggagatcct 1740
gccccggcac ttcgcccaat agcagccagt ccctttcccgc ttcagtgaca acgtcgagca 1800
cagctgcgca aggaacgccc gtcgtggcca gccacgatag ccgcgctgcc tcgtcctgca 1860
gttcattcag ggcaccggac aggtcggtct tgacaaaaag aaccgggcgc ccctgcgctg 1920
acagccggaa cacggcggca tcagagcagc cgattgtctg ttgtgcccag tcatagccga 1980
atagcctctc cacccaagcg gccggagaac ctgcgtgcaa tccatcttgt tcaatcatgc 2040
gaaacgatcc tcatcctgtc tcttgatcag atcttgatcc cctgcgccat cagatccttg 2100
gcggcaagaa agccatccag tttactttgc agggcttccc aaccttacca gagggcgccc 2160
cagctggcaa ttcc 2174

<210> 3
<211> 12707
<212> DNA

<213> Sequência artificial

<220>

<223> plasmídeo recombinante

<400> 3
gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg 60
ccctttcgtc ttcacctcga gaggctttac actttatgct tccggctcgt ataatgtgtg 120
gaattgtgag cggataacaa ttgaattcaa aggaggaaaa taaaatgaag aactgtgtga 180
ttgtttctgc ggtccgcacg gcgatcggca gctttaacgg ctctttagcg agcacctctg 240
caatcgatct gggtgcgacg gtcattaagg ccgccattga acgcgccaaa atcgacagcc 300
agcacgttga tgaggtgatc atgggcaatg tgttacaagc cggcctgggt caaaacccag 360
cgcgtcaagc actgttaaaa tctggtctgg ccgagaccgt gtgtggcttc accgtcaata 420
aggtttgcgg ctctggcctg aagagcgtgg ccctggcagc acaagcgatt caagccggtc 480
aggcacaaag catcgttgcg ggtggcatgg agaacatgtc tctggcgccg tacttattag 540
atgccaaagc ccgcagcggt tatcgcctgg gcgatggtca ggtgtacgac gtcatcttac 600
gcgatggctt aatgtgcgcg acccacggtt accacatggg tattacggcc gaaaacgtgg 660
cgaaagaata cggcattacg cgcgagatgc aggatgaatt agcactgcac tctcagcgca 720
aagcagcagc cgcgatcgag tctggtgcgt ttacggcgga aatcgtgcca gttaacgtgg 780
tcacgcgcaa gaagacgttc gttttcagcc aggacgagtt cccgaaggca aacagcaccg 840
cggaggcctt aggtgcctta cgcccagcct ttgacaaagc gggcacggtc accgccggta 900
atgcgagcgg catcaatgat ggtgcagcgg cactggtcat catggaagag agcgccgcat 960
tagcagcggg tctgacccca ttagcgcgca ttaaatctta tgccagcggc ggcgtcccac 1020
cagccctgat gggcatgggt ccggtcccag ccacgcaaaa agccctgcaa ttagcgggcc 1080
tgcaactggc cgacattgat ctgatcgagg cgaacgaggc gtttgcagcg cagttcctgg 1140
cggtgggtaa gaatctgggc ttcgacagcg agaaagtcaa tgtgaacggt ggcgcgattg 1200
cgttaggcca tccgattggt gcaagcggcg cacgcatctt agtgacgtta ctgcacgcca 1260
tgcaggcacg cgacaagacc ttaggcctgg cgaccttatg tattggtggc ggtcaaggta 1320
tcgccatggt gatcgaacgc ctgaactgaa gatctaggag gaaagcaaaa tgaaactgag 1380
caccaagctg tgctggtgtg gcatcaaggg tcgcctgcgc ccacaaaagc agcaacagct 1440
gcacaacacg aacctgcaaa tgaccgagct gaaaaagcag aagacggccg agcaaaagac 1500
ccgcccgcag aacgttggca tcaagggcat ccagatttat atcccgacgc agtgtgtcaa 1560
ccaatctgag ctggagaaat tcgatggcgt cagccagggt aagtacacca tcggcctggg 1620
ccagaccaac atgagcttcg tgaacgaccg tgaggacatc tattctatga gcctgacggt 1680
gctgtctaag ctgatcaaga gctacaacat cgacacgaat aagatcggtc gtctggaggt 1740
gggtacggag acgctgattg acaagagcaa aagcgtgaag tctgtcttaa tgcagctgtt 1800

```
cggcgagaac acggatgtcg agggtatcga caccctgaac gcgtgttacg gcggcaccaa 1860
cgcactgttc aatagcctga actggattga gagcaacgcc tgggatggcc gcgatgcgat 1920
cgtcgtgtgc ggcgatatcg ccatctatga caagggtgcg gcacgtccga ccggcggtgc 1980
aggcaccgtt gcgatgtgga ttggcccgga cgcaccaatt gtcttcgatt ctgtccgcgc 2040
gtcttacatg gagcacgcct acgactttta caagccggac ttcacgagcg aatacccgta 2100
cgtggacggc cacttctctc tgacctgcta tgtgaaggcg ctggaccagg tttataagtc 2160
ttatagcaaa aaggcgattt ctaagggcct ggtcagcgac ccggcaggca gcgacgccct 2220
gaacgtgctg aagtatttcg actacaacgt gttccatgtc ccgacctgca aattagtgac 2280
caaatcttat ggccgcctgt tatataatga tttccgtgcc aacccgcagc tgttcccgga 2340
ggttgacgcc gagctggcga cgcgtgatta cgacgagagc ctgaccgaca agaacatcga 2400
gaagaccttc gtcaacgtcg cgaagccgtt ccacaaagag cgtgtggccc aaagcctgat 2460
cgtcccgacc aacacgggca acatgtatac cgcgtctgtc tacgcggcat tcgcgagcct 2520
gctgaattac gtcggttctg acgacctgca gggcaagcgc gttggcctgt tcagctacgg 2580
tagcggctta gcggccagcc tgtatagctg caaaattgtc ggcgacgtcc agcacatcat 2640
caaggagctg gacatcacca acaagctggc gaagcgcatc accgagacgc cgaaagatta 2700
cgaggcagcg atcgagttac gcgagaatgc gcatctgaag aagaacttca agccgcaagg 2760
tagcatcgag cacctgcaga gcggcgtcta ctacctgacg aacattgacg acaagttccg 2820
ccgttcttat gacgtcaaaa agtaactagt aggaggaaaa catcatggtg ctgacgaaca 2880
aaaccgtcat tagcggcagc aaggtgaagt ctctgagcag cgcccaaagc tctagcagcg 2940
gcccgtctag cagcagcgag gaggacgaca gccgtgacat tgagtctctg gacaagaaga 3000
tccgcccgct ggaggagtta gaggccctgc tgagcagcgg caacaccaag cagctgaaga 3060
acaaggaagt tgcagcgctg gtgatccacg gtaagctgcc actgtatgcg ctggaaaaga 3120
aactgggcga tacgacgcgt gcggtcgcgg tgcgtcgcaa agccttaagc atcttagcgg 3180
aggccccggt gttagccagc gaccgcctgc cgtacaagaa ctacgactac gaccgcgtgt 3240
ttggcgcgtg ctgcgagaat gtcattggct acatgccgtt accggttggt gtgatcggcc 3300
cgctggtcat tgatggcacg agctatcaca ttccaatggc gaccacggaa ggttgcttag 3360
tcgccagcgc catgcgtggc tgtaaggcga ttaacgccgg cggtggcgcg acgaccgtgt 3420
taaccaagga tggtatgacg cgcggtccgg tcgtccgctt cccaacgctg aagcgcagcg 3480
gcgcgtgtaa gatttggctg gattctgagg agggccaaaa cgcgatcaag aaagccttca 3540
actctacgag ccgtttcgcg cgtttacagc atatccagac ctgcctggcc ggcgacctgc 3600
tgttcatgcg cttccgcacc accacgggcg atgcgatggg catgaacatg atcagcaagg 3660
gcgtcgaata tagcctgaaa caaatggtgg aagaatatgg ctgggaggac atggaggttg 3720
tctctgtgag cggcaactat tgcaccgaca agaagccggc agccattaac tggattgagg 3780
gtcgcggcaa aagcgtcgtg gcagaagcga ccatcccagg cgacgtggtc cgtaaggttc 3840
tgaagagcga cgtcagcgcc ctggttgagt taaatatcgc gaaaaacctg gtcggcagcg 3900
cgatggcggg cagcgtgggt ggctttaacg cacatgcagc gaatctggtt acggcggttt 3960
tcttagcctt aggtcaggac ccagcccaaa atgtcgagag cagcaactgc attaccttaa 4020
tgaaagaggt tgacggtgac ctgcgcatca gcgtttctat gccgtctatc gaggtcggca 4080
cgatcggcgg cggcaccgtt ttagaaccgc aaggtgcgat gctggatctg ctgggcgtgc 4140
gcggcccaca tgcaacggcc ccaggcacca atgcccgcca actggcccgt atcgtggcct 4200
gcgcggttct ggcgggtgag ctgagcctgt gcgccgcatt agccgcgggc catttagttc 4260
aatctcacat gacccacaac cgcaagccgg cagaaccaac caagccaaat aacctggacg 4320
caaccgacat taaccgtctg aaggatggca gcgtcacgtg cattaaaagc tgagcatgct 4380
actaagcttc gatcccatgg tacgcgtgct agaggcatca aataaaacga aaggctcagt 4440
cgaaagactg ggcctttcgt tttatctgtt gtttgtcggt gaacgctctc ctgagtagga 4500
caaatccgcc gccctagacc taggggatat attccgcttc ctcgctcact gactcgctac 4560
gctcggtcgt tcgactgcgg cgagcggaaa tggcttacga acggggcgga gatttcctgg 4620
aagatgccag gaagatactt aacagggaag tgagagggcc gcggcaaagc cgtttttcca 4680
taggctccgc cccctgaca agcatcacga aatctgacgc tcaaatcagt ggtggcgaaa 4740
cccgacagga ctataaagat accaggcgtt tccccctggc ggctccctcg tgcgctctcc 4800
tgttcctgcc tttcggttta ccggtgtcat tccgctgtta tggccgcgtt tgtctcattc 4860
cacgcctgac actcagttcc gggtaggcag ttcgctccaa gctggactgt atgcacgaac 4920
cccccgttca gtccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg 4980
aaagacatgc aaaagcacca ctggcagcag ccactggtaa ttgatttaga ggagttagtc 5040
ttgaagtcat gcgccggtta aggctaaact gaaaggacaa gttttggtga ctgcgctcct 5100
ccaagccagt tacctcggtt caaagagttg gtagctcaga gaaccttcga aaaaccgccc 5160
tgcaaggcgg ttttttcgtt ttcagagcaa gagattacgc gcagaccaaa acgatctcaa 5220
gaagatcatc ttattaatca gataaaatat ttctagattt cagtgcaatt tatctcttca 5280
aatgtagcac ctgaagtcag ccccatacga tataagttgt tactagtgct tggattctca 5340
ccaataaaaa acgcccggcg gcaaccgagc gttctgaaca aatccagatg gagttctgag 5400
gtcattactg gatctatcaa caggagtcca agcgagctcg atatcaaatt acgccccgcc 5460
ctgccactca tcgcagtact gttgtaattc attaagcatt ctgccgacat ggaagccatc 5520
acagacggca tgatgaacct gaatcgccag cggcatcagc accttgtcgc cttgcgtata 5580
```

