



(12) 发明专利

(10) 授权公告号 CN 101688191 B

(45) 授权公告日 2013. 05. 01

(21) 申请号 200880011310. 6

(22) 申请日 2008. 04. 02

(30) 优先权数据

0754235 2007. 04. 03 FR

(85) PCT申请进入国家阶段日

2009. 10. 09

(86) PCT申请的申请数据

PCT/FR2008/050576 2008. 04. 02

(87) PCT申请的公布数据

W02008/142318 FR 2008. 11. 27

(73) 专利权人 菲力普莫里斯产品股份公司

地址 瑞士纳沙泰尔

(72) 发明人 克里斯托弗·萨洛 丹尼斯·隆特因

阿兰·蒂西耶

(74) 专利代理机构 中原信达知识产权代理有限

责任公司 11219

代理人 张颖 樊卫民

(51) Int. Cl.

C12N 9/10 (2006. 01)

C12N 15/52 (2006. 01)

(56) 对比文件

US 20040249219 A1, 2004. 12. 09, 正文第 5

页右栏第 46 段 .

WO 2004111183 A2, 2004. 12. 23, 全文 .

STEVE R. ADAMS 等 . EVIDENCE FOR

TRANS-TRANS AND CIS-CIS FARNESYL

PYROPHOSPHATE SYNTHESIS IN GOSSYPIUM

HIRSUTUM. 《Phytochemist》. 1973, 第 12 卷

2167-2172.

P. F. HEINSTEIN 等 . Biosynthesis of

Gossypol INCORPORATION OF MEVALONATE-2-14C

AND ISOPRENYL PYROPHOSPHATES. 《THE JOURNAL

OF BIOLOGICAL CHEMISTRY》. 1970, 第 245 卷 (第

18 期), 4658-4665.

审查员 高宇

权利要求书1页 说明书25页

序列表16页 附图7页

(54) 发明名称

具有多种产物的 Z, Z- 法呢基二磷酸合酶和
倍半萜合酶的编码基因及其应用

(57) 摘要

本发明涉及参与 SB 类型的倍半萜 (α - 檀
香萜、表 - β - 檀香萜、顺式 - α - 香柠檬烯、反
式 - α - 香柠檬烯和内 - β - 香柠檬烯) 及其前体
Z, Z- 法呢基二磷酸 (ZZ-FPP) 的生物合成途径的
基因, 包括 Z, Z-FPP 合酶和倍半萜 SB 合酶, 及其在
生产 SB 类型的倍半萜化合物中的应用。

1. 在具有 DMAP 和 IPP 源的细胞中从 Z,Z-FPP 生产 SB 类型的倍半萜的方法,该方法包括:

a) 在所述细胞中导入具有含有第一个基因的表达盒的构建物,并导入具有含有第二个基因的表达盒的构建物,所述第一个基因编码 SEQ ID No. 2 所示的 zFPS,该 zFPS 具有 Z,Z-FPP 合酶活性,所述第二个基因编码 SEQ ID No. 4 所示的 SB 合酶,该 SB 合酶具有 SB 合酶活性;

b) 将转化的细胞培养在适当的条件下,以表达所述第一个和所述第二个基因;以及,

c) 任选地,收集所述细胞和/或培养基中包含的 SB 类型的倍半萜。

2. 分离或重组的蛋白,其由 SEQ ID No. 2 所示,并具有 Z,Z-FPP 合酶活性。

3. 核酸,其由编码权利要求 2 的蛋白的核苷酸序列构成。

4. 表达盒,其含有权利要求 3 的核酸。

5. 载体,其含有权利要求 3 的核酸或权利要求 4 的表达盒。

6. 细菌宿主细胞,其含有权利要求 3 的核酸、权利要求 4 的表达盒或权利要求 5 的载体。

7. 产生非人类转基因生物体的方法,包括:

a) 将具有含有编码权利要求 2 的 Z,Z-FPP 合酶的第一个基因的表达盒的构建物,以及具有含有编码 SEQ ID No. 4 所示的 SB 合酶并且具有 SB 合酶活性的第二个基因的表达盒的构建物,导入到生物体的细胞中;以及

b) 从所述细胞重构生物体,并筛选表达所导入的两个基因的非人类转基因生物体。

8. 权利要求 7 的方法,其特征在于所述生物体是具有植物腺毛的植物。

9. 权利要求 8 的方法,其特征在于所述生物体来自烟草属 (Nicotiana) 或茄属 (Solanum) 植物。

10. 在具有 IPP 和 DMAP 源的细胞中从 IPP 和 DMAP 生产 Z,Z-法呢基二磷酸 (Z,Z-FPP) 的方法,该方法包括:

a) 在所述细胞中导入具有表达盒的构建物,该表达盒带有编码权利要求 2 的蛋白的核苷酸序列;

b) 将转化的细胞培养在适当的条件下,以表达基因;以及

c) 任选地,收集所述细胞和/或培养基中包含的 Z,Z-FPP。

11. 权利要求 2 的蛋白、权利要求 3 的核酸、权利要求 6 的宿主细胞或权利要求 7 或 8 的方法制备得到的转基因生物体在制备 Z,Z-FPP、 α -檀香萜、表- β -檀香萜、 α -香柠檬烯、 β -香柠檬烯、内- β -香柠檬烯、 α -檀香醇、表- β -檀香醇、 α -香柠檬醇、 β -香柠檬醇和 Z,Z-法呢醇中的应用。

12. 微生物细胞,其特征在于权利要求 2 的蛋白的编码基因被失活。

13. 用于将编码 SEQ ID No. 2 所示的并具有 Z,Z-FPP 合酶活性的 zFPS 的 zFPS 基因和编码 SEQ ID No. 4 所示的并具有 SB 合酶活性的 SB 合酶的 SBS 基因,导入到与多毛番茄不是性亲和的物种中的方法,所述方法通过原生质体融合进行。

14. 权利要求 10 的方法,还包括通过磷酸酶将 Z,Z-FPP 去磷酸化用于生产 Z,Z-法呢醇。

具有多种产物的 Z, Z- 法呢基二磷酸合酶和倍半萜合酶的 编码基因及其应用

发明领域

[0001] 本发明涉及负责倍半萜混合物合成的两个基因,以及它们在活生物体例如细菌、酵母、动物细胞和植物中制备所述化合物中的用途。

背景技术

[0002] 萜类是存在于所有活生物体例如细菌、真菌、动物和植物中的分子。它们形成了生命世界中最大类别的天然产物。在植物中,这些分子作为激素(赤霉素,细胞分裂素,油菜素类固醇和脱落酸),或作为参与光合作用(类胡萝卜素,叶绿素,质体醌和叶绿醇)、蛋白异戊二烯化过程和膜结构(甾醇)的化合物,在初级代谢中发挥作用。然而,从次级代谢产生的类萜占该类分子结构多样性的大部分。所谓的次级类萜主要在环境相互作用中发挥作用,例如吸引对植物有益的昆虫(授粉以及植食性昆虫的捕食者),防御病原体和昆虫,以及针对光氧化胁迫的保护作用(Tholl, 2006)。

[0003] 萜类由多个 5 碳异戊二烯基单元构成。根据包含的异戊二烯基单元的数量,萜类被分类成单萜(具有 10 个碳原子的化合物或 C10)、倍半萜(C15)、二萜(C20)、两倍半萜(C25)、三萜(C30)、四萜(C40),等等。所有萜的通用前体是异戊烯二磷酸(IPP)和二甲基烯丙基二磷酸(DMAPP)。两个独立的途径导致 IPP 和 DMAPP 的形成。一种是,甲羟戊酸(MEV)途径,位于真核细胞的细胞质中,而另一个,甲基赤藓糖醇(MEP)途径,已知只存在于某些细菌和植物中,在植物中它位于叶绿体中(Rodriguez-Concepcion 和 Boronat, 2002)。在一定数量的生物体中,已经知道了这两个途径中的所有步骤以及编码所述步骤的酶的所有基因。

[0004] 从共同前体 IPP 和 DMAPP 生物合成萜的第一步由异戊二烯二磷酸合酶(或异戊二烯基转移酶)执行,它通过将高烯丙型前体 IPP 与烯丙型前体(DMAPP、香叶基二磷酸、法呢基二磷酸或香叶基香叶基二磷酸)缩合,产生长度不同的异戊二烯二磷酸链。有两种主要类别的异戊二烯二磷酸合酶:顺式异戊二烯基转移酶或 Z- 异戊二烯基转移酶,以及反式异戊二烯基转移酶或 E- 异戊二烯基转移酶。顺式异戊二烯基转移酶导致形成顺式构型的双键,而反式异戊二烯基转移酶导致形成反式构型的双键(Koyama 1999)。例如,通过反式法呢基二磷酸合酶或 E-FPS 向 DMAPP 连续添加两个单位的 IPP,产生了 E, E- 法呢基二磷酸(E, E-FPP, C15)。链延伸反应在消除二磷酸离子以形成烯丙型异戊二烯基后,通过烯丙位阳离子的形成开始。接下来,加入一个 IPP,立体特异性地消除了 2 位上的质子,在产物中形成了新的 C-C 键和新的双键。该 IPP 与含有烯丙型异戊二烯二磷酸的反复立体特异性缩合,导致合成了其链长和立体化学对于每种酶是特异性的异戊二烯二磷酸。具体来说,终产物的链长可以有相当大的变化,其范围从 C10(香叶基二磷酸)到聚异戊二烯二磷酸,后者是由几千个碳原子构成的天然乳胶的前体。

[0005] 正如以前提到的,异戊二烯基转移酶可以根据每次延伸环后形成的双键的立体化学(E 或 Z),分成两个不同的遗传家族(Poulter, 2006; Liang 等, 2002)。

[0006] 到目前为止已表征的 E- 异戊二烯基转移酶导致合成具有 E 构型的短链异戊二烯磷酸 (C10 到 C50)。例子包括香叶基二磷酸合酶 (GPS)、法呢基二磷酸合酶 (FPS)、香叶基香叶基二磷酸合酶 (GGPS)、八异戊二烯二磷酸合酶 (OPS)、茄呢基二磷酸合酶 (SPS) 和十异戊二烯二磷酸合酶 (DPS)，它们分别催化 (E)-GPP (C10)、(E, E)-FPP (C15)、(E, E, E)-GGPP (C20)、(E, E, E, E, E, E, E)-OPP (C40)、(E, E, E, E, E, E, E, E)-SPP (C45) 和 (E, E, E, E, E, E, E, E, E)-DPP (C50) 的合成。应该指出, 一般来说, (E, E)-FPP 是负责合成具有 20 个以上碳的链的异戊二烯基转移酶的底物。第一个被表征的编码 (E)- 异戊二烯基转移酶的基因是大鼠 FPS 的编码基因 (Clarke 等, 1987)。对 E- 异戊二烯基转移酶编码序列进行的比对显示出 DDXXD 类型的两个保守基序。通过 FPS 的 X- 射线晶体学首次确定的 E- 异戊二烯基转移酶的三维结构 (Tarshis 等, 1994), 以及位点定向诱变研究, 显示出两个 DDXXD 基序参与了底物结合和产生催化活性。最后, 已经显示, 少数位于第一个 DDXXD 基序上游的特定氨基酸, 决定了合成的异戊二烯链的长度 (Wang&Ohnuma, 1999)。

[0007] 第一个编码 Z- 异戊二烯基转移酶的基因在藤黄微球菌 (*Micrococcus luteus*) 中被克隆出来 (Shimizu 等, 1998)。该基因编码十一异戊二烯基转移酶 (UPS), 该酶使用 E-FPP 作为底物形成具有 Z, E 立体化学的 55 碳异戊二烯二磷酸。该分子在某些细菌中参与细胞壁的肽聚糖生物合成。在拟南芥 (*Arabidopsis*) 中表征的同源序列显示出负责合成具有 100 到 130 个碳原子的 Z, E- 异戊二烯二磷酸 (Oh 等, 2000)。这些酶的氨基酸序列分析没有揭示出与 E- 异戊二烯基转移酶的任何同源性 (Koyama, 1999)。具体来说, Z- 异戊二烯基转移酶不具有富含天冬氨酸的基序 (DDXXD)。已经鉴定了 7 个保守区, 它们都在其底物结合和催化中或多或少地发挥作用。所有到目前为止表征的 Z- 异戊二烯基转移酶都合成碳链长度大于或等于 55 个碳原子的异戊二烯二磷酸, 例外的是结核分枝杆菌 (*Mycobacterium tuberculosis*) 的 Z, E-FPS, 它使用 E-GPP 作为底物形成 Z, E-FPP (Schulbach 等, 2000)。最近对藤黄微球菌的 UPS 进行的位点定向诱变研究, 鉴定到了几个在合成的异戊二烯链的延伸环数中发挥关键作用的氨基酸 (Kharel 等, 2006)。

[0008] 异戊二烯二磷酸是一类酶——萜合酶的底物, 该酶催化 C10 (单萜)、C15 (倍半萜)、C20 (二萜) 和 C30 (三萜) 中基本环或无环萜骨架的形成 (Cane 1999; McMillan 和 Beale, 1999; Wise 和 Croteau, 1999)。这些酶的反应多样性构成了在自然界中发现的环萜骨架的多样性的基础。根据在分子的 2 位和 6 位上双键的立体化学构型, 理论上可能存在 4 种法呢基二磷酸的立体异构体 (E, E, Z, E, E, Z 或 Z, Z)。然而, 到目前为止, 所有基因已经被表征的倍半萜合酶都使用 E, E-FPP 作为底物。这种情况的例子是烟草的 5- 表马兜铃烯 (5-*epi-aristolochene*) 合酶 (Facchini&Chappell, 1992), 大冷杉 (*Abies grandis*) 的 (E)- α - 没药烯合酶 (Bohlmann 等, 1998), 菊苣的大根香叶烯 A 合酶 (Bouwmeester 等, 2002), 黄花蒿的紫穗槐-4, 11- 二烯合酶 (Wallaart 等, 2001), 番茄的大根香叶烯 C 合酶 (Colby 等, 1998), 黄花蒿的石竹烯合酶 (Cai 等, 2002) 和橙子的凡伦橘烯合酶 (Sharon-Asa 等, 2003)。然而, 最近的研究报道, 催化 14 种倍半萜混合物合成的玉米倍半萜合酶 (tps4) 接受 E, E-FPP 和 Z, E-FPP (Kollner 等, 2006), 从而显示出倍半萜合酶使用 E, E-FPP 之外的异构体作为底物。但是, 到目前为止, 还没有克隆到的倍半萜合酶使用 Z, Z-FPP。

[0009] 多毛番茄 (*Solanum habrochaites*) (以前被称为多毛番茄 (*Lycopersicon hirsutum*)) 的 VI 型腺毛分泌大量属于两种不同类别的倍半萜烯烃。I 类特别含有大根香叶

烯 B, II 类特别含有 α -檀香萜和 α -香柠檬烯。栽培的番茄基因型 (番茄 (*Lycopersicon esculentum* = *Solanum lycopersicum*)) 和野生型番茄基因型 (多毛番茄 (*Lycopersicon hirsutum* = *Solanum habrochaites*)) 之间的分离分析显示,这两种不同类别的倍半萜的生物合成由作图于两个不同染色体上的两个不同基因座 (Sst1A 和 Sst2) 控制 (van der Hoeven 等, 2000)。Sst1A 基因座位于 6 号染色体上,该基因座处的基因负责 I 类倍半萜的积累。多毛番茄的 Sst2 等位基因位于 8 号染色体上,控制 II 类倍半萜例如 α -檀香萜的积累,但是也控制顺式- α -香柠檬烯、反式- α -香柠檬烯、表- β -檀香萜和 β -香柠檬烯以及其它未鉴定的次要化合物的积累。尽管 Sst1A 基因座处基因的编码序列已经被鉴定 (van der Hoeven 等, 2000), 但是 Sst2 基因座处则没有鉴定。

