



(12)发明专利

(10)授权公告号 CN 103667370 B

(45)授权公告日 2016.08.10

(21)申请号 201310335874.8

(22)申请日 2006.06.29

(30)优先权数据

60/697,067 2005.07.05 US

(62)分案原申请数据

200680024666.4 2006.06.29

(73)专利权人 加利福尼亚大学董事会

地址 美国加利福尼亚州

(72)发明人 D·-K·罗 K·纽曼

E·M·帕若蒂斯 J·D·基斯林

M·奥莱特 R·伊切斯 K·霍

T·哈姆

(74)专利代理机构 上海专利商标事务所有限公

司 31100

代理人 余颖

(51)Int.Cl.

C12P 7/40(2006.01)

C12P 7/02(2006.01)

C12N 15/82(2006.01)

C12N 1/19(2006.01)

C12N 5/10(2006.01)

C12N 15/53(2006.01)

C12N 15/63(2006.01)

(56)对比文件

WO 93/21326 A2,1993.10.28,

WO 02/072758 A2,2002.09.19,

EP 1354955 A1,2003.10.22,

Jennewein Stefan et al.Coexpression
in yeast of taxus cytochrome P450
reductase with cytochrome P450 oxygenases
involved in taxol biosynthesis.

《Biotechnology and Bioengineering》.2005,
第89卷(第5期),588-598.

审查员 李影

权利要求书1页 说明书34页

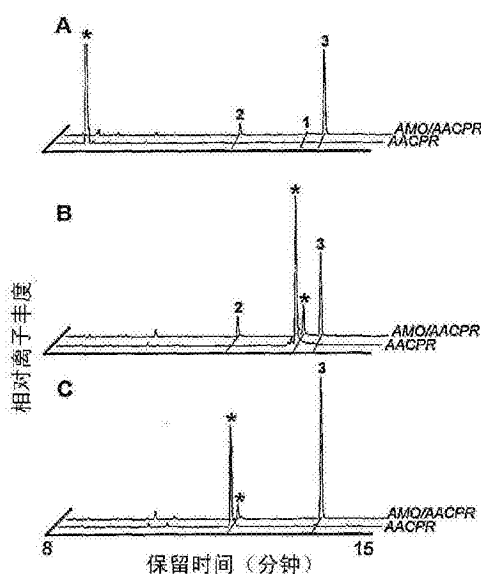
序列表12页 附图12页

(54)发明名称

编码类异戊烯修饰酶的多核苷酸和其使用
方法

(57)摘要

本发明提供了包含编码类异戊烯修饰酶的核苷酸序列的分离核酸,以及含有该核酸的重组载体。本发明还提供了包含本发明核酸或重组载体的遗传修饰的宿主细胞。本发明还提供了包含本发明核酸的转基因植物。本发明还提供了生产类异戊烯化合物的方法,所述方法通常包括在能够合成本发明核酸编码的类异戊烯化合物修饰酶的条件下培养本发明遗传修饰的宿主细胞。



1. 一种在宿主细胞内修饰类异戊烯化合物的方法,所述方法包括:

用合适培养基培养遗传修饰的宿主细胞,所述宿主细胞经编码细胞色素P450还原酶的核酸遗传修饰,所述核酸编码氨基酸序列如SEQ ID NO:4所示的细胞色素P450还原酶;

类异戊烯化合物存在下,所述细胞色素P450还原酶将电子由NADPH传递给倍半萜羟化酶,所述宿主细胞还经编码倍半萜羟化酶的核酸遗传修饰,所述倍半萜羟化酶的氨基酸序列如SEQ ID NO:6所示。

2. 如权利要求1所述的方法,所述宿主细胞是真核宿主细胞。

3. 如权利要求2所述的方法,所述宿主细胞是酵母细胞。

4. 如权利要求3所述的方法,所述酵母细胞是酿酒酵母。

5. 如权利要求2所述的方法,所述宿主细胞是植物细胞。

6. 一种制造转基因植物的方法,所述转基因植物经含有以下所述核苷酸序列的核酸遗传修饰,所述核苷酸序列编码氨基酸序列如SEQ ID NO:4所示的细胞色素P450还原酶,其中,该细胞色素P450还原酶将电子由NADPH传递给类异戊烯修饰酶。

7. 如权利要求6所述的方法,所述植物是单子叶植物。

8. 如权利要求6所述的方法,所述植物是双子叶植物。

9. 如权利要求6所述的方法,所述植物是烟草。

10. 一种载体,含有编码细胞色素P450还原酶的聚核苷酸,所述细胞色素P450还原酶的氨基酸序列如SEQ ID NO:4所示,所述载体还含有编码类异戊烯化合物修饰酶的聚核苷酸,所述类异戊烯化合物修饰酶的氨基酸序列如SEQ ID NO:2所示。

11. 一种宿主细胞,含有编码细胞色素P450还原酶的聚核苷酸,所述细胞色素P450还原酶的氨基酸序列如SEQ ID NO:4所示,所述细胞还含有编码类异戊烯化合物修饰酶的聚核苷酸,所述类异戊烯化合物修饰酶的氨基酸序列如SEQ ID NO:2所示,所述宿主细胞是:(i)原核宿主细胞;(ii)真核宿主细胞,但不是植物细胞;或者(iii)植物宿主细胞,所述植物宿主细胞不用作繁殖材料。

编码类异戊烯修饰酶的多核苷酸和其使用方法

[0001] 本申请是国际申请PCT/2006/025572,国际申请日2006年6月29日,中国国家阶段申请号200680024666.4的发明专利申请的分案申请。

[0002] 交叉参考

[0003] 本申请要求2005年7月5日提交的美国临时专利申请60/697,067的优先权,将该申请的全文纳入本文作参考。

技术领域

[0004] 本发明涉及类异戊烯(isoprenoid)化合物生产领域,具体是修饰类异戊烯化合物的酶。

背景技术

[0005] 类异戊烯构成了极大并且多样的一类天然产物,它们具有相同的生物合成来源,即一种代谢前体二磷酸异戊-1-烯酯(isopentenyl diphosphate, IPP)。已经描述了至少20,000种类异戊烯。由定义可以看出,类异戊烯由所谓的异戊烯(isoprene)(C₅)单元组成。类异戊烯中存在的C原子的数量一般可被5除尽(C₅、C₁₀、C₁₅、C₂₀、C₂₅、C₃₀和C₄₀),但也报道了不规则的类异戊烯和多萜。类异戊烯化合物也称为“萜”或“类萜”。类异戊烯的重要成员包括类胡萝卜素、倍半萜类、双萜类和单萜。类胡萝卜素包括例如:番茄红素,β-胡萝卜素等许多用作抗氧化剂的物质。倍半萜类包括例如:一种具有抗疟活性的化合物青蒿素。双萜类包括例如:一种癌症化疗药紫杉醇。

[0006] 类异戊烯包括许多结构各异的天然产物家族。在这个家族中,分离自植物和其它天然来源的萜类化合物用作商业性调味剂和香料化合物,以及药物化合物如抗疟药、抗病毒药和抗癌药。目前使用的大多数萜类化合物是天然产物或其衍生物。许多这些天然产物的来源生物体(如树、海洋无脊椎动物)既不适合生产商业上可行量所需的大规模培育,也不适合遗传操作以提高这些化合物的产量或衍生这些化合物。因此,必须由类似物半合成地产生,或用常规化学合成方法合成这些天然产物。而且,许多天然产物具有复杂结构,因此,目前合成这些天然产物不具经济性或不可能合成。这些天然产物必须由天然来源如树、海绵、珊瑚和海洋微生物提取;或由较丰富的前体合成或半合成地产生。由天然来源提取天然产物受限于天然来源的可用性;合成或半合成地生产天然产物可能遇到低产率和/或高成本的困扰。这些生产问题和天然来源的局限性可限制这种产物的商业和临床开发。

[0007] 倍半萜化合物的一个重要例子是青蒿素。青蒿素是非常有效的抗疟药,目前提取自植物(黄花蒿(*Artemisia annua*)),用于联合治疗药物疗法。植物衍生的青蒿素昂贵,并且其可获得性收到栽培这种植物的国家的天气和政治环境的影响。青蒿酸(artemisinic acid)是青蒿素生物合成中的关键中间体。通过传统化学方法将紫穗槐-4,11-二烯转化为青蒿醇是青蒿素制备中的重要步骤,该步骤难以进行并且成本高。

[0008] 本领域需要能避免一部分上述缺点的产生类异戊烯化合物的方法。本发明通过提供编码修饰类异戊烯化合物的酶的多核苷酸和经遗传修饰能生产这种酶的宿主细胞满足

了这种需求。

[0009] 文献

[0010] Berteau等(2005)Planta Med.71:40-47;deKraker等(2003)Tetradedron59:409-418;Martin等(2003)Nat.Biotechnol.21:796-802;W003/025193;美国专利公开号20050019882;美国专利公开号20030148479;美国专利公开号20040005678;美国专利公开号20030166255。

发明内容

[0011] 本发明提供了包含编码类异戊烯修饰酶的核苷酸序列的分离核酸,以及含有该核酸的重组载体。本发明还提供了包含本发明核酸或重组载体的遗传修饰的宿主细胞。本发明还提供了包含本发明核酸的转基因植物。本发明还提供了生产类异戊烯化合物的方法,所述方法通常包括在能够合成本发明核酸编码的酶的条件下培养本发明遗传修饰的宿主细胞,该酶能修饰类异戊烯化合物。

附图说明

- [0012] 图1描述了CYP71D-A4cDNA编码序列的核苷酸序列(SEQ ID NO:1)。
- [0013] 图2描述了紫穗槐二烯12-氧化酶氨基酸序列(SEQ ID NO:2)。
- [0014] 图3描述了黄花蒿细胞色素P450还原酶cDNA编码区的核苷酸序列(SEQ ID NO:3)。
- [0015] 图4描述了黄花蒿细胞色素P450还原酶氨基酸序列(SEQ ID NO:4)。
- [0016] 图5A-C描述了体内底物进料实验的结果。
- [0017] 图6A和6B描述了由GC-MS进行产物验证。
- [0018] 图7A-C描述了在酵母中从头生产青蒿酸。
- [0019] 图8A-C描述了体外紫穗槐二烯氧化酶实验。
- [0020] 图9描述了编码类异戊烯修饰酶的cDNA的核苷酸序列(克隆71D-B1)(SEQ ID NO:5)。
- [0021] 图10描述了类异戊烯修饰酶的氨基酸序列(71D-B1;SEQ ID NO:6)。
- [0022] 图11A-C描述了酶71D-B1的羟基化活性。
- [0023] 图12描述了编码类异戊烯修饰酶的基因组DNA的核苷酸序列(SEQ ID NO:7)。
- [0024] 图13是类异戊烯代谢途径的示意图,该代谢途径由二磷酸异戊-1-烯酯(IPP)和二磷酸二甲基烯丙酯(DMAPP)生产类异戊烯生物合成途径中间体二磷酸多异戊烯酯(polyprenyl diphosphate)、二磷酸牻牛儿基酯(GPP)、二磷酸法呢酯(FPP)和二磷酸牻牛儿基牻牛儿基酯(GGPPP)。
- [0025] 图14是产生IPP的甲羟戊酸(MEV)途径的示意图。
- [0026] 图15是用于产生IPP和焦磷酸二甲基烯丙酯(DMAPP)的DXP途径的示意图。

具体实施方式

[0027] 定义

[0028] 术语“类异戊烯”、“类异戊烯化合物”、“萜”、“萜化合物”、“类萜”和“类萜化合物”可互换使用。类异戊烯化合物由不同数量的所谓的异戊烯(C5)单位组成。类异戊烯中存在

的碳原子数目一般可被5整除(如C5、C10、C15、C20、C25、C30和C40)。曾报道过不规则的类异戊烯和多萜,它们也包括在“类异戊烯”的定义中。类异戊烯化合物包括但不限于:单萜、倍半萜、三萜、多萜和二萜。

[0029] 本文所用术语“二磷酸异戊烯酯(prenyl diphosphate)”与“焦磷酸异戊烯酯(prenyl pyrophosphate)”可互换使用,包括含有一个异戊烯基的二磷酸单异戊烯酯(如IPP和DMAPP),以及含有两个或多个异戊烯基的二磷酸多异戊烯酯。二磷酸单异戊烯酯包括焦磷酸异戊-1-烯酯(IPP)和其异构体焦磷酸二甲基烯丙酯(DMAPP)。

[0030] 本文所用术语“萜合酶”指能酶学修饰IPP、DMAPP或焦磷酸多异戊烯酯,以便产生萜类化合物的任何酶。术语“萜合酶”包括能催化二磷酸异戊烯酯转化为类异戊烯的酶。

[0031] 在本文中,术语“焦磷酸”可与“二磷酸”互换使用。因此,例如,术语“二磷酸异戊烯酯”和“焦磷酸异戊烯酯”可互换;术语“焦磷酸异戊-1-烯酯”和“二磷酸异戊-1-烯酯”可互换;术语“二磷酸法呢酯”和“焦磷酸法呢酯”可互换;等等。

[0032] 本文所用术语“甲羟戊酸途径”或“MEV途径”指将乙酰-CoA转变为IPP的生物合成途径。甲羟戊酸途径包括催化以下步骤的酶:(a)使两个乙酰-CoA分子缩合成乙酰乙酰CoA;(b)使乙酰乙酰CoA与乙酰CoA缩合形成HMG-CoA;(c)将HMG-CoA转化为甲羟戊酸;(d)将甲羟戊酸磷酸化为甲羟戊酸5-磷酸;(e)将甲羟戊酸5-磷酸转化为甲羟戊酸5-焦磷酸;和(f)将甲羟戊酸5-焦磷酸转化为焦磷酸异戊-1-烯酯。图14示意性说明了甲羟戊酸途径。甲羟戊酸途径的“上半部分”指负责通过MEV途径中间体将乙酰-CoA转变为甲羟戊酸的酶。

[0033] 本文所用术语“1-脱氧-D-木酮糖5-二磷酸途径”或“DXP途径”指通过DXP途径中间体将甘油醛-3-磷酸和丙酮酸转变为IPP和DMAPP的途径,其中DXP途径包括催化图15示意性说明的反应的酶。

[0034] 本文所用术语“异戊烯基转移酶”与术语“二磷酸异戊烯酯合酶”和“多异戊烯基合酶”(如“GPP合酶”、“FPP合酶”、“OPP合酶”等)可互换使用,指能催化二磷酸异戊-1-烯酯与烯丙基初始底物(primer substrate)的连续1'-4'缩合,导致形成各种链长度的二磷酸异戊烯酯的酶。

[0035] 术语“多核苷酸”和“核酸”在本文中可互换使用,指聚合形式的任何长度的核苷酸,包括核糖核苷酸或脱氧核苷酸。因此,该术语包括但不限于:单链、双链或多链DNA或RNA,基因组DNA,cDNA,DNA-RNA杂交体,或含有嘌呤和嘧啶碱基或其它天然的、化学或生化修饰的非天然的、或衍生的核苷酸碱基的聚合物。

[0036] 术语“肽”、“多肽”和“蛋白质”在本文中可互换使用,指聚合形式的任何长度的氨基酸,可包括编码和非编码的氨基酸,化学或生化修饰或衍生的氨基酸,以及含有修饰的肽主链的多肽。

[0037] 本文所用术语“天然产生”可应用于核酸、细胞或生物体,指在天然情况下发现的核酸、细胞或生物体。例如,可由天然来源分离并且未经实验室人员有意修饰的生物体(包括病毒)中存在的多肽或多核苷酸序列是天然产生的。

[0038] 本文所用术语“分离”指与多核苷酸、多肽或细胞天然存在环境不同的环境中的多核苷酸、多肽或细胞。分离的遗传修饰的宿主细胞可存在于遗传修饰的宿主细胞的混合群体中。

[0039] 本文所用术语“外源性核酸”指通常或天然情况下没有在给定的天然细菌、生物体

或细胞中发现和/或产生的核酸。本文所用术语“内源性核酸”指通常在给定的天然细菌、生物体或细胞中发现和/或产生的核酸。“内源性核酸”也称为“天然核酸”或对于给定的细菌、生物体或细胞“天然”的核酸。例如,编码HMGS、甲羟戊酸激酶和磷酸甲羟戊酸激酶的核酸代表大肠杆菌(*E.coli*)的外源性核酸。可由酿酒酵母(*Saccharomyces cerevisiae*)克隆这些甲羟戊酸途径核酸。在酿酒酵母中,染色体上编码HMGS、MK和PMK的基因序列为“内源性”核酸。

[0040] 本文所用术语“异源核酸”指至少满足以下一种条件的核酸:(a)核酸是给定的宿主微生物或宿主细胞以外的(“外源性”)(即不是天然发现的);(b)核酸包含在给定宿主微生物或宿主细胞中天然发现的(如“内源性”)核苷酸序列(如核酸包含宿主微生物或宿主细胞内源性核苷酸序列),但该核酸在细胞中的产量是非天然量(如大于预期产量或大于天然情况下的产量);或者该核酸的序列与内源性核苷酸序列不同,以致在细胞中产生非天然量(如大于预期产量或大于天然情况下的产量)的内源发现的同一编码蛋白(氨基酸序列相同或基本相同);(c)核酸包含在天然情况下相互关系不同的两种或多种核苷酸序列或节段,如核酸是重组核酸。

[0041] 本文所用术语“重组”指具体核酸(DNA或RNA)是克隆、限制性酶切和/或连接步骤的各种组合的产物,这些组合产生具有可与天然系统中发现的内源性核酸区别开的结构性编码或非编码序列的构建物。通常,编码结构性编码序列的DNA序列可由cDNA片段和短寡核苷酸接头、或由一系列合成寡核苷酸组装,以提供能够由细胞或无细胞转录和翻译系统所含的重组转录单元表达的合成核酸。可以不被内部非翻译序列或内含子(一般存在于真核基因中)中断的开放阅读框的形式提供这种序列。包含相关序列的基因组DNA也可用于形成重组基因或转录单元。非翻译DNA序列可以存在于开放阅读框的5'或3'端,这些序列不干扰编码区的操作或表达,并且实际上可用于以各种机制调节所需产物的产生(参见下面的“DNA调控序列”)。

[0042] 因此,例如,术语“重组”多核苷酸或核酸指不是天然产生的,例如通过人类介入人工组合两种分离的序列节段制备的多核苷酸或核酸。常常通过化学合成方法或通过人工操作分离的核酸节段(如遗传工程技术)实现此人工组合。通常这样做是为了用编码相同或保守性氨基酸的冗余密码子取代密码子,一般引入或去除序列识别位点。或者,进行该方法以连接具有所需功能的核酸节段,产生所需的功能组合。常常通过化学合成方法或通过人工操作分离的核酸节段(如遗传工程技术)实现此人工组合。

[0043] “构建物”指重组核酸,通常是重组DNA,产生重组DNA的目的是表达特定的核苷酸序列,或用其构建其它重组核苷酸序列。

[0044] 本文所用术语“操纵子”和“单转录单元”可互换使用,指受一个或多个控制元件(如启动子)协同调节的两个或多个毗连编码区(编码基因产物如RNA或蛋白质的核苷酸序列)。本文所用术语“基因产物”指DNA编码的RNA(反之亦然)或者RNA或DNA编码的蛋白质,其中基因一般包含编码蛋白质的一种或多种核苷酸序列,也可包括内含子和其它非编码核苷酸序列。

[0045] 本文中术语“DNA调控序列”、“控制元件”和“调控元件”可互换使用,指转录或翻译控制序列,如启动子、增强子、聚腺苷酸化信号、终止子、蛋白降解信号等,它们提供和/或调控宿主细胞中编码序列的表达和/或编码多肽的产生。

[0046] 本文中术语“转化”与“遗传修饰”可互换使用,指引入新核酸(即细胞外源性DNA)后永久或瞬时诱导的细胞遗传改变。可通过将新DNA掺入宿主细胞基因组,或通过新DNA瞬时或稳定地维持为外遗传元件实现遗传改变(“修饰”)。当细胞是真核细胞时,通常通过将DNA引入细胞基因组实现永久遗传改变。在原核细胞中,可将永久改变引入染色体或通过染色体外元件如质粒和表达载体引入,染色体外元件可包含一个或多个可选择标记以帮助将它们维持在重组宿主细胞中。遗传修饰的合适方法包括病毒感染、转染、接合、原生质体融合、电穿孔、基因枪技术、磷酸钙沉淀、直接显微注射等。对方法的选择通常取决于待转化的细胞类型和发生转化的环境(即体外、离体或体内)。对这些方法的总体讨论可参见Ausubel等,Short Protocols in Molecular Biology(分子生物学简单方法),第3版,维森出版集团(Wiley and Sons),1995。

[0047] “操作性连接”指并列,其中所述组件的相互关系允许它们以所需方式起作用。例如,如果启动子影响某编码序列转录或表达,那么该启动子操作性连接于该编码序列。本文所用术语“异源启动子”和“异源控制区”指正常情况下不与天然具体核酸相连的启动子和其它控制区。例如,“与编码区异源的转录控制区”是正常情况下不与天然编码区相连的转录控制区。

[0048] 本文所用“宿主细胞”指体内或体外真核细胞、原核细胞或来自多细胞生物体(如细胞系)但培养为单细胞实体的细胞,其中真核或原核细胞可用作或已用作核酸受体(如包含编码一种或多种生物合成途径基因产物如甲羟戊酸途径基因产物的核苷酸序列的表达载体),它包括已用核酸遗传修饰的原始细胞的后代。应理解,单个细胞的后代不一定与原始亲本的形态或基因组DNA或整套DNA完全相同,因为有天然、偶然或有意突变。“重组宿主细胞”(也称为“遗传修饰的宿主细胞”)是引入了异源核酸如表达载体的宿主细胞。例如,由于将异源核酸引入合适的原核宿主细胞,所述原核宿主细胞是遗传修饰的原核宿主细胞(如细菌),其中所述异源核酸是例如该原核宿主细胞以外(天然情况下未发现)的外源性核酸,或通常未在该原核宿主细胞中发现的重组核酸;由于将异源核酸引入了合适的真核宿主细胞,所述真核宿主细胞是遗传修饰的真核宿主细胞,其中异源核酸例如该真核宿主细胞以外的外源性核酸或通常未在该真核宿主细胞中发现的重组核酸。

[0049] 当在温度和溶液离子强度合适的条件下,核酸的单链形式可退火于另一核酸时,该核酸与另一核酸(如cDNA、基因组DNA或RNA)“可杂交”。杂交和洗涤条件是熟知的,参见例如Sambrook,J.,Fritsch,E.F.和Maniatis,T.《分子克隆:实验室手册》(Molecular Cloning:A Laboratory Manual),第二版,冷泉港实验室出版社(Cold Spring Harbor Laboratory Press),冷泉港(1989),特别是其中第11章和表11.1;和Sambrook,J.和Russell,W.,《分子克隆:实验室手册》(Molecular Cloning:A Laboratory Manual),第三版,冷泉港实验室出版社,冷泉港(2001)。温度和离子强度条件决定了杂交的“严谨性”。可调节严谨性条件以筛选中等类似的片段如相关性较少的生物体的同源序列,至高度相似的片段如复制密切相关生物体的功能酶的基因。杂交条件和杂交后洗涤可用于获得所需的杂交决定性严谨条件。一组说明性杂交后洗涤是一系列洗涤,开始是 $6\times\text{SSC}$ (SSC是 0.15M NaCl 和 15mM 柠檬酸盐缓冲液), $0.5\%\text{SDS}$,室温下15分钟,然后用 $2\times\text{SSC}$, $0.5\%\text{SDS}$, 45°C 重复洗涤30分钟,然后用 $0.2\times\text{SSC}$, $0.5\%\text{SDS}$, 50°C 洗涤30分钟,重复两次。用较高温度获得其它严谨条件,其中洗涤与上述洗涤相同,除了最后两次用 $0.2\times\text{SSC}$, $0.5\%\text{SDS}$ 洗涤30分钟的温

度提高到60℃。另一组高度严谨条件最后两次洗涤采用0.1×SSC,0.1%SDS,65℃。严谨杂交条件的另一个例子是在50℃或更高温度下和0.1×SSC(15mM氯化钠/1.5mM柠檬酸钠)中杂交。严谨杂交条件的另一个例子是在溶液中42℃孵育过夜,所用溶液是:50%甲酰胺、5×SSC(150mM NaCl、15mM柠檬酸三钠)、50mM磷酸钠(pH7.6)、5×Denhardt溶液、10%硫酸葡聚糖和20μg/ml变性剪切的鲑精DNA,然后用0.1×SSC在约65℃洗涤滤器。严谨杂交条件和杂交后洗涤条件是至少与上述代表性条件同样严谨的杂交条件和杂交后洗涤条件。

[0050] 杂交需要两种核酸含有互补序列,但根据杂交的严谨性,可能有碱基错配。核酸杂交的合适严谨性取决于核酸长度和互补程度,它们是本领域熟知的变量。两个核苷酸序列之间的相似性或同源性程度越大,具有这些序列的核酸杂交体的解链温度(T_m)越高。核酸杂交的相对稳定性(对应于较高 T_m)按以下顺序依次降低:RNA:RNA、DNA:RNA、DNA:DNA。对于长度大于100个核苷酸的杂交体,产生了计算 T_m 的等式(参见Sambrook等,同上,9.50-9.51)。对于较短核酸即寡核苷酸的杂交,错配位置变得更重要,寡核苷酸的长度决定其特异性(参见Sambrook等,同上,11.7-11.8)。一般地,可杂交核酸的长度至少约为10个核苷酸。可杂交核酸的示范性最小长度为:至少约15个核苷酸;至少约20个核苷酸;至少约30个核苷酸。而且,本领域技术人员知道,可根据诸如探针长度等因素调节温度和洗涤溶液盐浓度。

[0051] 术语“保守性氨基酸取代”指蛋白质中具有相似侧链的氨基酸残基可互换。例如,具有脂族侧链的一组氨基酸由甘氨酸、丙氨酸、缬氨酸、亮氨酸和异亮氨酸组成;具有脂族-羟基侧链的一组氨基酸由丝氨酸和苏氨酸组成;具有含酰胺侧链的一组氨基酸由天冬酰胺和谷氨酰胺组成;具有芳族侧链的一组氨基酸由苯丙氨酸、酪氨酸和色氨酸组成;具有碱性侧链的一组氨基酸由赖氨酸、精氨酸和组氨酸组成;具有含硫侧链的一组氨基酸由半胱氨酸和甲硫氨酸组成。示范性保守性氨基酸取代基是:缬氨酸-亮氨酸-异亮氨酸、苯丙氨酸-酪氨酸、赖氨酸-精氨酸、丙氨酸-缬氨酸和天冬酰胺-谷氨酰胺。