atatttgccc atggtgaaaa cgggggcgaa gaagttgtcc atattggcca cgtttaaatc 5640
aaaactggtg aaactcaccc agggattggc tgagacgaaa aacatattct caataaaccc 5700
tttagggaaa taggccaggt tttcaccgta acacgccaca tcttgcgaat atatgtgtag 5760
aaactgccgg aaatcgtcgt ggtattcact ccagagcgat gaaaacgttt cagtttgctc 5820
atggaaaacg gtgtaacaag ggtgaacact atcccatatc accagctcac cgtctttcat 5880
tgccatacgg aattccggat gagcattcat caggcgggca agaatgtgaa taaaggccgg 5940
ataaaacttg tgcttatttt tctttacggt ctttaaaaag gccgtaatat ccagctgaac 6000
ggtctggtta taggtacatt gagcaactga ctgaaatgcc tcaaaatgtt ctttacgatg 6060
ccattgggat atatcaacgg tggtatatcc agtgattttt ttctccattt tagcttcctt 6120
agctcctgaa aatctcgata actcaaaaaa tacgcccggt agtgatctta tttcattatg 6180
gtgaaagttg gaacctctta cgtgccgatc aacgtctcat tttcgccaga tatcgacgtc 6240
tttacacttt atgcttccgg ctcgtataat gagcacatca gcaggacgca ctgaccgaat 6300
taggaggggt accgggccca tgtcattacc gttcttaact tctgcaccgg gaaaggttat 6360
tatttttggt gaacactctg ctgtgtacaa caagcctgcc gtcgctgcta gtgtgtctgc 6420
gttgagaacc tacctgctaa taagcgagtc atctgcacca gatactattg aattggactt 6480
cccggacatt agctttaatc ataagtggtc catcaatgat ttcaatgcca tcaccgagga 6540
tcaagtaaac tcccaaaaat tggccaaggc tcaacaagcc accgatggct tgtctcagga 6600
actcgttagt cttttggatc cgttgttagc tcaactatcc gaatccttcc actaccatgc 6660
agcgttttgt ttcctgtata tgtttgtttg cctatgcccc catgccaaga atattaagtt 6720
ttctttaaag tctactttac ccatcggtgc tgggttgggc tcaagcgcct ctatttctgt 6780
atcactggcc ttagctatgg cctacttggg ggggttaata ggatctaatg acttggaaaa 6840
gctgtcagaa aacgataagc atatagtgaa tcaatgggcc ttcataggtg aaaagtgtat 6900
tcacggtacc ccttcaggaa tagataacgc tgtggccact tatggtaatg ccctgctatt 6960
tgaaaaagac tcacataatg gaacaataaa cacaaacaat tttaagttct tagatgattt 7020
cccagccatt ccaatgatcc taacctatac tagaattcca aggtctacaa aagatcttgt 7080
tgctcgcgtt cgtgtgttgg tcaccgagaa atttcctgaa gttatgaagc caattctaga 7140
tgccatgggt gaatgtgccc tacaaggctt agagatcatg actaagttaa gtaaatgtaa 7200
aggcaccgat gacgaggctg tagaaactaa taatgaactg tatgaacaac tattggaatt 7260
gataagaata aatcatggac tgcttgtctc aatcggtgtt tctcatcctg gattagaact 7320
tattaaaaat ctgagcgatg atttgagaat tggctccaca aaacttaccg gtgctggtgg 7380
cggcggttgc tctttgactt tgttacgaag agacattact caagagcaaa ttgacagctt 7440
caaaaagaaa ttgcaagatg attttagtta cgagacattt gaaacagact tgggtgggac 7500
tggctgctgt ttgttaagcg caaaaaattt gaataaagat cttaaaatca aatccctagt 7560
attccaatta tttgaaaata aaactaccac aaagcaacaa attgacgatc tattattgcc 7620
aggaaacacg aatttaccat ggacttcata ggaggcagat caaatgtcag agttgagagc 7680
cttcagtgcc ccagggaaag cgttactagc tggtggatat ttagttttag atacaaaata 7740
tgaagcattt gtagtcggat tatcggcaag aatgcatgct gtagcccatc cttacggttc 7800
attgcaaggg tctgataagt ttgaagtgcg tgtgaaaagt aaacaattta aagatgggga 7860
gtggctgtac catataagtc ctaaaagtgg cttcattcct gtttcgatag gcggatctaa 7920
gaacccttc attgaaaaag ttatcgctaa cgtatttagc tactttaaac ctaacatgga 7980
cgactactgc aatagaaact tgttcgttat tgatattttc tctgatgatg cctaccattc 8040
tcaggaggat agcgttaccg aacatcgtgg caacagaaga ttgagttttc attcgcacag 8100
aattgaagaa gttcccaaaa cagggctggg ctcctcggca ggtttagtca cagttttaac 8160
tacagctttg gcctcctttt ttgtatcgga cctggaaaat aatgtagaca aatatagaga 8220
agttattcat aatttagcac aagttgctca ttgtcaagct cagggtaaaa ttggaagcgg 8280
gtttgatgta gcggcggcag catatggatc tatcagatat agaagattcc cacccgcatt 8340
aatctctaat ttgccagata ttggaagtgc tacttacggc agtaaactgg cgcatttggt 8400
tgatgaagaa gactggaata ttacgattaa aagtaaccat ttaccttcgg gattaacttt 8460
atggatgggc gatattaaga atggttcaga aacagtaaaa ctggtccaga aggtaaaaaa 8520
ttggtatgat tcgcatatgc cagaaagctt gaaaatatat acagaactcg atcatgcaaa 8580
ttctagattt atggatggac tatctaaact agatcgctta cacgagactc atgacgatta 8640
cagcgatcag atatttgagt ctcttgagag gaatgactgt acctgtcaaa agtatcctga 8700
aatcacagaa gttagagatg cagttgccac aattagacgt tcctttagaa aaataactaa 8760
agaatctggt gccgatatcg aacctcccgt acaaactagc ttattggatg attgccagac 8820
cttaaaagga gttcttactt gcttaatacc tggtgctggt ggttatgacg ccattgcagt 8880
gattactaag caagatgttg atcttagggc tcaaaccgct aatgacaaaa gattttctaa 8940
ggttcaatgg ctggatgtaa ctcaggctga ctggggtgtt aggaaagaaa aagatccgga 9000
aacttatctt gataaatagg aggtaatact catgaccgtt tacacagcat ccgttaccgc 9060
acccgtcaac atcgcaaccc ttaagtattg ggggaaaagg gacacgaagt tgaatctgcc 9120
caccaattcg tccatatcag tgactttatc gcaagatgac ctcagaacgt tgacctctgc 9180
ggctactgca cctgagtttg aacgcgacac tttgtggtta aatggagaac cacacagcat 9240
cgacaatgaa agaactcaaa attgtctgcg cgacctacgc caattaagaa aggaaatgga 9300
atcgaaggac gcctcattgc ccacattatc tcaatggaaa ctccacattg tctccgaaaa 9360

```
taactttcct acagcagctg gtttagcttc ctccgctgct ggctttgctg cattggtctc 9420
tgcaattgct aagttatacc aattaccaca gtcaacttca gaaatatcta gaatagcaag 9480
aaaggggtct ggttcagctt gtagatcgtt gtttggcgga tacgtggcct gggaaatggg 9540
aaaagctgaa gatggtcatg attccatggc agtacaaatc gcagacagct ctgactggcc 9600
tcagatgaaa gcttgtgtcc tagttgtcag cgatattaaa aaggatgtga gttccactca 9660
gggtatgcaa ttgaccgtgg caacctccga actatttaaa gaaagaattg aacatgtcgt 9720
accaaagaga tttgaagtca tgcgtaaagc cattgttgaa aaagatttcg ccacctttgc 9780
aaaggaaaca atgatggatt ccaactcttt ccatgccaca tgtttggact ctttccctcc 9840
aatattctac atgaatgaca cttccaagcg tatcatcagt tggtgccaca ccattaatca 9900
gttttacgga gaaacaatcg ttgcatacac gtttgatgca ggtccaaatg ctgtgttgta 9960
ctacttagct gaaaatgagt cgaaactctt tgcatttatc tataaattgt ttggctctgt 10020
tcctggatgg gacaagaaat ttactactga gcagcttgag gctttcaacc atcaatttga 10080
atcatctaac tttactgcac gtgaattgga tcttgagttg caaaaggatg ttgccagagt 10140
gattttaact caagtcggtt caggcccaca agaaacaaac gaatctttga ttgacgcaaa 10200
gactggtcta ccaaaggaat aactgcagcc cgggaggagg attactatat gcaaacggaa 10260
cacgtcattt tattgaatgc acagggagtt cccacgggta cgctggaaaa gtatgccgca 10320
cacacggcag acacccgctt acatctcgcg ttctccagtt ggctgtttaa tgccaaagga 10380
caattattag ttacccgccg cgcactgagc aaaaaagcat ggcctggcgt gtggactaac 10440
tcggtttgtg ggcacccaca actgggagaa agcaacgaag acgcagtgat ccgccgttgc 10500
cgttatgagc ttggcgtgga aattacgcct cctgaatcta tctatcctga ctttcgctac 10560
cgcgccaccg atccgagtgg cattgtggaa aatgaagtgt gtccggtatt tgccgcacgc 10620
accactagtg cgttacagat caatgatgat gaagtgatgg attatcaatg gtgtgattta 10680
gcagatgtat tacacggtat tgatgccacg ccgtgggcgt tcagtccgtg gatggtgatg 10740
caggcgacaa atcgcgaagc cagaaaacga ttatctgcat ttacccagct taaataaccc 10800
gggggatcca ctagttctag agcggccgcc accgcggagg aggaatgagt aatggacttt 10860
ccgcagcaac tcgaagcctg cgttaagcag gccaaccagg cgctgagccg ttttatcgcc 10920
ccactgccct ttcagaacac tcccgtggtc gaaaccatgc agtatggcgc attattaggt 10980
ggtaagcgcc tgcgaccttt cctggtttat gccaccggtc atatgttcgg cgttagcaca 11040
aacacgctgg acgcacccgc tgccgccgtt gagtgtatcc acgcttactc attaattcat 11100
gatgatttac cggcaatgga tgatgacgat ctgcgtcgcg gtttgccaac ctgccatgtg 11160
aagtttggcg aagcaaacgc gattctcgct ggcgacgctt tacaaacgct ggcgttctcg 11220
attttaagcg atgccgatat gccggaagtg tcggaccgcg acagaatttc gatgatttct 11280
gaactggcga gcgccagtgg tattgccgga atgtgcggtg gtcaggcatt agatttagac 11340
gcggaaggca aacacgtacc tctggacgcg cttgagcgta ttcatcgtca taaaaccggc 11400
gcattgattc gcgccgccgt tcgccttggt gcattaagcg ccggagataa aggacgtcgt 11460
gctctgccgg tactcgacaa gtatgcagag agcatcggcc ttgccttcca ggttcaggat 11520
gacatcctgg atgtggtggg agatactgca acgttgggaa aacgccaggg tgccgaccag 11580
caacttggta aaagtaccta ccctgcactt ctgggtcttg agcaagcccg gaagaaagcc 11640
cgggatctga tcgacgatgc ccgtcagtcg ctgaaacaac tggctgaaca gtcactcgat 11700
acctcggcac tggaagcgct agcggactac atcatccagc gtaataaata agagctctcg 11760
aaccccagag tcccgctcag aagaactcgt caagaaggcg atagaaggcg atgcgctgcg 11820
aatcgggagc ggcgataccg taaagcacga ggaagcggtc agcccattcg ccgccaagct 11880
cttcagcaat atcacgggta gccaacgcta tgtcctgata gcggtccgcc acacccagcc 11940
ggccacagtc gatgaatcca gaaaagcggc cattttccac catgatattc ggcaagcagg 12000
catcgccatg ggtcacgacg agatcctcgc cgtcgggcat gcgcgccttg agcctggcga 12060
acagttcggc tggcgcgagc ccctgatgct cttcgtccag atcatcctga tcgacaagac 12120
cggcttccat ccgagtacgt gctcgctcga tgcgatgttt cgcttggtgg tcgaatgggc 12180
aggtagccgg atcaagcgta tgcagccgcc gcattgcatc agccatgatg gatactttct 12240
cggcaggagc aaggtgagat gacaggagat cctgccccgg cacttcgccc aatagcagcc 12300
agtcccttcc cgcttcagtg acaacgtcga gcacagctgc gcaaggaacg cccgtcgtgg 12360
ccagccacga tagccgcgct gcctcgtcct gcagttcatt cagggcaccg gacaggtcgg 12420
tcttgacaaa aagaaccggg cgcccctgcg ctgacagccg gaacacggcg gcatcagagc 12480
agccgattgt ctgttgtgcc cagtcatagc cgaatagcct ctccacccaa gcggccggag 12540
aacctgcgtg caatccatct tgttcaatca tgcgaaacga tcctcatcct gtctcttgat 12600
cagatcttga tcccctgcgc catcagatcc ttggcggcaa gaaagccatc cagtttactt 12660
tgcagggctt cccaacctta ccagagggcg ccccagctgg caattcc               12707
```

<210> 4
<211> 12707
<212> DNA

<213> Sequência artificial

<220>

<223> vetor recombinante

<400> 4
gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg 60
ccctttcgtc ttcacctcga gaggctttac actttatgct tccggctcgt ataatgtgtg 120
gaattgtgag cggataacaa ttgaattcaa aggaggaaaa taaaatgaag aactgtgtga 180
ttgtttctgc ggtccgcacg gcgatcggca gctttaacgg ctctttagcg agcacctctg 240
caatcgatct gggtgcgacg gtcattaagg ccgccattga acgcgccaaa atcgacagcc 300
agcacgttga tgaggtgatc atgggcaatg tgttacaagc cggcctgggt caaaacccag 360
cgcgtcaagc actgttaaaa tctggtctgg ccgagaccgt gtgtggcttc accgtcaata 420
aggtttgcgg ctctggcctg aagagcgtgg ccctggcagc acaagcgatt caagccggtc 480
aggcacaaag catcgttgcg ggtggcatgg agaacatgtc tctggcgccg tacttattag 540
atgccaaagc ccgcagcggt tatcgcctgg gcgatggtca ggtgtacgac gtcatcttac 600
gcgatggctt aatgtgcgcg acccacggtt accacatggg tattacggcc gaaaacgtgg 660
cgaaagaata cggcattacg cgcgagatgc aggatgaatt agcactgcac tctcagcgca 720
aagcagcagc cgcgatcgag tctggtgcgt ttacggcgga aatcgtgcca gttaacgtgg 780
tcacgcgcaa gaagacgttc gttttcagcc aggacgagtt cccgaaggca aacagcaccg 840
cggaggcctt aggtgcctta cgcccagcct ttgacaaagc gggcacggtc accgccggta 900
atgcgagcgg catcaatgat ggtgcagcgg cactggtcat catggaagag agcgccgcat 960
tagcagcggg tctgacccca ttagcgcgca ttaaatctta tgccagcggc ggcgtcccac 1020
cagccctgat gggcatgggt ccggtcccag ccacgcaaaa agccctgcaa ttagcgggcc 1080
tgcaactggc cgacattgat ctgatcgagg cgaacgaggc gtttgcagcg cagttcctgg 1140
cggtgggtaa gaatctgggc ttcgacagcg agaaagtcaa tgtgaacggt ggcgcgattg 1200
cgttaggcca tccgattggt gcaagcggcg cacgcatctt agtgacgtta ctgcacgcca 1260
tgcaggcacg cgacaagacc ttaggcctgg cgaccttatg tattggtggc ggtcaaggta 1320
tcgccatggt gatcgaacgc ctgaactgaa gatctaggag gaaagcaaaa tgaaactgag 1380
caccaagctg tgctggtgtg gcatcaaggg tcgcctgcgc ccacaaaagc agcaacagct 1440
gcacaacacg aacctgcaaa tgaccgagct gaaaaagcag aagacggccg agcaaaagac 1500
ccgcccgcag aacgttggca tcaagggcat ccagatttat atcccgacgc agtgtgtcaa 1560
ccaatctgag ctggagaaat tcgatggcgt cagccagggt aagtacacca tcggcctggg 1620
ccagaccaac atgagcttcg tgaacgaccg tgaggacatc tattctatga gcctgacggt 1680
gctgtctaag ctgatcaaga gctacaacat cgacacgaat aagatcggtc gtctggaggt 1740
gggtacggag acgctgattg acaagagcaa aagcgtgaag tctgtcttaa tgcagctgtt 1800
cggcgagaac acggatgtcg agggtatcga caccctgaac gcgtgttacg gcggcaccaa 1860
cgcactgttc aatagcctga actggattga gagcaacgcc tgggatggcc gcgatgcgat 1920
cgtcgtgtgc ggcgatatcg ccatctatga caagggtgcg gcacgtccga ccggcggtgc 1980
aggcaccgtt gcgatgtgga ttggcccgga cgcaccaatt gtcttcgatt ctgtccgcgc 2040
gtcttacatg gagcacgcct acgactttta caagccggac ttcacgagcg aatacccgta 2100
cgtggacggc cacttctctc tgacctgcta tgtgaaggcg ctggaccagg tttataagtc 2160
ttatagcaaa aaggcgattt ctaagggcct ggtcagcgac ccggcaggca gcgacgccct 2220
gaacgtgctg aagtatttcg actacaacgt gttccatgtc ccgacctgca aattagtgac 2280
caaatcttat ggccgcctgt tatataatga tttccgtgcc aacccgcagc tgttcccgga 2340
ggttgacgcc gagctggcga cgcgtgatta cgacgagagc ctgaccgaca agaacatcga 2400
gaagaccttc gtcaacgtcg cgaagccgtt ccacaaagag cgtgtggccc aaagcctgat 2460
cgtcccgacc aacacgggca acatgtatac cgcgtctgtc tacgcggcat tcgcgagcct 2520
gctgaattac gtcggttctg acgacctgca gggcaagcgc gttggcctgt tcagctacgg 2580
tagcggctta gcggccagcc tgtatagctg caaaattgtc ggcgacgtcc agcacatcat 2640
caaggagctg gacatcacca acaagctggc gaagcgcatc accgagacgc cgaaagatta 2700
cgaggcagcg atcgagttac gcgagaatgc gcatctgaag aagaacttca agccgcaagg 2760
tagcatcgag cacctgcaga gcggcgtcta ctacctgacg aacattgacg acaagttccg 2820
ccgttcttat gacgtcaaaa agtaactagt aggaggaaaa catcatggtg ctgacgaaca 2880
aaaccgtcat tagcggcagc aaggtgaagt ctctgagcag cgcccaaagc tctagcagcg 2940
gcccgtctag cagcagcgag gaggacgaca gccgtgacat tgagtctctg gacaagaaga 3000
tccgcccgct ggaggagtta gaggccctgc tgagcagcgg caacaccaag cagctgaaga 3060
acaaggaagt tgcagcgctg gtgatccacg gtaagctgcc actgtatgcg ctggaaaaga 3120
aactgggcga tacgacgcgt gcggtcgcgg tgcgtcgcaa agccttaagc atcttagcgg 3180
aggccccggt gttagccagc gaccgcctgc cgtacaagaa ctacgactac gaccgcgtgt 3240
ttggcgcgtg ctgcgagaat gtcattggct acatgccgtt accggttggt gtgatcggcc 3300
cgctggtcat tgatggcacg agctatcaca ttccaatggc gaccacggaa ggttgcttag 3360
tcgccagcgc catgcgtggc tgtaaggcga ttaacgccgg cggtggcgcg acgaccgtgt 3420
taaccaagga tggtatgacg cgcggtccgg tcgtccgctt cccaacgctg aagcgcagcg 3480

```
gcgcgtgtaa gatttggctg gattctgagg agggccaaaa cgcgatcaag aaagccttca 3540
actctacgag ccgtttcgcg cgtttacagc atatccagac ctgcctggcc ggcgacctgc 3600
tgttcatgcg cttccgcacc accacgggcg atgcgatggg catgaacatg atcagcaagg 3660
gcgtcgaata tagcctgaaa caaatggtgg aagaatatgg ctgggaggac atggaggttg 3720
tctctgtgag cggcaactat tgcaccgaca agaagccggc agccattaac tggattgagg 3780
gtcgcggcaa aagcgtcgtg gcagaagcga ccatcccagg cgacgtggtc cgtaaggttc 3840
tgaagagcga cgtcagcgcc ctggttgagt taaatatcgc gaaaaacctg gtcggcagcg 3900
cgatggcggg cagcgtgggt ggctttaacg cacatgcagc gaatctggtt acggcggttt 3960
tcttagcctt aggtcaggac ccagcccaaa atgtcgagag cagcaactgc attaccttaa 4020
tgaaagaggt tgacggtgac ctgcgcatca gcgtttctat gccgtctatc gaggtcggca 4080
cgatcggcgg cggcaccgtt ttagaaccgc aaggtgcgat gctggatctg ctgggcgtgc 4140
gcggcccaca tgcaacggcc ccaggcacca atgcccgcca actggcccgt atcgtggcct 4200
gcgcggttct ggcgggtgag ctgagcctgt gcgccgcatt agccgcgggc catttagttc 4260
aatctcacat gacccacaac cgcaagccgg cagaaccaac caagccaaat aacctggacg 4320
caaccgacat taaccgtctg aaggatggca gcgtcacgtg cattaaaagc tgagcatgct 4380
actaagcttc gatcccatgg tacgcgtgct agaggcatca aataaaacga aaggctcagt 4440
cgaaagactg ggcctttcgt tttatctgtt gtttgtcggt gaacgctctc ctgagtagga 4500
caaatccgcc gccctagacc taggggatat attccgcttc ctcgctcact gactcgctac 4560
gctcggtcgt tcgactgcgg cgagcggaaa tggcttacga acggggcgga gatttcctgg 4620
aagatgccag gaagatactt aacagggaag tgagagggcc gcggcaaagc cgtttttcca 4680
taggctccgc cccctgaca agcatcacga aatctgacgc tcaaatcagt ggtggcgaaa 4740
cccgacagga ctataaagat accaggcgtt tccccctggc ggctccctcg tgcgctctcc 4800
tgttcctgcc tttcggttta ccggtgtcat tccgctgtta tggccgcgtt tgtctcattc 4860
cacgcctgac actcagttcc gggtaggcag ttcgctccaa gctggactgt atgcacgaac 4920
cccccgttca gtccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg 4980
aaagacatgc aaaagcacca ctggcagcag ccactggtaa ttgatttaga ggagttagtc 5040
ttgaagtcat gcgccggtta aggctaaact gaaaggacaa gttttggtga ctgcgctcct 5100
ccaagccagt tacctcggtt caaagagttg gtagctcaga gaaccttcga aaaaccgccc 5160
tgcaaggcgg ttttttcgtt ttcagagcaa gagattacgc gcagaccaaa acgatctcaa 5220
gaagatcatc ttattaatca gataaaatat ttctagattt cagtgcaatt tatctcttca 5280
aatgtagcac ctgaagtcag ccccatacga tataagttgt tactagtgct tggattctca 5340
ccaataaaaa acgcccggcg gcaaccgagc gttctgaaca aatccagatg gagttctgag 5400
gtcattactg gatctatcaa caggagtcca agcgagctct tatttattac gctggatgat 5460
gtagtccgct agcgcttcca gtgccgaggt atcgagtgac tgttcagcca gttgtttcag 5520
cgactgacgg gcatcgtcga tcagatcccg ggctttcttc cgggcttgct caagacccag 5580
aagtgcaggg taggtacttt taccaagttg ctggtcggca ccctggcgtt ttcccaacgt 5640
tgcagtatct cccaccacat ccaggatgtc atcctgaacc tggaaggcaa ggccgatgct 5700
ctctgcatac ttgtcgagta ccggcagagc acgacgtcct ttatctccgg cgcttaatgc 5760
accaaggcga acggcggcgc gaatcaatgc gccggtttta tgacgatgaa tacgctcaag 5820
cgcgtccaga ggtacgtgtt tgccttccgc gtctaaatct aatgcctgac caccgcacat 5880
tccggcaata ccactggcgc tcgccagttc agaaatcatc gaaattctgt cgcggtccga 5940
cacttccggc atatcggcat cgcttaaaat cgagaacgcc agcgtttgta aagcgtcgcc 6000
agcgagaatc gcgtttgctt cgccaaactt cacatggcag gttggcaaac cgcgacgcag 6060
atcgtcatca tccattgccg gtaaatcatc atgaattaat gagtaagcgt ggatacactc 6120
aacggcggca gcgggtgcgt ccagcgtgtt tgtgctaacg ccgaacatat gaccggtggc 6180
ataaaccagg aaaggtcgca ggcgcttacc acctaataat gcgccatact gcatggtttc 6240
gaccacggga gtgttctgaa agggcagtgg ggcgataaaa cggctcagcg cctggttggc 6300
ctgcttaacg caggcttcga gttgctgcgg aaagtccatt actcattcct cctccgcggt 6360
ggcggccgct ctagaactag tggatccccc gggttattta agctgggtaa atgcagataa 6420
tcgttttctg gcttcgcgat ttgtcgcctg catcaccatc cacggactga acgcccacgg 6480
cgtggcatca ataccgtgta atacatctgc taaatcacac cattgataat ccatcacttc 6540
atcatcattg atctgtaacg cactagtggt gcgtgcggca aataccggac acacttcatt 6600
ttccacaatg ccactcggat cggtggcgcg gtagcgaaag tcaggataga tagattcagg 6660
aggcgtaatt tccacgccaa gctcataacg gcaacggcgg atcactgcgt cttcgttgct 6720
ttctcccagt tgtgggtgcc cacaaaccga gttagtccac acgccaggcc atgctttttt 6780
gctcagtgcg cggcgggtaa ctaataattg tcctttggca ttaaacagcc aactggagaa 6840
cgcgagatgt aagcgggtgt ctgccgtgtg tgcggcatac ttttccagcg tacccgtggg 6900
aactccctgt gcattcaata aaatgacgtg ttccgtttgc atatagtaat cctcctcccg 6960
ggctgcagtt attcctttgg tagaccagtc tttgcgtcaa tcaaagattc gtttgtttct 7020
tgtgggcctg aaccgacttg agttaaaatc actctggcaa catccttttg caactcaaga 7080
tccaattcac gtgcagtaaa gttagatgat tcaaattgat ggttgaaagc ctcaagctgc 7140
tcagtagtaa atttcttgtc ccatccagga acagagccaa acaatttata gataaatgca 7200
aagagtttcg actcattttc agctaagtag tacaacacag catttggacc tgcatcaaac 7260
```