发明内容

[0010] 本发明涉及负责倍半萜混合物合成的两个基因,该倍半萜混合物主要由 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯以及它们的前体 Z, Z-法呢基二磷酸 (Z, Z-FPP) 构成,及其在活生物体例如细菌、酵母、动物细胞和植物中制备所述化合物的用途。

[0011] 更具体来说,本发明描述了多毛番茄负责合成 II 类多毛番茄倍半萜 (特别是 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯) 的两个基因的鉴定和性质, II 类多毛番茄倍半萜在余下的文本中将被称为 SB 类型的倍半萜 (指檀香萜和香柠檬烯)。第一个基因编码 Z, Z-法呢基二磷酸合酶,第二个编码使用 Z, Z-法呢基二磷酸作为底物的多产物倍半萜合酶。相应的重组蛋白的生产显示出,在体外获得的倍半萜的分布 (profile) 与受番茄 Sst2 基因座控制的分布一致。

[0012] 一方面,本发明涉及参与从 IPP 和 DMAP 生物合成 Z, Z-FPP、以及从 Z, Z-FPP 生物合成主要由 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯构成的倍半萜混合物的酶活性,还涉及负责所述活性的肽序列以及编码所述肽序列的核酸序列。本发明还涉及使用所述酶活性生产所述化合物或其衍生物的方法。

[0013] 因此,本发明涉及在具有 IPP 和 DMAP 源的细胞中,从 Z, Z-FPP 生产主要由 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯构成的倍半萜混合物的方法,该方法包括:

[0014] a) 在所述细胞中导入带有含有编码本发明 Z, Z-FPS 的第一个基因的表达盒的构建物,以及带有含有编码本发明被称为 SBS 的倍半萜合酶的第二个基因的表达盒的构建物;

[0015] b) 将转化的细胞在适当的条件下培养,以表达所述第一个和所述第二个基因;以及,

[0016] c) 任选,收集所述细胞和/或培养基中包含的 Z, Z-FPP 或 SBS 酶的倍半萜产物或其衍生物。

[0017] 本发明涉及具有 Z, Z-FPS 活性的分离或重组蛋白。所述蛋白优选具有与 SEQ ID No. 2 有至少 80% 同一性的序列。本发明还涉及含有编码这样的蛋白的核苷酸序列、或在严格条件下能够与其杂交的序列的核酸。

[0018] 本发明还涉及具有 SB 合酶类型的活性、并具有与 SEQ ID No. 4 有至少 80% 同一性

的序列的分离或重组蛋白。本发明还涉及含有编码这样的蛋白的核苷酸序列、或在严紧条件下能够与其杂交的序列的核酸。

[0019] 本发明涉及含有本发明的核酸的表达盒,含有本发明的表达盒或核酸的载体,含有本发明的载体、表达盒或核酸的宿主细胞,以及含有本发明的细胞、载体、表达盒或核酸的非人类转基因生物体。

[0020] 本发明还涉及在具有 IPP 和 DMAP 源的细胞中,从 IPP 和 DMAP 生产 Z, Z- 法呢基二磷酸 (Z, Z-FPP) 或其衍生物的方法,该方法包括:

[0021] a) 在所述细胞中导入具有表达盒的构建物,该表达盒含有编码本发明的 Z, Z-FPS 的核酸序列;

[0022] b) 将转化的细胞在适当的条件下培养,以表达基因;以及,

[0023] c) 任选,收集所述细胞和 / 或培养基中包含的其衍生物。

[0024] 本发明涉及使用本发明的蛋白、核酸、宿主细胞或转基因生物体,制备 Z, Z-FPP 或 SB 类型的倍半萜,例如 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯,或其衍生物例如 α -檀香醇、表- β -檀香醇、顺式- α -香柠檬醇、反式- α -香柠檬醇和 β -香柠檬醇。

[0025] 本发明涉及细胞或非人类转基因生物体,其特征在于通过对编码本发明的 Z, Z-FPS 的基因、或编码本发明的 SB 合酶的基因、或者这两个基因进行失活,阻断 II 类倍半萜的合成途径。

[0026] 本发明还涉及鉴定分子标记的核酸的应用,该分子标记允许相应的基因组序列被导入到其它的物种或栽培番茄 (*Solanum lycopersicum*) 变种中。在本发明的情况下,核苷酸序列 SEQ ID No. 1 和 No. 2 或其变异体也可以用作将所述序列渐渗到其它物种或栽培番茄的变种中的分子标记。

[0027] 因此,本发明涉及含有与 SEQ ID No. 1 或 SEQ ID No. 3 具有至少 80% 同一性的核酸的全部或部分的分子标记,用于鉴定多毛番茄和与多毛番茄性亲和的茄属 (*Solanum*) 类型物种之间的基因组多态性,从而将相应的基因导入到所述物种中。本发明还涉及方法,在于通过原生质体融合将 zFPS 和 SBS 基因导入到与多毛番茄不是性亲和的物种中。

[0028] 本发明还涉及通过磷酸酶将 Z, Z-FPP 脱磷酸化来生产 Z, Z- 法呢醇的方法。

附图说明

[0029] 图 1 :FPP 可能的立体异构体。FPP 的所有立体异构体都从 IPP 和 DMAPP 生物合成。到目前为止,已知的只有在许多生物体中被表征的 E, E- 法呢基二磷酸合酶 (1) 和结核分枝杆菌 (*Mycobacterium tuberculosis*) 的 Z, E- 法呢基二磷酸合酶 (2) (Schulbach 等, 2000)。(3) :本发明的番茄 Z, Z- 法呢基二磷酸合酶 ;(4) :E, Z- 法呢基二磷酸合酶 :这种活性的酶尚未被描述。

[0030] 图 2 :携带 Sh-zFPS 和 Sh-SBS 转基因的 T-DNA 的图。Km :卡那霉素抗性基因 ;35S :CaMV 35S 基因启动子。CBTS-ter :CBTS 终止子基因 ;eCBTS1.0 :与 CaMV 35S 启动子的增强子融合的 CBTS 基因 1kb 启动子。Sh-zFPS :编码番茄 zFPS 的基因的编码框 (多毛番茄 LA1777) ;Sh-SBS :编码番茄 SB 合酶的基因的编码框 (多毛番茄 LA1777) ;LB :T-DNA 的左缘 ;RB :T-DNA 的右缘 ;35S-ter :CaMV 35S 基因终止子 ;nos-ter :根癌土壤杆菌

(*Agrobacterium tumefaciens*) 的胭脂碱合酶基因的终止子。图 2 :pLIBR064 二元载体的 T-DNA 片段 ;图 2B :pLIBR065 二元载体的 T-DNA 片段。

[0031] 图 3 :转基因烟草叶片中的 Sh-SBS 转基因表达分析。表达使用荧光探针 (®TAQ-MAN, ABI) 通过实时定量 PCR 来测定。一种探针特异性针对 Sh-SBS 转基因,另一种探针特异性针对烟草肌动蛋白基因 (对照基因)。纵坐标轴值表示使用 Sh-SBS 探针获得的值与使用肌动蛋白探针获得的值的 2 的指数效价比 (potency 2 ratios) ($2^{\text{SBS}}/2^{\text{肌动蛋白}}$)。该比率表示 Sh-SBS 转基因的表达与肌动蛋白的表达的比率。横坐标轴值表示转基因种系的号码。图 3A :整合有 Sh-zFPS 和 Sh-SBS 转基因 (pLIBR0-064) 的转基因烟草种系。图 3B :仅整合有 Sh-SBS 转基因 (pLIBR0-065) 的转基因烟草种系。

[0032] 图 4 :合并在法尼醇标准品 (异构体的混合物) 上的 (Z,Z)-法尼醇的 GC/MS 分布图。图 4A :灰色色谱图 (m/z :69),对应于在体外使用 Sh-zFPS-6His 重组蛋白获得的产物。将 Sh-FPS-6His 与 DMAPP 和 IPP 温育。将异戊二烯二磷酸产物在醇中去磷酸化,并通过液相色谱进行纯化。1 号峰对应于 Z, Z-法尼醇。黑色色谱图 (m/z :69) 对应于由 (Z, E)-、(E, Z)-和 (E, E)-法尼醇异构体的混合物构成的法尼醇标准品 (Fluka)。2 号峰对应于 (Z, E)-和 (E, Z)-法尼醇的混合物,3 号峰对应于 (E, E)-法尼醇。图 4B :对应于 1 号峰的 (Z, Z)-法尼醇的质谱图。图 4C :对应于 2 号峰的 (Z, E)-和 (E, Z)-法尼醇 (混合物) 的质谱图。图 4D :对应于 3 号峰的 (E, E)-法尼醇的质谱图。

[0033] 图 5 :使用 Sh-zFPS-6His 和 Sh-SBS-6His 重组蛋白进行的体外活性测试的 GC/MS 分布图。将两种重组酶 Sh-zFPS-6His 和 Sh-SBS-6His 与 IPP 和 DMAPP 一起进行温育。用戊烷萃取反应混合物,并通过 GC/MS 进行分析。图 5A :使用 Sh-zFPS 和 Sh-SBS 在体外获得的产物的色谱图。2 号峰是主要产物,对应于 α -檀香萜。1,顺式- α -香柠檬烯;2, α -檀香萜;3,反式- α -香柠檬烯;4,表- β -檀香萜;5,内- β -香柠檬烯。图 5B :对应于 2 号峰 (左侧) 和 α -檀香萜结构 (右侧) 的质谱图。

[0034] 图 6 :使用重组酶 Sh-zFPS-6His 和 Sh-SBS-6His、以及 TA517 同基因重组番茄种系的渗出物在体外获得的产物的合并的 GC 分布图。描迹线对应于提取的 m/z 94 离子。灰色 :通过 Sh-zFPS 和 Sh-KS 在体外生产的烯烃。黑色 :TA517 同基因种系的渗出物。1,顺式- α -香柠檬烯;2, α -檀香萜;3,反式- α -香柠檬烯;4,表- β -檀香萜;5,内- β -香柠檬烯。

[0035] 图 7 :由包涵 Sh-zFPS 和 Sh-SBS 转基因的转基因植物 (#3877) 散发出的挥发性分子的 GC/MS 分布图。将转基因植物 #3877 在培养箱中在受控的气氛下培养 24 小时。将由植物散发的挥发性分子截留在 Super® Q 基质 (Alltech) 上,并通过 GC/MS 进行分析。色谱图 (7A) 显示出几个具有 SB 类型倍半萜的特征性标志的峰。通过将它们的保留时间和质谱与 TA517 种系 (7B) 的进行比较,对峰进行鉴定。1,顺式- α -香柠檬烯;2, α -檀香萜;3,反式- α -香柠檬烯;4,表- β -檀香萜;5,内- β -香柠檬烯。

[0036] 图 8 :番茄 (*S. lycopersicum*(Sl)) 和多毛番茄 (*S. habrochaites*(Sh)) 之间的 zFPS 基因多态性 :通过 PCR,从番茄、多毛番茄和 TA517 的基因组 DNA 扩增完整的 zFPS 基因。将 PCR 产物在 0.8% 琼脂糖凝胶上分离,鉴定到了 3 条带 A、B、C,分别具有大约 3,000、2,500 和 2,000 个核苷酸的大小。对应于条带 B 的 PCR 产物是 S. h 基因组特异性的,并对应于本文中描述的 zFPS 基因。Kb :千碱基。

[0037] 缩写与定义

[0038] mRNA :信使核糖核酸

[0039] DNA :脱氧核糖核酸

[0040] cDNA :通过 mRNA 的反转录产生的互补 DNA。

[0041] CaMV :花椰菜花叶病毒

[0042] DMAPP :二甲基烯丙基二磷酸

[0043] CBT-ol :西柏三烯醇 (cembratrien-ol)

[0044] CBTS :西柏三烯醇合酶

[0045] GPP :香叶基二磷酸

[0046] FPP :法呢基二磷酸

[0047] GC :气相色谱

[0048] GGPP :香叶基香叶基二磷酸

[0049] ihpRNAi :干扰性内含子发夹 RNA

[0050] IPP :异戊二烯基二磷酸

[0051] MS :质谱

[0052] PCR :聚合酶链反应

[0053] RACE :cDNA 末端的快速扩增

[0054] A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V :按照通用命名法,氨基酸的标准代码 www3.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=t

[0055] A, C, G, T, B, D, H, K, M, N, R, S, V, W, Y :按照通用命名法,碱基的标准代码 www3.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=t

[0056] RFLP :限制性片段长度多态性

[0057] 本文使用的“Z,Z-法呢基二磷酸合酶”或 zFPS,是指能够从异戊二烯二磷酸 (IPP) 和二甲基烯丙基二磷酸 (DMAPP) 产生 Z,Z-法呢基二磷酸 [(2Z,6Z)-法呢基二磷酸或 cis, cis-FPP] 的酶。本文中使用的“Z,Z-FPS 活性”是指从 IPP 和 DMAPP 产生 Z,Z-法呢基二磷酸。Z,Z-FPS 活性可以通过测量 Z,Z-法呢基二磷酸的形成来评估。

[0058] 本文中使用的“SB 合酶”或 SBS,是指能够从 Z,Z-FPP 产生主要由 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯构成的倍半萜混合物的酶。“SB 类型的活性”是指从 Z,Z-法呢基二磷酸产生 SB 类型的倍半萜。SB 类型的活性可以通过测量一种或多种 SB 类型的倍半萜的形成来评估。测定 SB 类型活性的例子描述在实施例中。

[0059] 本文中使用的“SB 类型的倍半萜”,是指主要由 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯构成的倍半萜混合物。

[0060] “严紧杂交条件”:一般来说,对于给定大小和序列的多核苷酸,严紧条件是通过在比同样的反应混合物中由所述多核苷酸及其互补序列形成的杂交体的解链温度 (T_m) 低大约 5°C 到 10°C 的温度下操作获得的。给定多核苷酸的严紧杂交条件,可以由本技术领域的专业人员,根据所讨论的多核苷酸的大小和碱基组成、以及杂交混合物的组成 (特别是 pH 和离子强度) 来确定。高严紧条件包括在 65°C 使用 0.2xSSC 缓冲液的清洗步骤。

[0061] 对于大小更大或更小的多核苷酸,本技术领域的专业人员可以根据本技术领域的

专业人员已知的相关教导,,特别是 Sambrook 等,1989,《分子克隆 :实验室手册》(Molecular Cloning, A Laboratory Manual, Cold Spring Harbor);Maniatis 等,1982,《分子克隆 :实验室手册》(Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Lab. CSH, N. Y. USA) 或它们的再编版之一,以及 Ausubel 等主编,1995,《分子生物学现代方法》的第二章 (Current Protocols in Molecular Biology, Chapter 2 (Greene Publishing and Wiley-Interscience, N. Y.)) 中的描述,改变上面描述的严紧杂交条件。

[0062] 本文中使用的“异源”被理解为是指基因已经通过遗传工程被导入到细胞中。它可以以游离基因或染色体的形式存在。基因可以源自与它所导入的细胞不同的来源。然而,基因也可以来自与它所导入的细胞相同的物种,但是由于其环境不是天然的,它被当作是异源的。例如,由于基因在不是其天然启动子的启动子控制之下,它被导入到与其天然位置不同的位置,因此基因被称为是异源的。在导入异源基因之前,宿主细胞可以含有基因的内源拷贝,或者它可以不含内源拷贝。