[0052] “合成核酸”可由用本领域技术人员已知的方法化学合成的寡核苷酸结构块(building block)组装。连接和退火这些结构块,以形成基因节段,之后用酶组装这些节段构建整个基因。提到DNA序列时,“化学合成”指其组成氨基酸是在体外组装的。可以用已经建立的方法手工化学合成DNA,或可用许多市售机器之一自动化学合成DNA。可修饰核酸的核苷酸序列,以根据优化核苷酸序列以反映宿主细胞的密码子偏倚性(的原则)优化表达。本领域技术人员应理解,如果密码子使用偏向于适合宿主的密码子,就能提高成功表达的可能性。可根据对获自可获得序列信息的宿主细胞的基因研究,确定优选密码子。

[0053] 多核苷酸或多肽与另一种多核苷酸或多肽具有某百分数的“序列相同性”是指,当比对时,此百分数的碱基或氨基酸相同,以及在比较两个序列时,此百分数的碱基或氨基酸在相同的相对位置中。可用许多不同方式测定序列相似性。为了测定序列相同性,可用方法和计算机程序比对序列,计算机程序包括可由因特网址ncbi.nlm.nih.gov/BLAST获得的BLAST。参见例如,Altschul等(1990),J.Mol.Biol.215:403-10。另一种比对算法是FASTA,可从美国威斯康星州麦迪逊的遗传学计算组(GCG)包(牛津分子集团公司(Oxford Molecular Group,Inc.)的全资子公司)获得。其它比对技术参见《酶学方法》(Methods in Enzymology),第266卷:大分子序列分析的计算机方法(Computer Methods for Macromolecular Sequence Analysis)(1996),Doolittle编,学术出版社公司(Academic

Press, Inc.), 哈布公司(Harcourt Brace&Co.)的分公司, 美国加利福尼亚州圣地亚哥。尤其感兴趣的是允许序列有缺口的比对程序。Smith-Waterman是允许序列比对有缺口的一种算法类型。参见Meth.Mol.Biol.70:173-187(1997)。同时, 采用Needleman&Wunsch比对的GAP程序可用于比对序列。参见J.Mol.Biol.48:443-453(1970)。

[0054] 在进一步描述本发明之前, 应理解, 本发明不仅限于所述的具体实施方式, 当然它们也可以改变。也应理解, 本文所用术语仅仅为了描述具体实施方式, 不旨在限制, 因此, 本发明范围仅受所附权利要求书的限制。

[0055] 当提供值的范围时, 应理解, 本发明包括该范围上下限之间以下限单位十分之一为间隔的各居中值(除非文中明确指出不是这样)以及任何其它指出值或指出范围中的居中值。这些较小范围的上下限可独立地包括在较小范围内, 并且也属于本发明, 还可在指出范围中特别排除限值。当所述范围包括一个或两个限值时, 本发明也包括排除一个或两个限值的范围。

[0056] 除非另有说明, 本文所用的所有科技术语与本发明所属领域普通技术人员通常理解的含义相同。虽然实施或测试本发明也可采用类似或等效于本文所述内容的任何方法和材料, 但现在描述了优选的方法和材料。将本文所述所有发表物纳入本文作参考, 以便连同引用的发表物公开和描述方法和/或材料。

[0057] 必须注意, 本文和所附权利要求书中所用的单数形式“一”、“一种”以及“这种”包括复数含义, 除非文中明确指出不是这样。因此, 例如, 提到“一种类异戊烯修饰酶”包括多种这类酶, 提到“一种细胞色素P450还原酶”包括本领域技术人员已知的一种或多种细胞色素P450还原酶和其等效物等。还需注意, 可以使权利要求书排除任何任选元件。因此, 此声明旨在用作排除性术语如“单独”、“仅仅”等与权利要求部分联用的前提基础, 或采用“负”限制。

[0058] 提供本文所述发表物仅仅是因为它们在本申请的申请日之前公开。不应理解为承认由于在先发明而使本发明不具资格先于这种出版物。而且, 提供的公开日可能与实际公开日不同, 可能需要单独确认。

[0059] 发明详述

[0060] 本发明提供了含有编码类异戊烯修饰酶的核苷酸序列的分离核酸, 以及含有该核酸的重组载体。本发明还提供了包含本发明核酸或重组载体的遗传修饰的宿主细胞。本发明还提供了包含本发明核酸的转基因植物。本发明还提供了生产类异戊烯化合物的方法, 所述方法通常包括在能够合成本发明核酸编码的类异戊烯化合物修饰酶的条件下培养本发明遗传修饰的宿主细胞。

[0061] 核酸、载体和宿主细胞

[0062] 本发明提供了包含编码修饰类异戊烯化合物的酶的核苷酸序列的分离核酸, 在本文中将修饰类异戊烯化合物的酶称为“类异戊烯修饰酶”。包含编码类异戊烯修饰酶的核苷酸序列的本发明核酸称为“类异戊烯修饰酶核酸”。在具体实施方式中, 分离的本发明类异戊烯修饰酶核酸包含编码细胞色素P450加单氧酶的核苷酸序列。在具体实施方式中, 分离的本发明类异戊烯修饰酶核酸包含编码类异戊烯氧化酶的核苷酸序列。在一些实施方式中, 分离的本发明类异戊烯修饰酶核酸包含编码萜羟化酶的核苷酸序列。在一些实施方式中, 分离的本发明类异戊烯修饰酶核酸包含编码萜氧化酶的核苷酸序列。在一些实施方式

中,分离的本发明类异戊烯修饰酶核酸包含编码倍半萜氧化酶的核苷酸序列。在一些实施方式中,分离的本发明类异戊烯修饰酶核酸包含编码倍半萜羟化酶的核苷酸序列。

[0063] NADPH-细胞色素P450氧化还原酶(CPR,EC1.6.2.4)是许多P450-加单氧酶的氧化还原伴侣。本发明还提供了包含编码细胞色素P450还原酶(CPR)的核苷酸序列的分离核酸。包含编码CPR的核苷酸序列的本发明核酸称为“CPR核酸”。本发明CPR核酸编码的CPR将电子由NADPH传递给细胞色素P450。通常,本发明CPR核酸编码的CPR将电子由NADPH传递给本发明类异戊烯修饰酶编码核酸编码的类异戊烯修饰酶,如倍半萜氧化酶。

[0064] 编码类异戊烯修饰酶的核酸

[0065] 在一些实施方式中,本发明分离核酸包含编码具有类异戊烯羟化酶和/或类异戊烯氧化酶活性的多肽的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码细胞色素P450加单氧酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码类异戊烯羟化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码类异戊烯氧化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码进行连续的羟化和氧化反应的多肽的核苷酸序列,例如,所述多肽羟化萜化合物产生萜醇,氧化该萜醇产生萜醛,氧化该萜醛产生萜羧酸。在一些实施方式中,本发明分离核酸包含编码催化萜的异丙烯基的羟化和/或氧化的多肽的核苷酸序列,例如,该多肽催化单萜、二萜、三萜、倍半萜或多萜的异丙烯基的羟化。在一些实施方式中,本发明分离核酸包含编码单萜氧化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码单萜羟化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码多萜羟化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码多萜氧化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码二萜羟化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码二萜氧化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码三萜羟化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码三萜氧化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码倍半萜羟化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码倍半萜氧化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码倍半萜C12-羟化酶的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码能进行倍半萜的C12氧化的多肽的核苷酸序列。在一些实施方式中,本发明分离核酸包含编码紫穗槐二烯12-氧化酶的核苷酸序列。

[0066] 萜环化酶(也称为“萜合酶”)反应产物是所谓的“萜骨架”。在一些实施方式中,本发明分离核酸包含编码催化萜骨架或其下游产物的羟化和/或氧化的类异戊烯修饰酶的核苷酸序列。通常,本发明核酸编码的类异戊烯修饰酶的底物包括萜骨架或修饰的萜骨架。在许多实施方式中,本发明核酸编码的类异戊烯修饰酶的底物包括异丙烯基。

[0067] 本发明核酸编码的类异戊烯修饰酶的单萜底物包括但不限于:产生的氧化产物是单萜化合物或者是产生单萜化合物的生物合成途径的中间体的任何单萜底物。示范性单萜底物包括但不限于:属于以下家族的单萜底物:无环单萜、二甲基辛烷、薄荷烷、不规则的类单萜、桉油醇、茨烷、异茨烷、单环单萜、蒾烷、葑烷、苧烷、薹烷、紫罗酮、虹彩烷(Iridanes)和大麻(Cannabinoids)。示范性单萜底物、中间体和产物包括但不限于:苧烯、香茅醇(citranellol)、牻牛儿醇、薄荷醇、紫苏子醇、芳樟醇和宁酮。

[0068] 本发明核酸编码的类异戊烯修饰酶的二萜底物包括但不限于:产生的氧化产物是

二萜化合物或者是产生二萜化合物的生物合成途径的中间体的任何二萜底物。示范性二萜底物包括但不限于：属于以下家族的二萜底物：无环双萜类、双环双萜类、单环双萜类、半日花烷(Labdanes)、克罗登烷(Clerodanes)、紫杉烷、三环双萜类、四环双萜类、贝壳杉烯、贝叶烷(Beyerenes)、吡地烯(Atiserenes)、蚜肠毒素(Aphidicolins)、木藜芦毒素、赤霉素、大环二萜和伊丽莎白三烷(Elizabethatrianes)。示范性二萜底物、中间体和产物包括但不限于：甾麻素、艾榴塞洛素(eleutherobin)、紫杉醇、蔓生素(prostratin)和假蕨素。

[0069] 本发明核酸编码的类异戊烯修饰酶的三萜底物包括但不限于：产生的氧化产物是三萜化合物或者是产生三萜化合物的生物合成途径的中间体的任何三萜底物。示范性三萜底物、中间体和产物包括但不限于：阿布糖苷E(arbruside E)、鸦胆丁(bruceantin)、甾酮、孕酮、可的松和洋地黄毒苷。

[0070] 本发明核酸编码的类异戊烯修饰酶的倍半萜底物包括但不限于：产生的氧化产物是倍半萜化合物或者是产生倍半萜化合物的生物合成途径的中间体的任何倍半萜底物。示范性倍半萜底物包括但不限于：属于以下家族的倍半萜底物：法呢烷、单环法呢烷、单环倍半萜、双环倍半萜、双环法呢烷、双波烷(Bisbolanes)、檀香烷、卡波烷(Cupranes)、剪叶苔烷(Herbertanes)、瓣苔烷(Gymnomitranes)、单端孢霉烷、花柏烷(Chamigranes)、胡萝卜烷、菖蒲烷、安提萨汀(Antisatins)、杜松烷、倍半萜酮右旋日本刺参萜酮(Oplopananes)、胡椒烷(Copaanes)、苦毒烷(Picrotoxanes)、雪松烷、长叶蕨烷、长叶环烷(Longicyclanes)、丁香烷、莫得烷(Modhephanes)、斯非叶烷(Siphiperfolanes)、葎草烷、全叶烷(Intergrifolians)、丽皮叶烷(Lippifolians)、原伊鲁烷(Protoilludanes)、隐环伞烷(Illudanes)、多毛烷(Hirsutanes)、乳菇烷(Lactaranes)、斯蒂波烷(Sterpuranes)、富麻烷(Fomannosanes)、马拉烷(Marasmanes)、大根香叶烷、榄香烷、桉叶烷、贝克烷(Bakkanes)、绮罗烷(Chilosyphanes)、愈创木烷、假愈创木烷、三环倍半萜、广霍香烷、三烷(Trixanes)、香木兰烷(Aromadendranes)、高各烷(Gorgonananes)、纳多烷(Nardosinanes)、巴西烷(Brasilanes)、绿苔烷(Pinguisanes)、倍半蒎烷(Sequipinane)、倍半茨烷(Sequicamphane)、斧柏烷、双环葎草烷、葱烷(Alliacanes)、斯蒂波烷(Sterpuranes)、乳菇烷、亚夫丽烷(Africanes)、全叶烷、原伊鲁烷(Protoilludanes)、土青木香烷和纽兰烷(Neolemnanes)。示范性倍半萜底物包括但不限于：紫穗槐二烯、异长叶烯(alloisolongifolene)、(-)- α -反式-香柠檬烯(bergamotene)、(-)- β -榄香烯、(+)-大根香叶烯A、大根香叶烯B、(+)- γ -古芸烯、(+)-喇叭烯、十氢二甲基甲乙烯基萘酚(neointermedeol)、(+)- β -蛇床烯和(+)-朱栌倍半萜。

[0071] 不难采用合适底物以这些酶活性的标准测定方法来确定本发明核酸编码萜氧化酶或是萜羟化酶。通常用气相色谱-质谱分析酶修饰的产物。不难以这些酶活性的标准测定方法来确定本发明核酸编码倍半萜氧化酶或是倍半萜羟化酶。参见例如，美国专利公开号20050019882，将其内容纳入本文作参考。

[0072] 在一些实施方式中，本发明核酸包含图1所示和SEQ ID NO:1所列的核苷酸序列。在一些实施方式中，本发明核酸包含与SEQ ID NO:1所列核苷酸序列的核苷酸序列相同性为至少约45%、至少约50%、至少约55%、至少约57%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的核苷酸序列。在一些实施方式中，本发明核酸包含与SEQ ID NO:1所列核苷酸序列相比含有1、2、3、4、

5、6、7、8、9、10、约10-15、约15-20、约20-25、或者约25-50个核苷酸取代的核苷酸序列。

[0073] 在一些实施方式中,本发明核酸包含与SEQ ID NO:1所列核苷酸序列的核苷酸序列相同性为至少约45%、至少约50%、至少约55%、至少约57%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的核苷酸序列,其中该核酸编码具有萜羟化酶和/或萜氧化酶活性(如倍半萜氧化酶活性、倍半萜羟化酶活性等)的多肽。

[0074] 在一些实施方式中,本发明核酸包含与SEQ ID NO:1所列核苷酸序列中至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400或至少约1450个毗连核苷酸的部分的核苷酸序列相同性为至少约50%、至少约55%、至少约57%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的核苷酸序列。

[0075] 在一些实施方式中,本发明核酸包含SEQ ID NO:1所列核苷酸序列的至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400或至少约1450个毗连核苷酸。在一些实施方式中,本发明核酸包含SEQ ID NO:1所列核苷酸序列的至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400或至少约1450个毗连核苷酸,它编码具有萜羟化酶和/或萜氧化酶活性,如倍半萜羟化酶和/或氧化酶活性的多肽。

[0076] 在一些实施方式中,本发明核酸包含在严谨杂交条件下能与包含SEQ ID NO:1所列核苷酸序列或其互补物的核酸杂交的核苷酸序列。

[0077] 在一些实施方式中,本发明核酸包含编码含有图2所示和SEQ ID NO:2所列的氨基酸序列的多肽的核苷酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:2所列氨基酸序列的氨基酸序列相同性为至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的氨基酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:2所列氨基酸序列中至少约50、至少约75、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400、至少约450或至少约490个毗连氨基酸的部分的氨基酸序列相同性为至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的氨基酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:2所列氨基酸序列相比含有1、2、3、4、5、6、7、8、9、10、约10-15、约15-20、或者约20-25个保守性氨基酸取代的氨基酸序列。在一些实施方式中,所述编码多肽具有萜羟化酶和/或萜氧化酶活性。在一些实施方式中,所述编码多肽具有倍半萜氧化酶活性。在一些实施方式中,所述编码多肽能催化倍半萜底物的C12氧化。在其它实施方式中,所述编码多肽具有倍半萜羟化酶活性。

[0078] 在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:2所列氨基酸序列的氨基酸序列相同性为至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%、至少约99%或100%的氨基酸序列中的至少约50、至少约75、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400、至少约450或至少约490个毗连

氨基酸。在一些实施方式中,所述编码多肽具有萘羟化酶和/或萘氧化酶活性。在一些实施方式中,所述编码多肽具有倍半萘氧化酶活性。在一些实施方式中,所述编码多肽能催化倍半萘底物的C12氧化。在其它实施方式中,所述编码多肽具有倍半萘羟化酶活性。

[0079] 在一些实施方式中,本发明核酸包含图9所示和SEQ ID NO:5所列的核苷酸序列。在一些实施方式中,本发明核酸包含与SEQ ID NO:5所列核苷酸序列的核苷酸序列相同性为至少约45%、至少约50%、至少约55%、至少约57%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的核苷酸序列。在一些实施方式中,本发明核酸包含与SEQ ID NO:5所列核苷酸序列相比含有1、2、3、4、5、6、7、8、9、10、约10-15、约15-20、约20-25、或者约25-50个核苷酸取代的核苷酸序列。

[0080] 在一些实施方式中,本发明核酸包含与SEQ ID NO:5所列核苷酸序列的核苷酸序列相同性为至少约45%、至少约50%、至少约55%、至少约57%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的核苷酸序列,其中核酸编码具有萘羟化酶和/或萘氧化酶活性(如倍半萘氧化酶活性、倍半萘羟化酶活性等)的多肽。

[0081] 在一些实施方式中,本发明核酸包含与SEQ ID NO:5所列核苷酸序列中至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400或至少约1450个毗连核苷酸的部分的核苷酸序列相同性为至少约45%、至少约50%、至少约55%、至少约57%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的核苷酸序列。

[0082] 在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:6所列氨基酸序列的氨基酸序列相同性为至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%、至少约99%或100%的氨基酸序列的至少约50、至少约75、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400、至少约450或至少约480个毗连氨基酸。在许多实施方式中,所述编码多肽具有萘羟化酶和/或萘氧化酶活性。在许多实施方式中,所述编码多肽具有倍半萘氧化酶或倍半萘羟化酶活性。在许多实施方式中,所述编码多肽能催化倍半萘底物的羟化。

[0083] 在一些实施方式中,本发明核酸包含SEQ ID NO:5所列核苷酸序列的至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400或至少约1450个毗连核苷酸。在一些实施方式中,本发明核酸包含SEQ ID NO:5所列核苷酸序列的至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400或至少约1450个毗连核苷酸,其编码具有萘羟化酶和/或氧化酶活性,如倍半萘氧化酶活性、倍半萘羟化酶活性等的多肽。

[0084] 在一些实施方式中,本发明核酸包含在严谨杂交条件下能与包含SEQ ID NO:5所列核苷酸序列或其互补物的核酸杂交的核苷酸序列。

[0085] 在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含图9所示和SEQ ID NO:6所列的氨基酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:6所列氨基酸序列的氨基酸序列相同性为至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约

85%、至少约90%、至少约95%、至少约98%或至少约99%的氨基酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:6所列氨基酸序列中至少约50、至少约75、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400、至少约450或至少约480个毗连氨基酸的部分的氨基酸序列相同性为至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的氨基酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:6所列氨基酸序列相比含有1、2、3、4、5、6、7、8、9、10、约10-15、约15-20、或者约20-25个保守性氨基酸取代的氨基酸序列。在一些实施方式中,所述编码多肽具有萜羟化酶和/或萜氧化酶活性。在一些实施方式中,所述编码多肽具有倍半萜氧化酶活性。在一些实施方式中,所述编码多肽能催化倍半萜底物的羟化。在其它实施方式中,所述编码多肽具有倍半萜羟化酶活性。

[0086] 在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:6所列氨基酸序列的氨基酸序列相同性为至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约98%、至少约99%或100%的氨基酸序列中的至少约50、至少约75、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400、至少约450或至少约480个毗连氨基酸。在一些实施方式中,所述编码多肽具有萜羟化酶和/或萜氧化酶活性。在一些实施方式中,所述编码多肽具有倍半萜氧化酶活性。在一些实施方式中,所述编码多肽催化倍半萜底物的羟化。在其它实施方式中,所述编码多肽具有倍半萜羟化酶活性。

[0087] 在一些实施方式中,本发明核酸包含编码含有SEQ ID NO:2或SEQ ID NO:6所列氨基酸序列的多肽的变体的核苷酸序列。例如,在一些实施方式中,本发明核酸包含编码某种酶的核苷酸序列,与包含SEQ ID NO:2或SEQ ID NO:6所列氨基酸序列的酶相比,该酶具有以下一种或多种特性:1)酶活性提高;2)体外和/或体内稳定性提高;3)产率提高;4)蛋白质周转率改变;5)底物特异性改变(例如,使得变异酶修饰选择的底物);6)酶效率提高(例如将底物转变产生产物的效率提高);和7)溶解性提高(例如在胞质或胞浆中的溶解性)。

[0088] 编码细胞色素P450还原酶的核酸

[0089] 本发明提供了包含编码细胞色素P450还原酶(CPR)的核苷酸序列的分离核酸。在一些实施方式中,本发明CPR核酸包含编码CPR的核苷酸序列,CPR能将电子由NADPH传递给本发明类异戊烯修饰酶核酸编码的细胞色素P450氧化酶。

[0090] 在一些实施方式中,本发明核酸包含图3所示和SEQ ID NO:3所列的核苷酸序列。在一些实施方式中,本发明核酸包含与SEQ ID NO:3所列核苷酸序列的核苷酸序列相同性为至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的核苷酸序列。

[0091] 在一些实施方式中,本发明核酸包含在严谨杂交条件下能与包含SEQ ID NO:3所列核苷酸序列或其互补物的核酸杂交的核苷酸序列。

[0092] 在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含图4所示和SEQ ID NO:4所列的氨基酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:4所列氨基酸序列的氨基酸序列相同性为至少约80%、至少约85%、至少约90%、至少约95%、至少约98%或至少约99%的氨基酸序列。在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:4所列氨基酸序列

相比含有1、2、3、4、5、6、7、8、9、10、约10-15、约15-20、或者约20-25个保守性氨基酸取代的氨基酸序列。

[0093] 在一些实施方式中,本发明核酸包含编码一多肽的核苷酸序列,该多肽包含与SEQ ID NO:4所列氨基酸序列的氨基酸序列相同性为至少约80%、至少约85%、至少约90%、至少约95%、至少约98%、至少约99%或100%的氨基酸序列中的至少约50、至少约75、至少约100、至少约150、至少约200、至少约250、至少约300、至少约350、至少约400、至少约450、至少约500、至少约550、至少约600、至少约650或至少约700个毗连氨基酸。在一些实施方式中,所述编码多肽将电子由NADPH传递给本发明类异戊烯修饰酶核酸编码的多肽(如类异戊烯修饰酶)。

[0094] 在一些实施方式中,本发明核酸包含SEQ ID NO:3所列核苷酸序列中的至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400、至少约1500、至少约1600、至少约1700、至少约1800、至少约1900、至少约2000或至少约2100个毗连核苷酸。在一些实施方式中,本发明核酸包含SEQ ID NO:3所列核苷酸序列的至少约500、至少约600、至少约700、至少约800、至少约900、至少约1000、至少约1100、至少约1200、至少约1300、至少约1400、至少约1500、至少约1600、至少约1700、至少约1800、至少约1900、至少约2000或至少约2100个毗连核苷酸,其编码能将电子由NADPH传递给本发明类异戊烯修饰酶核酸编码的细胞色素P450氧化酶的多肽,例如,所述编码多肽能将电子由NADPH传递给本发明类异戊烯修饰酶核酸编码的多肽(如类异戊烯修饰酶)。

[0095] 在一些实施方式中,本发明核酸包含编码含有SEQ ID NO:4所列氨基酸序列的多肽的变体的核苷酸序列。例如,在一些实施方式中,本发明核酸包含编码某种酶的核苷酸序列,与包含SEQ ID NO:4所列氨基酸序列的酶相比,该酶具有以下一种或多种特性:1)酶活性提高;2)体外和/或体内稳定性提高;3)产率提高;4)蛋白质周转率改变;5)底物特异性改变(例如,使得变异酶修饰选择的底物);6)酶效率提高(例如将底物转变产生产物的效率提高);和7)溶解性提高(例如在胞质或胞浆中的溶解性)。

[0096] 在一些实施方式中,本发明核酸包含编码融合蛋白的核苷酸序列,所述融合蛋白包含融合于异源多肽(“融合伴侣”)的具有萜羟化酶和/或萜氧化酶活性的类异戊烯修饰酶(如上所述)的氨基酸序列,所述异源多肽是(例如)除上述类异戊烯修饰酶以外的多肽。在一些实施方式中,本发明核酸包含编码融合蛋白的核苷酸序列,所述融合蛋白包含上述CPR的氨基酸序列和异源多肽,所述异源多肽是(例如)除CPR以外的多肽。合适的融合伴侣包括但不限于:能提高类异戊烯修饰酶或CPR的溶解度的多肽;提供可检测信号的多肽(如荧光蛋白;产生可检测产物的酶如 β -半乳糖苷酶、荧光素酶、辣根过氧化物酶等);使类异戊烯修饰酶或CPR位于特定细胞区室(如胞浆、胞质等)中的多肽;等等。

[0097] 在一些实施方式中,本发明核酸包含编码类异戊烯修饰酶(如具有萜羟化酶和/或萜氧化酶活性的多肽)和CPR的核苷酸序列。在一些实施方式中,本发明核酸包含编码融合蛋白的核苷酸序列,所述融合蛋白包含融合于CPR多肽的具有萜羟化酶和/或萜氧化酶活性的类异戊烯修饰酶(如上所述)的氨基酸序列。在一些实施方式中,所述编码融合蛋白具有通式 $\text{NH}_2\text{-A-X-B-COOH}$,式中A是具有萜羟化酶和/或萜氧化酶活性的类异戊烯修饰酶,X是任选接头,B是CPR多肽。在一些实施方式中,所述编码融合蛋白具有通式 $\text{NH}_2\text{-A-X-B-COOH}$,式

中A是CPR多肽,X是任选接头,B是具有萆羟化酶和/或萆氧化酶活性的类异戊烯-修饰多肽。

[0098] 接头肽可具有各种氨基酸序列。可通过通常有弹性的间隔肽连接蛋白质,但不排除其它化学连接。接头可以是可切割接头。合适的接头序列通常是长度约为5-50个氨基酸,或者约为6-25个氨基酸的肽。通常采用有一定弹性的肽接头。连接肽几乎可以是任何氨基酸序列,记住优选接头具有产生通常有弹性的肽的序列。小氨基酸如甘氨酸和丙氨酸可用于产生弹性肽。本领域技术人员通常能够产生这种序列。可购得各种不同接头,它们也适用于本发明。