gtgtatgcaa cgattgtttc tccgtaaaac tgattaatgg tgtggcacca actgatgata 7320
cgcttggaag tgtcattcat gtagaatatt ggagggaaag agtccaaaca tgtggcatgg 7380
aaagagttgg aatccatcat tgtttccttt gcaaaggtgg cgaaatcttt ttcaacaatg 7440
gctttacgca tgacttcaaa tctctttggt acgacatgtt caattctttc tttaaatagt 7500
tcggaggttg ccacggtcaa ttgcataccc tgagtggaac tcacatcctt tttaatatcg 7560
ctgacaacta ggacacaagc tttcatctga ggccagtcag agctgtctgc gatttgtact 7620
gccatggaat catgaccatc ttcagctttt cccatttccc aggccacgta tccgccaaac 7680
aacgatctac aagctgaacc agacccctttt cttgctattc tagatatttc tgaagttgac 7740
tgtggtaatt ggtataactt agcaattgca gagaccaatg cagcaaagcc agcagcggag 7800
gaagctaaac cagctgctgt aggaaagtta ttttcggaga caatgtggag tttccattga 7860
gataatgtgg gcaatgaggc gtccttcgat tccatttcct ttcttaattg gcgtaggtcg 7920
cgcagacaat tttgagttct ttcattgtcg atgctgtgtg gttctccatt taaccacaaa 7980
gtgtcgcgtt caaactcagg tgcagtagcc gcagaggtca acgttctgag gtcatcttgc 8040
gataaagtca ctgatatgga cgaattggtg ggcagattca acttcgtgtc ccttttcccc 8100
caatacttaa gggttgcgat gttgacgggt gcggtaacgg atgctgtgta aacggtcatg 8160
agtattacct cctatttatc aagataagtt tccggatctt tttctttcct aacaccccag 8220
tcagcctgag ttacatccag ccattgaacc ttagaaaatc ttttgtcatt agcggtttga 8280
gccctaagat caacatcttg cttagtaatc actgcaatgg cgtcataacc accagcacca 8340
ggtattaagc aagtaagaac tccttttaag gtctggcaat catccaataa gctagtttgt 8400
acgggaggtt cgatatcggc accagattct ttagttattt ttctaaagga acgtctaatt 8460
gtggcaactg catctctaac ttctgtgatt tcaggatact tttgacaggt acagtcattc 8520
ctctcaagag actcaaatat ctgatcgctg taatcgtcat gagtctcgtg taagcgatct 8580
agtttagata gtccatccat aaatctagaa tttgcatgat cgagttctgt atatattttc 8640
aagctttctg gcatatgcga atcataccaa ttttttacct tctggaccag ttttactgtt 8700
tctgaaccat tcttaatatc gcccatccat aaagttaatc ccgaaggtaa atggttactt 8760
ttaatcgtaa tattccagtc ttcttcatca accaaatgcg ccagtttact gccgtaagta 8820
gcacttccaa tatctggcaa attagagatt aatgcgggtg ggaatcttct atatctgata 8880
gatccatatg ctgccgccgc tacatcaaac ccgcttccaa ttttaccctg agcttgacaa 8940
tgagcaactt gtgctaaatt atgaataact tctctatatt tgtctacatt attttccagg 9000
tccgatacaa aaaaggaggc caaagctgta gttaaaactg tgactaaacc tgccgaggag 9060
cccagccctg ttttgggaac ttcttcaatt ctgtgcgaat gaaaactcaa tcttctgttg 9120
ccacgatgtt cggtaacgct atcctcctga gaatggtagg catcatcaga gaaaatatca 9180
ataacgaaca agtttctatt gcagtagtcg tccatgttag gtttaaagta gctaaatacg 9240
ttagcgataa ctttttcaat gaaagggttc ttagatccgc ctatcgaaac aggaatgaag 9300
ccacttttag gacttatatg gtacagccac tccccatctt taaattgttt acttttcaca 9360
cgcacttcaa acttatcaga cccttgcaat gaaccgtaag gatgggctac agcatgcatt 9420
cttgccgata atccgactac aaatgcttca tattttgtat ctaaaactaa atatccacca 9480
gctagtaacg ctttccctgg ggcactgaag gctctcaact ctgacatttg atctgcctcc 9540
tatgaagtcc atggtaaatt cgtgtttcct ggcaataata gatcgtcaat ttgttgcttt 9600
gtggtagttt tattttcaaa taattggaat actagggatt tgattttaag atctttattc 9660
aaatttttg cgcttaacaa acagcagcca gtcccaccca agtctgtttc aaatgtctcg 9720
taactaaaat catcttgcaa tttctttttg aagctgtcaa tttgctcttg agtaatgtct 9780
cttcgtaaca aagtcaaaga gcaaccgccg ccaccagcac cggtaagttt tgtggagcca 9840
attctcaaat catcgctcag atttttaata agttctaatc caggatgaga aacaccgatt 9900
gagacaagca gtccatgatt tattcttatc aattccaata gttgttcata cagttcatta 9960
ttagtttcta cagcctcgtc atcggtgcct ttacatttac ttaacttagt catgatctct 10020
aagccttgta gggcacattc acccatggca tctagaattg gcttcataac ttcaggaaat 10080
ttctcggtga ccaacacacg aacgcgagca acaagatctt ttgtagacct tggaattcta 10140
gtataggtta ggatcattgg aatggctggg aaatcatcta agaacttaaa attgtttgtg 10200
tttattgttc cattatgtga gtctttttca aatagcaggg cattaccata agtggccaca 10260
gcgttatcta ttcctgaagg ggtaccgtga atacactttt cacctatgaa ggcccattga 10320
ttcactatat gcttatcgtt ttctgacagc ttttccaagt cattagatcc tattaacccc 10380
cccaagtagg ccatagctaa ggccagtgat acagaaatag aggcgcttga gcccaaccca 10440
gcaccgatgg gtaaagtaga ctttaaagaa aacttaatat tcttggcatg ggggcatagg 10500
caaacaaaca tatacaggaa acaaaacgct gcatggtagt ggaaggattc ggatagttga 10560
gctaacaacg gatccaaaag actaacgagt tcctgagaca agccatcggt ggcttgttga 10620
gccttggcca atttttggga gtttacttga tcctcggtga tggcattgaa atcattgatg 10680
gaccacttat gattaaagct aatgtccggg aagtccaatt caatagtatc tggtgcagat 10740
gactcgctta ttagcaggta ggttctcaac gcagacacac tagcagcgac ggcaggcttg 10800
ttgtacacag cagagtgttc accaaaaata ataacctttc ccggtgcaga agttaagaac 10860
ggtaatgaca tgggcccggt acccctccta attcggtcag tgcgtcctgc tgatgtgctc 10920
attatacgag ccggaagcat aaagtgtaaa gacgtcgata tctggcgaaa atgagacgtt 10980
gatcggcacg taagaggttc caactttcac cataatgaaa taagatcact accgggcgta 11040

```
tttttgagt tatcgagatt ttcaggagct aaggaagcta aaatggagaa aaaaatcact 11100
ggatatacca ccgttgatat atcccaatgg catcgtaaag aacattttga ggcatttcag 11160
tcagttgctc aatgtaccta taaccagacc gttcagctgg atattacggc ctttttaaag 11220
accgtaaaga aaaataagca caagttttat ccggccttta ttcacattct tgcccgcctg 11280
atgaatgctc atccggaatt ccgtatggca atgaaagacg gtgagctggt gatatgggat 11340
agtgttcacc cttgttacac cgttttccat gagcaaactg aaacgttttc atcgctctgg 11400
agtgaatacc acgacgattt ccggcagttt ctacacatat attcgcaaga tgtggcgtgt 11460
tacggtgaaa acctggccta tttccctaaa gggtttattg agaatatgtt tttcgtctca 11520
gccaatccct gggtgagttt caccagtttt gatttaaacg tggccaatat ggacaacttc 11580
ttcgcccccg ttttcaccat gggcaaatat tatacgcaag gcgacaaggt gctgatgccg 11640
ctggcgattc aggttcatca tgccgtctgt gatggcttcc atgtcggcag aatgcttaat 11700
gaattacaac agtactgcga tgagtggcag ggcggggcgt aatttgatat cgagctctcg 11760
aaccccagag tcccgctcag aagaactcgt caagaaggcg atagaaggcg atgcgctgcg 11820
aatcgggagc ggcgataccg taaagcacga ggaagcggtc agcccattcg ccgccaagct 11880
cttcagcaat atcacgggta gccaacgcta tgtcctgata gcggtccgcc acacccagcc 11940
ggccacagtc gatgaatcca gaaaagcggc cattttccac catgatattc ggcaagcagg 12000
catcgccatg ggtcacgacg agatcctcgc cgtcgggcat gcgcgccttg agcctggcga 12060
acagttcggc tggcgcgagc ccctgatgct cttcgtccag atcatcctga tcgacaagac 12120
cggcttccat ccgagtacgt gctcgctcga tgcgatgttt cgcttggtgg tcgaatgggc 12180
aggtagccgg atcaagcgta tgcagccgcc gcattgcatc agccatgatg gatactttct 12240
cggcaggagc aaggtgagat gacaggagat cctgccccgg cacttcgccc aatagcagcc 12300
agtcccttcc cgcttcagtg acaacgtcga gcacagctgc gcaaggaacg cccgtcgtgg 12360
ccagccacga tagccgcgct gcctcgtcct gcagttcatt cagggcaccg gacaggtcgg 12420
tcttgacaaa aagaaccggg cgcccctgcg ctgacagccg gaacacggcg gcatcagagc 12480
agccgattgt ctgttgtgcc cagtcatagc cgaatagcct ctccacccaa gcggccggag 12540
aacctgcgtg caatccatct tgttcaatca tgcgaaacga tcctcatcct gtctcttgat 12600
cagatcttga tcccctgcgc catcagatcc ttggcggcaa gaaagccatc cagtttactt 12660
tgcagggctt cccaacctta ccagagggcg ccccagctgg caattcc               12707
```