[0063] 本文中使用的两个核酸或氨基酸序列之间的“百分同一性”,是指两个待比较的序列之间在最佳比对后获得一致的核苷酸或氨基酸残基的百分率,所述百分率是纯粹统计学的,两个序列之间的差异随机分布并遍及它们的整个长度。最佳比对或最优比对是按照后文进行计算时,被比较的两个序列之间的百分同一性最高的比对。两个核酸或氨基酸序列之间的序列比较,通常是通过将所述序列最优比对后将它们进行比较来进行的,所述比较通过区段或比较窗口来进行,以便鉴定和比较序列局部区域的相似性。除了手工进行之外,序列也可以利用 Smith 和 Waterman 的局部同源性算法 (1981) (Ad. App. Math. 2 :482)、利用 Needleman 和 Wunsch 的局部同源性算法 (1970) (J. Mol. Biol. 48 :443)、利用 Pearson 和 Lipman 的同源性搜索方法 (1988) (Proc. Natl. Acad. Sci. USA 85 :2444)、利用运行所述算法的计算机软件 (威斯康辛遗传学软件包 (Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI) 中的 GAP、BESTFIT、FASTA 和 TFASTA) 进行最优比对,以用于比较。两个核酸或氨基酸序列之间的百分同一性通过使用比较窗口比较这两个最优比对的序列来确定,其中被比较的核酸或氨基酸序列区域,相对于用于这两个序列之间的最优比对的参比序列来说,可以包含添加或缺失。百分同一性的计算是通过确定两个序列之间同一的核苷酸或氨基酸残基的同一位置数量,将所述同一位置数量除以比较窗口中位置的总数,并将结果乘以 100 以获得这两个序列之间的百分同一性。

[0064] 发明详述

[0065] 因此,本发明第一次描述了参与许多倍半萜的合成途径的酶的编码基因。具体来说,本发明涉及具有多种产物的番茄倍半萜合酶 (SB 合酶) 的表征,该酶能够从 Z,Z-FPP 产生被称为 SB 类型化合物的倍半萜的混合物,该混合物主要由 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯构成。此外,本发明还涉及番茄的 Z,Z-法呢基二磷酸合酶 (zFPS) 的表征,该酶能够从 IPP 和 DMAP 产生 Z,Z-FPP。这些酶可用于在体外或在转基因生物体或重组微生物或细胞中体内生产 SB 类型的倍半萜或 Z,Z-FPP。考虑的有转基因生物体、细胞或微生物例如细菌、真菌、动物、昆虫或植物细胞和转基因动物或植物。它们也能够生产从 Z,Z-FPP 衍生的化合物,例如 Z,Z-法呢醇。更具体来说,本发明的方法可用于在微生物 (细菌,酵母) 中生产 Z,Z-FPP,以及在拥有分泌类型的

腺毛的植物中生产 SB 类型的化合物。也可以通过基因消除技术或通过诱变来减少或抑制植物例如番茄中 SB 类型的倍半萜混合物或其衍生物的产生。此外,在本发明中鉴定的序列可用于在其它生物物种、特别是植物中鉴定和 / 或克隆编码具有同样活性的酶的基因。最后,可以从编码 zFPS 和 SB 合酶的核酸中鉴定多态性分子标记,使得监测相应的功能性基因组序列在栽培番茄 (*Solanum lycopersicum*) 的其它物种或变种中的导入成为可能。

[0066] Z, Z-FPP 是倍半萜合酶的潜在底物,但是目前还不可商购。尽管到目前为止,绝大多数被表征的倍半萜合酶利用 E, E-FPP,但是显然缺乏可商购的 Z, Z-FPP 极大地限制了它用于实验目的。棉花的杜松烯合酶是描述了倍半萜合酶优选利用 Z, Z-FPP 的唯一例子 (Heinstein 等,1970)。在这种情况下, Z, Z-FPP 通过化学合成生产,在该专利中描述的 Z, Z-FPP 合酶提供了以容易和不昂贵的方式生产所述分子的方法。

[0067] 此外,SB 类型的倍半萜包含 α -檀香萜,它本身是 α -檀香醇的直接前体。 α -檀香醇是檀香木油的特征性成分之一。檀香木油在香水业有高度价值,并且由于檀香木主要在印度的过度开发,过去 10 年中其价格急剧升高。这种过度开发对天然檀香木资源的生存造成了威胁。 α -檀香醇可以通过简单的羟化从 α -檀香萜获得。

[0068] 因此,本发明涉及在具有 DMAP 和 IPPP 源的细胞中,从 Z, Z-FPP 生产 SB 类型的倍半萜或其衍生物的方法,该方法包括:

[0069] a) 在所述细胞中导入带有含有编码本发明的 zFPS 的第一个基因的表达盒的构建物,以及带有含有编码本发明的 SB 合酶的第二个基因的表达盒的构建物;

[0070] b) 将转化的细胞在适当的条件下培养,以表达所述第一个和所述第二个基因;以及,

[0071] c) 任选地,收集所述细胞和 / 或培养基中包含的 SB 类型的倍半萜或其衍生物。

[0072] 在具体实施方案中,在步骤 c) 中收集所产生的 SB 类型的倍半萜。在优选实施方案中,产生的 SB 类型的倍半萜选自 α -檀香萜、表- β -檀香萜、 α -香柠檬烯、 β -香柠檬烯和内- β -香柠檬烯。在另一个具体实施方案中,产生的 SB 类型的倍半萜是获得其它目标化合物例如 α -檀香醇的起始产物。在优选实施方案中,SB 类型的倍半萜衍生物选自 α -檀香醇、表- β -檀香醇、顺式- α -香柠檬醇、反式- α -香柠檬醇和内- β -香柠檬醇。

[0073] 在具体的实施方案中,本发明涉及在具有 DMAP 和 IPPP 源的细胞中,从 Z, Z-FPP 生产 SB 类型的倍半萜或其衍生物的方法,该方法包括:

[0074] a) 提供重组细胞,所述细胞含有编码本发明的 zFPS 的第一个异源基因以及编码本发明的 SB 合酶的第二个异源基因;

[0075] b) 将所述细胞在适当的条件下培养,以表达所述第一个和所述第二个基因;以及,

[0076] c) 任选,收集所述细胞和 / 或培养基中包含的 SB 类型的倍半萜或其衍生物。

[0077] 在优选实施方案中,所述细胞产生 Z, Z-FPP。在另一个实施方案中,向细胞供应 Z, Z-FPP。

[0078] 在具体的实施方案中,两个表达盒由同一个构建物携带。在另一个实施方案中,两个表达盒由两个独立的构建物携带。

[0079] 本发明涉及具有 Z, Z-FPP 合酶活性的分离或重组的蛋白。具体来说,本发明涉及具有 Z, Z-FPP 合酶活性的分离或重组多肽,并且其所具有的序列与 SEQ ID No. 2 有至少

80%、85%、90%、95%、97%、98%或99%的同一性。优选,该多肽具有与 SEQ ID No. 2 至少 95%同一性的序列。在具体的实施方案中,多肽包含序列 SEQ ID No. 2 或由其构成。所述多肽也可以包含其它便于酶的纯化的序列,例如含有几个连续的组氨酸氨基酸的标签序列。

[0080] 本发明还涉及含有核苷酸序列的分离核酸,该核苷酸序列编码的多肽具有 Z, Z-FPP 合酶活性,并且具有与 SEQ ID No. 2 有至少 80%、85%、90%、95%、97%、98%或 99% 同一性的序列,或在严紧条件下能够与其杂交的序列。核酸的例子是含有与 SEQ ID No. 1 具有至少 80%、85%、90%、95%、97%、98%或 99%同一性的序列或其互补序列的核酸,或由其构成的核酸。本发明涉及能够在高度严紧条件下与具有 Z, Z-FPP 合酶活性的多肽的编码核酸杂交的分离核酸,其中所述多肽具有与 SEQ ID No. 2 至少 95%或 98%同一性的序列。在具体的实施方案中,核酸包含序列 SEQ ID No. 1 或由其构成。本发明还涉及含有本发明的核酸的表达盒、载体、宿主细胞或转基因生物体。

[0081] 核酸可以是基因组 DNA、互补 DNA (cDNA) 或合成的 DNA。核酸可以是单链或双链体的形式,或二者的混合物。在本发明中,转录的核酸优选是无内含子的 cDNA。转录的核酸可以是合成的或半合成的、重组的分子,可以任选扩增或克隆到载体中,被化学修饰或包含非天然碱基。它们一般是通过本技术领域专业人员熟知的重组技术合成的分离的 DNA 分子。

[0082] 本发明还涉及具有 SB 合酶活性的分离或重组的多肽,并且其具有与 SEQ ID No. 4 至少 80%、85%、90%、95%、97%、98%或 99%同一性的序列。在具体实施方案中,多肽包含序列 SEQ ID No. 4 或由其构成。在这种情况下,本发明涉及具有 SB 合酶活性、并具有与 SEQ ID No. 4 至少 95%同一性的序列的分离或重组多肽。所述多肽也可以包含其它便于酶纯化的序列,例如含有几个连续的组氨酸氨基酸的标签序列。

[0083] 本发明还涉及分离的核酸,该核酸含有的核苷酸序列所编码的多肽具有 SB 合酶活性、并且序列与 SEQ ID No. 4 具有至少 80%、85%、90%、95%、97%、98%或 99%同一性,或能够在严紧条件下与其杂交的序列。这样的核酸的例子是包含与 SEQ ID No. 3 有至少 80%、85%、90%、95%、97%、98%或 99%同一性的序列或其互补序列的核酸,或由其构成的核酸。本发明涉及能够在高度严紧条件下与具有 SB 合酶活性的多肽的编码核酸杂交的分离核酸,其中所述多肽具有与 SEQ ID No. 4 有至少 95%或 98%同一性的序列。在具体的实施方案中,该核酸包含序列 SEQ ID No. 3 或由其构成。本发明还涉及含有本发明的核酸的表达盒、载体、宿主细胞或转基因生物体。

[0084] 更具体来说,本发明涉及含有两种核酸的载体、宿主细胞或转基因生物体,其中一种核酸含有异源序列,所述异源序列编码具有 SB 合酶活性、并且序列与 SEQ ID No. 4 具有至少 80%、85%、90%、95%、97%、98%或 99%同一性的多肽,和另一种核酸含有异源序列,所述异源序列编码具有 Z, Z-FPP 合酶活性、并且序列与 SEQ ID No. 2 具有至少 80%、85%、90%、95%、97%、98%或 99%同一性的多肽。在本发明的具体实施方案中,转基因生物体是植物,特别是植物毛状体。

[0085] 本发明还涉及含有核酸的载体、宿主细胞或转基因生物体,该核酸含有异源序列,所述异源序列编码具有 SB 合酶活性、并且序列与 SEQ ID No. 4 有至少 80%、85%、90%、95%、97%、98%或 99%同一性的多肽。

[0086] 本发明还涉及含有核酸的载体、宿主细胞或转基因生物体,该核酸含有异源序列,

所述异源序列编码具有 Z, Z-FPP 合酶活性、并优选序列与 SEQ ID No. 2 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性的多肽。

[0087] 本发明涉及从 IPP 和 DMAP 生产 SB 类型的倍半萜混合物或其衍生物的方法, 包括在适当条件下, 将 IPP 和 DMAP 与纯化或重组的 Z, Z-FPP 合酶和纯化或重组的 SB 合酶相接触, 所述 Z, Z-FPP 合酶与 SEQ ID No. 3 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性, 所述 SB 合酶与 SEQ ID No. 4 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性, 并收集获得的 SB 类型的倍半萜混合物或其衍生物。本发明涉及使用与 SEQ ID No. 2 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性的纯化或重组的 Z, Z-FPP 合酶, 以及与 SEQ ID No. 4 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性的纯化或重组的 SB 合酶, 从 IPP 和 DMAP 制备 SB 类型的倍半萜混合物或其衍生物。

[0088] 本发明涉及从 IPP 和 DMAP 生产 Z, Z-FPP 或其衍生物的方法, 包括在适当条件下将 IPP 和 DMAP, 与纯化或重组的 Z, Z-FPP 合酶相接触, 所述 Z, Z-FPP 合酶与 SEQ ID No. 2 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性, 并收集获得的 Z, Z-FPP 或其衍生物。任选, 方法还包括将获得的 Z, Z-FPP 与牛肠碱性磷酸酶温育, 并收集 Z, Z-法尼醇。本发明涉及使用与 SEQ ID No. 2 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性的纯化或重组的 Z, Z-FPP 合酶, 从 IPP 和 DMAP 制备 Z, Z-FPP 或其衍生物。

[0089] 本发明涉及从 Z, Z-FPP 生产 II 类倍半萜混合物或其衍生物的方法, 包括在适当条件下将 Z, Z-FPP 与纯化或重组的 SB 合酶相接触, 所述 SB 合酶与 SEQ ID No. 4 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性, 并收集获得的 SB 类型的倍半萜混合物或其衍生物。本发明涉及使用与 SEQ ID No. 4 具有至少 80%、85%、90%、95%、97%、98% 或 99% 同一性的纯化或重组的 SB 合酶, 从 Z, Z-FPP 制备 II 类倍半萜混合物或其衍生物。

[0090] 一般来说, 表达盒含有基因转录和翻译成蛋白所需的所有元件。具体来说, 它包含启动子、任选的增强子、转录终止子和翻译元件。

[0091] 启动子与宿主细胞相适应。例如, 如果细胞是原核的, 启动子可以选自下列启动子: LacI, LacZ, pLacT, ptac, pARA, pBAD, 噬菌体 T3 或 T7 的 RNA 聚合酶启动子, 多角体蛋白启动子, λ 噬菌体的 PR 或 PL 启动子。如果细胞是真核和动物的, 启动子可以选自下列启动子: 巨细胞病毒 (CMV) 早期启动子, 单纯性疱疹病毒 (HSV) 胸苷激酶启动子, 猿猴 40 病毒 (SV40) 早期或晚期启动子, 小鼠金属硫蛋白 -L 启动子, 以及某些反转录病毒的 LTR (长末端重复序列) 区。一般来说, 为了选择合适的启动子, 本技术领域专业人员可以优先参考 Sambrook 和 Russell (2000) 的工作, 或者参考在 Ausubel 等 (2006) 的工作中描述的方法。

[0092] 载体可以是质粒、噬菌体、噬粒、粘粒、病毒、YAC、BAC、土壤杆菌 (Agrobacterium) pTi 质粒, 等等。载体优选包含选自复制原点、多克隆位点和标记的一个或多个元件。标记可以是产生可检测信号的报告基因。可检测信号可以是荧光信号、染色、发光。因此, 标记可以是 GFP、EGFP、DsRed、 β -半乳糖苷酶、 β -葡萄糖苷酶、荧光素酶等。标记优选是赋予细胞抗生素或除草剂抗性的选择标记。例如, 基因可以是卡那霉素、新霉素等的抗性基因。在优选实施方案中, 载体是质粒。原核载体的例子包括但不限于下列: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540,

pBR322 和 pRIT5 (Pharmacia), pET (Novagen) 和 pQE-30 (QIAGEN)。真核载体的例子包括但不限于下列: pWLNE0, pSV2CAT, pPICZ, pcDNA3.1(+)-Hyg (Invitrogen), pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pCI-neo (Stratagene), pMSG, pSVL (Pharmacia)。病毒载体包括但不限于腺病毒、AAV、HSV、慢病毒等。优选, 表达载体是质粒或病毒载体。

[0093] 宿主细胞可以是原核生物, 例如大肠杆菌 (*Escherichia coli*), 枯草芽胞杆菌 (*Bacillus subtilis*), 链霉菌 (*Streptomyces* sp) 和假单胞菌 (*Pseudomonas* sp), 或是真核生物。真核生物可以是低等真核生物例如酵母 (例如酿酒酵母 (*Saccharomyces cerevisiae*)) 或丝状真菌 (例如来自曲霉属 (*Aspergillus*)), 或高等真核生物例如昆虫、哺乳动物或植物细胞。细胞可以是哺乳动物细胞, 例如 COS、CHO 细胞 (US4, 889, 803; US 5, 047, 335)。在具体实施方案中, 细胞是非人类的和非胚胎的。细胞可以是分离的, 例如在培养基中。细胞也可以包含在生物体中, 例如非人类转基因动物或转基因植物。

[0094] 因此, 本发明涉及在具有 IPP 和 DMAP 源的细胞中, 从 IPP 和 DMAP 生产 Z, Z-FPP 或其衍生物的方法, 该方法包括:

[0095] a) 将具有含有本发明的 Z, Z-FPP 合酶的编码基因的表达盒的构建物导入到所述细胞中;

[0096] b) 将转化的细胞在适当条件下进行培养, 以表达基因; 以及

[0097] c) 任选, 收集在所述细胞和 / 或培养基中包含的 Z, Z-FPP 或其衍生物。

[0098] 在具体的实施方案中, 在步骤 c) 中收集产生的 Z, Z-FPP。在另一个具体实施方案中, 产生的 Z, Z-FPP 是获得另一种目标化合物例如 Z, Z-法尼醇的起始产物。例如, Z, Z-FPP 通过磷酸酶转化成 Z, Z-法尼醇。因此, 方法还可以在步骤 a) 中包括导入具有含有磷酸酶的编码基因的表达盒的构建物, 从而允许在步骤 c) 中收集 Z, Z-法尼醇。因此, 本发明涉及在具有 IPP 和 DMAP 源、并含有编码本发明的 Z, Z-FPP 合酶的异源基因的重组细胞中, 从 IPP 和 DMAP 生产 Z, Z-FPP 或其衍生物的方法。因此, 本发明涉及在具有 IPP 和 DMAP 源的重组细胞中生产 Z, Z-FPP 的方法, 该方法包括:

[0099] a) 提供含有编码本发明 zFPS 的异源基因的重组细胞;

[0100] b) 将细胞在适当条件下进行培养, 以表达所述基因; 以及

[0101] c) 任选, 收集在所述细胞和 / 或培养基中包含的 Z, Z-FPP 或其衍生物。

[0102] 因此, 本发明涉及在具有 Z, Z-FPP 源的细胞中, 从 Z, Z-FPP 生产 SB 类型的倍半萜混合物或其衍生物的方法, 该方法包括:

[0103] a) 将具有含有本发明 SB 合酶的编码基因的表达盒的构建物导入到所述细胞中;

[0104] b) 将转化的细胞在适当条件下进行培养, 以表达基因; 以及

[0105] c) 任选, 收集在所述细胞和 / 或培养基中包含的 SB 类型的倍半萜混合物或其衍生物。

[0106] 在具体的实施方案中, 在步骤 c) 中收集产生的 SB 类型的倍半萜混合物。在另一个具体实施方案中, 产生的 SB 类型的倍半萜混合物是获得另一种目标化合物例如 α -檀香醇的起始产物。因此, 本发明涉及在具有 Z, Z-FPP 源、并含有编码本发明的 SB 合酶的异源基因的重组细胞中, 从 Z, Z-FPP 生产 SB 类型的倍半萜混合物或其衍生物的方法。因此, 本发明涉及在具有 Z, Z-FPP 源的重组细胞中从 Z, Z-FPP 生产 SB 类型的倍半萜或其衍生物的方法, 该方法包括:

- [0107] a) 提供含有编码本发明 SB 合酶的异源基因的重组细胞；
[0108] b) 将细胞在适当条件下进行培养，以表达所述基因；以及
[0109] c) 任选，收集在所述细胞和 / 或培养基中包含的 SB 类型的倍半萜混合物或其衍生物。

[0110] 细胞也可以是多细胞生物体例如植物或非人类动物的部分。在这种情况下，本发明涉及制备多细胞生物体的方法，包括：

- [0111] a) 将具有含有编码本发明 Z,Z-FPP 合酶的第一个基因的表达盒的构建物，以及具有含有编码本发明 SB 合酶的第二个基因的表达盒的构建物，导入到生物体的细胞中；以及
[0112] b) 从所述细胞重构生物体，并筛选表达所导入的两个基因的转基因生物体。

[0113] 在一个实施方案中，生物体是非人类动物。例如，生物体可以是小鼠、大鼠、豚鼠、兔等。在另一个优选实施方案中，生物体是植物，优选植物的腺毛。

[0114] 因此，本发明涉及生产 SB 类型的倍半萜混合物的方法，包括提供表达本发明 Z,Z-FPP 合酶和本发明 SB 合酶的转基因多细胞生物体，以及在所述转基因生物体中收集 SB 类型的倍半萜混合物或其衍生物体。

[0115] 该方法使得在不产生 SB 类型倍半萜混合物的生物体中生产这样的化合物、或增加已经产生它的生物体生产的 SB 类型倍半萜混合物的量，成为可能。

[0116] 因此，本发明涉及制备多细胞生物体的方法，包括：

- [0117] a) 将具有含有本发明 Z,Z-FPP 合酶的编码基因的表达盒的构建物导入到生物体的细胞中；以及
[0118] b) 从所述细胞重构生物体，并筛选表达导入基因的转基因生物体。

[0119] 本发明还涉及生产 Z,Z-FPP 或其衍生物的方法，包括提供表达本发明 Z,Z-FPP 合酶的转基因多细胞生物体，以及在所述转基因生物体中收集产生的 Z,Z-FPP 或其衍生物。

[0120] 在一个实施方案中，生物体是非人类动物。例如，生物体可以是小鼠、大鼠、豚鼠、兔等。在另一个优选实施方案中，生物体是植物，优选是具有腺毛的植物。该方法使得在不产生 Z,Z-FPP 的生物体中生产所述化合物、或增加生物体产生的 Z,Z-FPP 的量，成为可能。

[0121] 因此，本发明涉及制备多细胞生物体的方法，包括：

- [0122] a) 将具有含有本发明 SB 合酶的编码基因的表达盒的构建物导入到生物体的细胞中；以及
[0123] b) 从所述细胞重构生物体，并筛选表达导入基因的转基因生物体。

[0124] 本发明还涉及生产 SB 类型的倍半萜混合物或其衍生物的方法，包括提供表达本发明的 SB 合酶的转基因多细胞生物体，以及在所述转基因生物体中收集产生的 SB 类型倍半萜混合物或其衍生物。

[0125] 在一个实施方案中，生物体是非人类动物。例如，生物体可以是小鼠、大鼠、豚鼠、兔等。在另一个优选实施方案中，生物体是植物，优选植物的分泌性毛状体。该方法使得由不产生 SB 类型倍半萜混合物的生物体生产所述化合物、或增加生物体产生的 SB 类型倍半萜混合物的量，成为可能。

[0126] 具体来说，本发明适用于所有来自具有腺毛的科的植物，例如菊科 (Asteraceae) (向日葵等)、茄科 (Solanaceae) (番茄、烟草、土豆、胡椒、茄子等)、大麻科 (Cannabaceae) (例如大麻 (Cannabissativa)) 和唇形科 (Lamiaceae) (薄荷、罗勒、薰

衣草、百里香等)。具体来说,本发明适合茄科的植物,例如来自烟草属 (*Nicotiana*)、茄属 (*Solanum*)、辣椒属 (*Capsicum*)、矮牵牛属 (*Petunia*)、曼陀罗属 (*Datura*)、颠茄属 (*Atropa*) 等,特别是茄属和烟草属,例如栽培的番茄 (*Solanum lycopersicum*)、野生番茄 (多毛番茄 (*Solanum habrochaite*))、栽培的烟草 (*Nicotiana tabacum*)、林地烟草 (*Nicotiana sylvestris*)。非限制性地,本发明可应用于来自下列属的植物:杨属 (*Populus*)、烟草属 (*Nicotiana*)、大麻属 (*Cannabis*)、牵牛属 (*Pharbitis*)、水玉簪属 (*Apteria*)、九节木属 (*Psychotria*)、山黧属 (*Mercurialis*)、菊属 (*Chrysanthemum*)、水龙骨科属 (*Polypodium*)、天竺葵属 (*Pelargonium*)、沟酸浆属 (*Mimulus*)、母菊属 (*Matricaria*)、香蜂草属 (*Monarda*)、茄属 (*Solanum*)、蓍草属 (*Achillea*)、缬草属 (*Valeriana*)、罗勒属 (*Ocimum*)、苜蓿属 (*Medicago*)、七叶树属 (*Aesculus*)、白花丹属 (*Plumbago*)、粉叶蕨属 (*Pityrogramma*)、钟穗花属 (*Phacelia*)、海欖雌属 (*Avicennia*)、怪柳属 (*Tamarix*)、瓣鳞花属 (*Frankenia*)、补血草属 (*Limonium*)、茴香属 (*Foeniculum*)、百里香属 (*Thymus*)、鼠尾草属 (*Salvia*)、五味子属 (*Kadsura*)、大戟科 (*Beyeria*)、律草属 (*Humulus*)、薄荷属 (*Mentha*)、苦艾属 (*Artemisia*)、荆芥属 (*Nepta*)、大丁草属 (*Geraea*)、刺蕊草属 (*Pogostemon*)、牛至属 (*Majorana*)、醉蝶花属 (*Cleome*)、薊属 (*Cnicus*)、银胶菊属 (*Parthenium*)、*Ricinocarpos*、李叶豆属 (*Hymenaea*)、*Larrea*、报春花属 (*Primula*)、钟穗花属 (*Phacelia*)、鳞毛蕨属 (*Dryopteris*)、香茶菜属 (*Plectranthus*)、杓兰属 (*Cypripedium*)、矮牵牛属 (*Petunia*)、曼陀罗属 (*Datura*)、黧豆属 (*Mucuna*)、蓖麻属 (*Ricinus*)、金丝桃属 (*Hypericum*)、苦檻蓝属 (*Myoporum*)、金合欢属 (*Acacia*)、金钱槭属 (*Diplopeltis*)、车桑子属 (*Dodonaea*)、*Halgania*、马鞭草科 (*Cyanostegia*)、夏枯草属 (*Prostanthera*)、*Anthocercis*、树紫苑属 (*Olearia*)、*Viscaria*。优选,植物是来自菊科、大麻科、茄科或唇形科的植物。在更优选的实施方案中,植物属于茄属或烟草属,优选为番茄 (*Solanum esculentum*)、多毛番茄 (*Solanum habrochaite*)、普通烟草 (*Nicotiana tabacum*) 或林地烟草 (*Nicotiana sylvestris*)。

[0127] 在该实施方案中,基因处于允许表达的启动子、优选在植物毛状体中特异性的启动子控制之下。存在这样的启动子并已成为本技术领域专业人员所知 (Tissier 等, 2004)。

[0128] 在本发明中,“特异性”启动子是指主要在给定组织或细胞类群中有活性的启动子。应该了解,不能完全排除在其它组织或细胞中的残余的、通常较低的表达。本发明的具体特点是基于构建对腺毛分泌细胞具有特异性的启动子的能力,允许对植物叶分泌物的成分进行改变,特别是在其中表达本发明的允许制备 SB 类型倍半萜混合物的基因。因此,可以从所述植物的毛状体、特别是毛状体渗出物中,通过用溶剂萃取或通过蒸馏,来收集 Z、Z-FPP 和 SB 类型倍半萜混合物或其衍生物。

[0129] 例如,已经显示,位于 CYP71D16 基因的 ATG 上游的 1852bp 调控序列,指导了 *uidA* 报告基因在烟草毛状体的分泌细胞中特异性表达 (申请 US2003/0100050 A1, Wagner 等, 2003)。此外,已经鉴定到从不同物种提取的几个启动子序列,能够指导异源基因在烟草毛状体中的表达。

[0130]	基因缩写	基因名称	植物	参考文献
	LTP3	脂质转移蛋白	棉花	Liu 等, 2000, BBA, 1487:106111
	LTP6		棉花	Hsu 等, 1999, Plant Science, 143:6370
	wax9D		甘蓝 (<i>B. oleracea</i>)	Pyee 和 Kolattukudy, 1995, Plant J. 7:4559
	LTP1		拟南芥 (<i>Arabidopsis</i>)	Thoma 等, 1994, Plant Physiol. 105 3545
	CYC71D16	CBT 醇羟化酶	烟草	Wang 等, 2002, J. Exp. Bot. 18911897

[0131] 在本发明的优选实施方案中,在表达盒中使用的启动子源自林地烟草 (*Nicotiana sylvestris*) 种中与 CYC-2(CBT-ol 环化酶 ;NID :AF401234) 显示出强烈序列相似性的 NsTPS-02a、02b、03 和 04 基因。在专利申请 W02006040479(Tissier 等,2004) 中,对所述启动子序列进行了更全面的描述。

[0132] 在优选的终止子序列中,可以提及 NOS 终止子 (Bevan 等,1983) 和组蛋白基因终止子 (EPO 633 317)。

[0133] 在具体实施方案中,表达盒可以包含允许增加表达的序列 (“增强子”),例如 CaMV35S 启动子和章鱼碱合酶基因 (US 5 290 924) 的某些元件。优选,使用 CaMV35S 启动子的增强子元件。

[0134] 将具有本发明的一个或这两个基因的构建物导入到细胞或植物组织中,可以通过本技术领域的专业人员已知的任何方法进行,包括游离体或染色体形式。植物转基因技术在本领域中是众所周知的,包括例如使用细菌根癌土壤杆菌 (*Agrobacterium tumefaciens*),电穿孔,基因枪技术,特别是通过病毒载体转染,以及本技术领域的专业人员已知的任何其它方法。

[0135] 一种基于使用细菌根癌土壤杆菌的常用方法,主要在于将目标构建物 (核酸、表达盒、载体等) 导入到细菌根癌土壤杆菌中,然后将该转化的细菌与所选植物的叶盘相接触。表达盒通常通过使用 Ti 质粒 (或 T-DNA) 作为载体导入到细菌中,该 Ti 质粒 (或 T-DNA) 可以通过例如热冲击或通过电穿孔转移到细菌中。将转化的细菌与叶盘温育,使得 Ti 质粒转移到盘细胞的基因组中。任选,可以将它们在合适的条件下进行栽培,以便重构其细胞包含本发明的构建物的转基因植物。关于根癌土壤杆菌转化技术的进一步详细情况或不同的实施方案,可以参考例如 Horsch 等 (1985) 或 Hooykaas 和 Schilperoort (1992)。

[0136] 因此,在具体实施方案中,这样构建的表达盒被插入到卸甲 (disarmed) Ti 质粒的转移 DNA (T-DNA) 的左缘和右缘之间,用于通过根癌土壤杆菌转移到植物细胞中。T-DNA 还含有其表达赋予对抗生素或除草剂的抗性并且其使得能够进行转化体筛选的基因。

[0137] 一旦再生后,可以测试转基因植物中本发明基因的异源表达或毛状体中 SB 类型

的倍半萜混合物、Z, Z-FPP 或 Z, Z-法尼醇的产生。这可以通过收集叶片渗出物,并在所述渗出物中测试目标产物的存在来完成。分析植物散发出的挥发性化合物,也可以允许对该化合物进行鉴定。这也可以通过分析叶片中、更具体来说毛状体细胞中本发明的一种或多种异源酶的存在来进行(例如通过使用特异性引物或探针分析 mRNA 或基因组 DNA)。任选,可以对植物进行筛选、杂交、处理等,以便获得表达水平改进的植物。

[0138] 本发明的另一个目标也是基于含有例如上面定义的核酸、表达盒或载体的植物或种子。

[0139] 本发明的另一个目标涉及使用本发明的核酸作为分子标记,以便能够将多毛番茄 (*S. habrochaites*) 的相应基因组序列导入到其它性亲和的物种或栽培番茄 (*S. lycopersicum*) 的变种中。事实上,所述标志使得有可能控制多毛番茄 (*S. habrochaites*) 的相应基因组序列导入到其它性亲和的物种或栽培番茄 (*S. lycopersicum*) 的变种中,因此有可能例如鉴定和筛选从野生物种和栽培物种之间的杂交而产生的、其中已经发生了本发明核酸的渗入的个体。使用本发明的核酸作为分子标记,可以通过本技术领域专业人员已知的任何技术来进行。例如,常用的技术是基于通过用本发明的核酸进行分子杂交,来鉴定用不同限制性酶消化的两个基因组之间的条带多态性 (RFLP)。另一个例子是通过使用 PCR,在被比较的基因组上,从与本发明的核酸互补的引物扩增出本发明的核酸的全部或一部分,以便在 PCR 扩增后鉴定两个基因组之间的大小多态性,从而搜索多态性。也可以在使用在被比较的基因组之间产生不同大小的片段的一种或多种限制性酶,对所述 PCR 产物进行消化后,揭示出多态性。