[0099] 合适的接头肽常常包括富含丙氨酸和脯氨酸残基的氨基酸序列,已知这些残基能赋予蛋白质结构以弹性。示范性接头含有甘氨酸、丙氨酸、脯氨酸和甲硫氨酸残基的组合,如AAAGGM(SEQ ID NO:8);AAAGGMPPAAAGGM(SEQ ID NO:9);AAAGGM(SEQ ID NO:10);和PPAAAGGM(SEQ ID NO:11)。其它示范性接头肽包括IEGR(SEQ ID NO:12;和GGKGGK(SEQ ID NO:13)。然而,可采用通常长度约为5-50个氨基酸的任何弹性接头。接头几乎可以是能产生通常有弹性的肽的任何序列,包括上述类型的富含丙氨酸-脯氨酸的序列。

[0100] 构建物

[0101] 本发明还提供了包含本发明核酸的重组载体(“构建物”)。在一些实施方式中,本发明重组载体能够扩增本发明核酸。在一些实施方式中,本发明重组载体能够在真核细胞、原核细胞或无细胞转录/翻译系统中产生编码的类异戊烯修饰酶或编码的CPR。合适的表达载体包括但不限于:杆状病毒载体、噬菌体载体、质粒、噬菌粒、粘粒、F粘粒(fosmid)、细菌人工染色体、病毒载体(如基于牛痘病毒、脊髓灰质炎病毒、腺病毒、腺伴随病毒、SV40、单纯疱疹病毒等的病毒载体)、基于P1的人工染色体、酵母质粒、酵母人工染色体和对感兴趣的特定宿主(如大肠杆菌(E.coli)、酵母和植物细胞)特异的任何其它载体。

[0102] 在一些实施方式中,本发明重组载体包含本发明类异戊烯修饰酶-编码核酸和本发明CPR-编码核酸。在一些实施方式中,本发明重组载体是能够在真核细胞、原核细胞或无细胞转录/翻译系统中产生编码的类异戊烯修饰酶和编码的CPR的表达载体。

[0103] 某些载体类型能够扩增本发明表达盒。将本发明核酸有效引入细胞和引入后的表达需要其它载体类型。能够接受本发明核酸的任何载体均可考虑用作本发明的重组载体。载体可以是整合到宿主基因组中或保持附加体形式的环状或线性DNA。载体可能需要附加操作或具体条件才能有效掺入宿主细胞中(如许多表达质粒),或者可以是自身整合的细胞特异性系统的一部分(如重组病毒)。在一些实施方式中,载体在原核细胞中有功能,这种载体用于增殖重组载体和/或表达本发明核酸。在一些实施方式中,载体在真核细胞中有功能,在许多情况下这种载体是表达载体。

[0104] 本领域技术人员已知许多合适的表达载体,许多可从市场上购得。以举例方式提供以下载体;用于细菌宿主细胞:pBluescript(加州圣地亚哥的司查塔基公司(Stratagene)),pQE载体(凯杰公司(Qiagen)),pBluescript质粒,pNH载体、λ-ZAP载体(司查塔基公司);pTrc(Amann等,Gene,69:301-315(1988));pTrc99a、pKK223-3、pDR540和pRIT2T(法玛西亚公司(Pharmacia));用于真核宿主细胞:pXT1、pSG5(司查塔基公司)、pSVK3、pBPV、pMSG和pSVLSV40(法玛西亚公司)。然而,可采用任何其它质粒或其它载体,只要它与宿主细胞相容。

[0105] 在许多实施方式中,本发明重组载体含有一种或多种选择性标记基因,以提供用

于选择转化的宿主细胞的表型特征。合适的选择性标记物包括但不限于：二氢叶酸还原酶，用于真核细胞培养物的新霉素抗性；和用于原核宿主细胞如大肠杆菌的四环素或氨苄青霉素抗性。

[0106] 在许多实施方式中，本发明核酸包含编码类异戊烯修饰酶的核苷酸序列，其中编码类异戊烯修饰酶的核苷酸序列操作性连接于一种或多种转录和/或翻译控制元件。在许多实施方式中，本发明核酸包含编码CPR的核苷酸序列，其中编码CPR的核苷酸序列操作性连接于一种或多种转录和/或翻译控制元件。

[0107] 在一些实施方式中，如上所述，本发明重组载体包含本发明类异戊烯修饰酶-编码核酸和本发明CPR-编码核酸。在一些实施方式中，编码的类异戊烯修饰酶核苷酸序列和编码CPR的核苷酸序列操作性连接于不同的转录控制元件。在其它实施方式中，编码类异戊烯修饰酶的核苷酸序列和编码CPR的核苷酸序列操作性连接于同一个转录控制元件。在一些实施方式中，编码类异戊烯修饰酶的核苷酸序列和编码CPR的核苷酸序列均操作性连接于同一个诱导型启动子。在一些实施方式中，编码类异戊烯修饰酶的核苷酸序列和编码CPR的核苷酸序列均操作性连接于同一个组成型启动子。

[0108] 适用于原核宿主细胞的启动子包括但不限于：噬菌体T7RNA聚合酶启动子；trp启动子；lac操纵子启动子；杂合启动子，如lac/tac杂合启动子、tac/trc杂合启动子、trp/lac启动子、T7/lac启动子；trc启动子；tac启动子等；araBAD启动子；体内调节启动子，如ssaG启动子或相关启动子（参见例如，美国专利公开号20040131637）、pagC启动子（Pulkkinen和Miller, J. Bacteriol., 1991:173(1):86-93; Alpuche-Aranda等, PNAS, 1992;89(21):10079-83）、nirB启动子（Harborne等(1992) Mol. Micro. 6:2805-2813）等（参见例如Dunstan等(1999) Infect. Immun. 67:5133-5141; McKelvie等(2004) Vaccine 22:3243-3255; 和Chatfield等(1992) Biotechnol. 10:888-892）； $\sigma 70$ 启动子，如共有 $\sigma 70$ 启动子（参见例如，GenBank登录号AX798980、AX798961和AX798183）；稳定期启动子，如dps启动子、spv启动子等；衍生自致病岛SPI-2的启动子（参见例如W096/17951）；actA启动子（参见例如，Shetron-Rama等(2002) Infect. Immun. 70:1087-1096）；rpsM启动子（参见例如，Valdivia和Falkow(1996), Mol. Microbiol. 22:367-378）；tet启动子（参见例如，Hillen, W. 和Wissmann, A. (1989), 刊于Saenger, W. 和Heinemann, U所编的Topics in Molecular and Structural Biology, Protein-Nucleic Acid Interaction(分子和结构生物学专题, 蛋白质-核酸相互作用), 英国伦敦麦克米兰(Macmillan), 第10卷, 第143-162页）；SP6启动子（参见例如，Melton等(1984) Nucl. Acids Res. 12:7035-7056）；等等。

[0109] 合适的真核启动子的非限制性例子包括CMV立即早期、HSV胸苷激酶、早期和晚期SV40、来自逆转录病毒的LTR和小鼠金属硫蛋白-I。在一些实施方式中，例如在酵母细胞中表达时，合适启动子是组成型启动子如ADH1启动子、PGK1启动子、ENO启动子、PYK1启动子等；或者调节型启动子如GAL1启动子、GAL10启动子、ADH2启动子、PH05启动子、CUP1启动子、GAL7启动子、MET25启动子、MET3启动子等。本领域普通技术人员能够选择合适的载体和启动子。表达载体也可含有用于启动翻译的核糖体结合位点和转录终止子。表达载体还可含有扩大表达的合适序列。

[0110] 在许多实施方式中，编码类异戊烯修饰酶的核苷酸序列操作性连接于诱导型启动子。在许多实施方式中，编码CPR的核苷酸序列操作性连接于诱导型启动子。本领域熟知诱

导型启动子。合适的诱导型启动子包括但不限于： λ 噬菌体的pL;Plac;Ptrp;Ptac(Ptrp-lac杂合启动子);异丙基- β -D-硫代吡喃半乳糖苷(IPTG)-诱导型启动子,如lacZ启动子;四环素-诱导型启动子;阿拉伯糖诱导型启动子,如P_{BAD}(参见例如,Guzman等(1995) J.Bacteriol.177:4121-4130);木糖诱导型启动子,如P_{Xyl}(参见例如,Kim等(1996) Gene181:71-76);GAL1启动子;色氨酸启动子;lac启动子;醇诱导型启动子,如甲醇诱导型启动子、乙醇-诱导型启动子;棉子糖-诱导型启动子;热诱导型启动子,如热诱导型 λ PL启动子、由热敏阻抑物控制的启动子(如CI857-阻抑的 λ 基表达载体;参见例如,Hoffmann等(1999)FEMS Microbiol Lett.177(2):327-34);等等。

[0111] 在酵母中,可采用含有组成型或诱导型启动子的许多载体。综述参见Current Protocols in Molecular Biology(新编分子生物学实验指南),第2卷,1988,Ausubel等编,Greene Publish.Assoc.&Wiley Interscience(格林出版联合集团和韦氏科学出版社),第13章;Grant等,1987,Expression and Secretion Vector for Yeast(用于酵母的表达和分泌载体),Methods in Enzymology(酶学方法),Wu和Grossman编,31987,学术出版社(Acad.Press),纽约,第153卷,第516-544页;Glover,1986,DNA Cloning(DNA克隆),第II卷,IRL出版社,华盛顿特区,第3章;和Bitter,1987,Heterologous Gene Expression in Yeast(酵母中的异源基因表达),Methods in Enzymology(酶学方法),Berger和Kimmel编,学术出版社,纽约,第152卷,第673-684页;和The Molecular Biology of the Yeast Saccharomyces(酿酒酵母的分子生物学),1982,Strathern等编,冷泉港出版社,第I和II卷。可采用组成型酵母启动子如ADH或LEU2或诱导型启动子如GAL(Cloning in Yeast(在酵母中克隆),第3章,R.Rothstein,刊于DNA Cloning(DNA克隆)第11卷,A Practical Approach(实践方法),DM Glover编,1986,IRL出版社,华盛顿特区)。或者,可采用能促进外来DNA序列整合到酵母染色体中的载体。

[0112] 在一些实施方式中,本发明核酸或本发明载体包含用于在植物细胞中表达的启动子或其它调控元件。在植物细胞中有功能的合适的组成型启动子的非限制性例子是花椰菜花叶病毒35S启动子、串联35S启动子(Kay等,Science236:1299(1987))、花椰菜花叶病毒19S启动子、胭脂碱合酶基因启动子(Singer等,Plant Mol.Biol.14:433(1990);An,Plant Physiol.81:86(1986)、章鱼碱合酶基因启动子和泛素启动子。在植物细胞中有功能的合适的诱导型启动子包括但不限于:苯丙氨酸氨-裂合酶基因启动子、查耳酮合酶基因启动子、病理相关蛋白基因启动子、铜-诱导性调控元件(Mett等,Proc.Natl.Acad.Sci.USA90:4567-4571(1993);Furst等,Cell55:705-717(1988));四环素和氯四环素-诱导性调控元件(Gatz等,Plant J.2:397-404(1992);Röder等,Mol.Gen.Genet.243:32-38(1994);Gatz, Meth.Cell Biol.50:411-424(1995));蜕皮激素诱导性调控元件(Christopherson等, Proc.Natl.Acad.Sci.USA89:6314-6318(1992);Kreutzweiser等, Ecotoxicol.Environ.Safety28:14-24(1994));热激诱导性调控元件(Takahashi等,Plant Physiol.99:383-390(1992);Yabe等,Plant Cell Physiol.35:1207-1219(1994);Ueda等, Mol.Gen.Genet.250:533-539(1996));和lac操纵子元件,它们与组成性表达的lac阻抑物联用以产生(例如)IPTG-诱导性表达(Wilde等,EMBO J.11:1251-1259(1992));菠菜亚硝酸还原酶基因的硝酸-诱导性启动子(Back等,Plant Mol.Biol.17:9(1991));光诱导性启动子,如与RuBP羧化酶或LHCP基因家族的小亚基相关的启动子(Feinbaum等,

Mol.Gen.Genet.226:449(1991);Lam和Chua,Science248:471(1990));如美国专利公开号20040038400所述的光-反应性调控元件;水杨酸诱导性调控元件(Uknes等,Plant Cell15:159-169(1993);Bi等,Plant J.8:235-245(1995));植物激素诱导型调控元件(Yamaguchi-Shinozaki等,PlantMol.Biol.15:905(1990);Kares等,Plant Mol.Biol.15:225(1990));和人激素-诱导型调控元件如人糖皮质激素效应元件(Schena等,Proc.Natl.Acad.Sci.USA88:10421(1991))。

[0113] 本发明核酸或本发明载体中也可包含植物组织选择性调控元件。适用于在一种组织或有限数量的组织中异位表达核酸的组织选择性调控元件包括但不限于:木质部-选择性调控元件、管胞-选择性调控元件、纤维-选择性调控元件、毛状体-选择性调控元件(参见例如,Wang等(2002)J.Exp.Botany53:1891-1897)、腺毛-选择性调控元件等。

[0114] 本领域已知适用于植物细胞的载体,任何这种载体均可用于将本发明核酸引入植物宿主细胞。合适载体包括例如:根癌土壤杆菌(*Agrobacterium tumefaciens*)的Ti质粒或毛根土壤杆菌(*A.rhizogenes*)的Ri₁质粒。感染土壤杆菌(*Agrobacterium*)后,Ti或Ri₁质粒被输送到植物细胞中,并稳定整合到植物基因组中。J.Schell,Science,237:1176-83(1987)。适合采用的还有植物人工染色体,如美国专利号6,900,012所述。

[0115] 组合物

[0116] 本发明还提供了含有本发明核酸的组合物。本发明还提供了含有本发明重组载体的组合物。在许多实施方式中,含有本发明核酸或本发明表达载体的组合物包含以下一种或多种物质:盐,如NaCl、MgCl、KCl、MgSO₄等;缓冲剂,如Tris缓冲液、N-(2-羟乙基)哌嗪-N'-(2-乙磺酸)(HEPES)、2-(N-吗啉代)乙磺酸(MES)、2-(N-吗啉代)乙磺酸钠(MES)、3-(N-吗啉代)丙磺酸(MOPS)、N-三[羟甲基]甲基-3-氨基丙磺酸(TAPS)、等;增溶剂;去污剂,如非离子型去污剂如吐温-20等;核酸酶抑制剂;等等。在一些实施方式中,冻干本发明核酸或本发明重组载体。

[0117] 宿主细胞

[0118] 本发明提供了遗传修饰的宿主细胞,如用本发明核酸或本发明重组载体遗传修饰的宿主细胞。在许多实施方式中,本发明遗传修饰的宿主细胞是体外宿主细胞。在其它实施方式中,本发明遗传修饰的宿主细胞是体内宿主细胞。在其它实施方式中,本发明遗传修饰的宿主细胞是多细胞生物体的一部分。

[0119] 在许多实施方式中,宿主细胞是单细胞生物,或作为单细胞培养。在一些实施方式中,宿主细胞是真核细胞。合适的真核宿主细胞包括但不限于:酵母细胞、昆虫细胞、植物细胞、真菌细胞和藻类细胞。合适的真核宿主细胞包括但不限于:巴斯德毕赤酵母(*Pichia pastoris*)、芬兰毕赤酵母(*Pichia finlandica*)、*Pichia trehalophila*、*Pichia koclamae*、膜醭毕赤酵母(*Pichia membranaefaciens*)、*Pichia opuntiae*、耐热毕赤酵母(*Pichia thermotolerans*)、*Pichia salictaria*、*Pichia guercuum*、*Pichia pijperi*、树干毕赤酵母(*Pichia stiptis*)、甲醇毕赤酵母(*Pichia methanolica*)、毕赤酵母(*Pichia sp.*)、酿酒酵母(*Saccharomyces cerevisiae*)、酵母(*Saccharomyces sp.*)、多形汉逊酵母(*Hansenula polymorpha*)、克鲁维酵母(*Kluyveromyces sp.*)、乳糖克鲁维酵母(*Kluyveromyces lactis*)、白假丝酵母(*Candida albicans*)、构巢曲霉(*Aspergillus nidulans*)、黑曲霉(*Aspergillus niger*)、米曲霉(*Aspergillus oryzae*)、木霉

(*Trichoderma reesei*)、*Chrysosporium lucknowense*、镰刀菌(*Fusarium* sp.)、禾谷镰刀菌(*Fusarium gramineum*)、镰孢霉(*Fusarium venenatum*)、粗糙链胞菌(*Neurospora crassa*)、莱茵衣藻(*Chlamydomonas reinhardtii*)等。在一些实施方式中,宿主细胞是除植物细胞以外的真核细胞。

[0120] 在其它实施方式中,宿主细胞是植物细胞。植物细胞包括单子叶植物和双子叶植物的细胞。

[0121] 在其它实施方式中,宿主细胞是原核细胞。合适的原核细胞包括但不限于:大肠杆菌(*Escherichia coli*)、乳杆菌(*Lactobacillus* sp.)、沙门菌(*Salmonella* sp.)、志贺菌(*Shigella* sp.)等的各种实验室菌株。参见例如,Carrier等(1992)*J. Immunol.* 148:1176-1181;美国专利号6,447,784;和Sizemore等(1995)*Science* 270:299-302。可用于本发明的沙门菌菌株的例子包括但不限于:伤寒沙门菌(*Salmonella typhi*)和鼠伤寒沙门菌(*S. typhimurium*)。合适的志贺菌菌株包括但不限于:弗志贺菌(*Shigella flexneri*)、宋内志贺菌(*Shigella sonnei*)和*Shigella dysenteriae*。一般地,所述实验室菌株是非致病性菌株。其它合适细菌的非限制性例子包括但不限于:枯草芽孢杆菌(*Bacillus subtilis*)、普假单胞菌(*Pseudomonas putida*)、绿脓假单胞菌(*Pseudomonas aeruginosa*)、*Pseudomonas mevalonii*、类球红细菌(*Rhodobacter sphaeroides*)、荚膜红细菌(*Rhodobacter capsulatus*)、深红红螺菌(*Rhodospirillum rubrum*)、红球菌属(*Rhodococcus* sp.)等。在一些实施方式中,宿主细胞是大肠杆菌。

[0122] 为了产生本发明遗传修饰的宿主细胞,用已经建立的技术将包含编码类异戊烯修饰酶的核苷酸序列的本发明核酸稳定或瞬时地引入亲代宿主细胞中,这些技术包括但不限于:电穿孔、磷酸钙沉淀、DEAE-右旋糖苷介导的转染、脂质体介导的转染等。在稳定转化中,核酸一般还包括选择性标记物,例如几种熟知的选择性标记物如新霉素抗性、氨苄青霉素抗性、四环素抗性、氯霉素抗性、卡那霉素抗性等。

[0123] 在一些实施方式中,本发明遗传修饰的宿主细胞是植物细胞。本发明遗传修饰的植物细胞可用于在体外植物细胞培养物中产生选择的类异戊烯化合物。关于植物组织培养的指南可参见(例如):*Plant Cell and Tissue Culture*(植物细胞和组织培养),1994,Vasil和Thorpe编,克鲁维学术出版集团(Kluwer Academic Publishers);和*Plant Cell Culture Protocols*(植物细胞培养方法)(*Methods in Molecular Biology*(分子生物学方法)111),1999,Hall编,休曼出版社(Humana Press)。

[0124] 遗传修饰的宿主细胞

[0125] 在一些实施方式中,本发明遗传修饰的宿主细胞包含本发明表达载体,所述本发明表达载体包含编码类异戊烯修饰酶的核苷酸序列。在一些实施方式中,本发明遗传修饰的宿主细胞包含本发明表达载体,所述本发明表达载体包含编码具有萜羟化酶和/或萜氧化酶活性的多肽的核苷酸序列。

[0126] 在一些实施方式中,本发明遗传修饰的宿主细胞包含第一种本发明表达载体和第二种本发明表达载体,所述第一种本发明表达载体包含的本发明核酸含有编码具有萜羟化酶和/或萜氧化酶活性的多肽的核苷酸序列;所述第二种本发明表达载体包含的本发明核酸含有编码CPR的核苷酸序列。在其它实施方式中,本发明遗传修饰的宿主细胞包含本发明表达载体,其中所述本发明表达载体包含含有编码类异戊烯修饰酶的核苷酸序列的本发明

核酸和含有编码CPR的核苷酸序列的本发明核酸。在其它实施方式中,本发明遗传修饰的宿主细胞包含本发明表达载体,所述本发明表达载体包含的本发明核酸含有编码融合多肽(如包含类异戊烯修饰酶和CPR的多肽)的核苷酸序列。

[0127] 合适的CPR-编码核酸包括编码植物中发现的CPR的核酸。合适的CPR-编码核酸包括编码真菌中发现的CPR的核酸。合适的CPR编码核酸的例子包括:GenBank登录号AJ303373(普通小麦(*Triticum aestivum*)CPR);GenBank登录号AY959320(红豆杉(*Taxus chinensis*)CPR);GenBank登录号AY532374(大阿米(*Ammi majus*)CPR);GenBank登录号AG211221(水稻(*Oryza sativa*)CPR);和GenBank登录号AF024635(皱叶欧芹(*Petroselinum crispum*)CPR)。

[0128] 在一些实施方式中,本发明遗传修饰的宿主细胞是通常不通过甲羟戊酸途径合成焦磷酸异戊-1-烯酯(IPP)或甲羟戊酸的宿主细胞。甲羟戊酸途径包括:(a)将两个分子的乙酰-CoA缩合成乙酰乙酰-CoA;(b)使乙酰乙酰-CoA与乙酰-CoA缩合形成HMG-CoA;(c)将HMG-CoA转化为甲羟戊酸;(d)将甲羟戊酸磷酸化为甲羟戊酸5-磷酸;(e)将甲羟戊酸5-磷酸转化为甲羟戊酸5-焦磷酸;和(f)将甲羟戊酸5-焦磷酸转化为焦磷酸异戊-1-烯酯。产生IPP所需的甲羟戊酸途径酶可能因培养条件而变。

[0129] 如上所述,在一些实施方式中,本发明遗传修饰的宿主细胞是通常不通过甲羟戊酸途径合成焦磷酸异戊-1-烯酯(IPP)或甲羟戊酸的宿主细胞。在一些实施方式中,用本发明表达载体遗传修饰宿主细胞,所述本发明表达载体包含编码类异戊烯修饰酶的本发明核酸;并且用一种或多种异源核酸遗传修饰宿主细胞,所述异源核酸包含编码乙酰乙酰-CoA硫解酶、羟甲基戊二酰基-CoA合酶(HMGS)、羟甲基戊二酰基-CoA还原酶(HMGR)、甲羟戊酸激酶(MK)、磷酸甲羟戊酸激酶(PMK)和甲羟戊酸焦磷酸脱羧酶(MPD)(还有任选的IPP异构酶)的核苷酸序列。在许多这类实施方式中,用表达载体遗传修饰宿主细胞,所述表达载体包含编码CPR的核苷酸序列。在一些实施方式中,用本发明表达载体遗传修饰宿主细胞,所述本发明表达载体包含编码类异戊烯修饰酶的本发明核酸;用一种或多种异源核酸遗传修饰宿主细胞,所述异源核酸包含编码MK、PMK、MPD(还有任选的IPP异构酶)的核苷酸序列。在许多这类实施方式中,用表达载体遗传修饰宿主细胞,所述表达载体包含编码CPR的核苷酸序列。

[0130] 在一些实施方式中,本发明遗传修饰的宿主细胞是通常不通过甲羟戊酸途径合成IPP或甲羟戊酸的宿主细胞;用本发明表达载体遗传修饰宿主细胞,所述本发明表达载体包含编码类异戊烯修饰酶的本发明核酸;并且用一种或多种异源核酸遗传修饰宿主细胞,所述异源核酸包含编码乙酰乙酰-CoA硫解酶、HMGS、HMGR、MK、PMK、MPD、IPP异构酶和异戊烯基转移酶的核苷酸序列。在许多这类实施方式中,用表达载体遗传修饰宿主细胞,所述表达载体包含编码CPR的核苷酸序列。在一些实施方式中,本发明遗传修饰的宿主细胞是通常不通过甲羟戊酸途径合成IPP或甲羟戊酸的宿主细胞;用本发明表达载体遗传修饰宿主细胞,所述本发明表达载体包含编码类异戊烯修饰酶的本发明核酸;并且用一种或多种异源核酸遗传修饰宿主细胞,所述异源核酸包含编码MK、PMK、MPD、IPP异构酶和异戊烯基转移酶的核苷酸序列。在许多这类实施方式中,用表达载体遗传修饰宿主细胞,所述表达载体包含编码CPR的核苷酸序列。

[0131] 在一些实施方式中,本发明遗传修饰的宿主细胞是通常通过甲羟戊酸途径合成

IPP或甲羟戊酸的宿主细胞,例如,宿主细胞是包含内源性甲羟戊酸途径的宿主细胞。在一些实施方式中,宿主细胞是酵母细胞。在一些实施方式中,宿主细胞是酿酒酵母。

[0132] 在一些实施方式中,还用一种或多种核酸遗传修饰本发明遗传修饰的宿主细胞,所述核酸包含编码脱氢酶的核苷酸序列,所述脱氢酶进一步修饰类异戊烯化合物。所述编码的脱氢酶可以是在原核细胞或真核细胞中天然发现的脱氢酶,或者可以是这种脱氢酶的变体。在一些实施方式中,本发明提供了编码这种脱氢酶的核苷酸序列的分离核酸。

[0133] 甲羟戊酸途径核酸

[0134] 本领域已知编码MEV途径基因产物的核苷酸序列,可采用任何已知的编码MEV途径基因产物的核苷酸序列,以产生本发明遗传修饰的宿主细胞。例如,编码乙酰乙酰-CoA硫解酶、HMGS、HMGR、MK、PMK、MPD和IDI的核苷酸序列是本领域已知的。以下是编码MEV途径基因产物的已知核苷酸序列的非限制性例子,后面括号中的内容表示各MEV途径酶的GenBank登录号和生物体:乙酰乙酰-CoA硫解酶:(NC_000913区:2324131..2325315;大肠杆菌)、(D49362;脱氮副球菌(*Paracoccus denitrificans*))和(L20428;酿酒酵母);HMGS:(NC_001145互补物19061..20536;酿酒酵母)、(X96617;酿酒酵母)、(X83882;拟南芥(*Arabidopsis thaliana*))、(AB037907;灰北里孢菌(*Kitasatospora griseola*))和(BT007302;智人(*Homo sapiens*));HMGR:(NM_206548;黑腹果蝇(*Drosophila melanogaster*))、(NM_204485;红原鸡(*Gallus gallus*))、(AB015627;链霉菌(*Streptomyces* sp.)K0-3988)、(AF542543;烟草(*Nicotiana attenuata*))、(AB037907;灰北里孢菌(*Kitasatospora griseola*))、(AX128213,提供了编码截短的HMGR的序列;酿酒酵母)和(NC_001145:互补物(115734..118898;酿酒酵母));MK:(L77688;拟南芥)和(X55875;酿酒酵母);PMK:(AF429385;巴西橡胶树(*Hevea brasiliensis*))、(NM_006556;智人)、(NC_001145,互补物712315..713670;酿酒酵母);MPD:(X97557;酿酒酵母)、(AF290095;屎肠球菌(*Enterococcus faecium*))和(U49260;智人);和IDI:(NC_000913,3031087..3031635;大肠杆菌)和(AF082326;雨生红球藻(*Haematococcus pluvialis*))。