<210> 5
<211> 13055
<212> DNA

<213> Sequência artificial

<220>

<223> vetor recombinante

<400> 5

```
gaattccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt 60
gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt 120
ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga 180
tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagcttcct tagctcctga 240
aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt 300
ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc 360
ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat 420
ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt 480
gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg 540
acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact 600
ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa 660
aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc 720
actgactcgc tacgctcggt cgttcgactg cggcgagcgg aaatggctta cgaacggggc 780
ggagatttcc tggaagatgc caggaagata cttaacaggg aagtgagagg gccgcggcaa 840
agccgttttt ccataggctc cgccccctg acaagcatca cgaaatctga cgctcaaatc 900
agtggtggcg aaacccgaca ggactataaa gataccaggc gtttcccccc ggcggctccc 960
tcgtgcgctc tcctgttcct gcctttcggt ttaccggtgt cattccgctg ttatggccgc 1020
gtttgtctca ttccacgcct gacactcagt tccgggtagg cagttcgctc caagctggac 1080
tgtatgcacg aaccccccgt tcagtccgac cgctgcgcct tatccggtaa ctatcgtctt 1140
gagtccaacc cggaaagaca tgcaaaagca ccactggcag cagccactgg taattgattt 1200
agaggagtta gtcttgaagt catgcgccgg ttaaggctaa actgaaagga caagttttgg 1260
tgactgcgct cctccaagcc agttacctcg gttcaaagag ttggtagctc agagaacctt 1320
cgaaaaaccg ccctgcaagg cggtttttc gttttcagag caagagatta cgcgcagacc 1380
```

aaaacgatct caagaagatc atcttattaa tcagataaaa tatttctaga tttcagtgca 1440
atttatctct tcaaatgtag cacctgaagt cagccccata cgatataagt tgtaattctc 1500
atgtttgaca gcttatcatc gataagcttc cgatggcgcg ccgagaggct ttacacttta 1560
tgcttccggc tcgtataatg tgtggaattg tgagcggata acaattgaat tcaaaggagg 1620
ccatcctggc catgaagaac tgtgtgattg tttctgcggt ccgcacggcg atcggcagct 1680
ttaacggctc tttagcgagc acctctgcaa tcgatctggg tgcgacggtc attaaggccg 1740
ccattgaacg cgccaaaatc gacagccagc acgttgatga ggtgatcatg ggcaatgtgt 1800
tacaagccgg cctgggtcaa aacccagcgc gtcaagcact gttaaaatct ggtctggccg 1860
agaccgtgtg tggcttcacc gtcaataagg tttgcggctc tggcctgaag agcgtggccc 1920
tggcagcaca agcgattcaa gccggtcagg cacaaagcat cgttgcgggt ggcatggaga 1980
acatgtctct ggcgccgtac ttattagatg ccaaagcccg cagcggttat cgcctgggcg 2040
atggtcaggt gtacgacgtc atcttacgcg atggcttaat gtgcgcgacc cacggttacc 2100
acatgggtat tacggccgaa aacgtggcga aagaatacgg cattacgcgc gagatgcagg 2160
atgaattagc actgcactct cagcgcaaag cagcagccgc gatcgagtct ggtgcgttta 2220
cggcggaaat cgtgccagtt aacgtggtca cgcgcaagaa gacgttcgtt ttcagccagg 2280
acgagttccc gaaggcaaac agcaccgcgg aggccttagg tgccttacgc ccagcctttg 2340
acaaagcggg cacggtcacc gccggtaatg cgagcggcat caatgatggt gcagcggcac 2400
tggtcatcat ggaagagagc gccgcattag cagcgggtct gaccccatta gcgcgcatta 2460
aatcttatgc cagcggcggc gtcccaccag ccctgatggg catgggtccg gtcccagcca 2520
cgcaaaaagc cctgcaatta gcgggcctgc aactggccga cattgatctg atcgaggcga 2580
acgaggcgtt tgcagcgcag ttcctggcgg tgggtaagaa tctgggcttc gacagcgaga 2640
aagtcaatgt gaacggtggc gcgattgcgt taggccatcc gattggtgca agcggcgcac 2700
gcatcttagt gacgttactg cacgccatgc aggcacgcga caagacctta ggcctggcga 2760
ccttatgtat tggtggcggt caaggtatcg ccatggtgat cgaacgcctg aactgaagat 2820
ctaggaggaa agcaaaatga aactgagcac caagctgtgc tggtgtggca tcaagggtcg 2880
cctgcgccca caaaagcagc aacagctgca caacacgaac ctgcaaatga ccgagctgaa 2940
aaagcagaag acggccgagc aaaagacccg cccgcagaac gttggcatca agggcatcca 3000
gatttatatc ccgacgcagt gtgtcaacca atctgagctg gagaaattcg atggcgtcag 3060
ccagggtaag tacaccatcg gcctgggcca gaccaacatg agcttcgtga acgaccgtga 3120
ggacatctat tctatgagcc tgacggtgct gtctaagctg atcaagagct acaacatcga 3180
cacgaataag atcggtcgtc tggaggtggg tacggagacg ctgattgaca agagcaaaag 3240
cgtgaagtct gtcttaatgc agctgttcgg cgagaacacg gatgtcgagg gtatcgacac 3300
cctgaacgcg tgttacggcg gcaccaacgc actgttcaat agcctgaact ggattgagag 3360
caacgcctgg gatggccgcg atgcgatcgt cgtgtgcggc gatatcgcca tctatgacaa 3420
gggtgcggca cgtccgaccg gcggtgcagg caccgttgcg atgtggattg gcccggacgc 3480
accaattgtc ttcgattctg tccgcgcgtc ttacatggag cacgcctacg acttttacaa 3540
gccggacttc acgagcgaat acccgtacgt ggacggccac ttctctctga cctgctatgt 3600
gaaggcgctg gaccaggttt ataagtctta tagcaaaaag gcgatttcta agggcctggt 3660
cagcgacccg gcaggcagcg acgccctgaa cgtgctgaag tatttcgact acaacgtgtt 3720
ccatgtcccg acctgcaaat tagtgaccaa atcttatggc cgcctgttat ataatgattt 3780
ccgtgccaac ccgcagctgt tcccggaggt tgacgccgag ctggcgacgc gtgattacga 3840
cgagagcctg accgacaaga acatcgagaa gaccttcgtc aacgtcgcga agccgttcca 3900
caaagagcgt gtggcccaaa gcctgatcgt cccgaccaac acgggcaaca tgtataccgc 3960
gtctgtctac gcggcattcg cgagcctgct gaattacgtc ggttctgacg acctgcaggg 4020
caagcgcgtt ggcctgttca gctacggtag cggcttagcg gccagcctgt atagctgcaa 4080
aattgtcggc gacgtccagc acatcatcaa ggagctggac atcaccaaca agctggcgaa 4140
gcgcatcacc gagacgccga aagattacga ggcagcgatc gagttacgcg agaatgcgca 4200
tctgaagaag aacttcaagc cgcaaggtag catcgagcac ctgcagagcg gcgtctacta 4260
cctgacgaac attgacgaca agttccgccg ttcttatgac gtcaaaaagt aactagtagg 4320
aggaaaacat catggtgctg acgaacaaaa ccgtcattag cggcagcaag gtgaagtctc 4380
tgagcagcgc ccaaagctct agcagcggcc cgtctagcag cagcgaggag gacgacagcc 4440
gtgacattga gtctctggac aagaagatcc gcccgctgga ggagttagag gccctgctga 4500
gcagcggcaa caccaagcag ctgaagaaca aggaagttgc agcgctggtg atccacggta 4560
agctgccact gtatgcgctg gaaaagaaac tgggcgatac gacgcgtgcg gtcgcggtgc 4620
gtcgcaaagc cttaagcatc ttagcggagg ccccggtgtt agccagcgac cgcctgccgt 4680
acaagaacta cgactacgac cgcgtgtttg gcgcgtgctg cgagaatgtc attggctaca 4740
tgccgttacc ggttggtgtg atcggcccgc tggtcattga tggcacgagc tatcacattc 4800
caatggcgac cacggaaggt tgcttagtcg ccagcgccat gcgtggctgt aaggcgatta 4860
acgccggcgg tggcgcgacg accgtgttaa ccaaggatgg tatgacgcgc ggtccggtcg 4920
tccgcttccc aacgctgaag cgcagcggcg cgtgtaagat ttggctggat tctgaggagg 4980
gccaaaacgc gatcaagaaa gccttcaact ctacgagccg tttcgcgcgt ttacagcata 5040
tccagacctg cctggccggc gacctgctgt tcatgcgctt ccgcaccacc acgggcgatg 5100
cgatgggcat gaacatgatc agcaagggcg tcgaatatag cctgaaaacaa atggtggaag 5160