[0140] 在鉴定了多态性之后,该信息用于例如在从多毛番茄 (*S. habrochaites*) 和各种其中需要渗入本发明核酸的番茄之间杂交而产生的 F2 群体中,鉴定个体。

[0141] 本发明的另一个目标涉及通过细胞融合、特别是原生质体融合,将多毛番茄 (*S. habrochaites*) 编码 zFPS 和 SB 合酶的基因组序列导入到非性亲和的物种中。原生质体融合技术对于本技术领域专业人员来说是已知的 (Zimmermann 等, 1981, Bates 等, 1983), 它使得产生的杂合细胞含有属于两个融合物种的染色体成为可能。在融合过程中,两个物种的染色体或染色体片段被消除。然后可以通过 PCR,使用对编码 zFPS 的基因 (SEQ ID No. 1) 或编码 SB 合酶的基因 (SEQ ID No. 2) 或这两个基因特异的引物,立即筛选保留了含有本发明基因的染色体片段的杂合细胞。

[0142] 本发明还涉及细胞或非人类转基因生物体,其特征为通过失活本发明 Z, Z-FPP 合酶的编码基因或本发明 SB 合酶的编码基因或这两个基因,阻断了 SB 类型的倍半萜混合物的合成途径。所述基因的表达可以通过本技术领域专业人员已知的许多可用技术来阻断。基因可以被缺失、突变(例如通过用乙基甲烷磺酸化学诱变或通过电离辐射)、或打断(插入诱变)。此外,所述基因的表达也可以通过表达抑制性转录物进行基因消灭 (gene extinction) 来阻断。抑制性转录物是 RNA,它可以是双链 RNA、反义 RNA、核酶、能够形成三螺旋的 RNA 的形式,并与待阻断的基因的转录物具有一定的互补性或特异性。

[0143] 本发明还涉及能够降低或抑制 Z, Z-FPP 合酶的编码基因表达的核酸。事实上,可以在本发明教导的基础上制备抑制性 RNA (inhibitor RNA),所述抑制性 RNA 可以是双链 RNA、反义 RNA、核酶、能够形成三螺旋的 RNA 的形式,并与编码 Z, Z-FPP 合酶的基因具有一定互补性或特异性。例如,本发明涉及 Z, Z-FPP 合酶的编码基因的抑制性 RNA,含有 SEQ ID

No. 1 的至少 21、30、40、50、60、70、80、90 或 100 个连续的核苷酸,或与其互补的序列。同样,本发明涉及使用含有 SEQ ID No. 1 的至少 21、30、40、50、60、70、80、90 或 100 个连续的核苷酸或与其互补的序列的多核苷酸,来抑制 Z, Z-FPP 合酶的编码基因的表达。该酶的抑制可用于停止或减慢 SB 类型的倍半萜混合物的合成途径,并使 IPP 和 DMAP 可用于另一个需要提供 Z, Z-FPP 的目的合成途径。例如,该酶的抑制可以导致番茄叶片腺毛细胞中大根香叶烯 B 的生产水平的增加。因此,本发明涉及其中编码 Z, Z-FPP 合酶的基因已经被失活的转基因生物体或细胞。失活可以通过 RNA 干扰、基因缺失、化学突变技术、或通过同源重组失活来实现。因此,在所述转基因生物体或细胞中 Z, Z-FPP 或 SB 类型的倍半萜混合物的含量可以被降低。

[0144] 此外,本发明还涉及可以降低或抑制 SB 合酶的编码基因表达的核酸。事实上,可以在本发明教导的基础上制备抑制性 RNA,所述抑制性 RNA 可以是双链 RNA、反义 RNA、核酶、能够形成三螺旋的 RNA 的形式,并与编码 SB 合酶的基因具有一定互补性或特异性的抑制性 RNA。因此,本发明涉及编码 SB 合酶的基因的抑制性 RNA,含有 SEQ ID No. 3 的至少 30、40、50、60、70、80、90 或 100 个连续的核苷酸。同样,本发明涉及使用含有 SEQ ID No. 3 的至少 21、30、40、50、60、70、80、90 或 100 个连续核苷酸的多核苷酸,以便抑制编码 SB 合酶的基因的表达。该酶的抑制可用于停止 SB 类型的倍半萜混合物的合成途径,并使 Z, Z-FPP 可用于另一个需要提供所述化合物的目的合成途径,例如用于制备 Z, Z-FPP 衍生物,例如 Z, Z- 法尼醇,或是使用 Z, Z-FPP 作为底物的合酶的产物的其它倍半萜。

[0145] 因此,本发明涉及其中编码 SB 合酶的基因被失活的细胞或转基因生物体。失活可以通过 RNA 干扰技术、基因缺失、化学突变、或通过同源重组失活来实现。因此,在所述细胞或转基因生物体中,SB 类型的倍半萜混合物的含量可以被降低。

[0146] 本发明还涉及鉴定或克隆编码源自其它物种的 Z, Z-FPP 合酶或 SB 合酶的其它基因的方法,其中制备了分别具有序列 SEQ ID No. 1 或 3 的至少 15、30、50、75、100、200 或 300 个连续核苷酸的探针,并用于在样品中鉴定或筛选能够与所述探针杂交的核苷酸序列。样品可以是例如源自一种或多种生物体的基因组或 cDNA 文库。方法可以包含对鉴定或筛选到的序列进行表征的附加步骤。该附加步骤可以包括克隆、和 / 或测序、和 / 或序列比对、和 / 或测试酶活性。

[0147] 在下面的实施例中,本发明的其它特点和优点将变得了然,给出这些实施例是为了说明,而不是进行限制。

实施例

[0148] 番茄叶 cDNA 的合成 (多毛番茄 (*S. habrochaites*))

[0149] 使用总 RNA 提取试剂盒 (RNeasy Minikit, Qiagen) 从番茄叶 (多毛番茄) (*Solanum habrochaites* = *L. Hirsutum* LA1777) 提取 mRNA。使用反转录酶 (Roche, Cat. No. 03 531 317), 在制造商推荐的反应条件下,通过反转录从 polyT 引物合成相应的 cDNA。

[0150] 从多毛番茄 (*Solanum habrochaites*) 克隆 Z, Z-FPP 合酶 (Sh-zFPS) 和 SB 合酶 (Sh-SBS) 的编码序列

[0151] 在多毛番茄叶毛状体的 EST 文库中,从可公开获得的序列数据克隆两个基因 Sh-zFP 和 Sh-SBS 的完整和编码核苷酸序列 (Van der Hoeven 等, 2000, 未发表), 该序列数据

可以在 SGN 网站上获得 (“SOL 基因组学网络” (“SOL Genomics Network”), <http://www.sgn.cornell.edu/>)。

[0152] 通过 PCR 从番茄叶 cDNA 获得 Sh-zFPS (SEQ ID No. 1) 和 Sh-SBS (SEQ ID No. 3) 序列, 序列 ID No. 1 使用的引物是 5'-ACCATGGGTTCTTTGGTTCTTCAATGTTGGA-3' (SEQ ID No. 5) 和 5'-ACTCGAGATATGTGTGTCCACCAAAACGTCTATG-3' (SEQ ID No. 6), 序列 ID No. 3 使用的引物是 5'-TCCATGGTAGTTGGCTATAGAAGCACAATCA-3' (SEQ ID No. 7) 和 5'-TCTCGAGCATAAATTCAAA TTGAGGGATTAATGA-3' (SEQ ID No. 8)。在引物的 5' 末端加上限制性酶位点 (NcoI 和 XhoI), 以便于将 cDNA 克隆在启动子下游。使用 0.5 单位 Taq 聚合酶 (Eurogentec), 在制造商提供的缓冲液中进行 PCR 扩增。PCR 产物在 Qiaquick 柱 (Qiagen) 上纯化, 并通过用 T4DNA 连接酶 (NEB) 连接到 pGEM-T 克隆载体 (Promega) 中进行克隆。在对插入片段完全测序后, 选择具有预期序列的克隆。

[0153] Sh-zFPS 核苷酸序列 (SEQ ID No. 1) 长度为 912 个碱基, 编码 303 个氨基酸的蛋白 (SEQ ID No. 2)。使用 CCD 系统 (NCBI, Marchler-Bauer A, Bryant SH (2004)), 通过电脑运行 (“in silico”) 鉴定到了顺式 (Z)-异戊二烯二磷酸合酶 (cis-IPPS) 或者也被称为十一异戊二烯合酶 (UPPS) 类型的保守蛋白结构域。使用 BlastP 软件 (Altschul 等, 1997), 在 SwissProt 和 GenBank 蛋白数据库中搜索与 Sh-zFPS 相似的氨基酸序列。与十一聚异戊二烯二磷酸合酶显示出同源性的植物序列获得了最高的百分同一性, 但是它们的功能还没有被证实 (表 1)。例如, Sh-zFPS 分别与拟南芥 (*Arabidopsis thaliana*) BAA97345 和 AA063349 的推测 UPPS 具有 42% 和 39% 的同一性。也可以看到, Sh-zFPS 序列与结核微球菌 (*Micrococcus tuberculosis*) (053434) 的 Z, E-FPS 具有非常低的百分同一性 (24%)。

	生物体	PID	功能	% 同一性
	拟南芥 (<i>Arabidopsis thaliana</i>)	BAA97345	推测的 UPPS	42
	拟南芥 (<i>Arabidopsis thaliana</i>)	AAO63349	推测的 UPPS	39
	拟南芥 (<i>Arabidopsis thaliana</i>)	AAO42764	推测的 UPPS	39
	鱼腥藻 (<i>Anabaena</i> sp.) PCC 7120	DP58563	推测的 UPPS	37
	水稻 (<i>Oryza sativa</i>)	NP_915561	推测的 UPPS	36
[0154]	细长聚球藻 (<i>Synechococcus elongatus</i>)	Q8DI29	推测的 UPPS	35
	藤黄微球菌 (<i>Micrococcus luteus</i>)	O82827	推测的 UPPS	34
	<i>Gloeobacter violaceus</i>	Q7NPE7	推测的 UPPS	32
	结核微球菌 (<i>Micrococcus tuberculosis</i>)	P60479	UPPS	30
	结核微球菌 (<i>Micrococcus tuberculosis</i>)	O53434	E,Z-FPS	24

[0155] 表 1: 使用 BlastP 2.2.13 程序 (Altschul 等, 1997) 通过针对 SwissProt 和 GenBank 文库进行同源性搜索获得的与 Sh-zFPS 序列具有最高百分同一性 (%同一性) 的序列。

[0156] Sh-SBS 核苷酸序列 (SEQ ID No. 3) 长度为 2334 个碱基, 编码 777 个氨基酸的蛋白 (SEQ ID No. 4)。使用 CCD 系统 (NCBI, Marchler-BauerA, Bryant SH(2004)) 电脑运行, 鉴定到了保守的萜环化酶蛋白结构域。使用 BlastP 软件 (Altschul 等, 1997), 在 SwissProt 和 GenBank 蛋白文库中搜索与 Sh-SBS 相似的氨基酸序列。除了功能未知的烟草序列 (60%同一性) 之外, 植物的二萜合酶获得了最好的同源性 (表 2), 与编码不同植物 (拟南芥 (*Arabidopsis thaliana*)、笋瓜 (*Cucurbita maxima*)、甜菊 (*Stevia rebaudiana*)、水稻 (*Oryza sativa*)) 的贝壳杉烯合酶的序列具有大约 40 到 45% 的同一性。Sh-SBS 也显示出与已知功能的萜合酶例如松香二烯合酶和紫杉二烯合酶和 α -没药烯合酶的同源性较低 (26-33% 的同一性)。

	生物体	PID	功能	%同一性
	烟草 (<i>Nicotiana tabacum</i>)	AAS98912	推测的萜合酶	60
	拟南芥 (<i>Arabidopsis thaliana</i>)	AAC39443	贝壳杉烯合酶	44
	莴苣 (<i>Lactuca sativa</i>)	BAB12441	推测的贝壳杉烯合酶	43
	笋瓜 (<i>Cucurbita maxima</i>)	Q39548	贝壳杉烯合酶	43
[0157]	蒺藜苜蓿 (<i>Medicago truncatula</i>)	ABE87878	推测的萜合酶	43
	甜菊 (<i>Stevia rebaudiana</i>)	AAD34294	贝壳杉烯合酶	42
	水稻 (<i>Oryza sativa</i>)	AAQ72559	贝壳杉烯合酶	41
	大麦 (<i>Hordeum vulgare</i>)	AAT49066	贝壳杉烯合酶	36
	大冷杉 (<i>Abies grandis</i>)	AAC24192	α -没药烯合酶	33
	大冷杉 (<i>Abies grandis</i>)	AAB05407	松香二烯合酶	27
	欧洲紫杉 (<i>Taxus baccata</i>)	2211347A	紫杉二烯合酶	26

[0158] 表 2: 使用 BlastP 2.2.13 程序 (Altschul 等, 1997) 通过针对 SwissProt 和 GenBank 文库进行同源性搜索获得的与 Sh-SBS 序列具有最高百分同一性 (%同一性) 的序列。

[0159] Sh-zFPS 和 Sh-SBS 重组蛋白的生产

[0160] 将编码 Sh-zFPS 和 Sh-SBS 的核酸序列 (分别为 SEQ ID No. 1 和 3) 分别导入到 pET-30b 表达载体 (Novagen) 中。为了允许纯化重组蛋白, 删除了这些序列的终止密码子, 以便产生在 C-末端具有 6 个组氨酸尾巴的融合蛋白。将这样获得的 Sh-zFPS-pET30b 和 Sh-SBS-pET30b 载体通过电穿孔, 导入到修饰过的用于表达重组蛋白的大肠杆菌中 (BL21-CodonPlus, Stratagene)。为了生产蛋白, 在 37°C 开始进行大肠杆菌培养 (500ml), 直到 600nm 处的光密度达到 0.5。在这个阶段, 将培养温度降低到 16°C, 并向培养基中加入乙醇 (1% V/V)。温育 1 小时后, 以 1mM 的浓度向培养基中加入 IPTG, 以诱导重组蛋白的表

达。然后将培养物温育 18 小时,通过在 4℃ 离心细胞来停止培养。将细胞沉淀重新悬浮在 200mM 磷酸盐缓冲液中。使用 French 压力器 ($P = 19\text{bar}$) 裂解细胞,将裂解液以 15,000g 离心。将含有可溶性蛋白的上清液在 PD10 柱 (Amersham) 上脱盐,用 pH7 的 50mM 磷酸盐缓冲液平衡,并加样到镍-次氨基三乙酸亲和树脂 (Ni-NTA, Qiagen) 上,严格按照制造商的洗脱程序进行。通过 SDS-PAGE 鉴定具有最高量 Sh-zFPS-6His 和 Sh-SBS-6His 蛋白的级份,然后在用含有 100mM KCl 的 pH 7.8 的 50mM HEPES 缓冲液平衡的 PD10 柱上脱盐。对于两种蛋白来说,与粗蛋白级份相比,获得了大约 20 倍的富集。

[0161] Sh-zFPS 和 Sh-SBS 重组蛋白的体外酶活性测试

[0162] 活性测试在下面的缓冲液中进行:50mM HEPES, pH 7.8, 100mMKCl, 7.5mM MgCl_2 , 5% (w/v) 甘油,和 5mM DTT。为了进行活性测试,将 25 或 50 μg 来自富集级份的蛋白,与单独或组合的底物 IPP、DMAPP、GPP、FPP、GGPP (Sigma) 在 32℃ 温育 2 小时,每种底物的浓度为 65 μM 。