[0135] 在一些实施方式中,HMGR编码区编码缺少野生型HMGR的跨膜结构域的截短形式的HMGR(“tHMGR”)。HMGR的跨膜结构域含有该酶的调节部分,但没有催化活性。

[0136] 可以本领域已知的各种方式改变任何已知MEV途径酶的编码序列,以靶向改变编码酶的氨基酸序列。变异MEV途径酶的氨基酸通常与任何已知MEV途径酶的氨基酸序列基本相似,即其区别为至少一个氨基酸,区别可能是至少两个、至少5个、至少10个、或至少20个氨基酸,但一般不超过约50个氨基酸。序列改变可以是取代、插入或缺失。例如,如下所述,可根据具体宿主细胞的密码子偏倚性改变核苷酸序列。此外,可引入使编码蛋白发生保守性氨基酸改变的一个或多个核苷酸序列差异。

[0137] 异戊烯基转移酶

[0138] 在一些实施方式中,遗传修饰本发明遗传修饰的宿主细胞,以包含含有编码类异戊烯修饰酶的核苷酸序列的核酸;在一些实施方式中,还通过遗传修饰以包含含有编码一种或多种甲羟戊酸途径酶的核苷酸序列的一种或多种核酸(如上所述);和编码异戊烯基转移酶的核苷酸序列的核酸。

[0139] 异戊烯基转移酶构成了催化IPP连续缩合导致形成各种链长度的二磷酸异戊烯酯的一大类酶。合适的异戊烯基转移酶包括催化IPP与烯丙基初始底物缩合形成类异戊烯化

合物的酶,所述类异戊烯化合物具有约2个异戊烯单元-6000个异戊烯单元或更多,例如,2个异戊烯单元(焦磷酸牻牛儿酯合酶)、3个异戊烯单元(焦磷酸法呢酯合酶)、4个异戊烯单元(焦磷酸牻牛儿基牻牛儿酯合酶)、5个异戊烯单元、6个异戊烯单元(焦磷酸十六烷基酯合酶)、7个异戊烯单元、8个异戊烯单元(八氢番茄红素合酶、焦磷酸八异戊烯酯合酶)、9个异戊烯单元(焦磷酸九异戊烯酯合酶)、10个异戊烯单元(焦磷酸十异戊烯酯合酶)、约10个异戊烯单元-15个异戊烯单元、约15个异戊烯单元-20个异戊烯单元、约20个异戊烯单元-25个异戊烯单元、约25个异戊烯单元-30个异戊烯单元、约30个异戊烯单元-40个异戊烯单元、约40个异戊烯单元-50个异戊烯单元、约50个异戊烯单元-100个异戊烯单元、约100个异戊烯单元-250个异戊烯单元、约250个异戊烯单元-500个异戊烯单元、约500个异戊烯单元-1000个异戊烯单元、约1000个异戊烯单元-2000个异戊烯单元、约2000个异戊烯单元-3000个异戊烯单元、约3000个异戊烯单元-4000个异戊烯单元、约4000个异戊烯单元-5000个异戊烯单元、或约5000个异戊烯单元-6000个异戊烯单元或更多。

[0140] 合适的异戊烯基转移酶包括但不限于:二磷酸E-异戊烯酯合酶,包括但不限于:二磷酸牻牛儿酯(GPP)合酶、二磷酸法呢酯(FPP)合酶、二磷酸牻牛儿基牻牛儿酯(GGPP)合酶、二磷酸六异戊烯酯(HexPP)合酶、二磷酸七异戊烯酯(HepPP)合酶、二磷酸八异戊烯酯(OPP)合酶、二磷酸茄呢酯(SPP)合酶、二磷酸十异戊烯酯(DPP)合酶、糖胶树胶合酶和古塔胶合酶;以及二磷酸Z-异戊烯酯合酶,包括但不限于:二磷酸九异戊烯酯(NPP)合酶、二磷酸十一异戊烯酯(UPP)合酶、二磷酸脱氢多萜酯(dehydrodolichyl diphosphate)合酶、二磷酸二十异戊烯酯合酶、天然橡胶合酶和其它二磷酸Z-异戊烯酯合酶。

[0141] 已知各种物种的多种异戊烯基转移酶的核苷酸序列,可用于或修饰后用于产生所述遗传修饰的宿主细胞。本领域已知编码异戊烯基转移酶的核苷酸序列。参见例如,人焦磷酸法呢酯合成酶mRNA(GenBank登录号J05262;智人);二磷酸法呢酯合成酶(FPP)基因(GenBank登录号J05091;酿酒酵母);二磷酸单异戊烯酯:二磷酸二甲基烯丙酯异构酶基因(J05090;酿酒酵母);Wang和Ohnuma(2000)Biochim.Biophys.Acta1529:33-48;美国专利号6,645,747;拟南芥焦磷酸法呢酯合成酶2(FPS2)/FPP合成酶2/二磷酸法呢酯合酶2(At4g17190)mRNA(GenBank登录号NM_202836);银杏(Ginkgo biloba)二磷酸牻牛儿基牻牛儿酯合酶(ggpps)mRNA(GenBank登录号AY371321);拟南芥焦磷酸牻牛儿基牻牛儿酯合酶(GGPS1)/GGPP合成酶/法呢基转移酶(At4g36810)mRNA(GenBank登录号NM_119845);细长聚球藻(Synechococcus elongatus)的二磷酸法呢基、牻牛儿基牻牛儿基、牻牛儿基法呢基、六异戊烯基、七异戊烯基酯合酶基因(SeIF-HepPS)(GenBank登录号AB016095);等。

[0142] 萜合酶

[0143] 在一些实施方式中,遗传修饰本发明遗传修饰的宿主细胞,以包含含有编码萜合酶的核苷酸序列的核酸。在一些实施方式中,萜合酶是修饰FPP产生倍半萜的萜合酶。在其它实施方式中,萜合酶是修饰GPP产生单萜的萜合酶。在其它实施方式中,萜合酶是修饰GGPP产生二萜的萜合酶。

[0144] 本领域已知编码萜合酶的核苷酸序列,可用编码萜合酶的任何已知核苷酸序列遗传修饰宿主细胞。例如,已知并可采用以下编码萜合酶的核苷酸序列(后面是它们的GenBank登录号和鉴定到它们的生物):(-)-大根香叶烯D合酶mRNA(AY438099;小叶杨毛果亚种(Populus balsamifera subsp.trichocarpa)×三角杨(Populus deltoids));E,E- α -

法呢烯合酶mRNA(AY640154;黄瓜(*Cucumis sativus*));1,8-桉油素合酶mRNA(AY691947;拟南芥(*Arabidopsis thaliana*));萜合酶5(TPS5)mRNA(AY518314;玉米(*Zea mays*));萜合酶4(TPS4)mRNA(AY518312;玉米);香叶烯/罗勒烯合酶(TPS10)(At2g24210)mRNA(NM_127982;拟南芥);牻牛儿醇合酶(GES)mRNA(AY362553;罗勒(*Ocimum basilicum*));蒾烯合酶mRNA(AY237645;*Picea sitchensis*);香叶烯合酶1e20mRNA(AY195609;金鱼草(*Antirrhinum majus*));(E)- β -罗勒烯合酶(0e23)mRNA(AY195607;金鱼草(*Antirrhinum majus*));E- β -罗勒烯合酶mRNA(AY151086;金鱼草);萜合酶mRNA(AF497492;拟南芥);(-)-蒾烯合酶(AG6.5)mRNA(U87910;北美冷杉(*Abies grandis*));(-)-4S-蒾烯合酶基因(如基因组序列)(AF326518;北美冷杉); δ -蛇床烯合酶基因(AF326513;北美冷杉);紫穗槐-4,11-二烯合酶mRNA(AJ251751;黄花蒿(*Artemisia annua*));E- α -红没药烯合酶mRNA(AF006195;北美冷杉); γ -葎草烯合酶mRNA(U92267;北美冷杉); δ -蛇床烯合酶mRNA(U92266;北美冷杉);蒾烯合酶(AG3.18)mRNA(U87909;北美冷杉);香叶烯合酶(AG2.2)mRNA(U87908;北美冷杉)等。

[0145] 密码子使用

[0146] 在一些实施方式中,修饰用于产生本发明遗传修饰的宿主细胞的核苷酸序列,以使该核苷酸序列反映出具体的宿主细胞优选的密码子。例如,在一些实施方式中,根据酵母优选的密码子修饰核苷酸序列。参见例如,Bennetzen和Hall(1982)*J. Biol. Chem.* 257(6):3026-3031。作为另一个非限制性例子,在其它实施方式中,根据大肠杆菌优选的密码子修饰核苷酸序列。参见例如,Gouy和Gautier(1982)*Nucleic Acids Res.* 10(22):7055-7074;Eyre-Walker(1996)*Mol. Biol. Evol.* 13(6):864-872。也参见Nakamura等(2000)*Nucleic Acids Res.* 28(1):292。

[0147] 其它遗传修饰

[0148] 在一些实施方式中,本发明遗传修饰的宿主细胞指经过以下修饰的宿主细胞:遗传修饰以包含一种或多种含有编码类异戊烯修饰酶的核苷酸序列的核酸;以及进一步遗传修饰以提高萜生物合成途径中间体产生,和/或进一步遗传修饰以使内源性萜生物合成途径基因功能损伤。提到内源性萜生物合成途径基因时,本文所用术语“功能损伤”指遗传修饰萜生物合成途径基因,导致该基因编码的基因产物的产量低于正常水平,和/或没有功能。

[0149] 提高内源性萜生物合成途径中间体产量的遗传修饰包括但不限于:导致宿主细胞中磷酸转乙酰酶水平和/或活性降低的遗传修饰。通过增加乙酰辅酶A的细胞内浓度而增加萜生物合成途径中间体的细胞内浓度。大肠杆菌将大量的醋酸酯形式的细胞内乙酰辅酶A分泌入培养基中。编码磷酸转乙酰酶(pta,负责乙酰辅酶A转化形成醋酸酯的第一个酶)的基因的缺失将减少醋酸酯分泌。降低原核宿主细胞中磷酸转乙酰酶的水平,和/或活性的遗传修饰是特别有用的,其中,遗传修饰的宿主细胞是用包含编码一种或多种MEV途径基因产物的核苷酸序列的核酸遗传修饰的宿主细胞。

[0150] 在一些实施方式中,导致原核宿主细胞中磷酸转乙酰酶水平降低的遗传修饰是使原核宿主细胞编码磷酸转乙酰酶的内源性pta基因功能损伤的遗传突变。可以多种方式使pta基因功能损伤,包括:插入可动遗传元件(例如转座子等);缺失所有或部分基因,导致不形成基因产物、或产物被截短且在将乙酰辅酶A转化为乙酸酯的过程中没有功能;基因突变,导致不形成基因产物、产物被截短且在将乙酰辅酶A转化为乙酸酯的过程中没有功能;

缺失或突变一个或多个控制pta基因表达的控制元件,导致不产生基因产物;等等。

[0151] 在一些实施方式中,缺失了遗传修饰宿主细胞的内源性pta基因。可使用任何缺失基因的方法。pta基因缺失方法的一个非限制性例子是使用 λ Red重组系统。Datsenko和Wanner(2000)Proc Natl Acad Sci USA97(12):第6640-5页。在一些实施方式中,pta基因在用包含编码MK、PMK、MPD和IDI的核苷酸序列的核酸遗传修饰的宿主细胞(如大肠杆菌)中缺失。在一些实施方式中,pta基因在用包含编码MK、PMK、MPD和IPP的核苷酸序列的核酸遗传修饰的宿主细胞(如大肠杆菌)中缺失。在一些实施方式中,pta基因在用包含编码MK、PMK、MPD、IPP和异戊烯基转移酶的核苷酸序列的核酸遗传修饰的宿主细胞(如大肠杆菌)中缺失。

[0152] 在一些实施方式中,本发明遗传修饰的宿主细胞是经遗传修饰包含一种或多种核酸的宿主细胞,所述核酸包含编码MEV生物合成途径基因产物的核苷酸序列;和经进一步遗传修饰使内源性DXP生物合成途径基因功能损伤的核苷酸序列。在其它实施方式中,本发明遗传修饰的宿主细胞是经遗传修饰包含一种或多种核酸的宿主细胞,所述核酸包含编码DXP生物合成途径基因产物的核苷酸序列;和经进一步遗传修饰以使内源性MEV生物合成途径基因功能损伤的核苷酸序列。

[0153] 在一些实施方式中,当本发明遗传修饰的宿主细胞是用包含编码一种或多种MEV途径基因产物的核苷酸序列的核酸遗传修饰的原核宿主细胞时,进一步遗传修饰宿主细胞,以使一种或多种内源性DXP途径基因功能损伤。可被功能损伤的DXP途径基因包括编码任何以下DXP基因产物的一种或多种基因:1-脱氧-D-木酮糖-5-磷酸酯合酶、1-脱氧-D-木酮糖-5-磷酸酯还原酮异构酶(reductoisomerase)、4-二磷酸胞苷酰-2-C-甲基-D-赤藓醇合酶、4-二磷酸胞苷酰-2-C-甲基-D-赤藓醇激酶、2C-甲基-D-赤藓醇2,4-环二磷酸酯合酶和1-羟基-2-甲基-2-(E)-丁烯基4-二磷酸酯合酶。

[0154] 可以多种方式使内源性DXP途径基因功能损伤,包括插入可动遗传元件(例如,转座子等);缺失所有或部分基因,导致不形成基因产物、或产物被截短且没有酶活性;基因突变,导致不形成基因产物、产物被截短且没有酶功能;缺失或突变一个或多个控制基因表达的控制元件,导致不产生基因产物;等等。

[0155] 在其它实施方式中,当本发明遗传修饰的宿主细胞是用包含编码一种或多种DXP途径基因产物的核苷酸序列的核酸遗传修饰的原核宿主细胞时,进一步遗传修饰宿主细胞,以使一种或多种内源性MEV途径基因功能损伤。可被功能损伤的内源性MEV途径基因包括编码任何以下MEV基因产物的一种或多种基因:HMGS、HMGR、MK、PMK、MPD和IDI。可以多种方式使内源性MEV途径基因功能损伤,包括插入可动遗传元件(例如,转座子等);缺失所有或部分基因,导致不形成基因产物、或产物被截短且没有酶活性;基因突变,导致不形成基因产物、产物被截短且没有酶功能;缺失或突变一个或多个控制基因表达的控制元件,导致不产生基因产物;等等。

[0156] 包含本发明遗传修饰的宿主细胞的组合物

[0157] 本发明还提供包含本发明遗传修饰宿主细胞的组合物。本发明组合物包含本发明遗传修饰的宿主细胞;并且在一些实施方式中还包含一种或多种其它成分,这些成分的选择部分基于遗传修饰宿主细胞的特定用途。合适的成分包括但不限于:盐;缓冲剂;稳定剂;蛋白酶抑制剂;核酸酶抑制剂;细胞膜和/或细胞壁保存性化合物,例如甘油、二甲亚砜等;

适合细胞的营养培养基;等等。在一些实施方式中,冻干细胞。

[0158] 转基因植物

[0159] 在一些实施方式中,本发明核酸或本发明表达载体(如本发明类异戊烯修饰酶核酸或包含类异戊烯修饰酶核酸的本发明表达载体)用作转基因,以产生能生产编码的类异戊烯修饰酶的转基因植物。因此,本发明还提供了转基因植物,该植物包含含有本发明核酸的转基因,所述本发明核酸包含编码具有萜羟化酶和/或萜氧化酶活性的酶(如上所述)的核苷酸序列。在一些实施方式中,转基因植物的基因组包含本发明核酸。在一些实施方式中,转基因植物是遗传修饰纯合的。在一些实施方式中,转基因植物是遗传修饰杂合的。

[0160] 在一些实施方式中,与同一物种的对照植物如非转基因植物(不包含编码该多肽的转基因的植物)的多肽产量相比,本发明转基因植物产生的具有萜羟化酶和/或氧化酶活性的转基因编码多肽的产量高出至少约50%、至少约2倍、至少约5倍、至少约10倍、至少约25倍、至少约50倍或至少约100倍或更高。

[0161] 在一些实施方式中,本发明转基因植物是非转基因对照植物的转基因型式,该非转基因对照植物通常产生的类异戊烯化合物是具有萜羟化酶和/或氧化酶活性的转基因编码多肽产生的化合物或者是该转基因编码多肽的下游产物;与同一物种的对照植物如非转基因植物(不包含编码该多肽的转基因的植物)的类异戊烯化合物产量相比,该转基因植物的类异戊烯化合物产量高出至少约50%、至少约2倍、至少约5倍、至少约10倍、至少约25倍、至少约50倍或至少约100倍或更高。

[0162] 本领域熟知将外源性核酸引入植物细胞的方法。认为这种植物细胞被“转化”,如上所述。合适方法包括病毒感染(如双链DNA病毒)、转染、接合、原生质体融合、电穿孔、基因枪技术、磷酸钙沉淀、直接显微注射、碳化硅颈须(whisker)技术、土壤杆菌介导的转化等。方法的选择通常取决于转化的细胞类型和发生转化的环境(即体外、离体或体内)。

[0163] 基于土壤细菌根癌土壤杆菌(*Agrobacterium tumefaciens*)的转化方法特别适用于将外源性核酸分子引入维管植物。野生型土壤杆菌含有指导在宿主植物上产生致瘤性冠瘿生长的Ti(肿瘤诱导)质粒。将Ti质粒的肿瘤诱导型T-DNA区转移给植物基因组需要Ti质粒编码的毒力基因以及T-DNA边界,它们是勾划出待转移区域的一组同向DNA重复。基于土壤杆菌的载体是修饰的Ti质粒,其中肿瘤诱导功能被待引入植物宿主的感兴趣核酸序列所取代。

[0164] 土壤杆菌介导的转化通常采用共同整合载体,或者优选采用双元载体系统,其中Ti质粒的元件在辅助载体和穿梭载体之间分开,所述辅助载体永久地滞留在土壤杆菌宿主中并且携带毒力基因,所述穿梭载体含有以T-DNA序列为边界的感兴趣基因。本领域熟知各种双元载体,并且可购自(例如)克隆太克技术公司(Clontech)(加利福尼亚州帕洛阿尔托)。本领域熟知(例如)将土壤杆菌与培养的植物细胞或创伤组织如叶组织、根外植体、下胚轴(hypocotyledons)、茎块或块茎共同培养的方法。参见例如,Glick和Thompson(编),*Methods in Plant Molecular Biology and Biotechnology*(植物分子生物学和生物技术的方法),佛罗里达州伯卡拉顿(Boca Raton, Fla.):CRC出版社(CRC Press)(1993)。

[0165] 土壤杆菌介导的转化可用于产生各种转基因维管植物(Wang等,同上,1995)包括至少一种桉树属(*Eucalyptus*)和豆科牧草(forage legumes)如苜蓿(紫花苜蓿);牛角花、白三叶草、铅笔花属(*Stylosanthes*)、罗顿豆(*Lotononis bainesii*)和红豆草。

[0166] 也可采用微粒介导的转化产生本发明转基因植物。首先由Klein等(Nature327:70-73(1987))描述的这种方法依赖于通过氯化钙、亚精胺或聚乙二醇沉淀涂上所需核酸分子的微粒如金或钨。用一种装置如BIOLISTIC PD-1000(伯乐公司(Biorad);加利福尼亚州贺求斯(Hercules Calif.))使微粒高速射入被子植物组织中。

[0167] 将本发明核酸引入植物的方式应能使该核酸能够通过(例如)体内或离体方法进入植物细胞。“体内”指通过(例如)浸渗使核酸进入植物活体中。“离体”指在植株外修饰细胞或外植体,然后使该细胞或器官再生为植株。已经描述了适用于稳定转化植物细胞或建立转基因植物的许多载体,包括Weissbach和Weissbach(1989)Methods for Plant Molecular Biology(植物分子生物学方法),学术出版社;和Gelvin等,(1990)Plant Molecular Biology Manual(植物分子生物学手册),克鲁维学术出版集团所述的载体。具体例子包括衍生自根癌土壤杆菌的Ti质粒的载体,以及Herrera-Estrella等(1983)Nature303:209,Bevan(1984)Nucl Acid Res.12:8711-8721,Klee(1985)Bio/Technology3:637-642所述的载体。或者,可通过游离DNA递送技术采用非Ti载体将DNA转移到植物和细胞中。采用这些方法时,可产生转基因植物如小麦、水稻(Christou(1991)Bio/Technology9:957-962)和玉米(Gordon-Kamm(1990)Plant Cell12:603-618)。采用基因枪进行直接DNA递送技术时,不成熟的胚胎也可能是良好的单子叶植物靶组织(Weeks等(1993)Plant Physiol102:1077-1084;Vasil(1993)Bio/Technology10:667-674;Wan和Lemeaux(1994)Plant Physiol104:37-48;对于土壤杆菌介导的DNA转移参见(Ishida等(1996)Nature Biotech14:745-750)。将DNA引入叶绿体的示范性方法是生物射弹轰击、聚乙二醇转化原生质体和显微注射(Danieli等,Nat.Biotechnol16:345-348,1998;Staub等,Nat.Biotechnol18:333-338,2000;O'Neill等,Plant J.3:729-738,1993;Knoblauch等,Nat.Biotechnol17:906-909;美国专利5,451,513、5,545,817、5,545,818和5,576,198;国际申请W095/16783;以及Boynton等,Methods in Enzymology217:510-536(1993),Svab等,Proc.Natl.Acad.Sci.USA90:913-917(1993)和McBride等,Proc.Natl.Acad.Sci.USA91:7301-7305(1994))。适用于生物射弹轰击、聚乙二醇转化原生质体和显微注射方法的任何载体均可用作叶绿体转化的靶向载体。任何双链DNA载体均可用作转化载体,尤其是当引入方法不采用土壤杆菌时。

[0168] 可进行遗传修饰的植物包括谷类、牧草作物、水果、蔬菜、油籽作物、棕榈、林业植物和藤本植物。可修饰的植物的具体例子如下:玉米、香蕉、花生、紫花豌豆、向日葵、番茄、芸苔、烟草、小麦、大麦、燕麦、马铃薯、大豆、棉花、康乃馨、高粱、羽扇豆和稻。其它例子包括黄花蒿,或者已知能产生感兴趣的类异戊烯化化合物的其它植物。

[0169] 本发明也提供了转化的植物细胞,含有转化的植物细胞的组织、植株和产品。本发明转化细胞和含有该细胞的组织和产品的特征是本发明核酸整合到基因组中,和该植物细胞能产生具有萜羟化酶和/或萜氧化酶活性,如倍半萜氧化酶活性的多肽。本发明重组植物细胞可作为重组细胞群体使用,或作为组织、种子、整个植株、茎、果实、叶、根、花、茎、块茎、谷粒、动物饲料、大片植物等使用。

[0170] 本发明也提供了本发明转基因植物的繁殖材料,所述繁殖材料包括种子、后代植株和无性繁殖材料。

[0171] 产生类异戊烯化化合物的方法

[0172] 本发明提供了产生类异戊烯化合物的方法。在一些实施方式中,该方法通常包括用合适培养基培养遗传修饰的宿主细胞,其中所述宿主细胞是用本发明核酸遗传修饰的,本发明核酸包含编码类异戊烯修饰酶的核苷酸序列。在其它实施方式中,该方法通常包括在有利于产生编码的类异戊烯修饰酶的条件下维持本发明转基因植物。产生类异戊烯修饰酶则导致产生类异戊烯化合物。例如,在一些实施方式中,该方法通常包括用合适培养基培养遗传修饰的宿主细胞,其中所述宿主细胞是用本发明核酸遗传修饰的,本发明核酸包含编码萜氧化酶的核苷酸序列。产生萜氧化酶则导致产生类异戊烯化合物。该方法一般在体外进行,但也考虑到在体内产生类异戊烯化合物。在一些实施方式中,该宿主细胞是真核细胞,如酵母细胞。在其它实施方式中,该宿主细胞是原核细胞。在一些实施方式中,该宿主细胞是植物细胞。在一些实施方式中,在本发明转基因植物中进行该方法。

[0173] 细胞一般采用两种途径之一产生类异戊烯或类异戊烯前体(如IPP、二磷酸多异戊烯酯等)。图13-15用于说明细胞用以产生类异戊烯化合物或前体如二磷酸多异戊烯酯的途径。

[0174] 图13显示了类异戊烯途径,包括用异戊烯基转移酶修饰二磷酸异戊烯酯(IPP)和/或其异构体二磷酸二甲基烯丙酯(DMAPP),产生二磷酸多异戊烯酯二磷酸牻牛儿酯(GPP)、二磷酸法呢酯(FPP)和二磷酸牻牛儿基牻牛儿酯(GGPP)。GPP和FPP进一步由萜合酶修饰,分别产生单萜和倍半萜;GGPP进一步由萜合酶修饰,形成二萜和类胡萝卜素。IPP和DMAPP由以下两种途径中的一种产生:甲羟戊酸(MEV)途径和1-脱氧-D-木酮糖-5-磷酸酯(DXP)途径。

[0175] 图14示意性显示了MEV途径,其中,乙酰辅酶A通过一系列反应转化为IPP。

[0176] 图15示意性显示了DXP途径,其中,丙酮酸和D-甘油醛-3-磷酸酯通过一系列反应转化为IPP和DMAPP。除植物细胞之外的真核细胞仅利用MEV类异戊烯途径以使乙酰辅酶A(乙酰-CoA)转化为IPP,然后异构化为DMAPP。植物同时利用MEV和甲羟戊酸非依赖性、或DXP途径来合成类异戊烯。除一些例外情况,原核生物利用DXP途径通过分支点分别产生IPP和DMAPP。