aatatggctg ggaggacatg gaggttgtct ctgtgagcgg caactattgc accgacaaga 5220
agccggcagc cattaactgg attgagggtc gcggcaaaag cgtcgtggca gaagcgacca 5280
tcccaggcga cgtggtccgt aaggttctga agagcgacgt cagcgccctg gttgagttaa 5340
atatcgcgaa aaacctggtc ggcagcgcga tggcgggcag cgtgggtggc tttaacgcac 5400
atgcagcgaa tctggttacg gcggttttct tagccttagg tcaggaccca gcccaaaatg 5460
tcgagagcag caactgcatt accttaatga aagaggttga cggtgacctg cgcatcagcg 5520
tttctatgcc gtctatcgag gtcggcacga tcggcggcgg caccgtttta gaaccgcaag 5580
gtgcgatgct ggatctgctg ggcgtgcgcg gcccacatgc aacggcccca ggcaccaatg 5640
cccgccaact ggcccgtatc gtggcctgcg cggttctggc gggtgagctg agcctgtgcg 5700
ccgcattagc cgcgggccat ttagttcaat ctcacatgac ccacaaccgc aagccggcag 5760
aaccaaccaa gccaaataac ctggacgcaa ccgacattaa ccgtctgaag gatggcagcg 5820
tcacgtgcat taaaagctga gcatgctact aagcttggct gttttggcgg atgagagaag 5880
attttcagcc tgatacagat taaatcagaa cgcagaagcg gtctgataaa acagaatttg 5940
cctggcggca gtagcgcggt ggtcccacct gaccccatgc cgaactcaga agtgaaacgc 6000
cgtagcgccg atggtagtgt ggggtctccc catgcgagag tagggaactg ccaggcatca 6060
aataaaacga aaggctcagt cgaaagactg ggcctttcgt tttatctgtt gtttgtcggt 6120
gaacgctctc ctgagtagga caaatccgcc gggagcggat ttgaacgttg cgaagcaacg 6180
gcccggaggg tggcgggcag gacgcccgcc ataaactgcc aggcatcaaa ttaagcagaa 6240
ggccatcctg acggatggcc tttttgcgtt tctacaaact cttttgttta ttttctaaa 6300
tacattcaaa tatgtatccg ctcatgagac aataaccctg cgatcgccga gaggctttac 6360
actttatgct tccggctcgt ataatgtgtg gaattgtgag cggataacaa ttgaattcaa 6420
aggaggctcg agatgtcatt accgttctta acttctgcac cgggaaaggt tattattttt 6480
ggtgaacact ctgctgtgta caacaagcct gccgtcgctg ctagtgtgtc tgcgttgaga 6540
acctacctgc taataagcga gtcatctgca ccagatacta ttgaattgga cttcccggac 6600
attagcttta atcataagtg gtccatcaat gatttcaatg ccatcaccga ggatcaagta 6660
aactcccaaa aattggccaa ggctcaacaa gccaccgatg gcttgtctca ggaactcgtt 6720
agtcttttgg atccgttgtt agctcaacta tccgaatcct tccactacca tgcagcgttt 6780
tgtttcctgt atatgtttgt ttgcctatgc ccccatgcca agaatattaa gttttcttta 6840
aagtctactt tacccatcgg tgctgggttg ggctcaagcg cctctatttc tgtatcactg 6900
gccttagcta tggcctactt gggggggtta ataggatcta atgacttgga aaagctgtca 6960
gaaaacgata agcatatagt gaatcaatgg gccttcatag gtgaaaagtg tattcacggt 7020
acccccttcag gaatagataa cgctgtggcc acttatggta atgccctgct atttgaaaaa 7080
gactcacata atggaacaat aaacacaaac aattttaagt tcttagatga tttcccagcc 7140
attccaatga tcctaaccta tactagaatt ccaaggtcta caaaagatct tgttgctcgc 7200
gttcgtgtgt tggtcaccga gaaatttcct gaagttatga agccaattct agatgccatg 7260
ggtgaatgtg ccctacaagg cttagagatc atgactaagt taagtaaatg taaaggcacc 7320
gatgacgagg ctgtagaaac taataatgaa ctgtatgaac aactattgga attgataaga 7380
ataaatcatg gactgcttgt ctcaatcggt gtttctcatc ctggattaga acttattaaa 7440
aatctgagcg atgatttgag aattggctcc acaaaactta ccggtgctgg tggcggcggt 7500
tgctctttga ctttgttacg aagagacatt actcaagagc aaattgacag cttcaaaaag 7560
aaattgcaag atgattttag ttacgagaca tttgaaacag acttgggtgg gactggctgc 7620
tgtttgttaa gcgcaaaaaa tttgaataaa gatcttaaaa tcaaatccct agtattccaa 7680
ttatttgaaa ataaaactac cacaaagcaa caaattgacg atctattatt gccaggaaac 7740
acgaatttac catggacttc ataggaggca gatcaaatgt cagagttgag agccttcagt 7800
gccccaggga aagcgttact agctggtgga tatttagttt tagatacaaa atatgaagca 7860
tttgtagtcg gattatcggc aagaatgcat gctgtagccc atccttacgg ttcattgcaa 7920
gggtctgata agtttgaagt gcgtgtgaaa agtaaacaat ttaaagatgg ggagtggctg 7980
taccatataa gtcctaaaag tggcttcatt cctgtttcga taggcggatc taagaaccct 8040
ttcattgaaa aagttatcgc taacgtattt agctacttta aacctaacat ggacgactac 8100
tgcaatagaa acttgttcgt tattgatatt ttctctgatg atgcctacca ttctcaggag 8160
gatagcgtta ccgaacatcg tggcaacaga agattgagtt ttcattcgca cagaattgaa 8220
gaagttccca aaacagggct gggctcctcg gcaggtttag tcacagtttt aactacagct 8280
ttggcctcct tttttgtatc ggacctggaa aataatgtag acaaatatag agaagttatt 8340
cataatttag cacaagttgc tcattgtcaa gctcagggta aaattggaag cgggtttgat 8400
gtagcggcgg cagcatatgg atctatcaga tatagaagat tcccacccgc attaatctct 8460
aatttgccag atattggaag tgctacttac ggcagtaaac tggcgcattt ggttgatgaa 8520
gaagactgga atattacgat taaaagtaac catttacctt cgggattaac tttatggatg 8580
ggcgatatta agaatggttc agaaacagta aaactggtcc agaaggtaaa aaattggtat 8640
gattcgcata tgccagaaag cttgaaaata tatacagaac tcgatcatgc aaattctaga 8700
tttatggatg gactatctaa actagatcgc ttacacgaga ctcatgacga ttacagcgat 8760
cagatatttg agtctcttga gaggaatgac tgtacctgtc aaaagtatcc tgaaatcaca 8820
gaagttagag atgcagttgc cacaattaga cgttccttta gaaaaataac taaagaatct 8880
ggtgccgata tcgaacctcc cgtacaaact agcttattgg atgattgcca gaccttaaaa 8940

```
ggagttctta cttgcttaat acctggtgct ggtggttatg acgccattgc agtgattact 9000
aagcaagatg ttgatcttag ggctcaaacc gctaatgaca aaagattttc taaggttcaa 9060
tggctggatg taactcaggc tgactggggt gttaggaaag aaaaagatcc ggaaacttat 9120
cttgataaat aggaggtaat actcatgacc gtttacacag catccgttac cgcacccgtc 9180
aacatcgcaa cccttaagta ttgggggaaa agggacacga agttgaatct gcccaccaat 9240
tcgtccatat cagtgacttt atcgcaagat gacctcagaa cgttgacctc tgcggctact 9300
gcacctgagt ttgaacgcga cactttgtgg ttaaatggag aaccacacag catcgacaat 9360
gaaagaactc aaaattgtct gcgcgaccta cgccaattaa gaaaggaaat ggaatcgaag 9420
gacgcctcat tgcccacatt atctcaatgg aaactccaca ttgtctccga aaataacttt 9480
cctacagcag ctggtttagc ttcctccgct gctggctttg ctgcattggt ctctgcaatt 9540
gctaagttat accaattacc acagtcaact tcagaaatat ctagaatagc aagaaagggg 9600
tctggttcag cttgtagatc gttgtttggc ggatacgtgg cctgggaaat gggaaaagct 9660
gaagatggtc atgattccat ggcagtacaa atcgcagaca gctctgactg gcctcagatg 9720
aaagcttgtg tcctagttgt cagcgatatt aaaaaggatg tgagttccac tcagggtatg 9780
caattgaccg tggcaacctc cgaactattt aaagaaagaa ttgaacatgt cgtaccaaag 9840
agatttgaag tcatgcgtaa agccattgtt gaaaaagatt tcgccacctt tgcaaaggaa 9900
acaatgatgg attccaactc tttccatgcc acatgtttgg actctttccc tccaatattc 9960
tacatgaatg acacttccaa gcgtatcatc agttggtgcc acaccattaa tcagttttac 10020
ggagaaacaa tcgttgcata cacgtttgat gcaggtccaa atgctgtgtt gtactactta 10080
gctgaaaatg agtcgaaact ctttgcattt atctataaat tgtttggctc tgttcctgga 10140
tgggacaaga aatttactac tgagcagctt gaggctttca accatcaatt tgaatcatct 10200
aactttactg cacgtgaatt ggatcttgag ttgcaaaagg atgttgccag agtgatttta 10260
actcaagtcg gttcaggccc acaagaaaca aacgaatctt tgattgacgc aaagactggt 10320
ctaccaaagg aataactgca gcccgggagg aggattacta tatgcaaacg gaacacgtca 10380
ttttattgaa tgcacaggga gttcccacgg gtacgctgga aaagtatgcc gcacacacgg 10440
cagacacccg cttacatctc gcgttctcca gttggctgtt taatgccaaa ggacaattat 10500
tagttacccg ccgcgcactg agcaaaaaag catggcctgg cgtgtggact aactcggttt 10560
gtgggcaccc acaactggga gaaagcaacg aagacgcagt gatccgccgt tgccgttatg 10620
agcttggcgt ggaaattacg cctcctgaat ctatctatcc tgactttcgc taccgcgcca 10680
ccgatccgag tggcattgtg gaaaatgaag tgtgtccggt atttgccgca cgcaccacta 10740
gtgcgttaca gatcaatgat gatgaagtga tggattatca atggtgtgat ttagcagatg 10800
tattacacgg tattgatgcc acgccgtggg cgttcagtcc gtggatggtg atgcaggcga 10860
caaatcgcga agccagaaaa cgattatctg catttaccca gcttaaataa cccgggggat 10920
ccactagttc tagagcggcc gccaccgcgg aggaggaatg agtaatggac tttccgcagc 10980
aactcgaagc ctgcgttaag caggccaacc aggcgctgag ccgttttatc gccccactgc 11040
cctttcagaa cactcccgtg gtcgaaacca tgcagtatgg cgcattatta ggtggtaagc 11100
gcctgcgacc tttcctggtt tatgccaccg gtcatatgtt cggcgttagc acaaacacgc 11160
tggacgcacc cgctgccgcc gttgagtgta tccacgctta ctcattaatt catgatgatt 11220
taccggcaat ggatgatgac gatctgcgtc gcggtttgcc aacctgccat gtgaagtttg 11280
gcgaagcaaa cgcgattctc gctggcgacg ctttacaaac gctggcgttc tcgattttaa 11340
gcgatgccga tatgccggaa gtgtcggacc gcgacagaat ttcgatgatt tctgaactgg 11400
cgagcgccag tggtattgcc ggaatgtgcg gtggtcaggc attagattta gacgcggaag 11460
gcaaacacgt acctctggac gcgcttgagc gtattcatcg tcataaaacc ggcgcattga 11520
ttcgcgccgc cgttcgcctt ggtgcattaa gcgccggaga taaaggacgt cgtgctctgc 11580
cggtactcga caagtatgca gagagcatcg gccttgcctt ccaggttcag gatgacatcc 11640
tggatgtggt gggagatact gcaacgttgg gaaaacgcca gggtgccgac cagcaacttg 11700
gtaaaagtac ctaccctgca cttctgggtc ttgagcaagc ccggaagaaa gcccgggatc 11760
tgatcgacga tgcccgtcag tcgctgaaac aactggctga acagtcactc gatacctcgg 11820
cactggaagc gctagcggac tacatcatcc agcgtaataa ataagagctc caattcgccc 11880
tatagtgaga cgcgtgctag aggcatcaaa taaaacgaaa ggctcagtcg aaagactggg 11940
cctttcgttt tatctgttgt ttgtcggtga acgctctcct gagttaatta actccaggcc 12000
ggcctacgcg tttaaacttc cggttaacgc catgagcggc ctcatttctt attctgagtt 12060
acaacagtcc gcaccgctgc cggtagctcc ttccggtggg cgcggggcat gactatcgtc 12120
gccgcactta tgactgtctt ctttatcatg caactcgtag gacaggtgcc ggcagcgccc 12180
aacagtcccc cggccacggg gcctgccacc atacccacgc cgaaacaagc gccctgcacc 12240
attatgttcc ggatctgcat cgcaggatgc tgctggctac cctgtggaac acctacatct 12300
gtattaacga agcgctaacc gtttttatca ggctctggga ggcagaataa atgatcatat 12360
cgtcaattat tacctccacg gggagagcct gagcaaactg gcctcaggca tttgagaagc 12420
acacggtcac actgcttccg gtagtcaata aaccggtaaa ccagcaatag acataagcgg 12480
ctatttaacg accctgccct gaaccgacga ccgggtcgaa tttgctttcg aatttctgcc 12540
attcatccgc ttattatcac ttattcaggc gtagcaccag gcgtttaagg gcaccaataa 12600
ctgccttaaa aaaattacgc cccgccctgc cactcatcgc agtactgttg taattcatta 12660
agcattctgc cgacatggaa gccatcacag acggcatgat gaacctgaat cgccagcggc 12720
```

```
atcagcacct tgtcgccttg cgtataatat ttgcccatgg tgaaaacggg ggcgaagaag 12780
ttgtccatat tggccacgtt taaatcaaaa ctggtgaaac tcacccaggg attggctgag 12840
acgaaaaaca tattctcaat aaacccttta gggaaatagg ccaggttttc accgtaacac 12900
gccacatctt gcgaatatat gtgtagaaac tgccggaaat cgtcgtggta ttcactccag 12960
agcgatgaaa acgtttcagt ttgctcatgg aaaacggtgt aacaagggtg aacactatcc 13020
catatcacca gctcaccgtc tttcattgcc atacg                           13055
```

<210> 6
<211> 7120
<212> DNA

<213> Sequência artificial

<220>

<223> vetor recombinante

<400> 6

```
gtttgacagc ttatcatcga ctgcacggtg caccaatgct tctggcgtca ggcagccatc 60
ggaagctgtg gtatggctgt gcaggtcgta aatcactgca taattcgtgt cgctcaaggc 120
gcactcccgt tctggataat gtttttttgcg ccgacatcat aacggttctg gcaaatattc 180
tgaaatgagc tgttgacaat taatcatccg gctcgtataa tgtgtggaat tgtgagcgga 240
taacaatttc acacaggaaa cagaccatgg ccctgaccga agagaaaccg atccgcccga 300
tcgctaactt cccgccgtct atctggggtg accagttcct gatctacgaa aagcaggttg 360
agcagggtgt tgaacagatc gtaaacgacc tgaagaaaga agttcgtcag ctgctgaaag 420
aagctctgga catcccgatg aaacacgcta acctgctgaa actgatcgac gagatccagc 480
gtctgggtat cccgtaccac ttcgaacgcg aaatcgacca cgcactgcag tgcatctacg 540
aaacctacgg cgacaactgg aacggcgacc gttcttctct gtggtttcgt ctgatgcgta 600
aacagggcta ctacgttacc tgtgacgttt ttaacaacta caaggacaag aacggtgctt 660
tcaaacagtc tctggctaac gacgttgaag gcctgctgga actgtacgaa gcgacctcca 720
tgcgtgtacc gggtgaaatc atcctggagg acgcgctggg tttcacccgt tctcgtctgt 780
ccattatgac taaagacgct ttctctacta acccggctct gttcaccgaa atccagcgtg 840
ctctgaaaca gccgctgtgg aaacgtctgc cgcgtatcga agcagcacag tacattccgt 900
tttaccagca gcaggactct cacaacaaga ccctgctgaa actggctaag ctggaattca 960
acctgctgca gtctctgcac aaagaagaac tgtctcacgt ttgtaagtgg tggaaggcat 1020
ttgacatcaa gaaaaacgcg ccgtgcctgc gtgaccgtat cgttgaatgt tacttctggg 1080
gtctgggttc tggttatgaa ccacagtact cccgtgcacg tgtgttcttc actaaagctg 1140
tagctgttat caccctgatc gatgacactt acgatgctta cggcacctac gaagaactga 1200
agatctttac tgaagctgta gaacgctggt ctatcacttg cctggacact ctgccggagt 1260
acatgaaacc gatctacaaa ctgttcatgg atacctacac cgaaatggag gaattcctgg 1320
caaaagaagg ccgtaccgac ctgttcaact gcggtaaaga gtttgttaaa gaattcgtac 1380
gtaacctgat ggttgaagct aaatgggcta acgaaggcca tatcccgact accgaagaac 1440
atgacccggt tgttatcatc accggcggtg caaacctgct gaccaccact tgctatctgg 1500
gtatgtccga catctttacc aaggaatctg ttgaatgggc tgtttctgca ccgccgctgt 1560
tccgttactc cggtattctg ggtcgtcgtc tgaacgacct gatgacccac aaagcagagc 1620
aggaacgtaa acactcttcc tcctctctgg aatcctacat gaaggaatat aacgttaacg 1680
aggagtacgc acagactctg atctataaag aagttgaaga cgtatggaaa gacatcaacc 1740
gtgaatacct gactactaaa aacatcccgc gcccgctgct gatggcagta atctacctgt 1800
gccagttcct ggaagtacag tacgctggta aagataactt cactcgcatg ggcgacgaat 1860
acaaacacct gatcaaatcc ctgctggttt acccgatgtc catctgatcc cggggatcca 1920
ggaggaaata accatgtctc tgccattcct gacgtctgcg ccaggtaagg tgatcatctt 1980
cggcgagcac tctgcggtgt acaataagcc ggccgtcgcc gcctctgtgt ctgcgttacg 2040
cacctacctg ctgatcagcg aatcttctgc accggacacg atcgagctgg actttccgga 2100
catcagcttc aaccacaagt ggagcatcaa cgacttcaac gcgatcacgg aggaccaggt 2160
gaacagccaa aagctggcca aagcccagca agcaaccgac ggtctgtctc aggagctggt 2220
gtctctgctg gacccgctgt tagcgcagtt aagcgagagc ttccattacc acgccgcgtt 2280
ctgcttcctg tacatgttcg tttgcctgtg cccgcacgca aagaacatca agttcagcct 2340
gaagagcacg ctgccgattg gcgcaggctt aggctctagc gcatctatca gcgtgagcct 2400
ggcgctggcg atggcctatc tgggtggcct gattggcagc aacgacctgg agaaactgag 2460
cgaaaacgac aagcacatcg tgaaccagtg ggcctttatc ggcgagaagt gcattcatgg 2520
caccccgagc ggcattgaca acgcagttgc cacgtatggc aacgccctgc tgttcgagaa 2580
agacagccac aacggcacga tcaacacgaa caacttcaag ttcctggacg acttcccggc 2640
gatcccgatg attctgacct acacccgtat cccacgcagc accaaggatt tagtcgcccg 2700
```