[0163] 对于用 Sh-zFPS-6His 进行的测试,通过在 37℃ 加入 20 单位的牛肠碱性磷酸酶 (New England Biolabs) 1 小时,将反应产物脱磷酸化。然后用氯仿/甲醇/水混合物 (0.5/1/0.4) 提取磷酸化产物,萜醇。通过加入水 (0.5 倍体积) 和氯仿 (0.5 倍体积) 并以 2,000g 离心,分离了水相和有机相。收集有机相,在氮气流下干燥,溶解在戊烷中,通过质谱偶联的气相色谱进行分析 (GC-MS 5973N, Agilent Technologies)。通过将萜醇的质谱与国立标准与技术研究院 (National Institute of Standards and Technology (NIST)) 数据库中的质谱进行比较,并将其保留时间合并到香叶醇、法尼醇和香叶基香叶醇参比标准品 (Fluka) 的保留时间中,鉴定了萜醇。

[0164] 对于用 Sh-SBS-6His 进行的测试,使用戊烷 (体积对体积) 从反应混合物直接提取 (三次) 反应中产生的萜。将三份戊烷提取物合并,在氮气流下浓缩,通过 GC-MS 进行分析。通过查询 NIST 数据库并与多毛番茄 (*Solanum habrochaites*) 渗出物的提取物的色谱图进行比较,鉴定产生的萜。

[0165] 将 Sh-zFPS-6His 重组蛋白与底物 IPP+DMAPP 或 IPP+GPP 温育 (参见表 1)。在去磷酸化后,只有第一个组合 (IPP+DMAPP) 导致了形成具有 15 个碳原子 (C_{15}) 的单一萜醇,通过其质谱图与 NIST 数据库中标准谱图的比较,将其鉴定为法尼醇 (图 4)。保留时间与含有四种法尼醇异构体 (E, E- 法尼醇; E, Z- 法尼醇; Z, E- 法尼醇和 Z, Z- 法尼醇) 混合物的法尼醇标准品进行比较。Sh-zFPS-6His 产生的法尼醇具有与 Z, Z- 法尼醇相同的保留时间。

[0166] 将 Sh-SBS-6His 重组蛋白与底物 GPP、FPP 和 GGPP 温育。无论使用何种底物,没有检测到新的萜,表明酶对于反式含烯丙型底物没有活性 (表 1)。

	酶 1	酶 2	底物	IPP	产物
[0167]	Sh-zFPS	不存在	DMAP	IPP	A*
	Sh-zFPS	不存在	E-GPP	IPP	-
	不存在	Sh-SBS	E-GPP	No	-
	不存在	Sh-SBS	E,E-FPP	No	-
	不存在	Sh-SBS	E,E,E-GGPP	No	-
	Sh-zFPS	Sh-SBS	DMAPP	IPP	B,C,D,E,F
	Sh-zFPS	Sh-SBS	GPP	IPP	-

[0168] *:去磷酸化后

[0169] 表 3 :Sh-zFPS-6His 和 Sh-SBS-6His 重组蛋白对不同的类异戊二烯底物获得的体外酶活性测试汇总。A, Z, Z- 法尼醇 ;B, α - 檀香萜 ;C, 表 - β - 檀香萜 ;D, 内 - β - 香柠檬烯 ;E, 顺式 - α - 香柠檬烯 ;F, 反式 - α - 香柠檬烯。

[0170] 将 Sh-zFPS-6His 和 Sh-SBS-6His 酶与底物 DMAPP 和 IPP 或 GPP 和 IPP 一起温育。反应产物不用磷酸酶处理,而是用可以溶解所有烯烃的戊烷提取。使用 DMAPP 和 IPP 作为底物时,提取物的色谱分析表明存在几种萜化合物(图 5)。将 GC 分布图与从亲本栽培番茄和多毛番茄获得的几乎同基因种系——TA517 番茄种系的渗出物的 GC 分布图(Monforte 和 Tanksley, 2000) 进行比较。TA517 种系含有多毛番茄负责 II 类倍半萜生物合成的基因渗入片段(van der Hoeven 等, 2000)。两个 GC 分布图相同,证实了重组酶具有与番茄中存在的酶严格相同的活性(图 6)。通过将其保留时间和质谱图与 TA517 渗出物提取物的进行比较,主要的成分(图 5, 2 号峰)被鉴定为是 α - 檀香萜。按照量的降低,还鉴定到了 α - 檀香萜(4 号峰)、内 - β - 香柠檬烯(5 号峰)和 α - 香柠檬烯(1 号峰)。因此,Sh-SBS-6His 酶是其活性导致形成多种产物的酶。

[0171] 可以得出结论, Sh-zFPS 是作用于 IPP 和 DMAPP 以产生法呢基二磷酸类型的磷酸化化合物的酶。保留时间与 (Z, Z)- 法尼醇标准品相同,表明 Sh-zFPS 产生的法呢基二磷酸的特征为 Z, Z 构型类型。Sh-zFPS-6His 对 E- 香叶基二磷酸与 IPP 的组合没有活性的事实,表明该分子不是 Sh-zFPS-6His 的中间体,并通过结合一个 DMAPP 和一个 IPP 产生至少一个 Z 类型的键。Sh-SBS 只对 Sh-zFPS 产生的法呢基二磷酸化合物有活性,但是产生了至少 4 种被明确鉴定的倍半萜,主要的一种是 α - 檀香萜。最后,与原始植物的天然酶相比, C- 末端中 6 个组氨酸标签和间隔序列的存在对酶活性没有影响(图 3)。

[0172] Z, Z-FPP 合酶(Sh-zFPS) 和 SB 合酶(Sh-SBS) 基因在番茄中的表达

[0173] 将毛状体特异性烟草启动子(eCBTS02, Tissier 等, 2004) 克隆到 Sh-zFPS 和 Sh-SBS 编码序列的上游,以形成构建物 eCBTS1.0-Sh-zFPS[#1] 和 eCBTS1.0-Sh-SBS[#2]。将这两种构建物导入到同一个用于通过土壤杆菌转化、含有卡那霉素抗性基因的二元载体中,产生 pLIBR0-064 二元载体(图 2A)。也将构建物 #2 单独导入到 pLIBR0-065 二元载体中(图 2B)。通过电穿孔将载体导入土壤杆菌菌株 LBA4404。使用包涵不同载体的菌株通过叶盘方法进行林地烟草(N. sylvestris) 的遗传转化(Horsch 等, 1985)。

[0174] 通过遗传转化将携带转基因的 T-DNA 导入林地烟草转基因细胞系中(源种系 ihpCBTS),其中 CBTS 基因的表达被 RNA 干扰所抑制(Tissier et al.,2005)。在该种系中,叶片中 CBT- 二醇的量非常低。野生烟草,林地烟草(*N. sylvestris*) 不产生番茄中相应的倍半萜。对于每种构建物,获得了大约 25 株卡那霉素抗性植物。通过对叶片基因组 DNA 进行 PCR,证实了每株植物中转基因的存在。通过对来自烟草叶片的 mRNA 进行实时定量 PCR,检查了叶片中转基因的表达水平。将 Sh-SBS 转基因的表达水平与在烟草叶片中组成性表达的肌动蛋白基因的表达水平进行比较。结果示例在图 3 中。结果表明,选择的种系都以不同水平表达转基因,相对于肌动蛋白基因,其范围对于 pLIBR0-064 构建物来说是 1 到 50,对于 pLIBR0-065 构建物来说是 1 到 10。

[0175] 因为烯烃类倍半萜是挥发性分子,因此在培养箱中受控的气氛下分析了植物散发出的挥发性分子。将气氛样品捕获在含有吸收烯烃类萜的 SuperQ® 滤膜(Altech)的微分电路中。然后用 1ml 戊烷洗脱分子,通过 GC/MS 进行分析。图 7 显示了转基因植物散发出的挥发性分子的色谱分布图的例子。它显示,表达 Sh-zFP 和 Sh-SBS 基因种系 #3877 含有几个新的峰,它们具有该类倍半萜特征性的分子离子标志 93、94、121 和 204。将色谱图分布与从天然产生这些化合物的 TA517 番茄种系的渗出物获得的分布图进行比较(图 7)。两个分布图相同,每个峰的质谱图也与 TA517 种系的峰的质谱图相同。表达两个转基因 Sh-zFPS 和 Sh-SBS 的植物产生 α -檀香萜、表- β -檀香萜、顺式- α -香柠檬烯、反式- α -香柠檬烯和内- β -香柠檬烯,但是只整合了 Sh-SBS 基因的转基因种系不产生任何这些化合物。这些结果表明,这两个基因的存在对于合成 SB 类型的倍半萜是必需的。它们证实了在体外使用重组酶获得的数据。结论是,Sh-zFPS 和 Sh-SBS 基因在 van der Hoeven 等,2000 描述的 Sst2 基因座位置处,负责番茄倍半萜的生物合成。

[0176] 栽培番茄(*S. lycopersicum*)和多毛番茄(*S. habrochaites*)之间的 zFPS 基因多态性

[0177] 从栽培番茄、多毛番茄和 TA517 的基因组 DNA,使用引物 ACCATGGGTTCTTTGGTTCTTC AATGTTGGA(SEQ ID No. 9)和 ACTCGAGATATGTGTGTCCACCAAAACGTCTATG(SEQ ID No. 10),通过 PCR 扩增了完整的 zFPS 基因。在两个基因组中扩增到了两个产物,表明在两个基因组中存在基因的至少两个拷贝(图 8)。两个产物中具有大约 3000 个核苷酸大小的一个产物(条带 A),对于两个基因组是共同的。第二个产物在栽培番茄中具有大约 2000 个核苷酸的大小(条带 C),在多毛番茄中具有大约 2500 个核苷酸的大小(条带 B)。TA517 种系的分布图与多毛番茄的相同。这些结果表明,多毛番茄特异性的基因 B 已经导入到 TA517 种系中。产物 B 的测序证实它对应于本文描述的 zFPS 基因。这个结果也证明,通过使用 zFPS 基因在两个基因组之间存在的多态性,可以将多毛番茄的 zFPS 基因导入到栽培番茄中。

[0178] 参考文献

[0179] Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller 和 David J. Lipman. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. (带有间隙的 BLAST 和 PSI-BLAST: 新一代蛋白数据库搜索程序) *Nucleic Acids Res.* 25:3389-3402.

[0180] Bates GW, Gaynor JJ, Shekhawat NS. (1983) Fusion of Plant Protoplasts by Electric Fields. (通过电场融合植物原生质体) *Plant Physiol.* 72:1110-1113.

- [0181] Bohlmann J, Crock J, Jetter R, Croteau R. (1998) Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)- α -bisabolene synthase from grand fir (*Abies grandis*). (针叶树中基于类萜的防御: 来自大冷杉 (*Abies grandis*) 伤口诱导的 (E)- α -没药烯合酶的 cDNA 克隆、表征和功能性表达) *Proc. Natl. Acad. Sci. USA*. 95:6756-6761.
- [0182] Bouwmeester HJ, Kodde J, Verstappen FW, Altug IG, de Kraker JW, Wallaart TE. (2002) Isolation and characterization of two germacrene A synthase cDNA clones from chicory. (来自菊苣的两个大根香叶烯 A 合酶 cDNA 克隆的分离和表征) *Plant Physiol*. 29:134-144.
- [0183] Cai Y, Jia JW, Crock J, Lin ZX, Chen XY, Croteau R. (2002) A cDNA clone for beta-caryophyllene synthase from *Artemisia annua*. (来自黄花蒿 (*Artemisia annua*) 的 β -丁子香烯合酶的 cDNA 克隆) *Phytochemistry* 61:523-529.
- [0184] Cane DE. (1999) Sesquiterpene biosynthesis: Cyclization mechanisms. (倍半萜的生物合成: 环化机制) In *comprehensive natural products chemistry, isoprenoids including carotenoids and steroids, Vol. 2*, (在《综合天然产物化学, 包括类胡萝卜素和类固醇的类异戊二烯》第 2 卷中) D. D. Cane 主编, (Amsterdam, The Netherlands; Elsever), pp. 155-200.
- [0185] Carman RM, Duffield AR (1993) The Biosynthesis of Labdanoids—the Optical Purity of Naturally-Occurring Manool and Abienol. (类赖百当 (Labdanoids) 的生物合成——天然存在的泪杉醇和冷杉醇的光学纯度) *Aust. J. Chem.* 46:1105-1114.
- [0186] Clarke CF, Tanaka RD, Svenson K, Wamsley M, Fogelman AM, Edwards PA. (1987) Molecular cloning and sequence of a cholesterol-repressible enzyme related to prenyltransferase in the isoprene biosynthetic pathway. (异戊二烯生物合成途径中与异戊二烯基转移酶有关的胆固醇可抑制的酶的分子克隆和序列) *Mol. Cell. Biol.* 7:3138-3146.
- [0187] Colby SM, Crock J, Dowdle-Rizzo B, Lemaux PG, Croteau R. (1998) Germacrene C synthase from *Lycopersicon esculentum* cv. VFNT cherry tomato: cDNA isolation, characterization, and bacterial expression of the multiple product sesquiterpene cyclase. (来自番茄 (*Lycopersicon esculentum*) cv. VFNT 樱桃蕃茄的大根香叶烯 C 合酶: 多产物倍半萜环化酶的 cDNA 分离、表征以及细菌表达) *Proc. Natl. Acad. Sci. USA* 95:2216-2221.
- [0188] Facchini PJ, Chappell J. (1992) Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. (烟草中激发子诱导的倍半萜环化酶基因家族) *Proc. Natl. Acad. Sci. USA* 89(22):11088-11092.
- [0189] Heinstein PF, Herman DL, Tove SB, Smith FH. (1970) Biosynthesis of gossypol. Incorporation of mevalonate-2- 14 C and isoprenylpyrophosphates (棉子酚的生物合成。甲羟戊酸-2- 14 C 和异戊二烯焦磷酸的掺入) *J. Biol. Chem.* 245:4658-4665.
- [0190] Horsch RB, Rogers SG, Fraley RT. (1985) Transgenic plants. (《转基因植物》) *Cold Spring Harb Symp Quant Biol.* 50:433-437.

- [0191] Kharel Y, Takahashi S, Yamashita S, Koyama T. (2006) Manipulation of prenyl chain length determination mechanism of cis-prenyltransferases. (操纵顺式异戊二烯基转移酶的异戊二烯链长决定机制) FEBS J. 273 :647-657.
- [0192] Kollner TG, O' Maille PE, Gatto N, Boland W, Gershenzon J, Degenhardt J. (2006) Two pockets in the active site of maize sesquiterpene synthase TPS4 carry out sequential parts of the reaction scheme resulting in multiple products. (玉米倍半萜合酶 TPS4 活性位点中的两个小袋执行了产生多种产物的反应计划的连续部分) Arch. Biochem. Biophys. 448 :83-92.
- [0193] Koyama T. (1999) Molecular analysis of prenyl chain elongation enzymes. (异戊二烯基链延伸酶的分子分析) Biosci. Biotechnol. Biochem. 63 :1671-1676.
- [0194] Liang PH, Ko TP, Wang AH. (2002) Structure, mechanism and function of prenyltransferases. (异戊二烯基转移酶的结构、机理和功能) 269 :3339-3354.
- [0195] Marchler-Bauer A, Bryant SH. (2004) CD-Search : protein domain annotations on the fly. (CD- 搜索 : 高速蛋白结构域注解) Nucleic Acids Res. 32 (W) :327-331.
- [0196] McMillan J. 和 Beale M. H. (1999) Dipterene biosynthesis. (双萜生物合成) In comprehensive natural products chemistry, isoprenoids including carotenoids and steroids Vol. 2, (在《综合天然产物化学, 包括类胡萝卜素和类固醇的类异戊二烯》第 2 卷中) D. D. Cane 主编 (Amsterdam, The Netherlands ; Elsever), pp. 217-243.
- [0197] Monforte AJ, Tanksley SD. (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in an *L. esculentum* background : A tool for gene mapping and gene discovery. (在栽培番茄 (*L. esculentum*) 背景中含有大多数多毛番茄 (*Lycopersicon hirsutum*) 基因组的一组接近同基因和回交重组近交系的开发 : 用于基因作图和基因发现的工具) Genomics 43 :803-813.
- [0198] Oh SK, Han KH, Ryu SB, Kang H. (2000) Molecular cloning, expression, and functional analysis of a cis-prenyltransferase from *Arabidopsis thaliana*. Implications in rubber biosynthesis. (来自拟南芥 (*Arabidopsis thaliana*) 的顺异戊二烯基转移酶的分子克隆、表达和功能分析。在橡胶生物合成中的含义。) J. Biol. Chem. 275 :18482-18488.
- [0199] Poulter DC. (2006) Farnesyl Diphosphate Synthase. A Paradigm for Understanding Structure and Function Relationships in E-prenyl Diphosphate Synthases. (法呢基二磷酸合酶。理解 E- 聚异戊二烯二磷酸合酶中结构与功能关系的范例) Phytochem. Review 5 :17-26.
- [0200] Rodriguez-Concepcion M, Boronat A. (2002) Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids. A metabolic milestone achieved through genomics. (阐明细菌和质体中类异戊二烯化合物生物合成的甲基赤藓糖醇磷酸途径。通过基因组学完成的代谢里程碑) Plant Physiol. 130 (3) :1079-89.
- [0201] Sharon-Asa L, Shalit M, Frydman A, Bar E, Holland D, Or E, Lavi U, Lewinsohn

E, Eyal Y. (2003) Citrus fruit flavor and aroma biosynthesis: isolation, functional characterization, and developmental regulation of *Cstps1*, a key gene in the production of the sesquiterpene aroma compound valencene. (柑橘类水果香料和香味的生物合成: 倍半萜芳香化合物凡伦橘烯生产中的关键基因 *Cstps1* 的分离、功能表征和发育调控) *Plant J.* 36: 664-74.