[0177] 在一些实施方式中,用包含编码倍半萜氧化酶的核苷酸序列的本发明核酸遗传修饰宿主细胞,并且用包含倍半萜的培养基培养宿主细胞。倍半萜进入细胞,在细胞中被倍半萜氧化酶修饰。在许多实施方式中,倍半萜选自紫穗槐二烯、异长叶烯、(-)- α -反式-香柠檬烯、(-)- β -榄香烯、(+)-大根香叶烯A、大根香叶烯B、(+)- γ -古芸烯、(+)-喇叭烯、十氢二甲基甲乙烯基萘酚(neointermedeol)、(+)- β -蛇床烯和(+)-朱栾倍半萜。在一些实施方式中,倍半萜氧化酶是紫穗槐二烯氧化酶,用含有紫穗槐-4,11-二烯氧化酶的培养基培养宿主细胞。

[0178] 在其它实施方式中,还用含有编码萜合酶的核苷酸序列的核酸遗传修饰宿主细胞。因此,例如,用含有编码萜合酶和类异戊烯修饰酶(如倍半萜氧化酶)的核苷酸序列的一种或多种核酸遗传修饰宿主细胞。用合适培养基培养这种宿主细胞能产生萜合酶和类异戊烯修饰酶(如倍半萜氧化酶)。例如,萜合酶修饰焦磷酸法呢酯,以产生所述倍半萜氧化酶的倍半萜底物。

[0179] 根据培养宿主细胞的培养基和宿主细胞通过DXP途径或是甲羟戊酸途径合成IPP,在一些实施方式中,宿主细胞还可包含其它遗传修饰。例如,在一些实施方式中,宿主细胞是没有内源性甲羟戊酸途径的宿主细胞,例如,宿主细胞是通常不通过甲羟戊酸途径合成

IPP或甲羟戊酸的宿主细胞。例如,在一些实施方式中,宿主细胞是通常不通过甲羟戊酸途径合成IPP的宿主细胞,用包含编码甲羟戊酸途径中两种或多种酶、IPP异构酶、异戊烯基转移酶、萜合酶和类异戊烯修饰酶(如本发明核酸编码的类异戊烯修饰酶)的核苷酸序列的一种或多种核酸遗传修饰宿主细胞。培养这种宿主细胞能产生甲羟戊酸途径酶、IPP异构酶、异戊烯基转移酶、萜合酶和类异戊烯修饰酶(如倍半萜氧化酶)。产生甲羟戊酸途径酶、IPP异构酶、异戊烯基转移酶、萜合酶和类异戊烯修饰酶(如倍半萜氧化酶)则导致产生类异戊烯化合物。在许多实施方式中,异戊烯基转移酶是FPP合酶,它能产生本发明核酸编码的倍半萜氧化酶的倍半萜底物;产生倍半萜氧化酶则导致氧化宿主细胞中的倍半萜底物。适合采用编码甲羟戊酸途径酶、IPP异构酶、异戊烯基转移酶和萜合酶的任何核酸。例如,合适核酸参见例如Martin等(2003)同上。

[0180] 在一些上述实施方式中,用含有编码两种或多种甲羟戊酸途径酶的核苷酸序列的一种或多种核酸遗传修饰宿主细胞时,所述两种或多种甲羟戊酸途径酶包括MK、PMK和MPD,并在含有甲羟戊酸的培养基中培养宿主细胞。在其它实施方式中,所述两种或多种甲羟戊酸途径酶包括乙酰乙酰CoA硫解酶、HMGS、HMGR、MK、PMK和MPD。

[0181] 在一些实施方式中,宿主细胞是通常不通过甲羟戊酸途径合成IPP的宿主细胞,如上所述遗传修饰宿主细胞,宿主细胞还包含功能损伤的DXP途径。

[0182] 在一些实施方式中,用含有编码细胞色素P450还原酶(CPR)的核苷酸序列的核酸遗传修饰宿主细胞。已知各种CPR核苷酸序列,可采用任何已知的CPR编码核酸,只要编码的CPR具有由NADPH传递电子的活性。在一些实施方式中,CPR编码核酸编码能将电子由NADPH传递给由本发明类异戊烯修饰酶编码核酸编码的类异戊烯修饰酶,如倍半萜氧化酶的CPR。在一些实施方式中,CPR编码核酸是本发明CPR核酸。

[0183] 本发明方法可用于产生各种类异戊烯化合物,包括但不限于:青蒿酸(如倍半萜底物是紫穗槐-4,11-二烯时)、异长叶烯醇(如底物是异长叶烯时)、(E)-反式-香柠檬-2,12-二烯-14-醇(如底物是(-)- α -反式-香柠檬烯时)、(-)-榄香-1,3,11(13)-三烯-12-醇(如底物是(-)- β -榄香烯时)、大牻牛儿三烯醇(germacra)-1(10),4,11(13)-三烯-12-醇(如底物是(+)-大根香叶烯A时)、大根香叶烯B醇(如底物是大根香叶烯B时)、5,11(13)-愈创木二烯-12-醇(如底物是(+)- γ -古芸烯时)、喇叭烯醇(如底物是(+)-喇叭烯时)、4 β -H-桉叶-11(13)-烯-4,12-二醇(如底物是十氢二甲基甲乙烯基萘酚(neointermedeol)时)、(+)- β -广木香醇(如底物是(+)- β -蛇床烯等时);和任何上述物质的衍生物。

[0184] 在一些实施方式中,用合适培养基(如Luria-Bertoni肉汤,任选补充有一种或多种其它物质,例如诱导物(例如编码类异戊烯修饰酶的核苷酸序列在诱导型启动子控制下)等)培养本发明遗传修饰的宿主细胞;用有机溶剂如十二烷覆盖培养基,形成有机层。遗传修饰的宿主细胞产生的类异戊烯化合物分配到有机层中,由有机层纯化该化合物。在一些实施方式中,当编码类异戊烯修饰酶的核苷酸序列操作性连接于诱导型启动子时,将诱导物加入培养基中;适当时间后,由覆盖在培养基上的有机层分离类异戊烯化合物。

[0185] 在一些实施方式中,将类异戊烯化合物与有机层中可能存在的其它产物分离开。不难采用(如)标准的色谱方法将类异戊烯化合物与有机层中可能存在的其它产物分离开。

[0186] 在一些实施方式中,在无细胞反应中用化学方法进一步修饰本发明方法合成的类异戊烯化合物。例如,在一些实施方式中,由培养基和/或细胞裂解物分离青蒿酸,在无细胞

反应中用化学方法进一步修饰青蒿酸,以产生青蒿素。

[0187] 在一些实施方式中,类异戊烯化合物是纯净的,例如,纯度为至少约40%、至少约50%、至少约60%、至少约70%、至少约80%、至少约90%、至少约95%、至少约98%、或高于98%,提到类异戊烯化合物时,“纯净”指不含其它类异戊烯化合物、大分子、污染物等的类异戊烯化合物。

[0188] 实施例

[0189] 提出以下实施例,以便向本领域普通技术人员提供如何制备和使用本发明的完整公开和描述,但并不旨在限制发明人所认定的发明范围,它们也不代表以下实验是进行的所有或仅有实验。努力确保所用数值(如量、温度等)的准确性,但应该允许一些实验误差和偏差。除非另有说明,份数是重量份数,分子量是重均分子量,温度是摄氏度,压力是大气压或近大气压。可采用标准缩写,例如:bp,碱基对;kb,千碱基;pl,皮升;s或sec,秒;min,分钟;h或hr,小时;aa,氨基酸;kb,千碱基;bp,碱基对;nt,核苷酸;i.m.,肌肉内;i.p.,腹膜内;s.c.,皮下;等等。

[0190] 实施例1:类异戊烯修饰酶的克隆和测序

[0191] 已知能羟化萜的大多数酶是细胞色素P450。将萜羟化酶的所有可得氨基酸序列与向日葵和莴苣的细胞色素P450的氨基酸序列进行比对。这两种植物属于菊科(Asteraceae),黄花蒿也属于这一科。类异戊烯修饰酶如CYP71D家族聚集在一起,这表明它们有共同的祖先。设计简并式聚合酶链反应(PCR)引物,这些引物用于扩增菊科CYP71D家族的基因。

[0192] 克隆CYP71AV1(也称为CYP71D-A4或AMO)和CPR cDNA。通过Super SMART PCR cDNA合成试剂盒(BD生物科学公司(BD Bioscience))用由黄花蒿富毛状体(trichome-enriched)细胞纯化的50ng总RNA制备cDNA库。由莴苣和向日葵CYP71亚家族的保守氨基酸基序设计简并P450引物:[Y/Q]G[E/D][H/Y]WR的引物1(正向)和FIPERF的引物2(反向)(表I提供了引物的序列信息)。

[0193]

表 I. 用于构建质粒的引物

引物 编号	序列 (5'-3')
1	TCCGACCA(C/T)ANGGNGAN(C/T)A(C/T)TGGAG; SEQ ID NO: 14
2	TCCGACCAAANC(G/T)(C/T)TCNGG(A/G/T)AT(A/G)AA; SEQ ID NO: 15
3	CCAGCACA(A/G)TA(C/T)GA(A/G)CA(C/T)TT(C/T)AA(C/T)AA(A/G)AT; SEQ ID NO: 16
4	CCAGCAGCCATNCC(C/T)TTNGC(A/G)TCNCC(A/G)CA; SEQ ID NO: 17
5	ACGCTCTAGAATGAAGAGTATACTAAAAGCAATG; SEQ ID NO: 18
6	ACGCTCTAGAGCGAACTTGGAACGAGTAACAACT; SEQ ID NO: 19
7	ATGGATCCTATGCAATCAACAACTTCCGTTAAGTTAT; SEQ ID NO: 20
8	TATGTCGACCCATACATCACGGAGATATCTTCCT SEQ ID NO: 21
9	GGACTAGTAAAACAATGGCCCTGACCGAAGAG; SEQ ID NO: 22
10	CCAAGCTTTCAGATGGACATCGGGTAAAC; [0194] SEQ ID NO: 23
11	CTGCCGCGGGGCGCAAATTAAAGCCTTC; SEQ ID NO: 24
12	CTGCCGCGGTAGTACGGATTAGAAGCCGC; SEQ ID NO: 25
13	CGGGATCCAAAACAATGGCTGCAGACCAATTGGTG; SEQ ID NO: 26
14	GCGTCGACTTAGGATTTAATGCAGGTGACG; SEQ ID NO: 27
15	CGGGATCCAAAACAATGAGCGAAGTCGGTATACAG; SEQ ID NO: 28
16	GCGTCGACTCATAACGAAAAATCAGAGAAATTTG; SEQ ID NO: 29
17	GGACTAGTAAAACAATGGCTTCAGAAAAAGAAATTAG; SEQ ID NO: 30
18	TCCCCCGGGCTATTTGCTTCTCTTGTAAC; SEQ ID NO: 31

[0195] 采用这些引物和黄花蒿cDNA的聚合酶链反应(PCR)产生1-kb DNA片段。所用PCR程序是7个退火温度为48℃的循环和27个退火温度为55℃的循环。扩增基因片段的推定氨基酸与向日葵(QH_CA_Contig1442)和莴苣(QG_CA_Contig7108)毗连群的氨基酸相同性分别

为85%和88%。可以在cgpdb.ucdavis.edu上找到菊科EST-数据库。采用由保守性QYEHFNKI (SEQ ID NO:32)和CGDAKGMA(SEQ ID NO:33)基序分别设计的正向引物(引物3)和反向引物(引物4)分离黄花蒿CPR片段。所用的PCR程序是30个退火温度为50℃的循环。用RLM-RACE试剂盒(安必昂公司(Ambion))确定CYP71AV1(“CYP71D-A4”)和CPR的5'-和3'-端序列,然后从黄花蒿叶cDNA中回收全长cDNA。在FLAG和cMyc标记过程中,用PCR扩增CYP71AV1和CPR的开放阅读框并分别连接到pESC-URA(司查塔基公司)的SpeI和BamHI/SaII位点中。在CYP71AV1的PCR扩增中,采用引物5和6;在CPR的PCR扩增中,采用引物7和8。所用PCR程序是35个退火温度为55℃的循环。对所有克隆进行测序以验证序列。

[0196] 植物提取物分析。在混有5.8μM十八烷作为内标的1mL己烷中剧烈振荡黄花蒿叶(100-200mg鲜重)2小时。将己烷提取物浓缩至200μL,采用DB-XLB柱(0.25mm i.d.x0.25μm x30m,J和W科学公司(J&W Scientific))对1μL样品进行GC-MS分析,以确定14个植物样品的青蒿素含量,如上所述。Woerdenbag等(1991)Phytochem. Anal., 2, 215-219。所用的GC炉程序是以5℃/分钟的升温速度从100℃至250℃。通过装有DB5柱的GC-FID,用TMS-重氮甲烷衍生植物己烷提取物,以确定青蒿酸含量(n=8)。所用的GC炉程序是80℃(保持2分钟),以20℃/分钟的升温速度升温至140℃,通过以5℃/分钟的升温速度升温至220℃分离产物。青蒿素标样购自密苏里州圣路易斯的西格马-奥尔德里奇公司(Sigma-Aldrich, St. Louis, MO)。

[0197] 合成青蒿醇。将青蒿酸(100.0mg, 0.43mmol)溶解于THF(10.0mL),加入LiAlH₄(17.0mg, 0.45mmol)。将该不均一混合物保持回流(70℃)15小时。冷却后,用水3.0mL和15% NaOH水溶液(3.0mL)终止该反应,搅拌10分钟,通过硅藻土过滤。分离有机相,用MgSO₄干燥,用旋转蒸发器浓缩。用柱色谱(2:1己烷/EtOAc)纯化产物,得到61.0mg(65%产率)无色油状醇。通过中性氧化铝的柱色谱(布鲁克门(Brockman)活性1)进一步去除少量青蒿酸污染物。鉴定数据与文献值一致。

[0198] 合成青蒿醛。按照文献报道的方法将青蒿醇氧化成青蒿醛。Sharpless等Tetrahedron Letters 17, 2503-2506(1976)。氩气氛下在含有RuCl₂(PPh₃)₃(17.0mg, 0.018mmol)和N-甲基吗啉N-氧化物(60.0mg, 0.51mmol)的火焰干燥的10-mL烧瓶中加入丙酮(4.0mL)。通过注射器向该溶液中加入溶解于丙酮(1.0mL)的青蒿醇(55.0mg, 0.25mmol)。23℃搅拌该混合物2小时,真空浓缩。用柱色谱(4:1己烷/EtOAc)纯化粗产物,得到32.0mg(59%产率)无色油状青蒿醛。鉴定数据与文献报道一致。

[0199] EPY菌株的产生和鉴定

[0200] 化学品。十二烷和石竹烯购自密苏里州圣路易斯的西格马-奥尔德里奇公司。5-氟乳清酸(5-FOA)购自加利福尼亚州奥伦奇的发酵研究公司(Zymo Research, Orange, CA)。用于配制合成确定(SD)培养基的完全补充混合物购自加利福尼亚州欧文的Q生物基因公司(Qbiogene, Irvine, CA)。所有其它培养基组分均购自西格马-奥尔德里奇公司或新泽西州富兰克林湖的贝克顿迪金森公司(Becton, Dickinson, Franklin Lakes, NJ)。

[0201] 菌株和培养基。用大肠杆菌菌株DH10B和DH5α进行细菌转化和质粒扩增,以构建用于本研究的表达质粒。用含有100mg L⁻¹氨苄青霉素的LB(Luria-Bertani)培养基37℃培养该菌株,除了用DH5α在含有50mg L⁻¹氨苄青霉素的培养基中培育pδ-UB基质粒。

[0202] 酿酒酵母菌株BY4742(Brachmann等, Yeast 14, 115-132(1998))是S288C衍生物,将其用作所有酵母菌株的亲代菌株。用富YPD培养基培养该菌株。Burke等, Methods in yeast

Genetics:a Cold Spring Harbor laboratory course manual(酵母遗传学方法:冷泉港实验室教程手册)(纽约普莱恩维尤的冷泉港实验室出版社(Cold Spring Harbor Laboratory Press, Plainview, NY), 2000)。用适当省去亮氨酸、尿嘧啶、组氨酸和/或甲硫氨酸的SD培养基(Burke等, 同上)培养工程改造的酵母菌株。为了诱导由GAL1启动子表达的基因, 用2%半乳糖作为唯一碳源培养酿酒酵母菌株。

[0203] 质粒构建。为了建立用GAL1启动子表达ADS的质粒pRS425ADS, 用引物对9和10(表1)由pADS(Martin等Nat. Biotechnol. 21: 第796-802页(2003))PCR扩增ADS。采用这些引物, 将核苷酸序列5'-AAAACA-3'恰好克隆入ADS的起始密码子上游。此共有序列用于有效翻译ADS和本研究所用的其它半乳糖诱导型基因。用SpeI和HindIII切割扩增产物, 并克隆入SpeI和HindIII消化的pRS425GAL1中(Mumberg等, Nucleic Acids Research 22, 5767-5768 (1994))。

[0204] 为了整合tHMGR的表达盒, 构建质粒p δ -HMGR。首先, 将SacII限制性位点引入pRS426GAL1(Mumberg等, 同上)中GAL1启动子的5'端和CYC1终止子的3'端。为了实现这一目的, 用引物对11和12对pRS426GAL1的启动子-多克隆位点-终止子盒进行PCR扩增。将扩增产物直接克隆入PvuII消化的pRS426GAL1中, 以构建载体pRS426-SacII。用引物对13和14由质粒pRH127-3(Donald等, Appl. Environ. Microbiol. 63: 3341-3344(1997))PCR扩增HMG1的催化结构域。用BamHI和SalI切割扩增产物, 并克隆入BamHI和XhoI消化的pRS426-SacII。用SacII切割pRS-HMGR, 凝胶提取表达盒片段并克隆入SacII消化的p δ -UB(Lee等, Biotechnol. Prog. 13, 368-373(1997))中。

[0205] 用引物对15和16由质粒pBD33PCR扩增UPC2的upc2-1等位基因。用BamHI和SalI切割扩增产物并克隆入BamHI和XhoI消化的pRS426-SacII中, 产生质粒pRS-UPC2。为了整合upc2-1, 通过用SacII消化pRS-UPC2和将合适片段移至p δ -UB, 以相同方式产生p δ -UPC2。

[0206] 为了用MET3启动子取代ERG9启动子, 构建质粒pRS-ERG9。质粒pRH973(Gardner等, J. Biol. Chem. 274, 31671-31678(1999))含有位于MET3启动子之后的ERG9的截短的5'节段。用ApaI和ClaI酶切pRH973, 并克隆入ApaI和ClaI消化的pRS403(含有HIS3选择性标记物)(Sikorski等, Genetics 122, 19-27(1989))中。

[0207] 为了表达ERG20, 构建质粒p δ -ERG20。先用SalI和XhoI消化质粒pRS-SacII, 产生相容性粘性末端。然后质粒自身连接, 消除SalI和XhoI位点, 产生质粒pRS-SacII-DX。用引物对17和18由BY4742的基因组DNA PCR扩增ERG20。用SpeI和SmaI酶切扩增的产物, 并克隆入SpeI和SmaI消化的pRS-SacII-DX。然后用SacII酶切pRS-ERG20, 凝胶抽提表达盒片段, 并克隆入SacII消化的p δ -UB中。

[0208] 酵母转化和菌株构建。酿酒酵母菌株BY4742(Brachmann等, 同上)是S288C的衍生物, 将其用作所有酿酒酵母菌株的亲代菌株。通过标准的乙酸锂法转化所有酿酒酵母菌株。Gietz, R.D. 和 Woods, R.A. Guide to Yeast Genetics and Molecular and Cell Biology (酵母遗传学和分子细胞生物学的指南), B部分, 87-96(圣地亚哥的学术出版社公司(Academic Press Inc, San Diego), 2002)。筛选各转化的3-10个菌落, 以选择紫穗槐二烯产量最高的转化子。用质粒pRS425ADS转化菌株BY4742并在SD-LEU平板上选择, 从而构建菌株EPY201。用XhoI消化质粒p δ -HMGR, 然后将DNA转化到菌株EPY201中。在SD-LEU-URA平板上初步选择后, 培养转化子, 并接种于含有 1g L^{-1} 5-FOA的SD-LEU平板上, 以选择URA3标记物的

缺失。然后,用XhoI-消化的p δ -UPC2质粒DNA转化得到的尿嘧啶营养缺陷型EPY208。在SD-LEU-URA平板上初步选择后,培养转化子,并接种于含有 1g L^{-1} 5-FOA的SD-LEU平板上,以构建EPY210。用HindIII切割质粒pRS-ERG9,以将P_{MET3}-ERG9融合物整合到EPY208和EPY210的ERG9基因座上,以分别构建EPY213和EPY225。在SD-LEU-HIS-MET平板上选择这些菌株。然后,用XhoI消化的p δ -HMGR质粒DNA转化EPY213。在SD-LEU-URA-HIS-MET平板上初步选择后,培养转化子,并接种于含有 1g L^{-1} 5-FOA的SD-LEU-HIS-MET平板上,以构建EPY219。用XhoI消化的p δ -ERG20质粒DNA转化EPY219。在SD-LEU-URA-HIS-MET平板上初步选择后,培养转化子,并接种于含有 1g L^{-1} 5-FOA的SD-LEU-HIS-MET平板上,以构建EPY224。

[0209] 用两组引物通过PCR分析鉴定pRS-ERG9的整合。各组引物含有能结合于插入的DNA的一种寡核苷酸和能结合于围绕该插入物的基因组DNA的一种寡核苷酸。用结合于GAL1启动子的5'-端和融合基因的3'-端的引物验证所有其它整合物中的全长插入物。

[0210] 酵母培养。用贝克曼(Beckman)DU-640分光光度计测定600nm处的光密度(OD₆₀₀)。为了检测紫穗槐二烯产量,用感兴趣菌株接种含有5mL SD(2%半乳糖)培养基(如上所述适当省去某些氨基酸)的培养管。这些接种体在30℃生长至OD₆₀₀为1-2。用这些种子培养物接种含有50mL SD培养基的无挡板的培养瓶(250mL),使OD₆₀₀为0.05。生长6天后测定紫穗槐二烯产量。各培养物中存在1mM甲硫氨酸,以抑制ERG9基因座上的P_{MET3}-ERG9融合物。所有烧瓶中还含有5mL十二烷。对此十二烷层采样,用乙酸乙酯稀释,用GC-MS测定紫穗槐二烯产量。

[0211] 结果

[0212] 在特化的植物细胞一腺毛中产生青蒿素。从黄花蒿分离腺毛细胞;由这些细胞提取RNA。采用简并引物,分离称为CYP71D-A4的新基因的部分cDNA。通过快速扩增cDNA末端(RACE)回收全长基因。cDNA编码区的核苷酸序列见图1(SEQ ID NO:1);翻译的氨基酸序列见图2(SEQ ID NO:2)。

[0213] 在酵母细胞中表达全长CYP71D-A4cDNA。为了测定紫穗槐二烯氧化酶活性,将CYP71D-A4置于pESC-URA(司查塔基公司)主链质粒的Gal10启动子的转录控制下,其中由Gal1启动子表达黄花蒿的CPR基因(AACPR;图3;编码蛋白的氨基酸序列见图4)。用上述简并引物PCR和RACE法由黄花蒿腺毛mRNA获得AACPR基因。

[0214] 为了进行紫穗槐-4,11-二烯氧化酶活性的体内测定,将该质粒(p71D-A4/CPR::pESC-URA)和缺少CYP71D-A4基因的对照质粒转化到经工程改造能产生紫穗槐-4,11-二烯的酿酒酵母细胞中。简言之,这些细胞是携带编码截短的HMGCoA还原酶的整合基因的菌株BY4742,该酶在酵母中可溶。这些细胞携带pRS425ADS,其中密码子优化的ADS基因在GAL1启动子控制下。用省去合成亮氨酸和尿嘧啶的培养基培养转化细胞,用2%半乳糖诱导29小时,用醚提取该培养基。浓缩该提取物,通过装有EXL柱的气相色谱-质谱(GC-MS)分析1 μ L,其采用的温度程序为以5℃/分钟的升温速度从50℃升温至250℃。用青蒿酸标样合成青蒿醇和青蒿醛,它们用作标准品。以此方法,用表达CPR和CYP71D-A4的细胞检测到两个峰,但用仅表达CPR的对照细胞无法检测到两个峰。通过将保留时间和质谱与标样作比较,确定这些峰对应于青蒿醇和青蒿醛。未检测到青蒿酸;由于挥发性低,估计不经衍生就无法用GC检测到青蒿酸。

[0215] 进行紫穗槐-4,11-二烯氧化酶活性的体内给料实验,其中将相同的两个质粒分别转化到野生型酿酒酵母菌株YPH499中。用50mL 2%右旋糖和省去尿嘧啶的培养基培养酵母细

胞,用2%半乳糖诱导24小时。离心收集5mL诱导的酵母细胞,用含有150 μ M紫穗槐-4,11-二烯、青蒿醇或青蒿醛的新鲜培养基重悬酵母细胞。然后于30 $^{\circ}$ C培养酵母细胞5小时。用醚提取该培养基,然后用N-(叔丁基二甲基甲硅烷基)-N-甲基三氟乙酰胺衍生,以用GC-MS检测任何青蒿酸。也用类似方法衍生青蒿醇和青蒿酸标样。用GC-MS分析衍生的对照和样品各1 μ L。所用温度程序是以5 $^{\circ}$ C/分钟的升温速度从50 $^{\circ}$ C升温至250 $^{\circ}$ C。

[0216] 给予细胞紫穗槐-4,11-二烯时,仅在既表达CPR又表达CYP71D-A4的酵母细胞中检测到青蒿酸的明显累积和少量青蒿醇和醛化合物(图5A)。给予细胞青蒿醇或青蒿醛时,CPR/CYP71D-A4转化的酵母细胞的培养基中青蒿酸的相对累积量高于仅用CPR转化的对照菌株(图5B和5C)。

[0217] 图5A-C。以150 μ M的浓度将紫穗槐二烯(图5A)和两种其它青蒿素中间体—青蒿醇(图5B)和青蒿醛(图5C)—加入培养基中,该培养基中培养有仅用CPR转化(上方的色谱图)和用CPR和CYP71D-A4转化(下方的色谱图)的酵母细胞,这些酵母细胞已用2%半乳糖诱导。箭头表示紫穗槐二烯(1)、青蒿醇(2)、青蒿醛(3)和青蒿酸(4)。用N-(叔丁基二甲基甲硅烷基)-N-甲基三氟乙酰胺衍生后检测青蒿醇(2)和青蒿酸(4)。星号表示加入培养基的底物。