cgtgcgtgtt ttagtcaccg aaaagttccc ggaggtgatg aagccgatcc tggacgcgat 2760
gggcgagtgc gcgctgcagg gtctggagat catgaccaag ctgagcaagt gcaagggcac 2820
cgacgatgag gcggtggaga ccaacaatga gctgtacgag cagctgctgg agctgatccg 2880
tatcaatcac ggcctgctgg tctctatcgg tgtgtctcac ccgggcctgg aactgatcaa 2940
aaacctgagc gacgacctgc gcattggctc tacgaaatta acgggtgcag gtggcggtgg 3000
ctgctcttta acgctgctgc gccgtgacat tacgcaggag caaatcgaca gcttcaagaa 3060
gaagctgcag gacgacttca gctacgagac gttcgagacg gacctgggcg gcacgggctg 3120
ttgcctgctg agcgccaaaa atctgaacaa ggacctgaag atcaaaagcc tggtgttcca 3180
gctgttcgaa aacaagacga ccacgaagca gcagatcgac gacctgttac tgccgggtaa 3240
caccaatctg ccgtggacgt cttaaagctt ggctgttttg gcggatgaga gaagattttc 3300
agcctgatac agattaaatc agaacgcaga agcggtctga taaaacagaa tttgcctggc 3360
ggcagtagcg cggtggtccc acctgacccc atgccgaact cagaagtgaa acgccgtagc 3420
gccgatggta gtgtggggtc tccccatgcg agagtaggga actgccaggc atcaaataaa 3480
acgaaaggct cagtcgaaag actgggcctt tcgttttatc tgttgtttgt cggtgaacgc 3540
tctcctgagt aggacaaatc cgccgggagc ggatttgaac gttgcgaagc aacggcccgg 3600
agggtggcgg gcaggacgcc cgccataaac tgccaggcat caaattaagc agaaggccat 3660
cctgacggat ggccttttg cgtttctaca aactcttttt gtttattttt ctaaatacat 3720
tcaaatatgt atccgctcat gagacaataa ccctgataaa tgcttcaata atattgaaaa 3780
aggaagagta tgagtattca acatttccgt gtcgccctta ttccctttt tgcggcattt 3840
tgccttcctg tttttgctca cccagaaacg ctggtgaaag taaaagatgc tgaagatcag 3900
ttgggtgcac gagtgggtta catcgaactg gatctcaaca gcggtaagat ccttgagagt 3960
tttcgccccg aagaacgttt tccaatgatg agcacttta aagttctgct atgtggcgcg 4020
gtattatccc gtgttgacgc cgggcaagag caactcggtc gccgcataca ctattctcag 4080
aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta 4140
agagaattat gcagtgctgc cataaccatg agtgataaca ctgcggccaa cttacttctg 4200
acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta 4260
actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga cgagcgtgac 4320
accacgatgc ctacagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt 4380
actctagctt cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca 4440
cttctgcgct cggcccttcc ggctggctgg tttattgctg ataaatctgg agccggtgag 4500
cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta 4560
gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgctgag 4620
ataggtgcct cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt 4680
tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat cctttttgat 4740
aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta 4800
gaaaagatca aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa 4860
acaaaaaaac caccgctacc agcggtggtt tgtttgccgg atcaagagct accaactctt 4920
tttccgaagg taactggctt cagcagagcg cagataccaa atactgtcct tctagtgtag 4980
ccgtagttag gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta 5040
atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca 5100
agacgatagt taccggataa ggcgcagcgg tcgggctgaa cgggggttc gtgcacacag 5160
cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa 5220
agcgccacgc ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga 5280
acaggagagc gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc 5340
gggtttcgcc acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc 5400
ctatggaaaa acgccagcaa cgcggccttt ttacggttcc tggccttttg ctggcctttt 5460
gctcacatgt tctttcctgc gttatcccct gattctgtgg ataaccgtat taccgccttt 5520
gagtgagctg ataccgctcg ccgcagccga acgaccgagc gcagcgagtc agtgagcgag 5580
gaagcggaag agcgcctgat gcggtatttt ctccttacgc atctgtgcgg tatttcacac 5640
cgcatatggt gcactctcag tacaatctgc tctgatgccg catagttaag ccagtataca 5700
ctccgctatc gctacgtgac tgggtcatgg ctgcgccccg acacccgcca acacccgctg 5760
acgcgccctg acgggcttgt ctgctcccgg catccgctta cagacaagct gtgaccgtct 5820
ccgggagctg catgtgtcag aggttttcac cgtcatcacc gaaacgcgcg aggcagcaga 5880
tcaattcgcg cgcgaaggcg aagcggcatg catttacgtt gacaccatcg aatggtgcaa 5940
aacctttcgc ggtatggcat gatagcgccc ggaagagagt caattcaggg tggtgaatgt 6000
gaaaccagta acgttatacg atgtcgcaga gtatgccggt gtctcttatc agaccgtttc 6060
ccgcgtggtg aaccaggcca gccacgtttc tgcgaaaacg cgggaaaaag tggaagcggc 6120
gatggcggag ctgaattaca ttcccaaccg cgtggcacaa caactggcgg gcaaacagtc 6180
gttgctgatt ggcgttgcca cctccagtct ggccctgcac gcgccgtcgc aaattgtcgc 6240
ggcgattaaa tctcgcgccg atcaactggg tgccagcgtg gtggtgtcga tggtagaacg 6300
aagcggcgtc gaagcctgta aagcggcggt gcacaatctt ctcgcgcaac gcgtcagtgg 6360
gctgatcatt aactatccgc tggatgacca ggatgccatt gctgtggaag ctgcctgcac 6420
taatgttccg gcgttatttc ttgatgtctc tgaccagaca cccatcaaca gtattatttt 6480

ctcccatgaa gacggtacgc gactgggcgt ggagcatctg gtcgcattgg gtcaccagca 6540
aatcgcgctg ttagcgggcc cattaagttc tgtctcggcg cgtctgcgtc tggctggctg 6600
gcataaatat ctcactcgca atcaaattca gccgatagcg gaacgggaag gcgactggag 6660
tgccatgtcc ggttttcaac aaaccatgca aatgctgaat gagggcatcg ttcccactgc 6720
gatgctggtt gccaacgatc agatggcgct gggcgcaatg cgcgccatta ccgagtccgg 6780
gctgcgcgtt ggtgcggata tctcggtagt gggatacgac gataccgaag acagctcatg 6840
ttatatcccg ccgtcaacca ccatcaaaca ggattttcgc ctgctggggc aaaccagcgt 6900
ggaccgcttg ctgcaactct ctcagggcca ggcggtgaag ggcaatcagc tgttgcccgt 6960
ctcactggtg aaaagaaaaa ccaccctggc gcccaatacg caaaccgcct ctccccgcgc 7020
gttggccgat tcattaatgc agctggcacg acaggtttcc cgactggaaa gcgggcagtg 7080
agcgcaacgc aattaatgtg agttagcgcg aattgatctg 7120

<210> 7
<211> 36
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 7
ttctgggccc atgtcattac cgttcttaac ttctgc 36

<210> 8
<211> 35
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 8
gggcacgcgt ctcactatag ggcgaattgg agctc 35

<210> 9
<211> 34
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 9
agtgtcgaca ggaggaatta accatgtcat tacc 34

<210> 10
<211> 32
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 10
tacaagcttc tatgaagtcc atggtaaatt cg 32

<210> 11
<211> 35
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 11
agtgaattca ggaggcagat caaatgtcag agttg 35

<210> 12
<211> 30
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 12
tacgtcgacc tatttatcaa gataagtttc 30

<210> 13
<211> 35
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 13
agtgaattca ggaggtaata ctcatgaccg tttac 35

<210> 14
<211> 31
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 14
tacgtcgact tattcctttg gtagaccagt c 31

<210> 15
<211> 36
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 15
tcagaattca ggaggattac tatatgcaaa cggaac 36

<210> 16
<211> 38
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 16
agagtcgact tatttaagct gggtaaatgc agataatc 38

<210> 17
<211> 34
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 17
tcagaattca ggaggaatga gtaatggact ttcc 34

<210> 18
<211> 36
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 18
agagtcgact tatttattac gctggatgat gtagtc 36

<210> 19
<211> 34
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 19
cgcggatcca ggaggaaata accatgtctc tgcc 34

<210> 20
<211> 34
<212> DNA

<213> Sequência artificial

<220>

<223> Iniciador

<400> 20
ccgcaagctt taagacgtcc acggcagatt ggtg 34

<210> 21
<211> 93
<212> DNA

<213> Sequência artificial

<220>

<223> Promotor

<400> 21
caacgcaatt aatgtgagtt agctcactca ttaggcaccc caggctttac actttatgct 60
tccggctcgt atgttgtgtg gaattgtgag cgg 93

<210> 22
<211> 120
<212> DNA

<213> Sequência artificial

<220>

<223> Promotor

<400> 22
gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg tatcacgagg 60
ccctttcgtc ttcacctcga gaggctttac actttatgct tccggctcgt ataatgtgtg 120

REIVINDICAÇÕES

1. Célula hospedeira capaz de produzir um isoprenóide, a dita célula hospedeira caracterizada pelo fato de que compreende:

pelo menos duas regiões de codificação que compreendem sequências de nucleotídeo que codificam uma mevalonato quinase;

um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto mevalonato quinase; e

um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma terpeno sintase.

2. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de que a uma ou mais enzimas do caminho do mevalonato são uma pluralidade de enzimas da caminho do mevalonato.

3. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de que a uma ou mais enzimas do caminho do mevalonato são selecionadas a partir de uma acetoacetil-CoA tiolase, uma hidroximetilglutaril-CoA sintase (HMGS), uma hidroximetilglutaril-CoA redutase (HMGR), uma fosfomevalonato quinase e uma mevalonato pirofosfato descarboxilase.

4. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de que a terpeno sintase é selecionada a partir de uma amorfa-4,11-dieno sintase; uma beta-cariofileno sintase; uma germacreno A sintase; uma 8-epicedrol sintase; uma valenceno sintase; uma (+)-delta-cadineno sintase; uma germacreno C sintase; uma (E)-beta-farneseno sintase; uma casbeno sintase; uma vetispiradieno sintase; uma 5-epi-aristoloqueno sintase; uma aristolqueno sintase alfa-humuleno sintase; uma (E,E)-alfa-farneseno sintase; uma (-)-beta-pineno sintase; uma gama-terpineno sintase; uma limoneno ciclase; uma linalool sintase; uma 1,8-cineol sintase; uma (+)-sabineno sintase; uma E-alfa-bisaboleno sintase; uma (+)-difosfato de bornila sintase; uma levopimaradieno sintase; uma abietadieno sintase; uma isopimaradieno sintase; uma (E)-gama-bisaboleno sintase; uma taxadieno sintase; uma pirofosfato de copalila sintase; uma kaureno sintase;

uma longifoleno sintase; uma gama-humuleno sintase; uma Delta-selineno sintase; uma beta-felandreno sintase; uma limoneno sintase; uma mirceno sintase, terpinoleno sintase; uma (-)-canfeno sintase; uma (+)-3-careno sintase; uma difosfato de sin-copalila sintase; uma alfa-terpineol sintase; uma sin-pimara-7,15-dieno sintase; uma ent-sanda-aracopimaradieno sintase; uma estemer-13-eno sintase; uma E-beta-ocimeno; uma S-linalool sintase; uma geraniol sintase; uma gama-terpineno sintase; uma linalool sintase; uma E-beta-ocimeno sintase; uma epi-cedrol sintase; uma alfa-zingibereno sintase; uma guaiadieno sintase; uma cascariladieno sintase; uma cis-muroladieno sintase; uma afidicolina-16b-ol sintase; uma elizabethatrieno sintase; uma sandalol sintase; uma patchoulol sintase; uma zinzanol sintase; uma cedrol sintase; uma escareol sintase; uma copalol sintase e uma manool sintase.

5. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de que adicionalmente compreende uma sequência de nucleotídeo que codifica uma pirofosfato de isopentenila (IPP) isomerase.

6. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de que adicionalmente compreende um ácido nucléico heterólogo que compreende uma sequência de nucleotídeo que codifica uma preniltransferase selecionada a partir de uma difosfato de farnesila sintase, uma difosfato de geranila sintase e uma difosfato de geranilgeranila sintase.

7. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de que o isoprenóide é um diterpeno, um triterpeno ou um sesquiterpeno.

8. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de que uma ou mais das sequências de nucleotídeo que codificam a mevalonato quinase ou enzimas do caminho do mevalonato é códon-otimizada para expressão na dita célula hospedeira.

9. Célula hospedeira, de acordo com a reivindicação 1, caracterizada pelo fato de ser uma célula procariótica.

10. Célula hospedeira, de acordo com a reivindicação 1,

caracterizada pelo fato de que a célula hospedeira é uma célula eucariótica,

11. Célula hospedeira capaz de produzir um isoprenóide, caracterizada pelo fato de que compreende:

uma região de codificação que codifica uma mevalonato quinase (MK), em que a região de codificação de MK está sob controle de um primeiro promotor;

um ácido nucléico que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho do mevalonato, exceto mevalonato quinase, sob controle de um segundo promotor, em que o segundo promotor é mais fraco do que o primeiro promotor; e

uma sequência de nucleotídeo heteróloga que codifica uma terpeno sintase.

12. Célula hospedeira, de acordo com a reivindicação 11, caracterizada pelo fato de que o primeiro promotor é um promotor constitutivamente ativo.

13. Célula hospedeira, de acordo com a reivindicação 11, caracterizada pelo fato de que o primeiro promotor é um promotor induzível.

14. Célula hospedeira, de acordo com a reivindicação 11, caracterizada pelo fato de que o primeiro promotor é mais forte do que um promotor Lac nativo de *Escherichia coli* ou um promotor *PLac*.

15. Célula hospedeira, de acordo com a reivindicação 11, caracterizada pelo fato de que o primeiro promotor é um promotor *lac*UV5.

16. Célula hospedeira, de acordo com a reivindicação 11, caracterizada pelo fato de que o primeiro promotor é um promotor trc.

17. Célula hospedeira de *Escherichia coli* capaz de produzir um isoprenóide, caracterizada pelo fato de que compreende:

um primeiro plasmídeo de expressão que compreende uma sequência de nucleotídeo que codifica uma mevalonato quinase;

um segundo plasmídeo de expressão que compreende uma sequência de nucleotídeo que codifica uma ou mais enzimas do caminho MEV, exceto mevalonato quinase, em que o número de cópia do segundo plasmídeo de expressão é mais baixo do que aquele do primeiro plasmídeo

de expressão; e

uma sequência de nucleotídeo heteróloga que codifica uma terpeno sintase.

18. Célula hospedeira, de acordo com a reivindicação 17, caracterizada pelo fato de que o primeiro plasmídeo de expressão compreende uma origem p15A de replicação.

19. Célula hospedeira, de acordo com a reivindicação 17, caracterizada pelo fato de que a terpeno sintase é selecionada dentre uma amorfa-4,11-dieno sintase; uma beta-cariofileno sintase; uma germacreno A sintase; uma 8-epicedrol sintase; uma valenceno sintase; uma (+)-delta-cadineno sintase; uma germacreno C sintase; uma (E)-beta- farneseno sintase; uma casbeno sintase; uma vetispiradieno sintase; uma 5-epi-aristoloqueno sintase; uma aristolqueno sintase alfa-humuleno sintase; uma (E,E)-alfa-farneseno sintase; uma (-)-beta-pineno sintase; uma gama-terpineno sintase; uma limoneno ciclase; uma linalool sintase; uma 1,8-cineol sintase; uma (+)-sabineno sintase; uma E-alfa-bisaboleno sintase; uma (+)-difosfato de bornila sintase; uma levopimaradieno sintase; uma abietadieno sintase; uma isopimaradieno sintase; uma (E)-gama-bisaboleno sintase; uma taxadieno sintase; uma pirofosfato de copalila sintase; uma kaureno sintase; uma longifoleno sintase; uma gama-humuleno sintase; uma Delta-selineno sintase; uma beta-felandreno sintase; uma limoneno sintase; uma mirceno sintase, terpinoleno sintase; uma (-)-canfeno sintase; uma (+)-3-careno sintase; uma difosfato de sin-copalila sintase; uma alfa-terpineol sintase; uma sin-pimara-7,15-dieno sintase; uma ent-sanda-aracopimaradieno sintase; uma estemer-13-eno sintase; uma E-beta-ocimeno; uma S-linalool sintase; uma geraniol sintase; uma gama-terpineno sintase; uma linalool sintase; uma E-beta-ocimeno sintase; uma epi-cedrol sintase; uma alfa-zingibereno sintase; uma guaiadieno sintase; uma cascariladieno sintase; uma cis-muroladieno sintase; uma afidicolina-16b-ol sintase; uma elizabethatrieno sintase; uma sandalol sintase; uma patchoulol sintase; uma zinzanol sintase; uma cedrol sintase; uma escareol sintase; uma copalol sintase e uma manool sintase.

20. Célula hospedeira, de acordo com a reivindicação 17, caracterizada pelo fato de que o isoprenóide é um diterpeno, um triterpeno ou um sesquiterpeno.