[0202] Shimizu N, Koyama T, Ogura K. (1998) Molecular cloning, expression, and purification of undecaprenyl diphosphate synthase. No sequence similarity between E- and Z-prenyl diphosphate synthases. (十一异戊二烯二磷酸合酶的分子克隆、表达和纯化。在 E- 和 Z- 异戊二烯二磷酸合酶之间没有序列相似性) *J. Biol. Chem.* 273: 19476-19481.

[0203] Schulbach MC, Brennan PJ, Crick DC. (2000) Identification of a short (C15) chain Z-isoprenyl diphosphate synthase and a homologous long (C50) chain isoprenyl diphosphate synthase in *Mycobacterium tuberculosis*. (在结核分枝杆菌 (*Mycobacterium tuberculosis*) 中鉴定短链 (C15) Z- 异戊二烯二磷酸合酶和同源的长链 (C50) 异戊二烯二磷酸合酶) *J. Biol. Chem.* 275: 22876-22881.

[0204] Tarshis LC, Yan M, Poulter CD, Sacchettini JC. (1994) Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-Å resolution. (重组法呢基二磷酸合酶的 2.6-Å 分辨率的晶体结构) *Biochemistry* 33: 10871-10877.

[0205] Tholl D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. (萜合酶以及萜代谢的调控、多样性和生物学功能) *Curr. Opin. Plant Biol.* 9: 297-304.

[0206] Thompson, JD, Gibson, TJ, Plewniak, F, Jeanmougin, F 和 Higgins, DG. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. (ClustalX 窗口界面: 质量分析工具辅助的多序列比对灵活策略) *Nucleic Acids Res.* 24: 4876-4882.

[0207] Tissier A, Sallaud C, Rontein D. (2004) Promoteurs vegetaux et utilisations (利用的植物启动子) PCT/FR 05/02530.

[0208] Tissier A, Sallaud C, Rontein D. (2005) Systeme de production de terpenes dans les plantes (在植物中生产萜的体系). French patent application FR 05 00855 (法国专利申请 FR 05 00855)。

[0209] van Der Hoeven RS, Monforte AJ, Breeden D, Tanksley SD, Steffens JC. (2000) Free in PMC (PMC 中可自由得到) Genetic control and evolution of sesquiterpene biosynthesis in *Lycopersicon esculentum* and *L. hirsutum*. (栽培番茄 (*Lycopersicon esculentum*) 和多毛番茄 (*L. hirsutum*) 中倍半萜生物合成的遗传控制和进化) *Plant Cell.* 12: 2283-2294.

[0210] Wallaart TE, Bouwmeester HJ, Hille J, Poppinga L, Maijers NC. (2001) Amorpho-4,11-diene synthase: cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel antimalarial drug artemisinin. (紫穗槐-4,11-二烯合酶: 新的抗疟疾药物青蒿素的生物合成途径中关键酶的克隆和功能表达)

Planta. 212 :460-465.

[0211] Wang K, Ohnuma SI. (1999) Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. (异戊二烯二磷酸合酶的链长决定机制以及对分子进化的含义) 24 :445-451.

[0212] Wise ML, Croteau RA. (1999) Monoterpene biosynthesis. (单萜的生物合成) In comprehensive natural products chemistry, isoprenoids including carotenoids and steroids Vol. 2, (在《综合天然产物化学, 包括类胡萝卜素和类固醇的类异戊二烯》第2卷中) D. D. Cane 主编, (Amsterdam, The Netherlands ; Elsevier), pp. 97-153.

[0213] Zimmermann U, Scheurich P. (1981) High frequency fusion of plant protoplasts by electric fields. (使用电场的植物原生质体高频率融合) Planta 151 :26-32.

[0001]

序列表

<110> 菲利普莫里斯产品股份公司

<120> 具有多种产物的 Z, Z-法呢基二磷酸合酶和倍半萜合酶的编码基因及其应用

<130> SCT093674-90

<160> 10

<170> PatentIn version 3.3

<210> 1

<211> 912

<212> DNA

<213> Solanum abutiloides

<220> -

<221> CDS

<222> (1)..(912)

<400> 1

atg agt tct ttg gtt ctt caa tgt tgg aaa tta tca tct cca tct ctg	48
Met Ser Ser Leu Val Leu Gln Cys Trp Lys Leu Ser Ser Pro Ser Leu	
1 5 10 15	

att tta caa caa aat aca tca ata tcc atg ggt gca ttc aaa ggt att	96
Ile Leu Gln Gln Asn Thr Ser Ile Ser Met Gly Ala Phe Lys Gly Ile	
20 25 30	

cat aaa ctt caa atc cca aat tca cct ctg aca gtg tct gct cgt gga	144
His Lys Leu Gln Ile Pro Asn Ser Pro Leu Thr Val Ser Ala Arg Gly	
35 40 45	

ctc aac aag att tca tgc tca ctc agc tta caa acc gaa aaa ctt tgt	192
Leu Asn Lys Ile Ser Cys Ser Leu Ser Leu Gln Thr Glu Lys Leu Cys	
50 55 60	

tat gag gat aat gat aat gat ctt gat gaa gaa ctt atg cct aaa cac	240
Tyr Glu Asp Asn Asp Asn Asp Leu Asp Glu Glu Leu Met Pro Lys His	
65 70 75 80	

att gct ttg ata atg gat ggt aat agg aga tgg gca aag gat aag ggt	288
---	-----

[0002]

Ile Ala Leu Ile Met Asp Gly Asn Arg Arg Trp Ala Lys Asp Lys Gly	
85 90 95	
tta gac gta tcc gaa ggt cac aaa cat ctc ttt cca aaa tta aaa gag	336
Leu Asp Val Ser Glu Gly His Lys His Leu Phe Pro Lys Leu Lys Glu	
100 105 110	
att tgt gac att tct tct aaa ttg gga ata caa gtt atc act gct ttt	384
Ile Cys Asp Ile Ser Ser Lys Leu Gly Ile Gln Val Ile Thr Ala Phe	
115 120 125	
gca ttc tct act gaa aat tgg aaa cga gcc aag ggg gag gtt gat ttc	432
Ala Phe Ser Thr Glu Asn Trp Lys Arg Ala Lys Gly Glu Val Asp Phe	
130 135 140	
ttg atg caa atg ttc gaa gaa ctc tat gat gag ttt tcg agg tct gga	480
Leu Met Gln Met Phe Glu Glu Leu Tyr Asp Glu Phe Ser Arg Ser Gly	
145 150 155 160	
gta aga gtg tct att att ggt tgt aaa acc gac ctc cca atg aca tta	528
Val Arg Val Ser Ile Ile Gly Cys Lys Thr Asp Leu Pro Met Thr Leu	
165 170 175	
caa aaa tgc ata gca tta aca gaa gag act aca aag gga aac aaa gga	576
Gln Lys Cys Ile Ala Leu Thr Glu Glu Thr Thr Lys Gly Asn Lys Gly	
180 185 190	
ctt cac ctt gtg att gca cta aac tat ggt ggc tat tat gac ata ttg	624
Leu His Leu Val Ile Ala Leu Asn Tyr Gly Gly Tyr Tyr Asp Ile Leu	
195 200 205	
caa gca aca aaa agc att gtt aat aaa gca atg aat ggt tta tta gat	672
Gln Ala Thr Lys Ser Ile Val Asn Lys Ala Met Asn Gly Leu Leu Asp	
210 215 220	
gta gaa gat atc aac aag aat tta ttt gat caa gaa ctt gaa agc aag	720
Val Glu Asp Ile Asn Lys Asn Leu Phe Asp Gln Glu Leu Glu Ser Lys	
225 230 235 240	
tgt cca aat cct gat tta ctt ata agg aca gga ggt gat caa aga gtt	768
Cys Pro Asn Pro Asp Leu Leu Ile Arg Thr Gly Gly Asp Gln Arg Val	
245 250 255	
agt aac ttt ttg ttg tgg caa ttg gct tat act gaa ttt tac ttc acc	816

[0003]

Ser Asn Phe Leu Leu Trp Gln Leu Ala Tyr Thr Glu Phe Tyr Phe Thr
260 265 270

aaa aca ttg ttt cct gat ttt gga gag gaa gat ctt aaa gag gca ata 864
Lys Thr Leu Phe Pro Asp Phe Gly Glu Glu Asp Leu Lys Glu Ala Ile
275 280 285

ata aac ttt caa caa agg cat aga cgt ttt ggt gga cac aca tat tga 912
Ile Asn Phe Gln Gln Arg His Arg Arg Phe Gly Gly His Thr Tyr
290 295 300

<210> 2

<211> 303

<212> PRT

<213> Solanum abutiloides

<400> 2

Met Ser Ser Leu Val Leu Gln Cys Trp Lys Leu Ser Ser Pro Ser Leu
1 5 10 15

Ile Leu Gln Gln Asn Thr Ser Ile Ser Met Gly Ala Phe Lys Gly Ile
20 25 30

His Lys Leu Gln Ile Pro Asn Ser Pro Leu Thr Val Ser Ala Arg Gly
35 40 45

Leu Asn Lys Ile Ser Cys Ser Leu Ser Leu Gln Thr Glu Lys Leu Cys
50 55 60

Tyr Glu Asp Asn Asp Asn Asp Leu Asp Glu Glu Leu Met Pro Lys His
65 70 75 80

Ile Ala Leu Ile Met Asp Gly Asn Arg Arg Trp Ala Lys Asp Lys Gly
85 90 95

Leu Asp Val Ser Glu Gly His Lys His Leu Phe Pro Lys Leu Lys Glu

[0004]

100	105	110
Ile Cys Asp Ile Ser Ser Lys Leu Gly Ile Gln Val Ile Thr Ala Phe 115	120	125
Ala Phe Ser Thr Glu Asn Trp Lys Arg Ala Lys Gly Glu Val Asp Phe 130	135	140
Leu Met Gln Met Phe Glu Glu Leu Tyr Asp Glu Phe Ser Arg Ser Gly 145	150	155 160
Val Arg Val Ser Ile Ile Gly Cys Lys Thr Asp Leu Pro Met Thr Leu 165	170	175
Gln Lys Cys Ile Ala Leu Thr Glu Glu Thr Thr Lys Gly Asn Lys Gly 180	185	190
Leu His Leu Val Ile Ala Leu Asn Tyr Gly Gly Tyr Tyr Asp Ile Leu 195	200	205
Gln Ala Thr Lys Ser Ile Val Asn Lys Ala Met Asn Gly Leu Leu Asp 210	215	220
Val Glu Asp Ile Asn Lys Asn Leu Phe Asp Gln Glu Leu Glu Ser Lys 225	230	235 240
Cys Pro Asn Pro Asp Leu Leu Ile Arg Thr Gly Gly Asp Gln Arg Val 245	250	255
Ser Asn Phe Leu Leu Trp Gln Leu Ala Tyr Thr Glu Phe Tyr Phe Thr 260	265	270
Lys Thr Leu Phe Pro Asp Phe Gly Glu Glu Asp Leu Lys Glu Ala Ile		

[0005]

275	280	285	
Ile Asn Phe Gln Gln Arg His Arg Arg Phe Gly Gly His Thr Tyr			
290	295	300	
<210> 3			
<211> 2334			
<212> DNA			
<213> Solanum habrochaites			
<220>			
<221> CDS			
<222> (1)..(2334)			
<400> 3			
atg ata gtt ggc tat aga agc aca atc ata acc ctt tct cat cct aag			48
Met Ile Val Gly Tyr Arg Ser Thr Ile Ile Thr Leu Ser His Pro Lys			
1 5 10 15			
cta ggc aat ggg aaa aca att tca tcc aat gca att ttc cag aga tca			96
Leu Gly Asn Gly Lys Thr Ile Ser Ser Asn Ala Ile Phe Gln Arg Ser			
20 25 30			
tgt aga gta aga tgc agc cac agt acc cct tca tca atg aat ggt ttc			144
Cys Arg Val Arg Cys Ser His Ser Thr Pro Ser Ser Met Asn Gly Phe			
35 40 45			
gaa gat gca agg gat aga ata agg gaa agt ttt ggg aaa gta gag tta			192
Glu Asp Ala Arg Asp Arg Ile Arg Glu Ser Phe Gly Lys Val Glu Leu			
50 55 60			
tct cct tct tcc tat gac aca gca tgg gta gct atg gtc cct tca aaa			240
Ser Pro Ser Ser Tyr Asp Thr Ala Trp Val Ala Met Val Pro Ser Lys			
65 70 75 80			
cat tca cta aat gag cca tgt ttt cca caa tgt ttg gat tgg att att			288
His Ser Leu Asn Glu Pro Cys Phe Pro Gln Cys Leu Asp Trp Ile Ile			
85 90 95			
gaa aat caa aga gaa gat gga tct tgg gga cta aac cct agc cat cca			336
Glu Asn Gln Arg Glu Asp Gly Ser Trp Gly Leu Asn Pro Ser His Pro			

[0006]