[0218] 用青蒿酸标样验证样品中是否真实存在衍生的青蒿酸(图6A和6B)。这些数据说明,CYP71D-A4克隆中编码的细胞色素P450酶催化第一次羟化,随后青蒿醇氧化转变为青蒿醛和青蒿醛氧化转变为青蒿酸很可能是由CYP71D-A4重组酶以及酵母内源性氧化活性催化的。

[0219] 图6A和6B。将紫穗槐二烯给予CPR/71D-A4转化的酵母细胞后产生的新化合物的质谱和保留时间见图6A,青蒿酸标样的质谱和保留时间见图6B。使基础分子量增加114个质量单位的衍生反应后,用GC-MS检测产物和标样。

[0220] 用携带CPR(“AACPR”)和AMO(“CYP17D-A4”)的pESC-URA(pESC-URA::AACPR/AMO)遗传修饰EPY224,从而在工程改造的酵母中由简单糖如半乳糖从头合成青蒿酸。将编码截短的酵母HMGCoA还原酶的构建物整合到酵母菌株BY4742中两次。过度表达转录因子upc2-1,以提高麦角固醇生物合成途径中几种基因的转录水平。用甲硫氨酸阻抑型启动子MET3下调鲨烯合酶基因(ERG9)。用Gal1启动子过度表达FPP合酶,也用pRS425主链中的Gal1启动子过度表达ADS。在含有1.8%半乳糖和0.2%葡萄糖的合成培养基中30 $^{\circ}$ C培养携带pESC-URA::AACPR/AMO的酵母菌株EPY2245天。离心酵母细胞,用碱性缓冲液(Tris缓冲液pH9)洗涤。通过加入HCl将该缓冲液酸化至pH2;用乙酸乙酯萃取酸化的缓冲液。将TMS-重氮甲烷和甲醇加入乙酸乙酯组分中,以衍生青蒿酸。用GC-MS检测甲酯形式的青蒿酸。

[0221] 图7A-7C描述了在表达AACPR和AMO的酵母中从头产生青蒿酸。相反,在仅表达AACPR的对照酵母菌株中检测不到青蒿酸。13.62分钟的新峰(图7A,峰1)显示的质量碎裂(fragmentation)图案与来自黄花蒿的真正青蒿酸(图7B和C)相同。

[0222] 图8A-8C描述了体外AMO酶试验。由表达AACPR或CPR/AMO的酿酒酵母YPH499分离微粒体。所用底物的色谱峰用星号表示。各酶试验中,采用10 μ M紫穗槐二烯(a)、25 μ M青蒿醇(b)或25 μ M青蒿醛(c)。衍生醚-萃取组分,用GC-MS以选择离子模式(m/z:121、189、204、218、220和248)进行分析。酶产物为:1,青蒿醇[保留时间(Rt)=13.20];2,青蒿醛(Rt=11.79);3,青蒿酸(Rt=13.58,以甲酯形式检测)。

[0223] 图9描述了编码羟化酶的称为71D-B1(用于紫穗槐二烯羟化酶时也称为“AMH”)

的cDNA克隆的核苷酸序列。

[0224] 图10描述了71D-B1(AMH)编码的蛋白质的氨基酸序列。

[0225] 图11A-C描述了AMH克隆(71D-B1)中编码的重组酶的羟化活性。在过度表达HMGC_oA的酵母中表达AM₀时,A图中16.82分钟的峰是青蒿酸;在过度表达HMGC_oA的酵母中过度表达AMH和AACPR时,B图中18.50分钟的峰是羟化的紫穗槐二烯。羟化的紫穗槐二烯的质量碎裂图案见图11C。显示了羟化紫穗槐二烯的母离子(220)的峰,也显示了倍半萜和萜的其它典型的离子碎裂图案(如93、119、132、145、159和177)。

[0226] 图12描述了编码萜羟化酶/氧化酶的基因组DNA的核苷酸序列。

[0227] 虽然参照具体实施方式描述了本发明,但本领域技术人员应理解,可以在不背离本发明的真实构思和范围的情况下进行各种改变,并可用等同物取代。此外,可进行许多修改以使具体情况、材料、物质组成、方法、工艺步骤适应本发明的目的、构思和范围。所有这些修改应属于所附权利要求书的范围。

序 列 表

<110> D.-K. 罗(RO, DAE-KYUN)
K. 纽曼(NEWMAN, KARYN)
E.M. 帕若蒂斯(PARADISE, ERIC M.)
J.D. 基斯林(KEASLING, JAY D.)
M. 奥莱特(OUELLET, MARIO)
R. 伊切斯(EACHUS, RACHEL)
K. 霍(HO, KIMBERLY)
T. 哈姆(HAM, TIMOTHY)

<120> 编码类异戊烯修饰酶的多核苷酸和其使用方法

<130> BERK-049W0

<150> 60/697,067

<151> 2005-07-05

<160> 33

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 1488

<212> DNA

<213> 黄花蒿(Artemisia annua)

<400> 1

[0001]

```
atgaagagta tactaaaagc aatggcactc tcaactgacca ctccattgc tcttgcaacg 60
atccttttgt tcgtttacaa gttegtact cgttccaaat ccaccaaaaa aagccttcct 120
gagccatggc ggettccecat tattggtcac atgcatact tgattggtac aacgccacat 180
cgtgggggtta gggatttagc cagaaagtat ggatcttga tgcatttaca gcttggtgaa 240
gttccaacaa tcgtgggtgc atctccgaaa tgggctaagg agattttgac aacgtacgac 300
attacctttg ctaacaggcc cgagacttta aciggtgaga ttgtttata tcacaatacg 360
gatgttggtc ttgcacccta tggtaatac tggaggcaat tacgtaaaaa ttgcacattg 420
gagcttttga gtgttaagaa agtaaagtca tticagtac ttcgtgaaga ggagtgttg 480
aatllgggtc aagagattaa agcttcaggt tcaggggagac cggttaacct ttcagagaat 540
gttttcaagt tgattgcaac gatacttagt agagccgcat ttgggaaagg gatcaaggac 600
cagaaagagt taacggagat tgtgaaagag atactgaggc aaactggtgg ttttgatgtg 660
gcagatatct ttccctcaaa gaaatttctt catcatctt cgggcaagag agctcgttta 720
actagccttc gcaaaaagat cgataattta atcgataacc ttgtagctga gcatactgtt 780
aacacctcca gtaaaactaa cgagacactc ctcgatgttc ttttaaggct caaagacagt 840
getgaattcc cattaaacac tgataacatt aaagccatca ttttgatgat gtttgaggac 900
ggcacagaca cttcctcacc cacaatcgaa tgggcgattt cgggaactcat aaagtgtccg 960
aaagcaatgg agaaagtaca agcgggaattg aggaaagcat tgaacggaaa agaaaagatc 1020
catgaggaag acattcaaga actaagctac ttgaacatgg taatcaaaga aacattgagg 1080
ttgcaccctc cactaccctt ggttctgcca agagagtgcg gccaacagct caatttggct 1140
ggatacaaca tacccaataa gaccaaacct attgtcaacg tctttgcgat aaatagggac 1200
cctgaatatt ggaaagacgc tgaagcttc atccctgaac gatttgaaaa tagttctgca 1260
actgtcatgg gtgcagaata cgagtatctt ccgtttggag ctgggagaag gatgtgtcct 1320
ggagccgcac ttggttttagc taacgtgcag ctcccgtcgc ctaatatact atatcatttc 1380
aactggaaac tccccaatgg tgtgagctat gaccagatcg acatgaccga gagctctgga 1440
gccacgatgc aaagaaagac tgagttgita ctegttccaa gtttctag 1488
```

<210> 2

<211> 495

<212> PRT

<213> 黄花蒿(Artemisia annua)

<400> 2

```
Met Lys Ser Ile Leu Lys Ala Met Ala Leu Ser Leu Thr Thr Ser Ile
1           5           10           15
Ala Leu Ala Thr Ile Leu Leu Phe Val Tyr Lys Phe Ala Thr Arg Ser
20           25           30
Lys Ser Thr Lys Lys Ser Leu Pro Glu Pro Trp Arg Leu Leu Pro Ile Ile
35           40           45
Gly His Met His His Leu Ile Gly Thr Thr Pro His Arg Gly Val Arg
50           55           60
Asp Leu Ala Arg Lys Tyr Gly Ser Leu Met His Leu Gln Leu Gly Glu
```

65 Val Pro Thr Ile Val Val Ser Ser Pro Lys Trp Ala Lys Glu Ile Leu
 70 85 90 95
 Thr Thr Tyr Asp Ile Thr Phe Ala Asn Arg Pro Glu Thr Leu Thr Gly
 100 105 110
 Glu Ile Val Leu Tyr His Asn Thr Asp Val Val Leu Ala Pro Tyr Gly
 115 120 125
 Glu Tyr Trp Arg Gln Leu Arg Lys Ile Cys Thr Leu Glu Leu Leu Ser
 130 135 140
 Val Lys Lys Val Lys Ser Phe Gln Ser Leu Arg Glu Glu Glu Cys Trp
 145 150 155 160
 Asn Leu Val Gln Glu Ile Lys Ala Ser Gly Ser Gly Arg Pro Val Asn
 165 170 175
 Leu Ser Glu Asn Val Phe Lys Leu Ile Ala Thr Ile Leu Ser Arg Ala
 180 185 190
 Ala Phe Gly Lys Gly Ile Lys Asp Gln Lys Glu Leu Thr Glu Ile Val
 195 200 205
 Lys Glu Ile Leu Arg Gln Thr Gly Gly Phe Asp Val Ala Asp Ile Phe
 210 215 220
 Pro Ser Lys Lys Phe Leu His His Leu Ser Gly Lys Arg Ala Arg Leu
 225 230 235 240
 Thr Ser Leu Arg Lys Lys Ile Asp Asn Leu Ile Asp Asn Leu Val Ala
 245 250 255
 Glu His Thr Val Asn Thr Ser Ser Lys Thr Asn Glu Thr Leu Leu Asp
 260 265 270
 Val Leu Leu Arg Leu Lys Asp Ser Ala Glu Phe Pro Leu Thr Ser Asp
 275 280 285
 Asn Ile Lys Ala Ile Ile Leu Asp Met Phe Gly Ala Gly Thr Asp Thr
 290 295 300
 Ser Ser Ser Thr Ile Glu Trp Ala Ile Ser Glu Leu Ile Lys Cys Pro
 305 310 315 320
 Lys Ala Met Glu Lys Val Gln Ala Glu Leu Arg Lys Ala Leu Asn Gly
 325 330 335
 Lys Glu Lys Ile His Glu Glu Asp Ile Gln Glu Leu Ser Tyr Leu Asn
 340 345 350
 Met Val Ile Lys Glu Thr Leu Arg Leu His Pro Pro Leu Pro Leu Val
 355 360 365
 Leu Pro Arg Glu Cys Arg Gln Pro Val Asn Leu Ala Gly Tyr Asn Ile
 370 375 380
 Pro Asn Lys Thr Lys Leu Ile Val Asn Val Phe Ala Ile Asn Arg Asp
 385 390 395 400
 Pro Glu Tyr Trp Lys Asp Ala Glu Ala Phe Ile Pro Glu Arg Phe Glu
 405 410 415
 Asn Ser Ser Ala Thr Val Met Gly Ala Glu Tyr Glu Tyr Leu Pro Phe
 420 425 430
 Gly Ala Gly Arg Arg Met Cys Pro Gly Ala Ala Leu Gly Leu Ala Asn
 435 440 445
 Val Gln Leu Pro Leu Ala Asn Ile Leu Tyr His Phe Asn Trp Lys Leu
 450 455 460
 Pro Asn Gly Val Ser Tyr Asp Gln Ile Asp Met Thr Glu Ser Ser Gly
 465 470 475 480
 Ala Thr Met Gln Arg Lys Thr Glu Leu Leu Leu Val Pro Ser Phe
 485 490 495

[0002]

<210> 3
 <211> 2157
 <212> DNA
 <213> 黄花蒿 (*Artemisia annua*)

<400> 3
 atgcaatcaa caacttcggt taagttatct cccttcgacg taatgacggc gttacttaac 60
 ggcaaggatg cgttcgacac atcaaacaca tcggatacga atattccggt agcgggtgtt 120
 atggagaatc gtgagctttt gatgatttta actacttcgg ttgcgggtgtt gatcggatgc 180
 gttgtggtgc ttgtgtggag acggctcgct tcggcggcga agaaagcggc ggagtcgccg 240
 gtgattgttg tgcgaagaa agtgacggag gatgaggttg atgacggacg gaagaaagt 300
 actgtgtttt ttggaactca gactgttact gctgaagggt ttgctaaggc gcttgttgaa 360
 gaagctaaag cgcgatatga aaaggcgggtg tttaaagtga ttgatttgga tgattatgct 420
 gctgaagatg atgagtatga ggagaagtta aagaaagaat ctcttgcttt tttcttttta 480
 gctacgtatg gagatgggtg gccgacagat aatgctgcta gattctataa atggtttacc 540

```

gagggggaag agaaagggtga atggcttgac aagcttcaat acgcagtggt tggacttggt 600
aacagacagt atgagcattt caacaagatt gcgaagggtg tcgatgaaaa acctgtggag 660
caggggtgcaa agcccttgtt tccgttggc atgggagacg atgatcaatg tatcgaagac 720
gacttcactg catggaaaga gttggtgtgg cctgagttgg atcaattact tctgatgag 780
gatgatacat ctgttgccac tccatacaca gctgctgttg gagaataccg tgttgtgttc 840
catgacaaac cagagacata tgatcaggat caactgacaa atggccatgc tgttcatgat 900
getcaacatc catgcagatc caatgtcget gtcaaaaagg agctccattc cctctatct 960
tggatgaaga atgcccgtgc tatgacagaa agccaggatt gcagttgggc cccaatttat 1020
ggggaccatg ttggagtcta cgttgagaat ctaagtgaag ttgtggacga agctgaaaaa 1080
ttaatagggt taccgcccga cacttatttc tcaatgacata ctgataacga agacgggaca 1140
ccacttggtg gagcctcttt gccacctcct ttccctccat gcactttaag aaaagcattg 1200
gcttctcatg ccgatgtttt gagctctcct aaaaagtcag ctttgcttgc tttagctgct 1260
catgctactg attctactga agctgataga ctgaaatttt ttgcgtctcc tgcgtgaaag 1320
gatgaatatg ctcatgtgat agttgcaagc cacagaagtc tccitgaggt catggaggcc 1380
ttcccatcag ctaagcctcc gcttgggtgt tttttgcat ctgtgcgcc acgtttgcag 1440
ccgagatact attccatttc ttcttcccca aagtttgcgc caaataggat tcatgtaact 1500
tgtgcattag tgtatgagca aacaccatca ggcccgttc acaaggaggt ctgttcaaca 1560
tgatgaaga atgcccgtgc tatgacagaa agccaggatt gcagttgggc cccaatttat 1620
gttagaacat ccaatttcag acttccctct gatectaagg tcccagttat catgattggc 1680
ccaggcactg gattggctcc atttagaggt ttccctcagg aaaggtagc tcagaaggaa 1740
gctgggactg agctcggaac agccatctta ttcttcggat gcaggaatcg caaagtggat 1800
ttcatatatg aggacgagct taataatttc gtggagacgg gggtctttc cgagcttgtt 1860
acggccttct ctggtgaagg tgccactaag gactacgtgc aacacaagat gactcagaag 1920
gcttcggata tctggaattt actctctgag ggagcatatt tgtatgtttg cggtgatgcc 1980
aaaggcatgg ccaaagatgt acatcggaact ctgcacacta ttgtgcaaga acagggatct 2040
ctagactcct caaaggcgga gctctacgtg aagaatctac aaatggcagg aagatatctc 2100
cgtgatgtat gggtcgacat ggaacagaag ttgatttccg aagaagacct cgagtaa 2157

```

<210> 4

<211> 718

<212> PRT

<213> 黄花蒿 (*Artemisia annua*)

<400> 4

[0003]

```

Met Gln Ser Thr Thr Ser Val Lys Leu Ser Pro Phe Asp Leu Met Thr
1      5      10      15
Ala Leu Leu Asn Gly Lys Val Ser Phe Asp Thr Ser Asn Thr Ser Asp
20      25      30
Thr Asn Ile Pro Leu Ala Val Phe Met Glu Asn Arg Glu Leu Leu Met
35      40      45
Ile Leu Thr Thr Ser Val Ala Val Leu Ile Gly Cys Val Val Val Leu
50      55      60
Val Trp Arg Arg Ser Ser Ser Ala Ala Lys Lys Ala Ala Glu Ser Pro
65      70      75      80
Val Ile Val Val Pro Lys Lys Val Thr Glu Asp Glu Val Asp Asp Gly
85      90      95
Arg Lys Lys Val Thr Val Phe Phe Gly Thr Gln Thr Gly Thr Ala Glu
100      105      110
Gly Phe Ala Lys Ala Leu Val Glu Glu Ala Lys Ala Arg Tyr Glu Lys
115      120      125
Ala Val Phe Lys Val Ile Asp Leu Asp Asp Tyr Ala Ala Glu Asp Asp
130      135      140
Glu Tyr Glu Glu Lys Leu Lys Lys Glu Ser Leu Ala Phe Phe Phe Leu
145      150      155      160
Ala Thr Tyr Gly Asp Gly Glu Pro Thr Asp Asn Ala Ala Arg Phe Tyr
165      170      175
Lys Trp Phe Thr Glu Gly Glu Glu Lys Gly Glu Trp Leu Asp Lys Leu
180      185      190
Gln Tyr Ala Val Phe Gly Leu Gly Asn Arg Gln Tyr Glu His Phe Asn
195      200      205
Lys Ile Ala Lys Val Val Asp Glu Lys Leu Val Glu Gln Gly Ala Lys
210      215      220
Arg Leu Val Pro Val Gly Met Gly Asp Asp Asp Gln Cys Ile Glu Asp
225      230      235      240
Asp Phe Thr Ala Trp Lys Glu Leu Val Trp Pro Glu Leu Asp Gln Leu
245      250      255
Leu Arg Asp Glu Asp Asp Thr Ser Val Ala Thr Pro Tyr Thr Ala Ala
260      265      270
Val Gly Glu Tyr Arg Val Val Phe His Asp Lys Pro Glu Thr Tyr Asp
275      280      285

```

Gln Asp Gln Leu Thr Asn Gly His Ala Val His Asp Ala Gln His Pro
 290 295 300
 Cys Arg Ser Asn Val Ala Val Lys Lys Glu Leu His Ser Pro Leu Ser
 305 310 315 320
 Asp Arg Ser Cys Thr His Leu Glu Phe Asp Ile Ser Asn Thr Gly Leu
 325 330 335
 Ser Tyr Glu Thr Gly Asp His Val Gly Val Tyr Val Glu Asn Leu Ser
 340 345 350
 Glu Val Val Asp Glu Ala Glu Lys Leu Ile Gly Leu Pro Pro His Thr
 355 360 365
 Tyr Phe Ser Val His Thr Asp Asn Glu Asp Gly Thr Pro Leu Gly Gly
 370 375 380
 Ala Ser Leu Pro Pro Pro Phe Pro Pro Cys Thr Leu Arg Lys Ala Leu
 385 390 395 400
 Ala Ser Tyr Ala Asp Val Leu Ser Ser Pro Lys Lys Ser Ala Leu Leu
 405 410 415
 Ala Leu Ala Ala His Ala Thr Asp Ser Thr Glu Ala Asp Arg Leu Lys
 420 425 430
 Phe Phe Ala Ser Pro Ala Gly Lys Asp Glu Tyr Ala Gln Trp Ile Val
 435 440 445
 Ala Ser His Arg Ser Leu Leu Glu Val Met Glu Ala Phe Pro Ser Ala
 450 455 460
 Lys Pro Pro Leu Gly Val Phe Phe Ala Ser Val Ala Pro Arg Leu Gln
 465 470 475 480
 Pro Arg Tyr Tyr Ser Ile Ser Ser Ser Pro Lys Phe Ala Pro Asn Arg
 485 490 495
 Ile His Val Thr Cys Ala Leu Val Tyr Glu Gln Thr Pro Ser Gly Arg
 500 505 510
 Val His Lys Gly Val Cys Ser Thr Trp Met Lys Asn Ala Val Pro Met
 515 520 525
 Thr Glu Ser Gln Asp Cys Ser Trp Ala Pro Ile Tyr Val Arg Thr Ser
 530 535 540
 Asn Phe Arg Leu Pro Ser Asp Pro Lys Val Pro Val Ile Met Ile Gly
 545 550 555 560
 Pro Gly Thr Gly Leu Ala Pro Phe Arg Gly Phe Leu Gln Glu Arg Leu
 565 570 575
 Ala Gln Lys Glu Ala Gly Thr Glu Leu Gly Thr Ala Ile Leu Phe Phe
 580 585 590
 Gly Cys Arg Asn Arg Lys Val Asp Phe Ile Tyr Glu Asp Glu Leu Asn
 595 600 605
 Asn Phe Val Glu Thr Gly Ala Leu Ser Glu Leu Val Thr Ala Phe Ser
 610 615 620
 Arg Glu Gly Ala Thr Lys Glu Tyr Val Gln His Lys Met Thr Gln Lys
 625 630 635 640
 Ala Ser Asp Ile Trp Asn Leu Leu Ser Glu Gly Ala Tyr Leu Tyr Val
 645 650 655
 Cys Gly Asp Ala Lys Gly Met Ala Lys Asp Val His Arg Thr Leu His
 660 665 670
 Thr Ile Val Gln Glu Gln Gly Ser Leu Asp Ser Ser Lys Ala Glu Leu
 675 680 685
 Tyr Val Lys Asn Leu Gln Met Ala Gly Arg Tyr Leu Arg Asp Val Trp
 690 695 700
 Val Asp Met Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Glu
 705 710 715

[0004]

<210> 5
 <211> 1467
 <212> DNA
 <213> 黄花蒿 (*Artemisia annua*)

<400> 5
 atggcacttt caetgaccac ctccattgct cttgccacga tccttttctt cgtaatttac 60
 aagttcgcta ctggttccaa atccacaaaa aacagccttc ctgagccatg ggcacttccc 120
 attattggtc acatgcatca cttgattggt acaataccac atcgtgggct tatggattta 180
 gccagaaagt atggatcttt aatgcatlta cagcttgggtg aagtttcaac aatcgtgggtg 240
 tcattctcga aatgggctaa agagattttg acaacgtacg acattgcctt tgetaacagg 300
 ccctggactt tgctgggtga gattgttgta tatcgcaata caaatattgc tgetgcacct 360
 tatggtgaat actggaggcg attacgtaaa ctttgacat cggagcttat gagtgttaag 420
 aaagtaaagt catatcagtc gcttcgtgaa gaggagtgtt ggaatttggt tcaagagatt 480

```

aaagcttcag gttcagggat accggttaac ctttcagaga acattttcaa gttgattgca 540
acgatacttt gttagcgcg gtttggaaaa ggagtcaagg accagaagga gtgtacggag 600
attatgaaag agatgttgag ggaagtiggt ggttttgatg tggcagatat ctttccttcg 660
aagaaatttc ttcattatct ttcaggcaag agagccaggt taactagcat tcataagaag 720
ctcgataatt ttatcaataa ccttgttgct gaggcatactt tcaaaacttc aagtaaaact 780
gaggagacac ttcttgatgt tcttctaagg ctcaaagata gcgctgaatt cccattaaca 840
gctgacaatg ttaaagccat cttttggat atatttgcag cagggacaga cacttcacat 900
accacaatcg aatgggcatg ttcggaactc ataaagtgtc cgagagcgat ggagaaagta 960
caagcagaac tgaggaaagc acttaacgga aaagaaaaga tccatgagga agatattcaa 1020
ggactaagct acttaaacct ggtaatacaa gaaacattaa ggttgcaccc tccactacce 1080
ttgttgccaa gagagtgcgg tgaaccagtc aatttggctg gatacgacat acccaataag 1140
acaagactta ttgtcaacgt ctttgcgata aatagggacc cagaatactg gaaagacgct 1200
gaaattttca tcccgaacg atttgaaaat agttctacaa ctctcatggg tgcagaatat 1260
gagtatcttc cgtttggagc tgggagaagg atgtgtcctg gagccgcact tggtttagcc 1320
aacgtgcagc taccgctgc taatatacta tatcatttca actggaaact cccaacgggt 1380
gcgagctatg atcagatcga catgaccgag aggtttggaa tctcggttga aagaaagact 1440
cagttgttac tcgtaccaag ttctag 1467

```

<210> 6

<211> 488

<212> PRT

<213> 黄花蒿(*Artemisia annua*)

<400> 6

[0005]

```

Met Ala Leu Ser Leu Thr Thr Ser Ile Ala Leu Ala Thr Ile Leu Phe
1      5      10      15
Phe Val Ile Tyr Lys Phe Ala Thr Arg Ser Lys Ser Thr Lys Asn Ser
20     25     30
Leu Pro Glu Pro Trp Arg Leu Pro Ile Ile Gly His Met His His Leu
35     40     45
Ile Gly Thr Ile Pro His Arg Gly Leu Met Asp Leu Ala Arg Lys Tyr
50     55     60
Gly Ser Leu Met His Leu Gln Leu Gly Glu Val Ser Thr Ile Val Val
65     70     75     80
Ser Ser Pro Lys Trp Ala Lys Glu Ile Leu Thr Thr Tyr Asp Ile Ala
85     90     95
Phe Ala Asn Arg Pro Trp Thr Leu Ala Gly Glu Ile Val Val Tyr Arg
100    105    110
Asn Thr Asn Ile Ala Ala Ala Pro Tyr Gly Glu Tyr Trp Arg Arg Leu
115    120    125
Arg Lys Leu Cys Thr Ser Glu Leu Met Ser Val Lys Lys Val Lys Ser
130    135    140
Tyr Gln Ser Leu Arg Glu Glu Glu Cys Trp Asn Leu Val Gln Glu Ile
145    150    155    160
Lys Ala Ser Gly Ser Gly Ile Pro Val Asn Leu Ser Glu Asn Ile Phe
165    170    175
Lys Leu Ile Ala Thr Ile Leu Cys Arg Ala Ala Phe Gly Lys Gly Val
180    185    190
Lys Asp Gln Lys Glu Cys Thr Glu Ile Met Lys Glu Met Leu Arg Glu
195    200    205
Val Gly Gly Phe Asp Val Ala Asp Ile Phe Pro Ser Lys Lys Phe Leu
210    215    220
His His Leu Ser Gly Lys Arg Ala Arg Leu Thr Ser Ile His Lys Lys
225    230    235    240
Leu Asp Asn Phe Ile Asn Asn Leu Val Ala Glu His Thr Phe Lys Thr
245    250    255
Ser Ser Lys Thr Glu Glu Thr Leu Leu Asp Val Leu Leu Arg Leu Lys
260    265    270
Asp Ser Ala Glu Phe Pro Leu Thr Ala Asp Asn Val Lys Ala Ile Ile
275    280    285
Leu Asp Ile Phe Ala Ala Gly Thr Asp Thr Ser Ser Thr Thr Ile Glu
290    295    300
Trp Ala Ile Ser Glu Leu Ile Lys Cys Pro Arg Ala Met Glu Lys Val
305    310    315    320
Gln Ala Glu Leu Arg Lys Ala Leu Asn Gly Lys Glu Lys Ile His Glu
325    330    335
Glu Asp Ile Gln Gly Leu Ser Tyr Leu Asn Leu Val Ile Lys Glu Thr
340    345    350
Leu Arg Leu His Pro Pro Leu Pro Leu Leu Pro Arg Glu Cys Arg Glu
355    360    365