FIG. 1

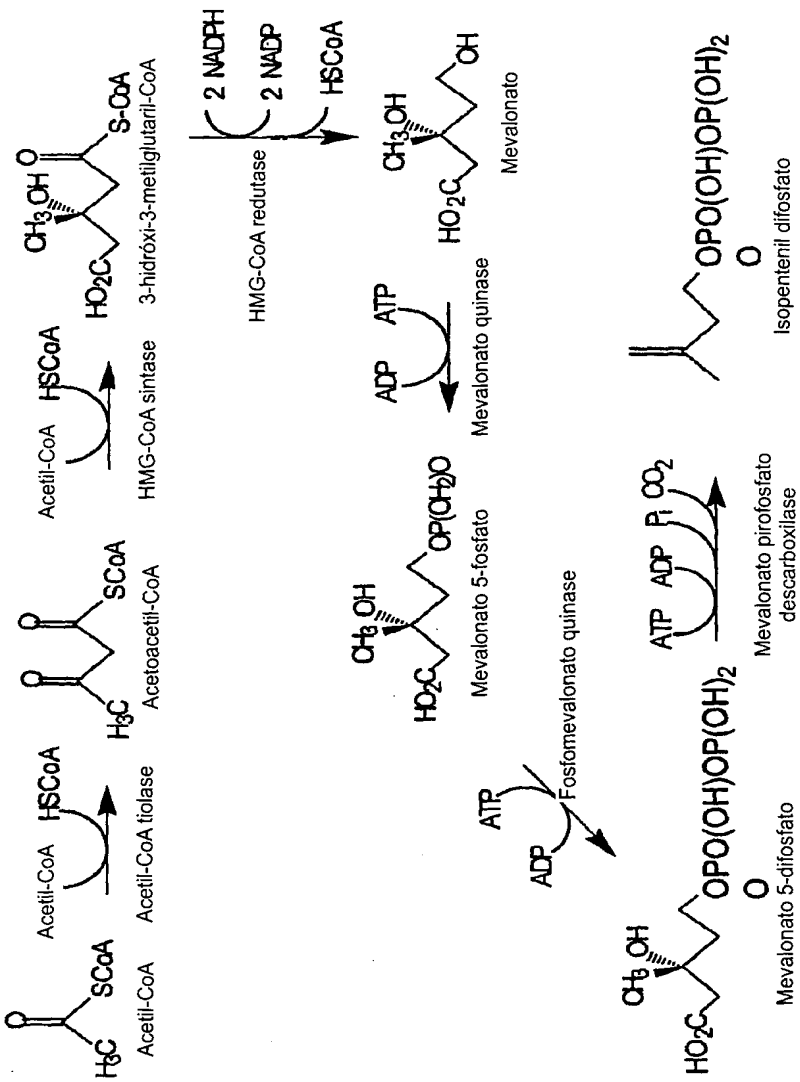

Acetil-CoA
Acetil-CoA tiolase
Acetoacetil-CoA
HMG-CoA sintase
3-hidróxi-3-metilglutaril-CoA
HMG-CoA redutase
2 NADPH
2 NADP
HSCoA
Mevalonato
Mevalonato quinase
ATP
ADP
Mevalonato 5-fosfato
Fosfomevalonato quinase
Mevalonato 5-difosfato
Mevalonato pirofosfato descarboxilase
Isopentenil difosfato

FIG. 2

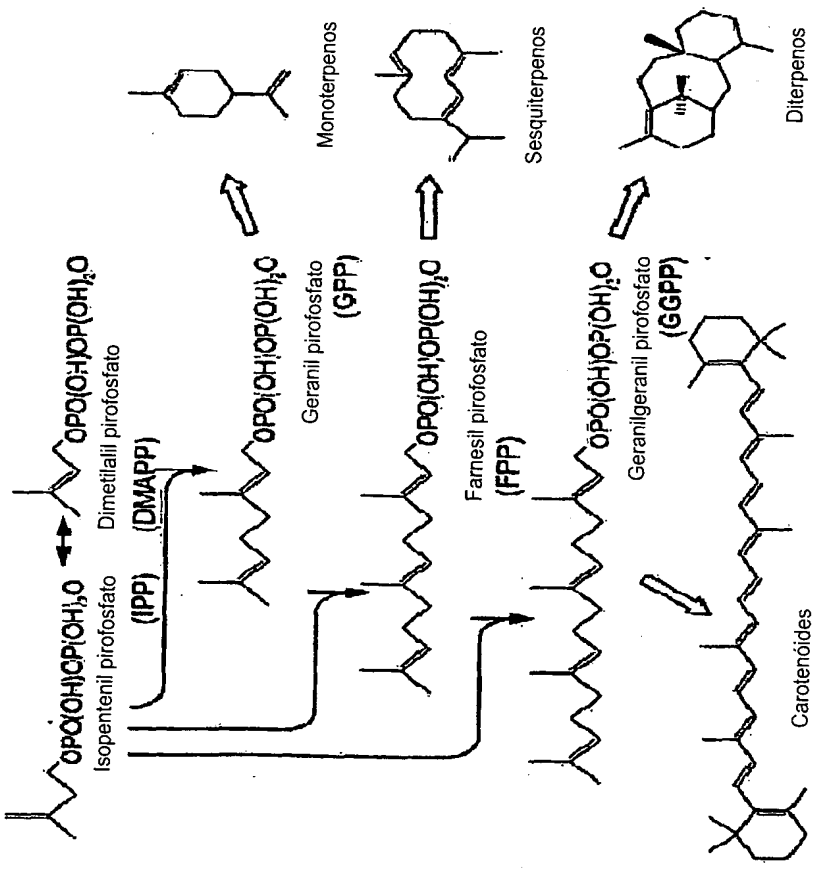

Isopentenil pirofosfato
(IPP)
Dimetilalil pirofosfato
(DMAPP)
Geranil pirofosfato
(GPP)
Farnesil pirofosfato
(FPP)
Geranilgeranil pirofosfato
(GGPP)
Monoterpenos
Sesquiterpenos
Diterpenos
Carotenóides

FIG. 3

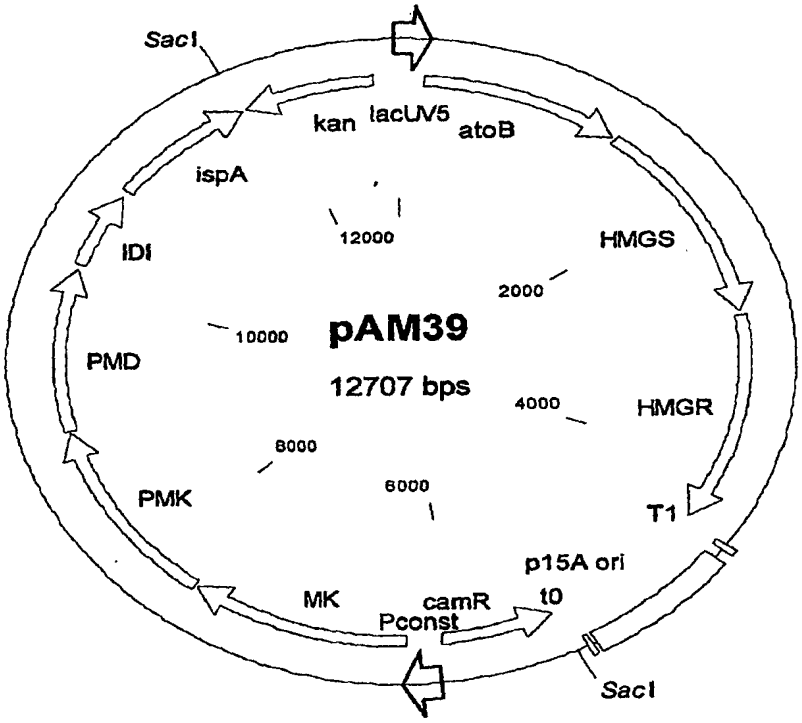

SacI
kan
lacUV5
atoB
ispA
IDI
12000
HMGS
2000
10000
pAM39
12707 bps
PMD
4000
HMGR
8000
6000
PMK
T1
p15A ori
MK
camR
t0
Pconst
SacI

FIG. 4

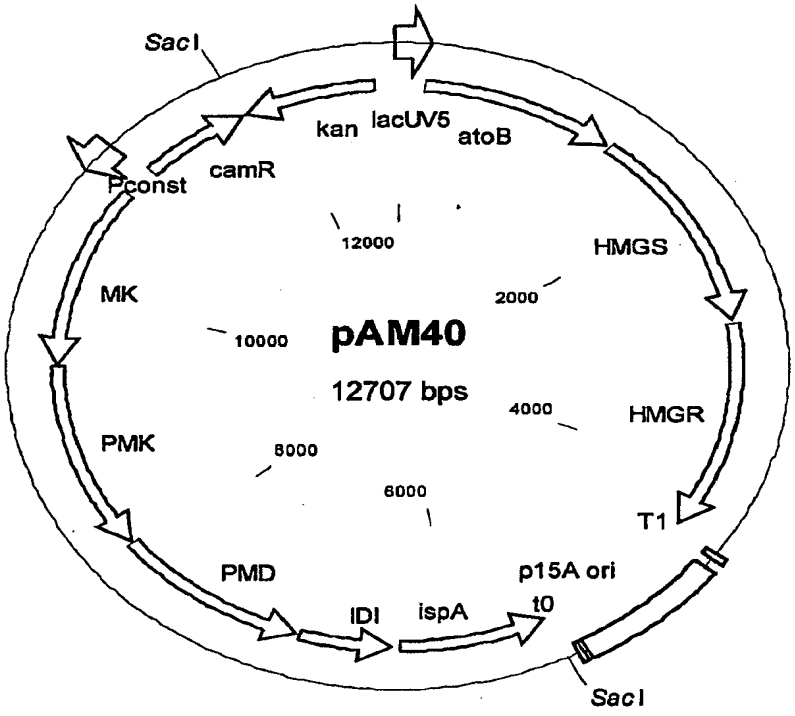

SacI
kan
lacUV5
atoB
camR
Pconst
12000
HMGS
2000
MK
10000
pAM40
12707 bps
4000
HMGR
PMK
8000
6000
T1
PMD
p15A ori
IDI
ispA
t0
SacI

FIG. 5

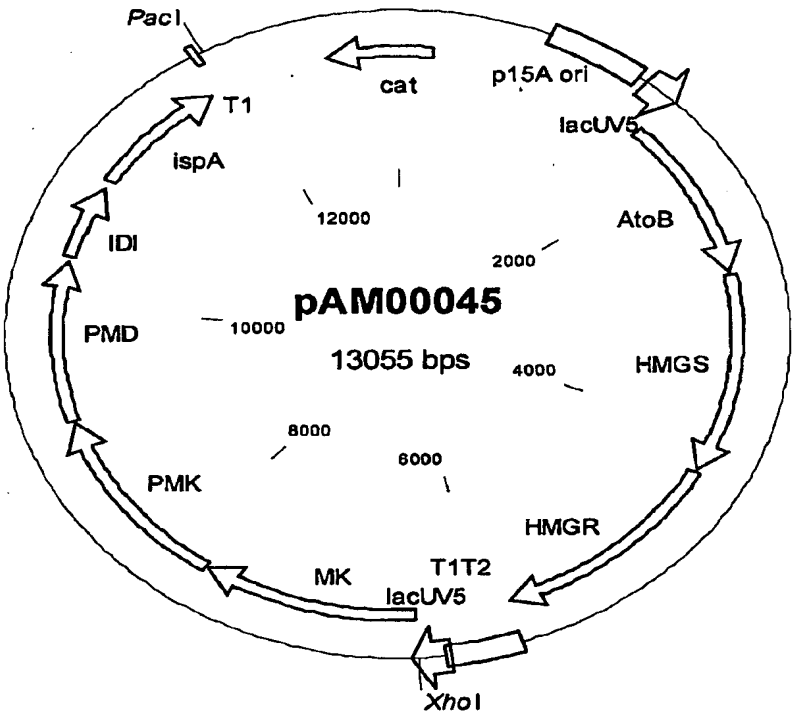

PacI
cat
p15A ori
T1
lacUV5
ispA
12000
AtoB
IDI
2000
pAM00045
PMD
10000
13055 bps
4000
HMGS
8000
6000
PMK
HMGR
MK
T1T2
lacUV5
XhoI

FIG. 6

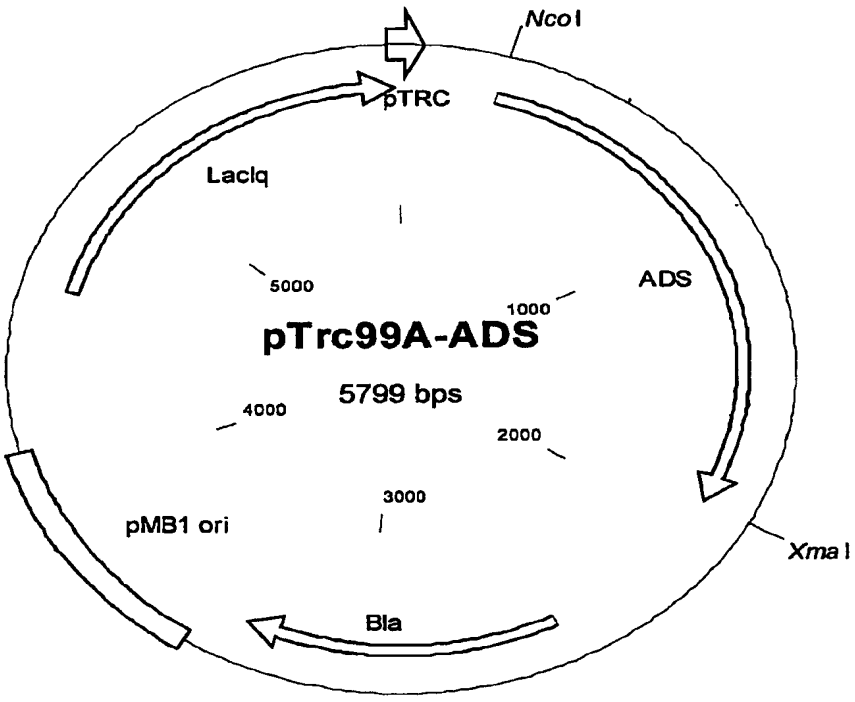

NcoI
pTRC
LacIq
5000
1000
ADS
pTrc99A-ADS
5799 bps
4000
2000
3000
pMB1 ori
XmaI
Bla

FIG. 7

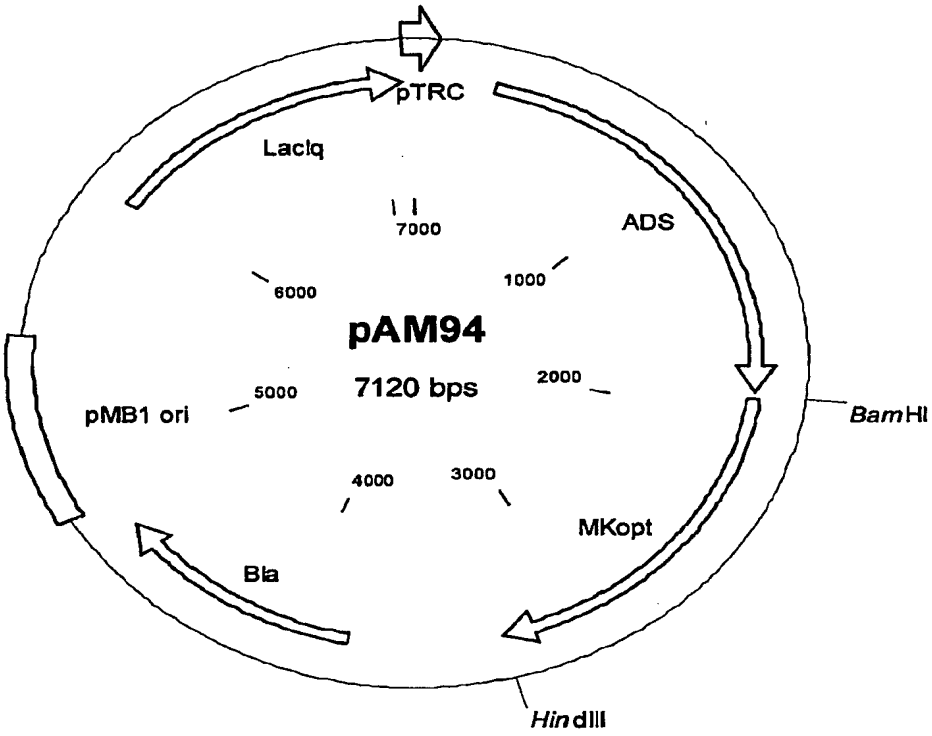

pTRC
LacIq
7000
ADS
6000
1000
pAM94
7120 bps
5000
2000
pMB1 ori
BamHI
4000
3000
MKopt
Bla
HindIII