```

Pro Val Asn Leu Ala Gly Tyr Asp Ile Pro Asn Lys Thr Arg Leu Ile
 370 375 380
 Val Asn Val Phe Ala Ile Asn Arg Asp Pro Glu Tyr Trp Lys Asp Ala
 385 390 395 400
 Glu Ile Phe Ile Pro Glu Arg Phe Glu Asn Ser Ser Thr Thr Leu Met
 405 410 415
 Gly Ala Glu Tyr Glu Tyr Leu Pro Phe Gly Ala Gly Arg Arg Met Cys
 420 425 430
 Pro Gly Ala Ala Leu Gly Leu Ala Asn Val Gln Leu Pro Leu Ala Asn
 435 440 445
 Ile Leu Tyr His Phe Asn Trp Lys Leu Pro Asn Gly Ala Ser Tyr Asp
 450 455 460
 Gln Ile Asp Met Thr Glu Arg Phe Gly Ile Ser Val Glu Arg Lys Thr
 465 470 475 480
 Gln Leu Leu Leu Val Pro Ser Phe
 485

<210> 7
 <211> 1038
 <212> DNA
 <213> 黄花蒿(*Artemisia annua*)

[0006] <400> 7
 gcccttcgag ccgtatgggg attactggcg gcaattacgt aaactttgca cattggagct 60
 tttagtgct aagaaagtag agtcatatca gtcgcttcgt gaagaggagt gttggaattt 120
 agttcaagag attaaagctt caggttcagg gataccggtt aacctttcag agaataatta 180
 caagttaggt gcaatgatac ttagtagagc tgcgtttggg aaaagaatca aggaccataa 240
 ggagtttacg gagcttggg aacagatgtt gagggaaactt ggtggttttg atgtggcaga 300
 tatctttctc tcgcagaaat ttctacatca tatctcgggc aagagatcta ggttaactag 360
 cattcacaaa aagctcgata atttaatcaa taaccttggt gctgagcata ttgttgaagc 420
 ctcaagtaaa actaaggaga cgtctcttga tgttctteta aggcacaaag atagccttga 480
 attccattg acagctgata acgttaaagc catcattttg gtatgaatta atccaatata 540
 tttttttttt caaaaggcca taatagtgtt aaacaagctt gaaatttttt ataactaagt 600
 acatgcacta actttagtac tcgtgaaaat ataattgagtc atcatagggg ttccatgaaa 660
 tatacaggac atgtttacag caggcacaga cacttcgtca accacaatcg aatgggtgat 720
 ttcggaactc ataaagtgtc cgagagctat ggagaaaata caagcggaac tgaggaaagc 780
 acttaacgga aaagaaaaga tccacgagga agacatccaa gaactaagct acttaaacct 840
 ggtaatcaaa gaaacattaa ggttgacccc tccactaccc ttggttttgc cagcagagtg 900
 ccgtcaacca gtcaatttgg ctggatatga cataccat aagaccaaac ttattgtcaa 960
 cgtctttgcg ataaataggg accctgaata ctggaaagac gctgaatctt tcatcccaga 1020
 gcgcttctta actctggt 1038

<210> 8
 <211> 6
 <212> PRT
 <213> 人工序列

<220>
 <223> 接头肽

<400> 8
 Ala Ala Ala Gly Gly Met
 1 5

<210> 9
 <211> 14
 <212> PRT
 <213> 人工序列

<220>
 <223> 接头肽

<400> 9
 Ala Ala Ala Gly Gly Met Pro Pro Ala Ala Ala Gly Gly Met
 1 5 10

<210> 10

<211> 6
<212> PRT
<213> 人工序列

<220>
<223> 接头肽

<400> 10
Ala Ala Ala Gly Gly Met
1 5

<210> 11
<211> 8
<212> PRT
<213> 人工序列

<220>
<223> 接头肽

<400> 11
Pro Pro Ala Ala Ala Gly Gly Met
1 5

<210> 12
<211> 4
<212> PRT
<213> 人工序列

<220>
<223> 接头肽

[0007] <400> 12
Ile Glu Gly Arg
1

<210> 13
<211> 6
<212> PRT
<213> 人工序列

<220>
<223> 接头肽

<400> 13
Gly Gly Lys Gly Gly Lys
1 5

<210> 14
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 合成引物

<220>
<221> misc_feature
<222> 9, 18, 20
<223> N = C 或 T

<220>
<221> misc_feature
<222> 11, 14, 17
<223> n=A、T、C 或 G

<400> 14

	tccgaccana nggngannan tggag	25
	<210> 15	
	<211> 25	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<220>	
	<221> misc_feature	
	<222> (13)... (13)	
	<223> n=G 或 T	
	<220>	
	<221> misc_feature	
	<222> (14)... (14)	
	<223> N=C 或 T	
	<220>	
	<221> misc_feature	
	<222> (20)... (20)	
	<223> N=A、G 或 T	
	<220>	
	<221> misc_feature	
	<222> (23)... (23)	
	<223> N=A 或 G	
	<220>	
	<221> misc_feature	
	<222> 11, 17	
	<223> n=A、T、C 或 G	
[0008]	<400> 15	
	tccgaccaaa ncnntcnggn atnaa	25
	<210> 16	
	<211> 29	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<220>	
	<221> misc_feature	
	<222> (9)... (9)	
	<223> N=A 或 G	
	<220>	
	<221> misc_feature	
	<222> (12)... (12)	
	<223> N=C 或 T	
	<220>	
	<221> misc_feature	
	<222> (15)... (15)	
	<223> N=A 或 G	
	<220>	
	<221> misc_feature	
	<222> (18)... (18)	
	<223> N=C 或 T	
	<220>	
	<221> misc_feature	
	<222> (21)... (21)	
	<223> N=C 或 T	

	<220>	
	<221> misc_feature	
	<222> (24)... (24)	
	<223> N=C 或 T	
	<220>	
	<221> misc_feature	
	<222> (27)... (27)	
	<223> N=A 或 G	
	<400> 16	
	ccagcacant angancantt naanaanat	29
	<210> 17	
	<211> 29	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<220>	
	<221> misc_feature	
	<222> (15)... (15)	
	<223> N=C 或 T	
	<220>	
	<221> misc_feature	
	<222> (21)... (21)	
	<223> N=A 或 G	
	<220>	
	<221> misc_feature	
	<222> (27)... (27)	
[0009]	<223> N=A 或 G	
	<220>	
	<221> misc_feature	
	<222> 12, 18, 24	
	<223> n=A、T、C 或 G	
	<400> 17	
	ccagcagcca tncnnttngc ntncncca	29
	<210> 18	
	<211> 33	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<400> 18	
	acgtctagaa tgaagagtat actaaaagca atg	33
	<210> 19	
	<211> 34	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<400> 19	
	acgtctagag cgaaacttgg aacgagtaac aact	34
	<210> 20	
	<211> 37	
	<212> DNA	

	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<400> 20	
	atggatccta tgcaatcaac aacttccgtt aagttat	37
	<210> 21	
	<211> 34	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<400> 21	
	taigtgcacc catacatcac ggagataatc tcct	34
	<210> 22	
	<211> 32	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<400> 22	
	ggactagtaa aacaatggcc ctgaccgaag ag	32
	<210> 23	
	<211> 29	
	<212> DNA	
	<213> 人工序列	
[0010]	<220>	
	<223> 合成引物	
	<400> 23	
	ccaagctttc agatggacat cgggtaaac	29
	<210> 24	
	<211> 29	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<400> 24	
	ctgccgcggg gccgcaaatt aaagccttc	29
	<210> 25	
	<211> 29	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	
	<400> 25	
	ctgccgcggt agtacggatt agaagccgc	29
	<210> 26	
	<211> 35	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 合成引物	

	<400> 26 cgggatccaa aacaatggct gcagaccaat tgggtg	35
	<210> 27 <211> 30 <212> DNA <213> 人工序列	
	<220> <223> 合成引物	
	<400> 27 gcgtcgactt aggatttaat gcaggtgacg	30
	<210> 28 <211> 35 <212> DNA <213> 人工序列	
	<220> <223> 合成引物	
	<400> 28 cgggatccaa aacaatgagc gaagtcggta tacag	35
	<210> 29 <211> 34 <212> DNA <213> 人工序列	
	<220> <223> 合成引物	
[0011]	<400> 29 gcgtcgactc ataacgaaaa atcagagaaa ttg	34
	<210> 30 <211> 37 <212> DNA <213> 人工序列	
	<220> <223> 合成引物	
	<400> 30 ggactagtaa aacaatggct tcagaaaaag aaattag	37
	<210> 31 <211> 30 <212> DNA <213> 人工序列	
	<220> <223> 合成引物	
	<400> 31 tcccccgggc tatttgcttc tcttgtaaac	30
	<210> 32 <211> 8 <212> PRT <213> 人工序列	
	<220> <223> 基序	
	<400> 32 Gln Tyr Glu His Phe Asn Lys Ile 1 5	

<210> 33
<211> 8
<212> PRT
<213> 人工序列

[0012] <220>
<223> 基序

<400> 33
Cys Gly Asp Ala Lys Gly Met Ala
1 5

CYP71D-A4 cDNA 序列.**CYP71D-A4 cDNA 编码序列**

ATGAAGAGTATACTAAAAGCAATGGCACTCTCACTGACCACTTCCATT
 GCTCTTGCAACGATCCTTTTGTTCGTTTACAAGTTCGCTACTCGTTCC
 AAATCCACCAAAAAAGCCTTCTGAGCCATGGCGGCTTCCATTATT
 GGTACATGCATCACTTGATTGGTACAACGCCACATCGTGGGGTTAGG
 GATTTAGCCAGAAAGTATGGATCTTTGATGCATTTACAGCTTGGTGAA
 GTTCCAACAATCGTGGTGTCTATCTCCGAAATGGGCTAAAGAGATTTTG
 ACAACGTACGACATTACCTTTGCTAACAGGCCCGAGACTTTAACTGGT
 GAGATTGTTTTATATCACAATACGGATGTTGTTCTTGACACCTTATGGT
 GAATACTGGAGGCAATTACGTAAAATTTGCACATTGGAGCTTTTGAGT
 GTTAAGAAAGTAAAGTCATTTTCAGTCACTTCGTGAAGAGGAGTGTTGG
 AATTTGGTTCAAGAGATTAAAGCTTCAGGTTTCAGGGAGACCGGTAAAC
 CTTTCAGAGAATGTTTTCAAGTTGATTGCAACGATACTTAGTAGAGCC
 GCATTTGGGAAAGGGATCAAGGACCAGAAAGAGTTAACGGAGATTGTG
 AAAGAGATACTGAGGCAAACTGGTGGTTTTGATGTGGCAGATATCTTT
 CCTTCAAAGAAATTTCTTCATCATCTTTCGGGCAAGAGAGCTCGGTTA
 ACTAGCCTTCGCAAAAAGATCGATAATTTAATCGATAACCTTGTAGCT
 GAGCATACTGTTAACACCTCCAGTAAACTAACGAGACACTCCTCGAT
 GTTCTTTTAAGGCTCAAAGACAGTGCTGAATTTCCATTAAACATCTGAT
 AACATTAAAGCCATCATTTTTGGATATGTTTGGAGCAGGCACAGACACT
 TCCTCATCCACAATCGAATGGGCGATTTCCGAACTCATAAAGTGTCGG
 AAAGCAATGGAGAAAGTACAAGCGGAATGAGGAAAGCATTGAACGGA
 AAAGAAAAGATCCATGAGGAAGACATTCAAGAACTAAGCTACTTGAAC
 ATGGTAATCAAAGAAACATTGAGGTTGCACCCTCCACTACCTTGGTT
 CTGCCAAGAGAGTGCCGCCAACCACTCAATTTGGCTGGATACAACATA
 CCCAATAAGACCAAACCTTATTGTCAACGTCTTTGCGATAAAATAGGGAC
 CCTGAATATTGGAAAGACGCTGAAGCTTTCATCCCTGAACGATTTGAA
 AATAGTTCTGCAACTGTCATGGGTGCAGAATACGAGTATCTTCCGTTT
 GGAGCTGGGAGAAGGATGTGTCCTGGAGCCGCACTTGGTTTAGCTAAC
 GTGCAGCTCCCGCTCGCTAATATACTATATCATTTCAACTGGAACTC
 CCCAATGGTGTGAGCTATGACCAGATCGACATGACCGAGAGCTCTGGA
 GCCACGATGCAAAGAAAGACTGAGTTGTTACTCGTTCGAAGTTTCTAG
 (SEQ ID NO:1)

图1

紫穗槐二烯12-氧化酶氨基酸序列

MKSILKAMALSLTTSIALATILLFVYKFATRSKSTKKSLPEPWRLPII
 GHMHHLIGTTPHRGVRDLARKYGSLMHLQLGEVPTIVVSSPKWAKEIL
 TTYDITFANRPETLTGEIVLYHNTDVLAPYGEYWRQLRKICTLELLS
 VKKVKSFQSLREEECWNLVQEIKASGSGRPVNLSENVFKLIATILSRA
 AFGKGIKDQKELTEIVKEILRQTGGFDVADIFPSKKFLHHLSGKRRL
 TSLRKKIDNLDNLVAEHTVNTSSKTNETLLDVLLRLKDSAEPPLTSD
 NIKAIILDMFGAGTDTSSSTIEWAISELIKCPKAMEKVQAE LRKALNG
 KEKIHEEDIQELSYLNMVIKETLRLHPPLPLVLPRECRQPVNLAGYNI
 PNKTKLIVNVFAINRDPEYWKDAEAFIPERFENSATVMGA EYEYLPF
 GAGRRMCPGAALGLANVQLPLANILYHFNWKL PNGVSYDQIDMTESG
 ATMQRKTELLLVPSF (SEQ ID NO:2)

图2

黄花蒿细胞色素P450还原酶DNA

AACPR cDNA 编码序列

```
ATGcAATCAACAACCTCCGTTAAGTTATCTCCCTTCGATCTAATGA
CGGCGTTACTTAACGGCAAGGTATCGTTCGACACATCAAACACATC
GGATACGAATATTCGTTAGCGGTGTTTATGGAGAATCGTGAGCTT
TTGATGATTTTAATACTTTCGGTTGCGGTGTTGATCGGATGCGTTG
TGGTGCTTGTGTGGAGACGGTCGTCGTCGGCGGCGAAGAAAGCGGC
GGAGTCGCGCGGTGATTGTTGTGCCGAAGAAAGTGACGGAGGATGAG
GTTGATGACGGACGGAAGAAAGTTACTGTGTTTTTTTGGAACTCAGA
CTGGTACTGCTGAAGGTTTTGCTAAGGCGCTTGTTGAAGAAGCTAA
AGCGCGATATGAAAAGGCGGTGTTTAAAGTGATTGATTTGGATGAT
TATGCTGCTGAAGATGATGAGTATGAGGAGAAGTTAAAGAAAGAAT
CTCTTGCTTTTTTCTTTTAGCTACGTATGGAGATGGTGAGCCGAC
AGATAATGCTGCTAGATTCTATAAATGGTTTACCGAGGGTGAAGAG
AAAGGTGAATGGCTTGACAAGCTTCAATACGCAGTGTTTGGACTTG
GTAAACAGACAGTATGAGCATTTCAACAAGATTGCGAAGGTGGTCTGA
TGAAAAACTTGTGGAGCAGGGTGCAAAGCGCCTTGTTCTGTGGC
ATGGGAGACGATGATCAATGTATCGAAGACGACTTCACTGCATGGA
AAGAGTTGGTGTGGCCTGAGTTGGATCAATTACTTCGTGATGAGGA
TGATACATCTGTTGCCACTCCATACACAGCTGCTGTTGGAGAATAC
CGTGTTGTGTTCCATGACAAACCAGAGACATATGATCAGGATCAAC
TGACAAATGGCCATGCTGTTTCATGATGCTCAACATCCATGCAGATC
CAATGTGCTGCTCAAAAAGGAGCTCCATTCCCCTCTATCTGACCGG
TCTTGCACTCATTGGAATTTGATATCTCTAATACTGGATTATCGT
ATGAAACTGGGGACCATGTTGGAGTCTACGTTGAGAATCTAAGTGA
AGTTGTGGACGAAGCTGAAAAATTAATAGGTTTACCGCCGCACACT
TATTTCTCAGTACATACTGATAACGAAGACGGGACACCACTTGGTG
GAGCCTCTTTGCCACCTCCTTTCCCTCCATGCACTTTAAGAAAAGC
ATTGGCTTCCCTATGCCGATGTTTTGAGCTCTCCTAAAAAGTCAGCT
TTGCTTGCTTTAGCTGCTCATGCTACTGATTCTACTGAAGCTGATA
GACTGAAATTTTTTGCGTCTCCTGCTGGAAAGGATGAATATGCTCA
GTGGATAGTTGCAAGCCACAGAAGTCTCCTTGAGGTCATGGAGGCC
TTCCCATCAGCTAAGCCTCCGCTTGGTGTTTTTTTTGCATCTGTCTG
CCCCACGTTTGCAGCCGAGATACTATTCATTTCTTCTTCCCCAAA
GTTTGCGCCAAATAGGATTCATGTAACCTGTGTCATTAGTGTATGAG
CAAACACCATCAGGCCGCGTTTCAAAAGGAGTCTGTTCAACATGGA
TGAAGAATGCCGTGCCTATGACAGAAAGCCAGGATTGCAGTTGGGC
CCCAATTTATGTTAGAATCAATTTTCAGACTTCCTTCTGATCCT
AAGGTCCCAGTTATCATGATTGGCCAGGCACTGGATTGGCTCCAT
TTAGAGGTTTCCCTTCAGGAAAGGTTAGCTCAGAAGGAAGCTGGGAC
TGAGCTCGGAACAGCCATCTTATTCTTCGGATGCAGGAATCGCAA
GTGGATTTTCATATATGAGGACGAGCTTAATAATTTTCGTGGAGACGG
GGGCTCTTTCCGAGCTTGTACGGCCTTCTCTCGTGAAGGTGCCAC
TAAGGAGTACGTGCAACACAAGATGACTCAGAAGGCTTCGGATATC
TGGAATTTACTCTCTGAGGGAGCATAATTTGTATGTTTGGGTGATG
CCAAAGGCATGGCCAAAGATGTACATCGGACTCTGCACACTATTGT
GCAAGAACAGGGATCTCTAGACTCCTCAAAGGCGGAGCTCTACGTG
AAGAATCTACAAATGGCAGGAAGATATCTCCGTGATGTATGGGTCTG
ACATGGAACAGAAGTTGATTTCCGAAGAAGACCTCGAGTAA (SEQ
ID NO:3)
```

图3

黄花蒿细胞色素P450还原酶氨基酸序列

MQSTTSVKLSFPDLMTALLNGKVSFDTSNTSDTNIP LAVFMENREL
LMILTTSVAVLIGCVVVLVWRRSSSAKKAAESPVIVVPKKVTEDE
VDDGRKKVTVFFGTQTGTAE GFAKALVEEAKARYEKAVFKVIDLDD
YAAEDDEYEEKLKKE SLAFFFLATYGDGEPTDNAARFYKWFTEGEE
KGEWLDKLQYAVFGLGNRQYEHFNKIAKV VDEKLVEQGA KRLVPVG
MGDDDQCIEDDFTAWKELVWPELDQLLRDEDDTSVATPYTAAVGEY
RVVFHDKPETYDQDQLTNGHAVHDAQHPCRSNVAVKKELHSPLSDR
SCTHLEFDISNTGLSYETGDHVG VYVENLSEVVDEAEK LIGLP PHT
YFSVHTDNEDGTPLGGASLPPFP PCTLRKALASYADVLSSPKKSA
LLALAAHATDSTEADRLKFFASPAGKDEYAQWIVASHRSLLEVMEA
FPSAKPPLGVFFASVAPRLQPRYYSISSSPKFAPNRIHVT CALVYE
QTPSGRVHKGVCSTWMKNAVPMTESQDCSWAPIYVRTSNFRLPSDP
KVPVIMIGPGTGLAPFRGFLQERLAQKEAGTELGTAILFFGCRNRK
VDFIYEDELNNFVETGALSELVTAFSREGATKEYVQHKMTQKASDI
WNLLSEGAYLYVCGDAKGMAKDVHRTLHTIVQE QGSLDSSKAELYV
KNLQ MAGRYLRDVWVDMEQKLISEEDLE (SEQ ID NO:4)

图4

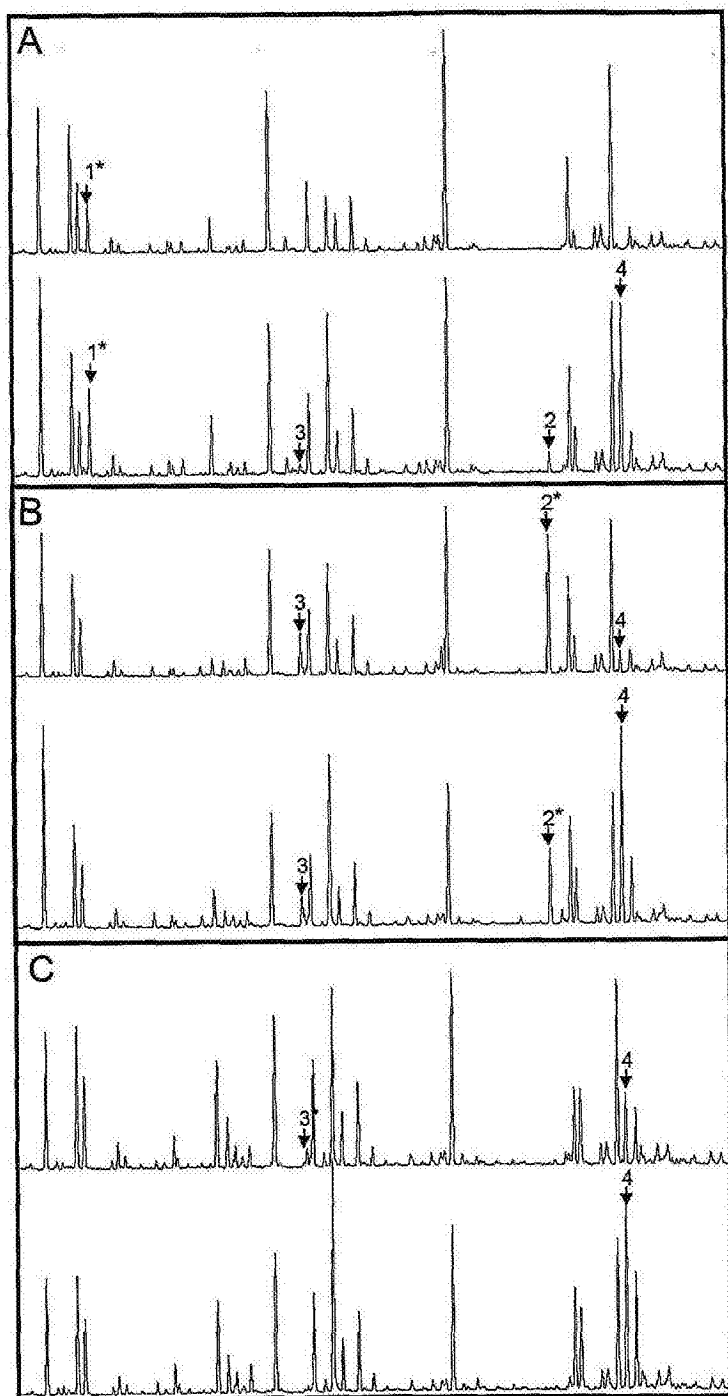


图5

丰度 $R_t=34.53$, 样品, 与标样99%MS匹配

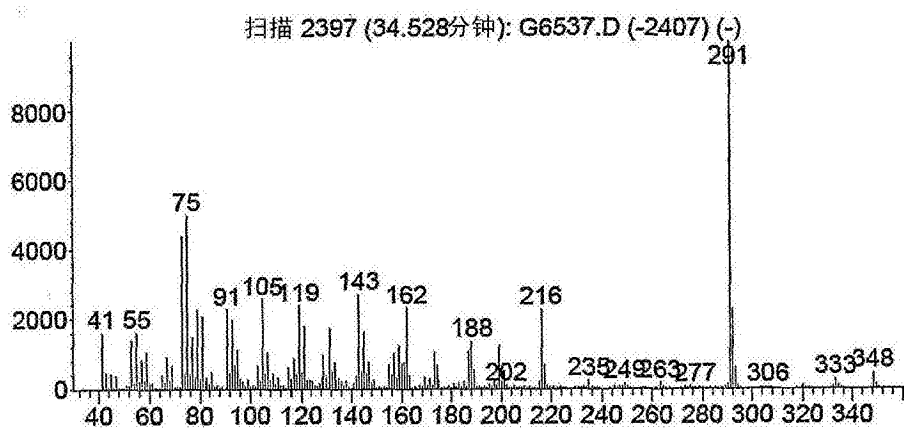
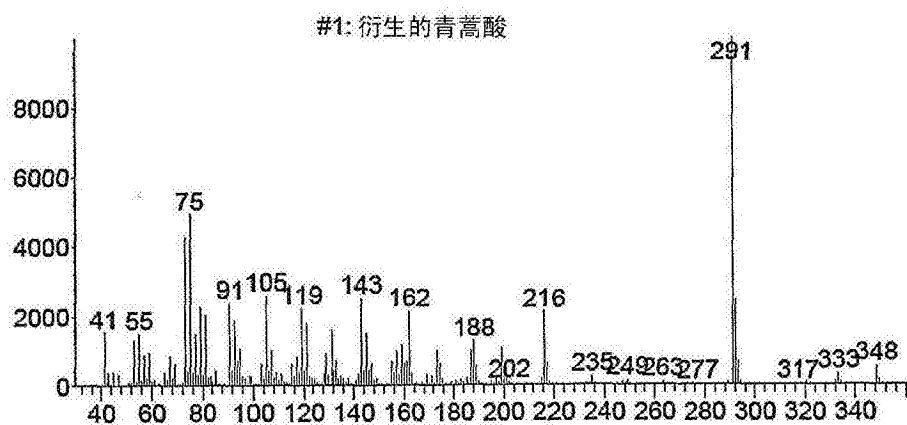


图6A

$m/z \rightarrow$

丰度 $R_t=34.57$, 标样



$m/z \rightarrow$

图6B

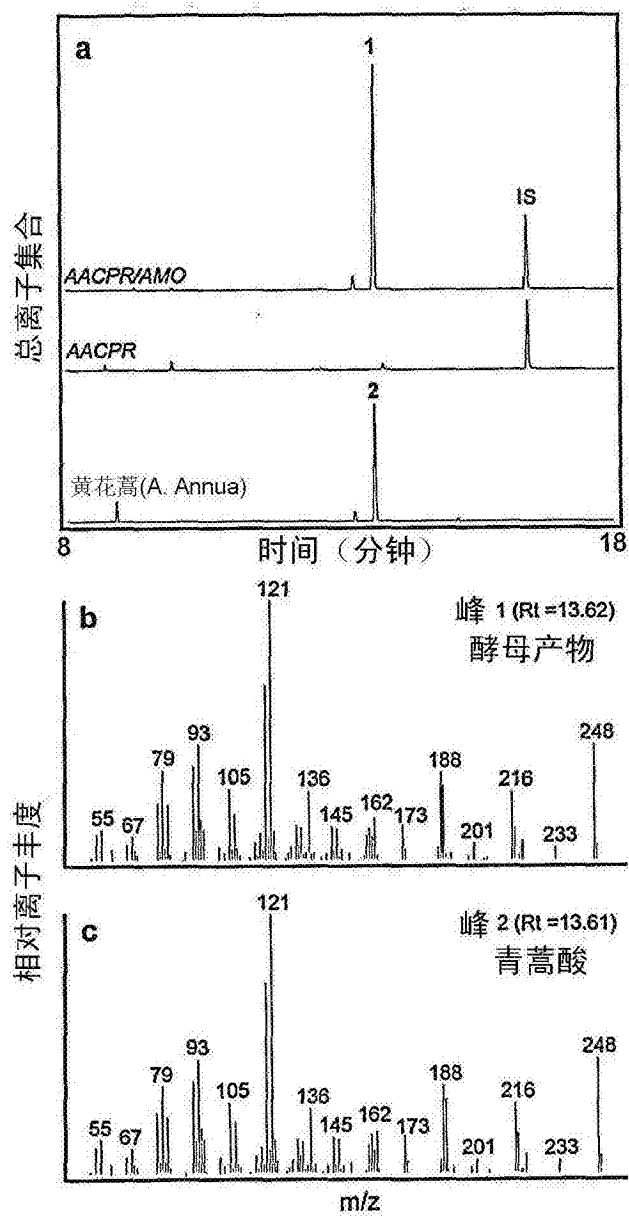


图7

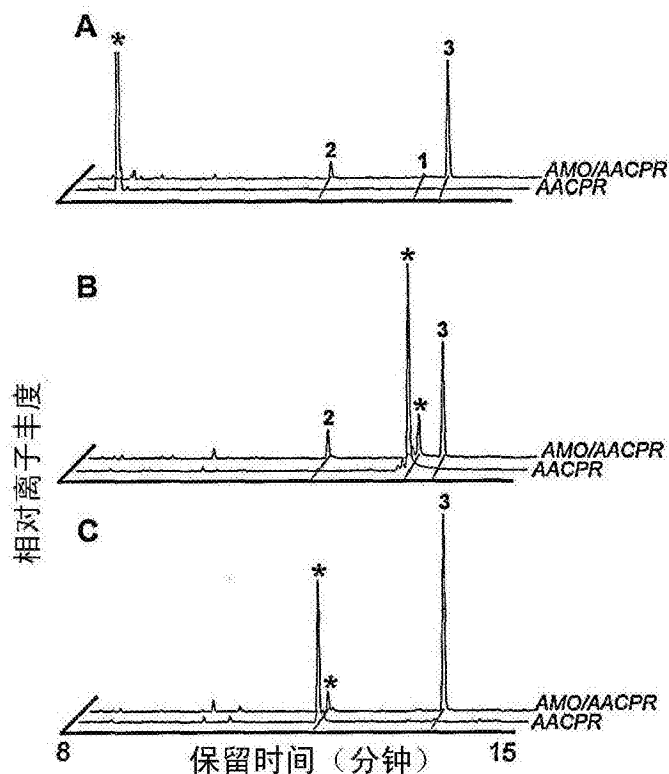


图8

>71D-B1-nt

ATGGCACTTTCACCTGACCACCTCCATTGCTCTTGCCACGATCCTTTTCTTCGTAATTTACAAGTTCGCTAC
 TCGTTCCAAATCCACAAAAACAGCCTTCCTGAGCCATGGCGACTTCCCATTATTGGTCACATGCATCACT
 TGATTGGTACAATACCACATCGTGGGCTTATGGATTAGCCAGAAAGTATGGATCTTAAATGCATTTACAG
 CTTGGTGAAGTTTCAACAATCGTGGTGTCTATCCGAAATGGGCTAAAGAGATTTTGACAACGTACGACAT
 TGCCTTTGCTAACAGGCCCTGGACTTTGGCTGGTGAGATTGTTGTATATCGCAATACAAATATTGCTGCTG
 CACCTTATGGTGAATACTGGAGGCGATTACGTAACTTTGCACATCGGAGCTTATGAGTGTAAAGAAAGTA
 AAGTCATATCAGTCGCTTCGTGAAGAGGAGTGTGGAAATTTGGTTCAAGAGATTAAAGCTTCAGGTTTCAGG
 GATACCGGTTAACCTTTCAGAGAACATTTCAAGTTGATTGCAACGATACTTTGTAGAGCCGCGTTTGGA
 AAGGAGTCAAGGACCAGAAGGAGTGTACGGAGATTATGAAAGAGATGTTGAGGGAAGTTGGTGGTTTTGAT
 GTGGCAGATATCTTTCCCTTCGAAGAAATTTCTTCATCATCTTTTCAGGCAAGAGAGCCAGGTTAACTAGCAT
 TCATAAGAAGCTCGATAATTTTATCAATAACCTTGTGCTGAGCATACTTTCAAACTTCAAGTAAACTG
 AGGAGACACTTCTTGATGTTCTTCTAAGGCTCAAAGATAGCGCTGAATTTCCATTACAGCTGACAATGTT
 AAAGCCATCATTTTGGATATATTTGCAGCAGGGACAGACACTTCATCAACCACAATCGAATGGGCGATTTT
 GGAATCATAAAGTGTCCGAGAGCGATGGAGAAAGTACAAGCAGAACTGAGGAAAGCACTTAACGGAAAAG
 AAAAGATCCATGAGGAAGATATTCAAGGACTAAGCTACTTAACTTGGTAATCAAAGAAACATTAAGGTTG
 CACCCTCCACTACCTTGTGCTCAAGAGAGTGCCGTGAACCAAGTCAATTTGGCTGGATACGACATACCCAA
 TAAGACAAGACTTATTGTCAACGTCTTTGCGATAAATAGGGACCCAGAATACTGGAAAGACGCTGAAATTT
 TCATCCCCGAACGATTTGAAAATAGTTCTACAACCTCATGGGTGCAGAATATGAGTATCTTCCGTTTGGA
 GCTGGGAGAAGGATGTGCTCGAGCCGCACTTGGTTTAGCCAACGTGCAGCTACCGCTCGCTAATATACT
 ATATCATTTCAACTGGAACTCCCCAACGGTGCGAGCTATGATCAGATCGACATGACCGAGAGGTTTGGA
 TCTCGGTTGAAAGAAAGACTCAGTTGTTACTCGTACCAAGTTTCTAG (SEQ ID NO:5)

图9

>71D-B1-aa

MALSLTTSIALATILFFVIYKFATRISKSTKNSLPEPWRLPIIGHMHHLIGTIPHRGLMDLARKYGSIMHLQ
 LGEVSTIVVSSPKWAKEILTYYDIAFANRPWTLAGEIVVYRNTNIAAAPYGEYWRRLRKLTSELMSVKKV
 KSYQSLREEECWNLVQEIKASGSGIPVNLSENIFKLIATILCRAAFGKGVDQKECTEIMKEMLRVGGFD
 VADIFPSKKFLHHLGKRRLTSIHKKLDNFINNVAEHTFKTSSKTEETLLDVLLRLKDSAEFFLTADNV
 KAIILDIFAAGTDTSSSTTIEWAISELIKCPAMEKVQAE LRKALNGKEKIHEEDIQGLSYLNLVIKETRL
 HPPLPLLPRECREPVNLGYDIPNKTRLIVNVFAINRDPEYWKDAEIFIPERFENSSTLMGAIEYELPFG
 AGRRMCPGAALGLANVQLPLANILYHFNWKL PNGASYDQIDMTERFGISVERKTQLLLVPSE (SEQ ID
 NO:6)

图10

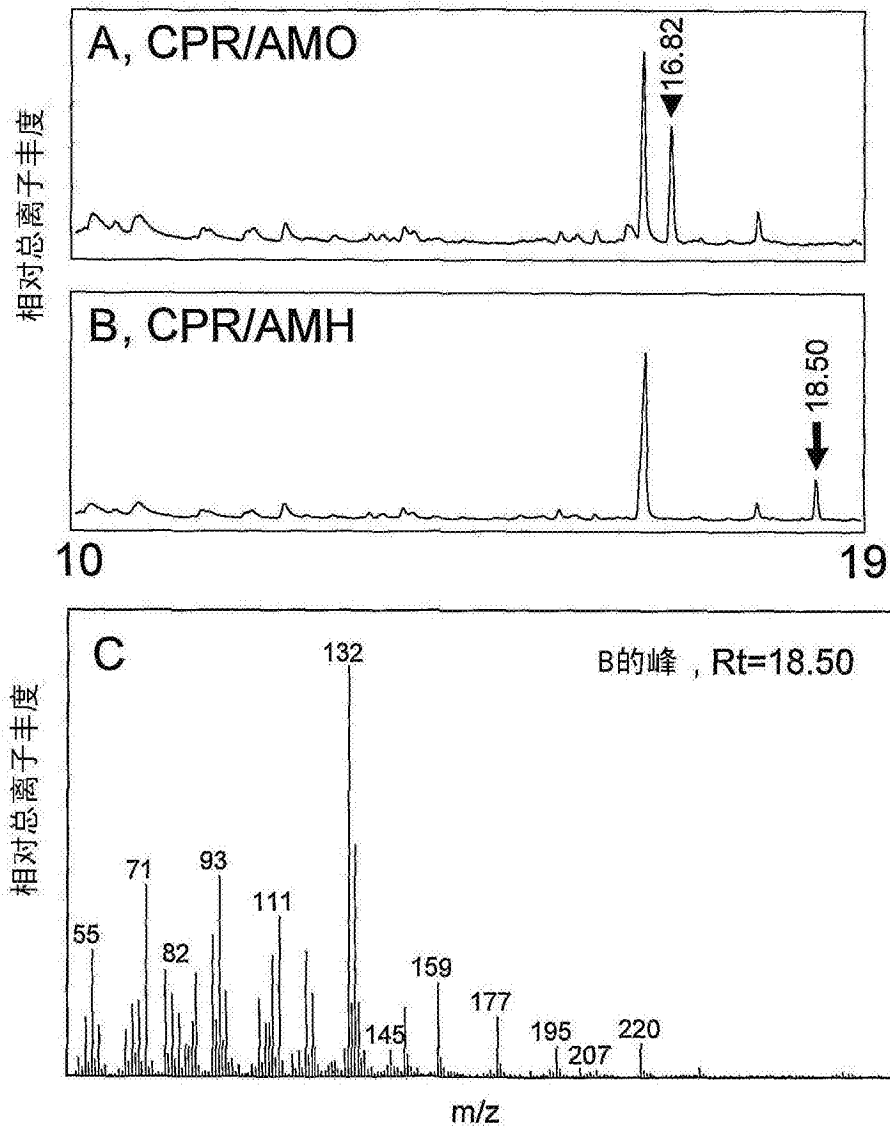


图11

>71D-C1-nt

GCCCTTCGAGCCGTATGGGGATTACTGGCGGCAATTACGTAAACTTTGCACATTGGAGCTTTTGAGTGCTA
AGAAAGTAGAGTCATATCAGTCGCTTCGTGAAGAGGAGTGTTGGAATTTAGTTCAAGAGATTAAAGCTTCA
GGTTCAGGGATACCGGTTAACCTTTTCAGAGAATATTTACAAGTTGGTTGCAATGATACTTAGTAGAGCTGC
GTTTGGGAAAAGAAATCAAGGACCATAAGGAGTTTACGGAGCTTGTGGAACAGATGTTGAGGGAACTTGGTG
GTTTTGATGTGGCAGATATCTTTTCCTTCGCAGAAATTTCTACATCATATTTCCGGCAAGAGATCTAGGTTA
ACTAGCATTCACAAAAAGCTCGATAATTTAATCAATAACCTTGTTGCTGAGCATATTGTTGAAGCCTCAAG
TAAACTAAGGAGACGCTCCTTGATGTTCTTCTAAGGCACAAAGATAGCCTTGAATTCCCATTGACAGCTG
ATAACGTTAAAGCCATCATTTTGGTATGAATTAATCCAATATATTTTTTTTTTCAAAAGGCCATAATAGTG
TTAAACAAGCTTGAAATTTTTTATAACTAAGTACATGCACTAAGTTAGTACTCGTGAAAAATATAATGAGT
CATCATAGGGGTTCCATGAAATATACAGGACATGTTTACAGCAGGCACAGACACTTCGTCAACCACAATCG
AATGGGTGATTTCCGAACTCATAAAGTGTCCGAGAGCTATGGAGAAAATACAAGCGGAACTGAGGAAAGCA
CTTAACGGAAAAGAAAAGATCCACGAGGAAGACATCCAAGAACTAAGCTACTTAAACTTGGTAATCAAAGA
AACATTAAGGTTGCACCCCTCCACTACCCTTGGTTTTGCCACGAGAGTGCCGTCAACCAGTCAATTTGGCTG
GATATGACATACCCAATAAGACCAAACCTTATGTCAACGTCTTTGCGATAAATAGGGACCCTGAATACTGG
AAAGACGCTGAATCTTTCATCCCAGAGCGCTTCTTAACCTCTGGT (SEQ ID NO:7)

图12

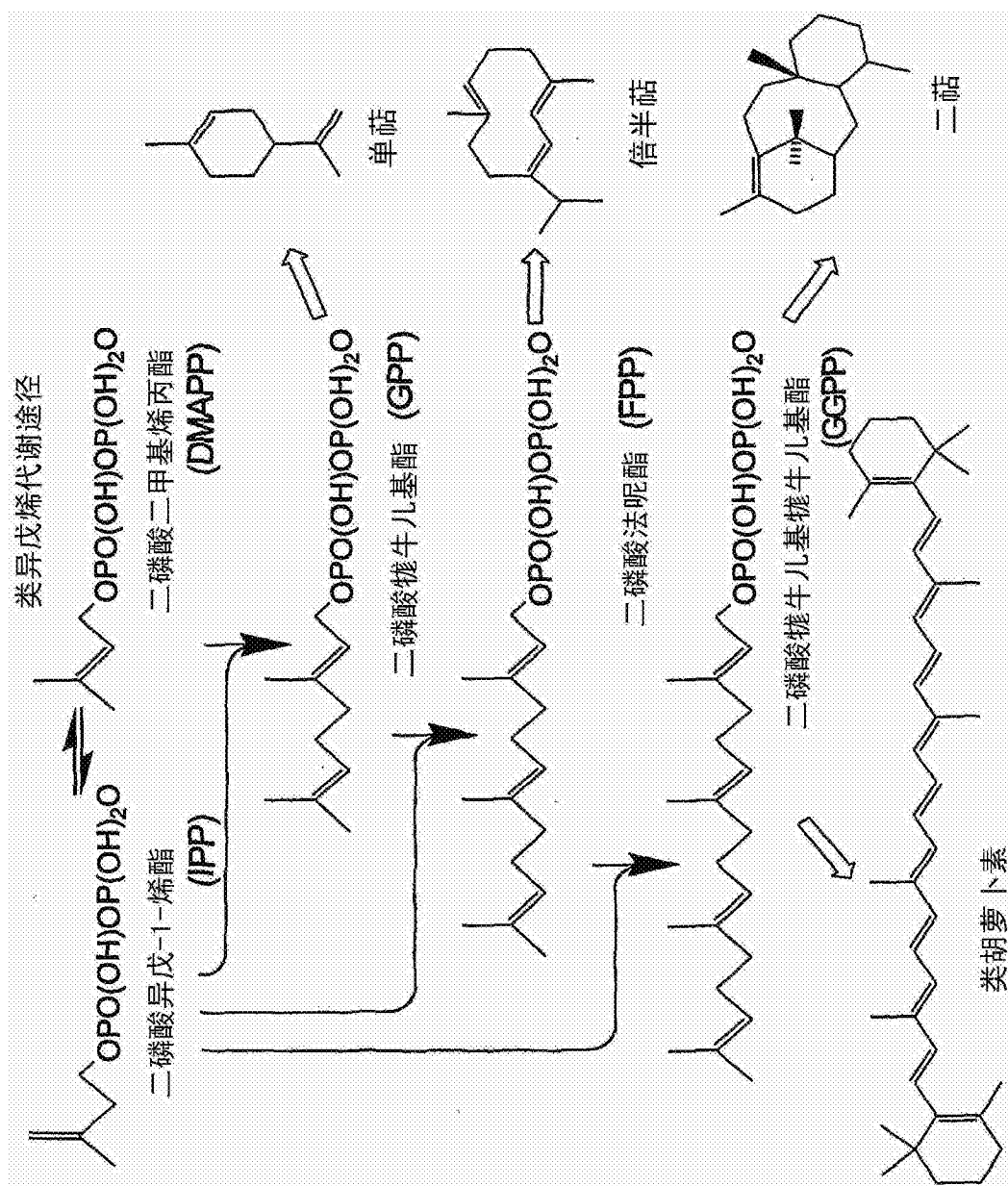


图13

甲羟戊酸途径

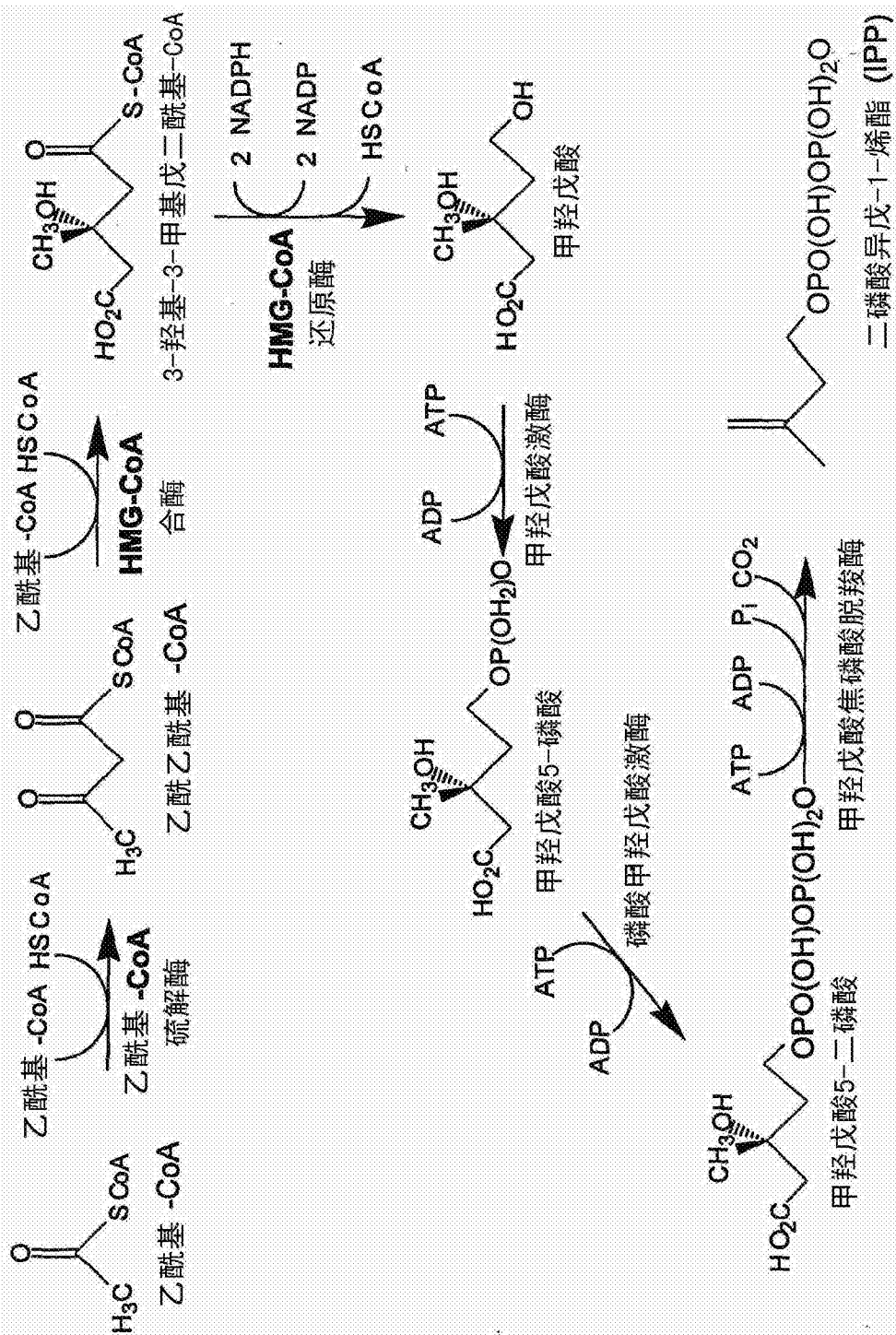


图14

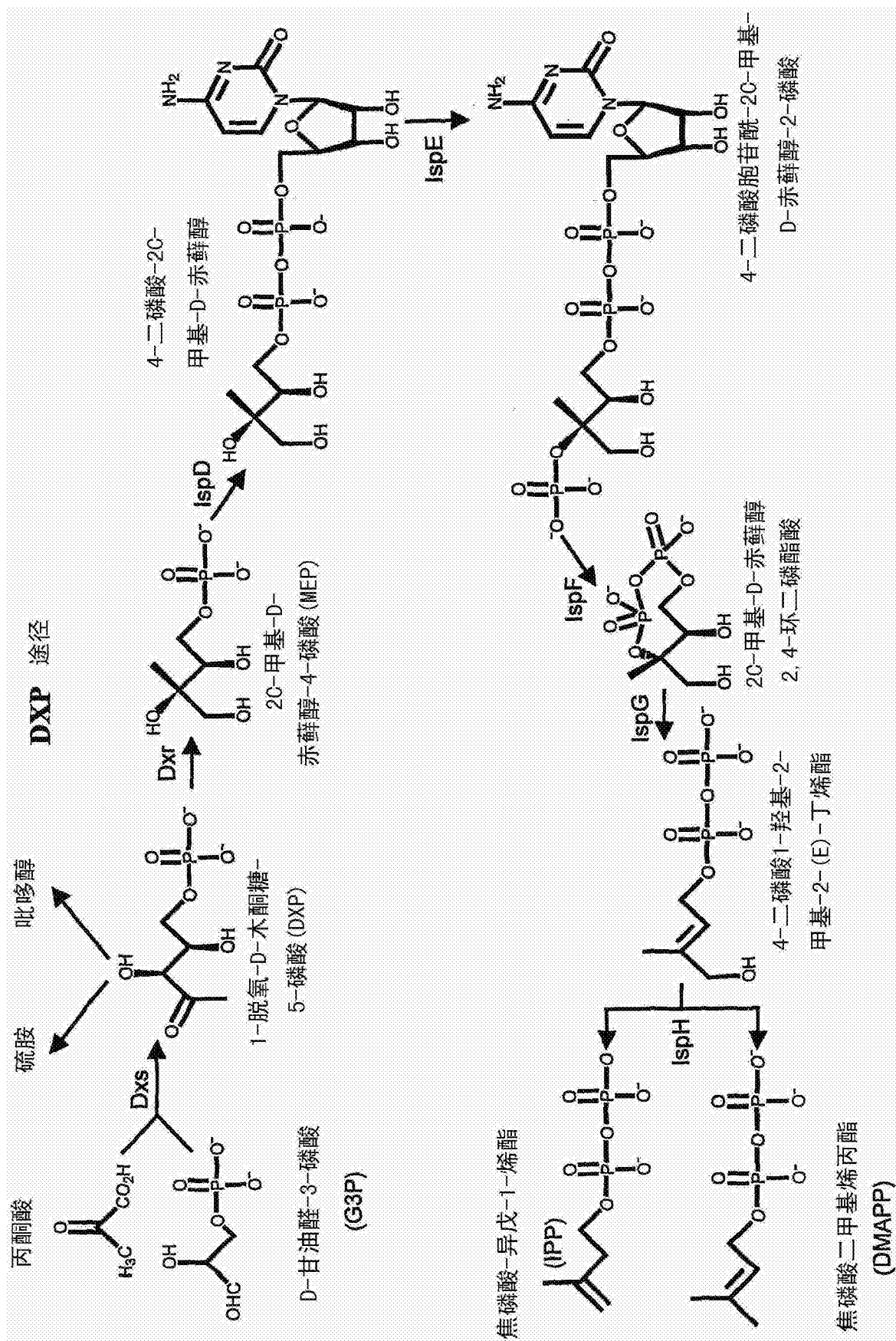


图15