FIG. 8A

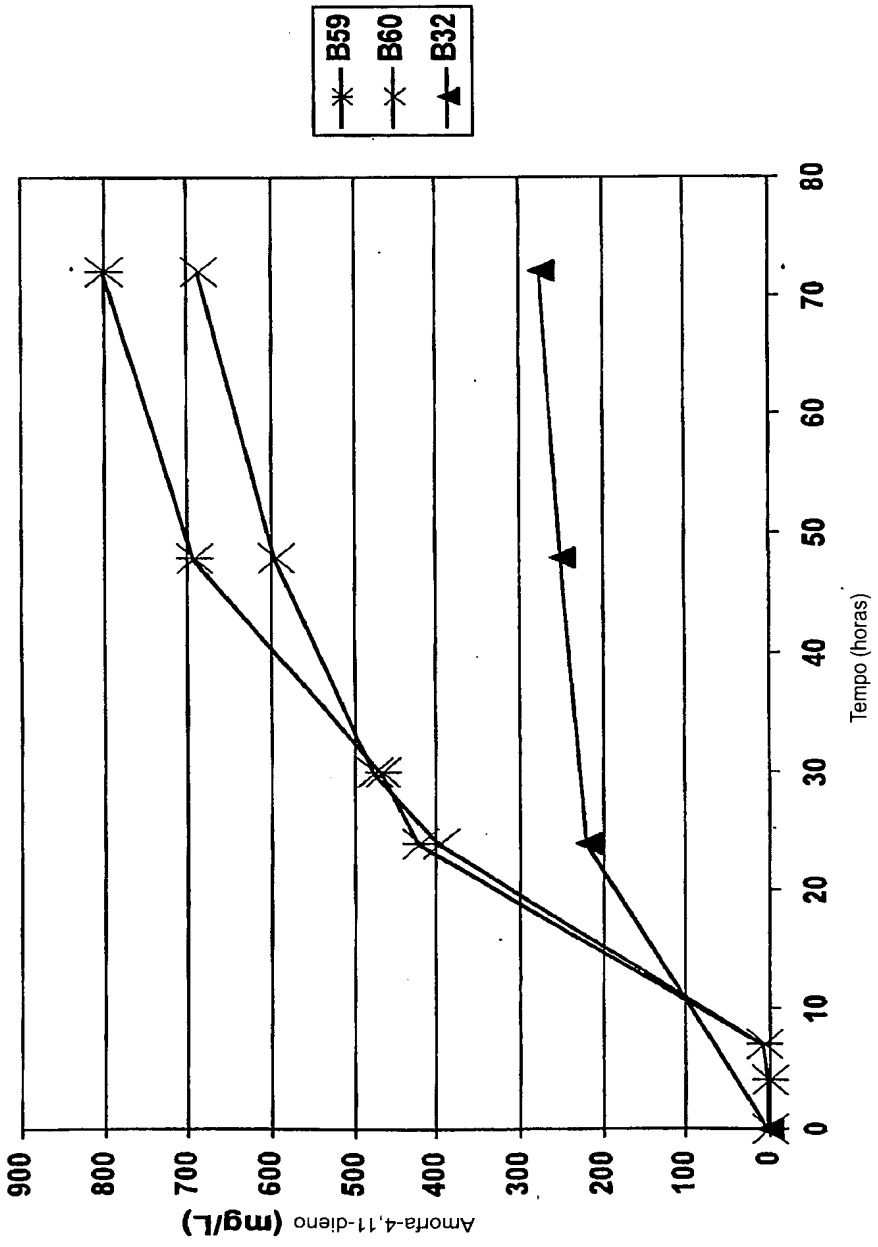

B59
B60
B32
Amorfa-4,11-dieno (mg/L)
900
800
700
600
500
400
300
200
100
0
0
10
20
30
40
50
60
70
80
Tempo (horas)

FIG. 8B

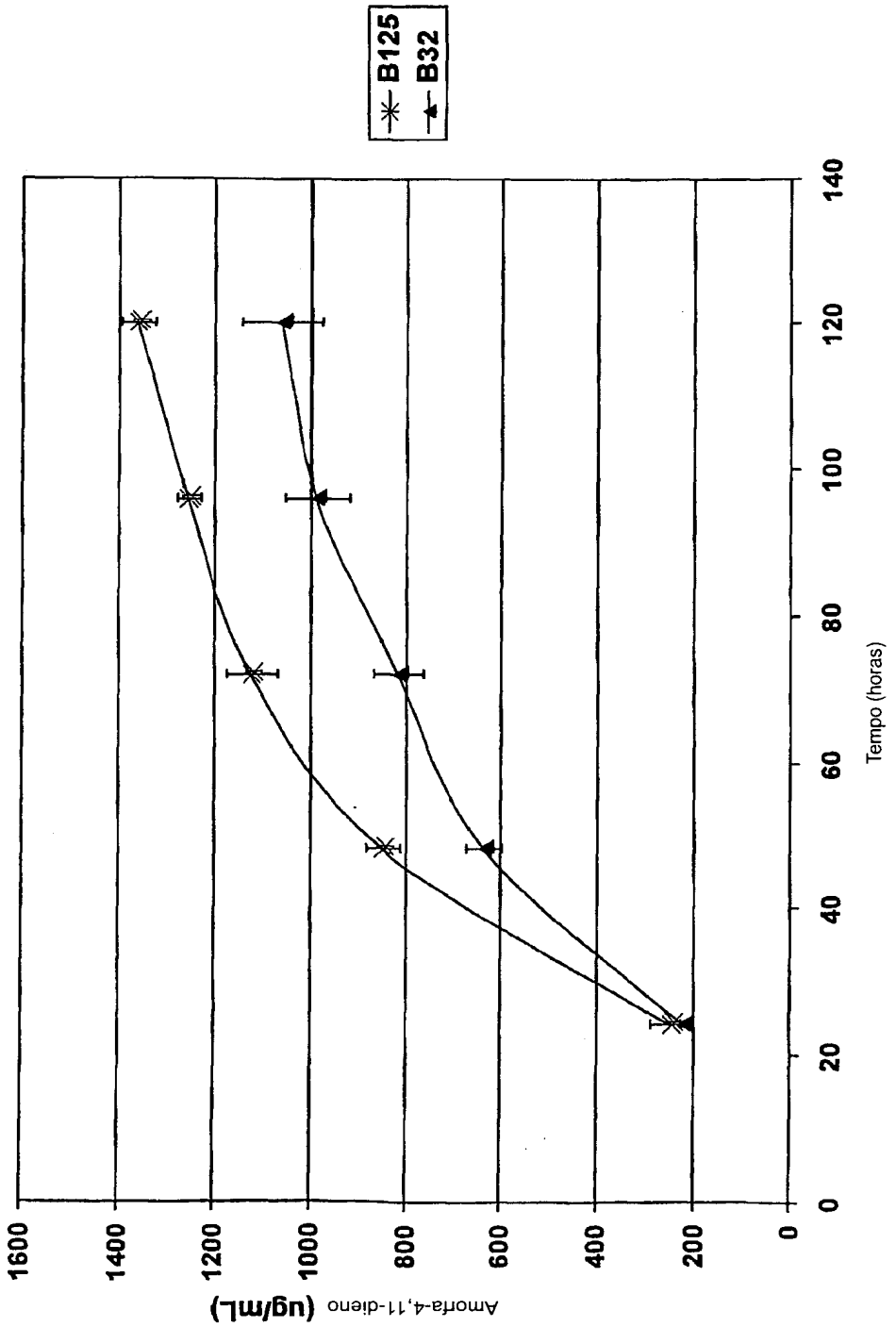

B125
B32
Amorfa-4,11-dieno (ug/mL)
0
200
400
600
800
1000
1200
1400
1600
0
20
40
60
80
100
120
140
Tempo (horas)

FIG. 9

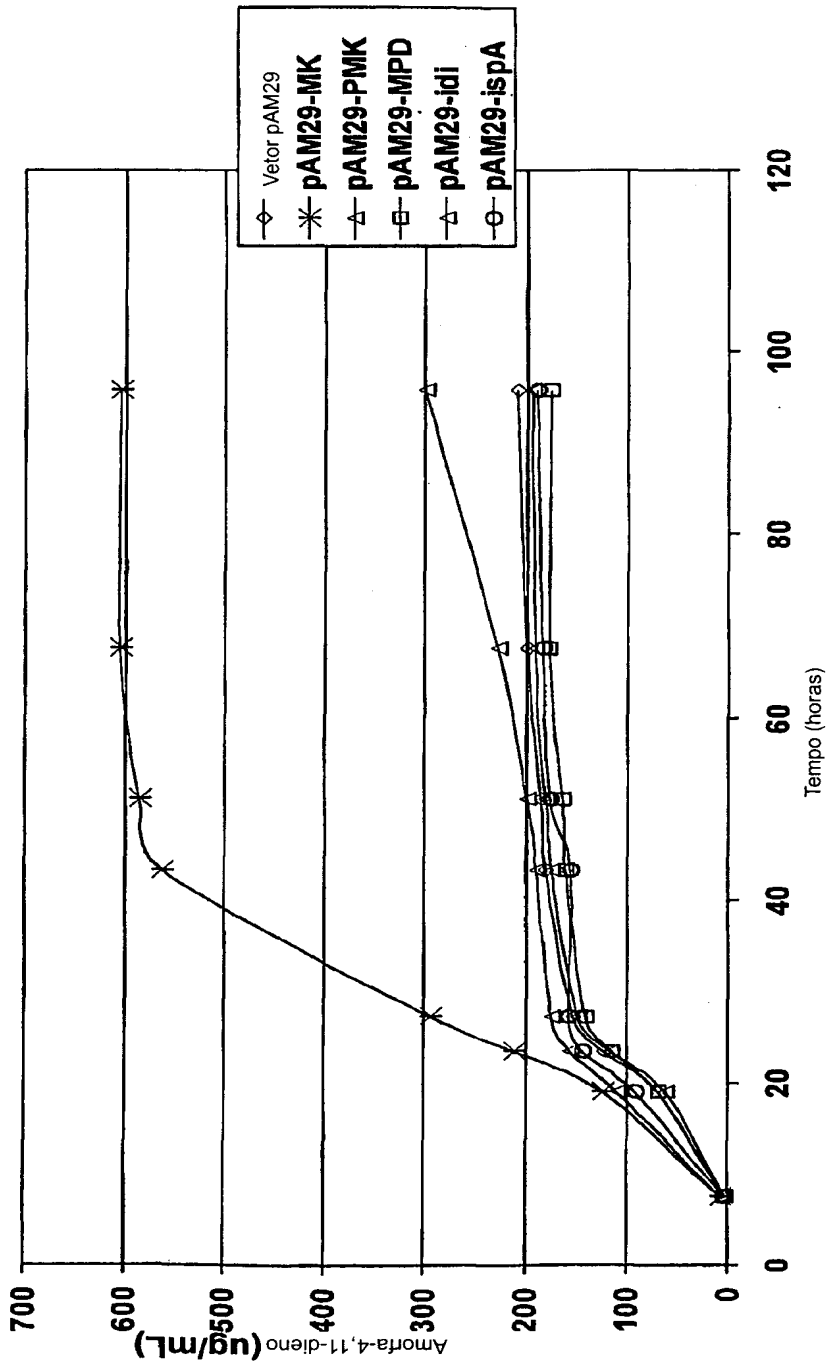
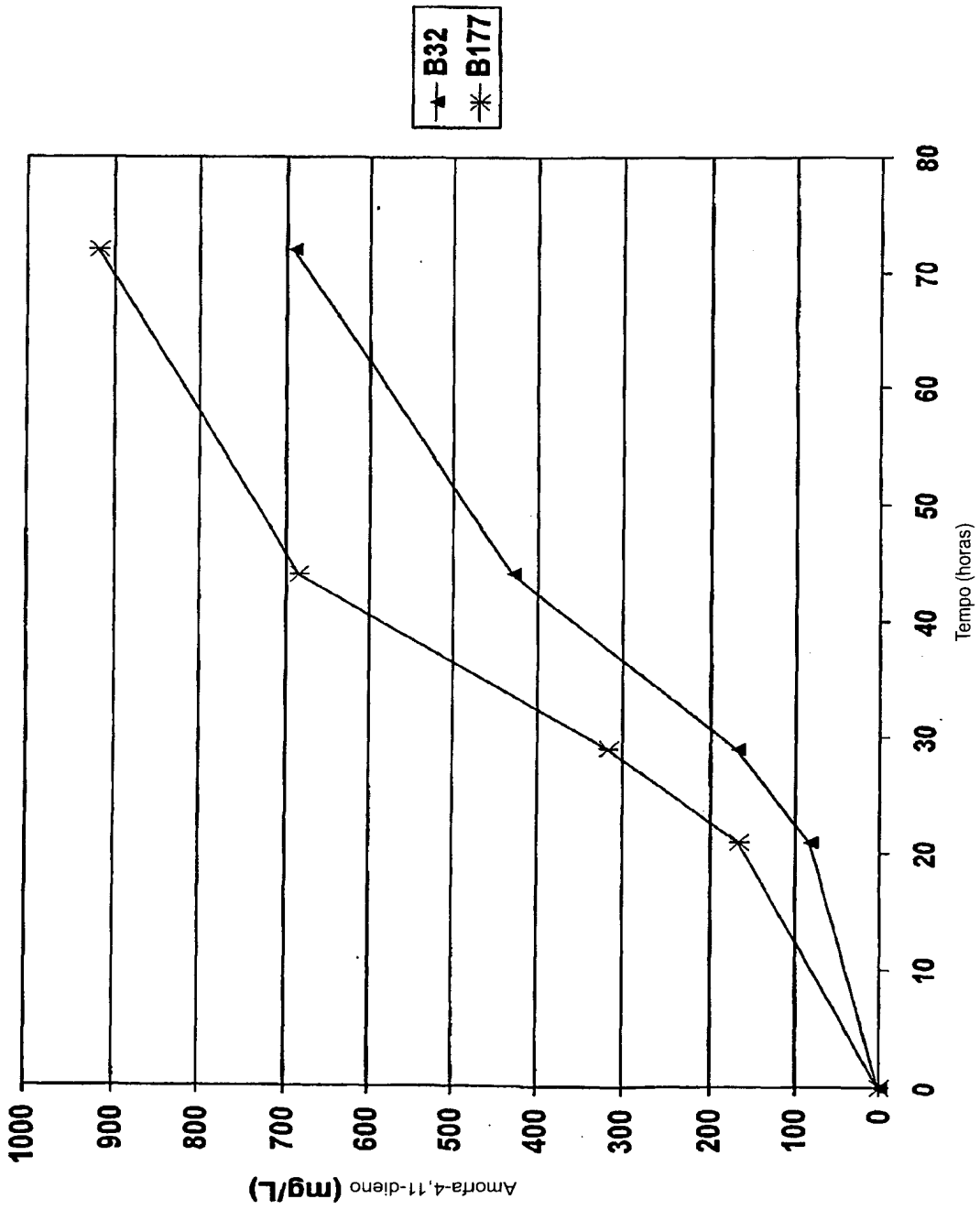

Vetor pAM29
pAM29-MK
pAM29-PMK
pAM29-MPD
pAM29-idi
pAM29-ispA
Amorfa-4,11-dieno (ug/mL)
0
100
200
300
400
500
600
700
0
20
40
60
80
100
120
Tempo (horas)

FIG. 10


B32
B177
Amorfa-4,11-dieno (mg/L)
1000
900
800
700
600
500
400
300
200
100
0
0
10
20
30
40
50
60
70
80
Tempo (horas)

RESUMO

Patente de Invenção: **"CÉLULAS HOSPEDEIRAS CAPAZES DE PRODUZIR ISOPRENÓIDES".**

A presente invenção refere-se a células hospedeiras geneticamente modificadas e ao uso das mesmas para a produção de compostos isoprenóides.