100	105	110	
ttg ctt ctc aag gac tca ctt tct tcc act ctt gca tgt ttg ctt gca			384
Leu Leu Leu Lys Asp Ser Leu Ser Ser Thr Leu Ala Cys Leu Leu Ala			
115	120	125	
cta acc aaa tgg aga gtt gga gat gag caa atc aaa aga ggc ctt ggc			432
Leu Thr Lys Trp Arg Val Gly Asp Glu Gln Ile Lys Arg Gly Leu Gly			
130	135	140	
ttt att gaa acc cag agt tgg gca att gat aac aag gat caa att tca			480
Phe Ile Glu Thr Gln Ser Trp Ala Ile Asp Asn Lys Asp Gln Ile Ser			
145	150	155	160
cct cta gga ttt gaa att ata ttt ccc agt atg atc aag tct gca gaa			528
Pro Leu Gly Phe Glu Ile Ile Phe Pro Ser Met Ile Lys Ser Ala Glu			
165	170	175	
aaa cta aac tta aat cta gca att aac aaa aga gat tca aca att aaa			576
Lys Leu Asn Leu Asn Leu Ala Ile Asn Lys Arg Asp Ser Thr Ile Lys			
180	185	190	
aga gca tta cag aat gag ttc acg agg aat att gaa tat atg agt gaa			624
Arg Ala Leu Gln Asn Glu Phe Thr Arg Asn Ile Glu Tyr Met Ser Glu			
195	200	205	
gga gtt ggt gaa tta tgt gat tgg aag gaa ata ata aag tta cat caa			672
Gly Val Gly Glu Leu Cys Asp Trp Lys Glu Ile Ile Lys Leu His Gln			
210	215	220	
agg caa aat ggt tca tta ttt gat tca cca gcc act act gca gct gcc			720
Arg Gln Asn Gly Ser Leu Phe Asp Ser Pro Ala Thr Thr Ala Ala Ala			
225	230	235	240
ttg att tac cat cag cat gat aaa aaa tgc tat gaa tat ctt aat tca			768
Leu Ile Tyr His Gln His Asp Lys Lys Cys Tyr Glu Tyr Leu Asn Ser			
245	250	255	
atc ttg caa caa cac aaa aat tgg gtt ccc act atg tat cca aca aag			816
Ile Leu Gln Gln His Lys Asn Trp Val Pro Thr Met Tyr Pro Thr Lys			
260	265	270	
ata cat tca ttg ctt tgc ttg gtt gat aca ctt caa aat ctt gga gta			864
Ile His Ser Leu Leu Cys Leu Val Asp Thr Leu Gln Asn Leu Gly Val			

[0007]

275	280	285	
cat cgg cat ttt aaa tca gaa ata aag aaa gct cta gat gaa ata tac			912
His Arg His Phe Lys Ser Glu Ile Lys Lys Ala Leu Asp Glu Ile Tyr			
290	295	300	
agg cta tgg caa caa aag aat gaa caa att ttc tca aat gtc acc cat			960
Arg Leu Trp Gln Gln Lys Asn Glu Gln Ile Phe Ser Asn Val Thr His			
305	310	315	320
tgt gct atg gct ttt cga ctt cta agg atg agc tac tat gat gtc tcc			1008
Cys Ala Met Ala Phe Arg Leu Leu Arg Met Ser Tyr Tyr Asp Val Ser			
325	330	335	
tca gat gaa cta gca gaa ttt gtg gat gaa gaa cat ttt ttt gca ata			1056
Ser Asp Glu Leu Ala Glu Phe Val Asp Glu Glu His Phe Phe Ala Ile			
340	345	350	
agt ggg aaa tat aca agt cat gtt gaa att ctt gaa ctc cac aaa gca			1104
Ser Gly Lys Tyr Thr Ser His Val Glu Ile Leu Glu Leu His Lys Ala			
355	360	365	
tca caa ttg gct att gat cat gag aaa gat gac att ttg gat aag att			1152
Ser Gln Leu Ala Ile Asp His Glu Lys Asp Asp Ile Leu Asp Lys Ile			
370	375	380	
aac aat tgg aca aga aca ttt atg gag caa aaa ctc tta aac aat ggc			1200
Asn Asn Trp Thr Arg Thr Phe Met Glu Gln Lys Leu Leu Asn Asn Gly			
385	390	395	400
ttc ata gat agg atg tca aaa aag gag gtg gaa ctt gct ttg agg aag			1248
Phe Ile Asp Arg Met Ser Lys Lys Glu Val Glu Leu Ala Leu Arg Lys			
405	410	415	
ttt tat acc ata tct gat cta gca gaa aat aga aga tgt ata aag tca			1296
Phe Tyr Thr Ile Ser Asp Leu Ala Glu Asn Arg Arg Cys Ile Lys Ser			
420	425	430	
tac gaa gag aac aat ttt aaa atc tta aaa gca gct tat agg tca cct			1344
Tyr Glu Glu Asn Asn Phe Lys Ile Leu Lys Ala Ala Tyr Arg Ser Pro			
435	440	445	
aac att tac aat aag gac ttg ttt ata ttt tca ata cgc aac ttt gaa			1392
Asn Ile Tyr Asn Lys Asp Leu Phe Ile Phe Ser Ile Arg Asn Phe Glu			

[0008]

450	455	460	
tta tgc caa gct caa cac caa gaa gaa ctt caa caa ttc aag agg tgg			1440
Leu Cys Gln Ala Gln His Gln Glu Glu Leu Gln Gln Phe Lys Arg Trp			
465	470	475	480
ttt gaa gat tat aga ttg gac caa ctc gga att gca gaa cga tat ata			1488
Phe Glu Asp Tyr Arg Leu Asp Gln Leu Gly Ile Ala Glu Arg Tyr Ile			
	485	490	495
cat gat act tac tta tgt gct gtt att gtt gtc ccc gag cct gaa tta			1536
His Asp Thr Tyr Leu Cys Ala Val Ile Val Val Pro Glu Pro Glu Leu			
	500	505	510
tcc gat gct cgt ctc ttg tac gcg aaa tac gtc ttg ctc ctg act att			1584
Ser Asp Ala Arg Leu Leu Tyr Ala Lys Tyr Val Leu Leu Leu Thr Ile			
	515	520	525
gtc gat gat cag ttc gac agt ttt gca tct aca gat gaa tgt ctc aac			1632
Val Asp Asp Gln Phe Asp Ser Phe Ala Ser Thr Asp Glu Cys Leu Asn			
	530	535	540
atc att gaa tta gta gaa agg tgg gat gac tat gca agt gta ggt tat			1680
Ile Ile Glu Leu Val Glu Arg Trp Asp Asp Tyr Ala Ser Val Gly Tyr			
	545	550	555
aaa tct gag aag gtt aaa gtt ttc ttt tca act ttg tac aaa tca ata			1728
Lys Ser Glu Lys Val Lys Val Phe Phe Ser Thr Leu Tyr Lys Ser Ile			
	565	570	575
gag gag ctt gta aca att gct gaa att aaa caa gga cga tct gtc aaa			1776
Glu Glu Leu Val Thr Ile Ala Glu Ile Lys Gln Gly Arg Ser Val Lys			
	580	585	590
aat cac ctt ctt aat ttg tgg ctt gaa tlg glg aag ttg atg ttg atg			1824
Asn His Leu Leu Asn Leu Trp Leu Glu Leu Val Lys Leu Met Leu Met			
	595	600	605
gaa cga gta gag tgg ttt tct ggc aag aca atc cca agc ata gaa gag			1872
Glu Arg Val Glu Trp Phe Ser Gly Lys Thr Ile Pro Ser Ile Glu Glu			
	610	615	620
tat ttg tat gtt aca tct ata aca ttt ggt gca aga ttg att cct ctc			1920
Tyr Leu Tyr Val Thr Ser Ile Thr Phe Gly Ala Arg Leu Ile Pro Leu			

[0009]

625	630	635	640	
aca aca caa tat ttt ctt gga ata aaa ata tcc gaa gat att tta gaa				1968
Thr Thr Gln Tyr Phe Leu Gly Ile Lys Ile Ser Glu Asp Ile Leu Glu				
	645	650	655	
agt gat gaa ata tat ggt tta tgc aac tgt acc ggt aga gtc ctt cga				2016
Ser Asp Glu Ile Tyr Gly Leu Cys Asn Cys Thr Gly Arg Val Leu Arg				
	660	665	670	
atc ctt aat gat tta caa gat tcc aag aaa gaa caa aag gag gac tca				2064
Ile Leu Asn Asp Leu Gln Asp Ser Lys Lys Glu Gln Lys Glu Asp Ser				
	675	680	685	
gta act ata gtc aca tta cta atg aaa agt atg tct gag gaa gaa gct				2112
Val Thr Ile Val Thr Leu Leu Met Lys Ser Met Ser Glu Glu Glu Ala				
	690	695	700	
ata atg aag ata aag gaa atc ttg gaa atg aat aga gag tta ttg				2160
Ile Met Lys Ile Lys Glu Ile Leu Glu Met Asn Arg Arg Glu Leu Leu				
	705	710	715	720
aaa atg gtt tta gtt caa aaa aag gga agc caa ttg cct caa ata tgc				2208
Lys Met Val Leu Val Gln Lys Lys Gly Ser Gln Leu Pro Gln Ile Cys				
	725	730	735	
aaa gat ata ttt tgg agg aca agc aac tgg gct gat ttc att tat tta				2256
Lys Asp Ile Phe Trp Arg Thr Ser Asn Trp Ala Asp Phe Ile Tyr Leu				
	740	745	750	
caa act gat gga tat aga att gca gag gaa atg aag aat cac att gat				2304
Gln Thr Asp Gly Tyr Arg Ile Ala Glu Glu Met Lys Asn His Ile Asp				
	755	760	765	
gaa gtc ttt tac aaa cca ctc aat cat taa				2334
Glu Val Phe Tyr Lys Pro Leu Asn His				
	770	775		

<210> 4

<211> 777

<212> PRT

<213> Solanum habrochaites

[0010]

<400> 4

Met Ile Val Gly Tyr Arg Ser Thr Ile Ile Thr Leu Ser His Pro Lys
1 5 10 15

Leu Gly Asn Gly Lys Thr Ile Ser Ser Asn Ala Ile Phe Gln Arg Ser
20 25 30

Cys Arg Val Arg Cys Ser His Ser Thr Pro Ser Ser Met Asn Gly Phe
35 40 45

Glu Asp Ala Arg Asp Arg Ile Arg Glu Ser Phe Gly Lys Val Glu Leu
50 55 60

Ser Pro Ser Ser Tyr Asp Thr Ala Trp Val Ala Met Val Pro Ser Lys
65 70 75 80

His Ser Leu Asn Glu Pro Cys Phe Pro Gln Cys Leu Asp Trp Ile Ile
85 90 95

Glu Asn Gln Arg Glu Asp Gly Ser Trp Gly Leu Asn Pro Ser His Pro
100 105 110

Leu Leu Leu Lys Asp Ser Leu Ser Ser Thr Leu Ala Cys Leu Leu Ala
115 120 125

Leu Thr Lys Trp Arg Val Gly Asp Glu Gln Ile Lys Arg Gly Leu Gly
130 135 140

Phe Ile Glu Thr Gln Ser Trp Ala Ile Asp Asn Lys Asp Gln Ile Ser
145 150 155 160

Pro Leu Gly Phe Glu Ile Ile Phe Pro Ser Met Ile Lys Ser Ala Glu
165 170 175

[0011]

Lys Leu Asn Leu Asn Leu Ala Ile Asn Lys Arg Asp Ser Thr Ile Lys
180 185 190

Arg Ala Leu Gln Asn Glu Phe Thr Arg Asn Ile Glu Tyr Met Ser Glu
195 200 205

Gly Val Gly Glu Leu Cys Asp Trp Lys Glu Ile Ile Lys Leu His Gln
210 215 220

Arg Gln Asn Gly Ser Leu Phe Asp Ser Pro Ala Thr Thr Ala Ala Ala
225 230 235 240

Leu Ile Tyr His Gln His Asp Lys Lys Cys Tyr Glu Tyr Leu Asn Ser
245 250 255

Ile Leu Gln Gln His Lys Asn Trp Val Pro Thr Met Tyr Pro Thr Lys
260 265 270

Ile His Ser Leu Leu Cys Leu Val Asp Thr Leu Gln Asn Leu Gly Val
275 280 285

His Arg His Phe Lys Ser Glu Ile Lys Lys Ala Leu Asp Glu Ile Tyr
290 295 300

Arg Leu Trp Gln Gln Lys Asn Glu Gln Ile Phe Ser Asn Val Thr His
305 310 315 320

Cys Ala Met Ala Phe Arg Leu Leu Arg Met Ser Tyr Tyr Asp Val Ser
325 330 335

Ser Asp Glu Leu Ala Glu Phe Val Asp Glu Glu His Phe Phe Ala Ile
340 345 350

[0012]

Ser Gly Lys Tyr Thr Ser His Val Glu Ile Leu Glu Leu His Lys Ala
355 360 365

Ser Gln Leu Ala Ile Asp His Glu Lys Asp Asp Ile Leu Asp Lys Ile
370 375 380

Asn Asn Trp Thr Arg Thr Phe Met Glu Gln Lys Leu Leu Asn Asn Gly
385 390 395 400

Phe Ile Asp Arg Met Ser Lys Lys Glu Val Glu Leu Ala Leu Arg Lys
405 410 415

Phe Tyr Thr Ile Ser Asp Leu Ala Glu Asn Arg Arg Cys Ile Lys Ser
420 425 430

Tyr Glu Glu Asn Asn Phe Lys Ile Leu Lys Ala Ala Tyr Arg Ser Pro
435 440 445

Asn Ile Tyr Asn Lys Asp Leu Phe Ile Phe Ser Ile Arg Asn Phe Glu
450 455 460

Leu Cys Gln Ala Gln His Gln Glu Glu Leu Gln Gln Phe Lys Arg Trp
465 470 475 480

Phe Glu Asp Tyr Arg Leu Asp Gln Leu Gly Ile Ala Glu Arg Tyr Ile
485 490 495

His Asp Thr Tyr Leu Cys Ala Val Ile Val Val Pro Glu Pro Glu Leu
500 505 510

Ser Asp Ala Arg Leu Leu Tyr Ala Lys Tyr Val Leu Leu Leu Thr Ile
515 520 525

[0013]

Val Asp Asp Gln Phe Asp Ser Phe Ala Ser Thr Asp Glu Cys Leu Asn
530 535 540

Ile Ile Glu Leu Val Glu Arg Trp Asp Asp Tyr Ala Ser Val Gly Tyr
545 550 555 560

Lys Ser Glu Lys Val Lys Val Phe Phe Ser Thr Leu Tyr Lys Ser Ile
565 570 575

Glu Glu Leu Val Thr Ile Ala Glu Ile Lys Gln Gly Arg Ser Val Lys
580 585 590

Asn His Leu Leu Asn Leu Trp Leu Glu Leu Val Lys Leu Met Leu Met
595 600 605

Glu Arg Val Glu Trp Phe Ser Gly Lys Thr Ile Pro Ser Ile Glu Glu
610 615 620

Tyr Leu Tyr Val Thr Ser Ile Thr Phe Gly Ala Arg Leu Ile Pro Leu
625 630 635 640

Thr Thr Gln Tyr Phe Leu Gly Ile Lys Ile Ser Glu Asp Ile Leu Glu
645 650 655

Ser Asp Glu Ile Tyr Gly Leu Cys Asn Cys Thr Gly Arg Val Leu Arg
660 665 670

Ile Leu Asn Asp Leu Gln Asp Ser Lys Lys Glu Gln Lys Glu Asp Ser
675 680 685

Val Thr Ile Val Thr Leu Leu Met Lys Ser Met Ser Glu Glu Glu Ala
690 695 700

[0014]

Ile Met Lys Ile Lys Glu Ile Leu Glu Met Asn Arg Arg Glu Leu Leu
705 710 715 720

Lys Met Val Leu Val Gln Lys Lys Gly Ser Gln Leu Pro Gln Ile Cys
725 730 735

Lys Asp Ile Phe Trp Arg Thr Ser Asn Trp Ala Asp Phe Ile Tyr Leu
740 745 750

Gln Thr Asp Gly Tyr Arg Ile Ala Glu Glu Met Lys Asn His Ile Asp
755 760 765

Glu Val Phe Tyr Lys Pro Leu Asn His
770 775

<210> 5
<211> 31
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400> 5
accatgggtt ctttggttct tcaatgttg a 31

<210> 6
<211> 34
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400> 6
actcgagata tgttgtcca caaaacgtc tatg 34

[0015]

<210> 7
<211> 31
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400> 7
tccatggtag ttggctatag aagcacaatc a

31

<210> 8
<211> 34
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400> 8
tctcgagcat aaattcaaat tgagggatta atga

34

<210> 9
<211> 31
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400> 9
accatgggtt ctttggttct tcaatgttgg a

31

<210> 10
<211> 34
<212> DNA
<213> artificial sequence

<220>

[0016]

<223> primer

<400> 10

actcgagata tgtgtgtcca ccaaaacgtc tatg

34

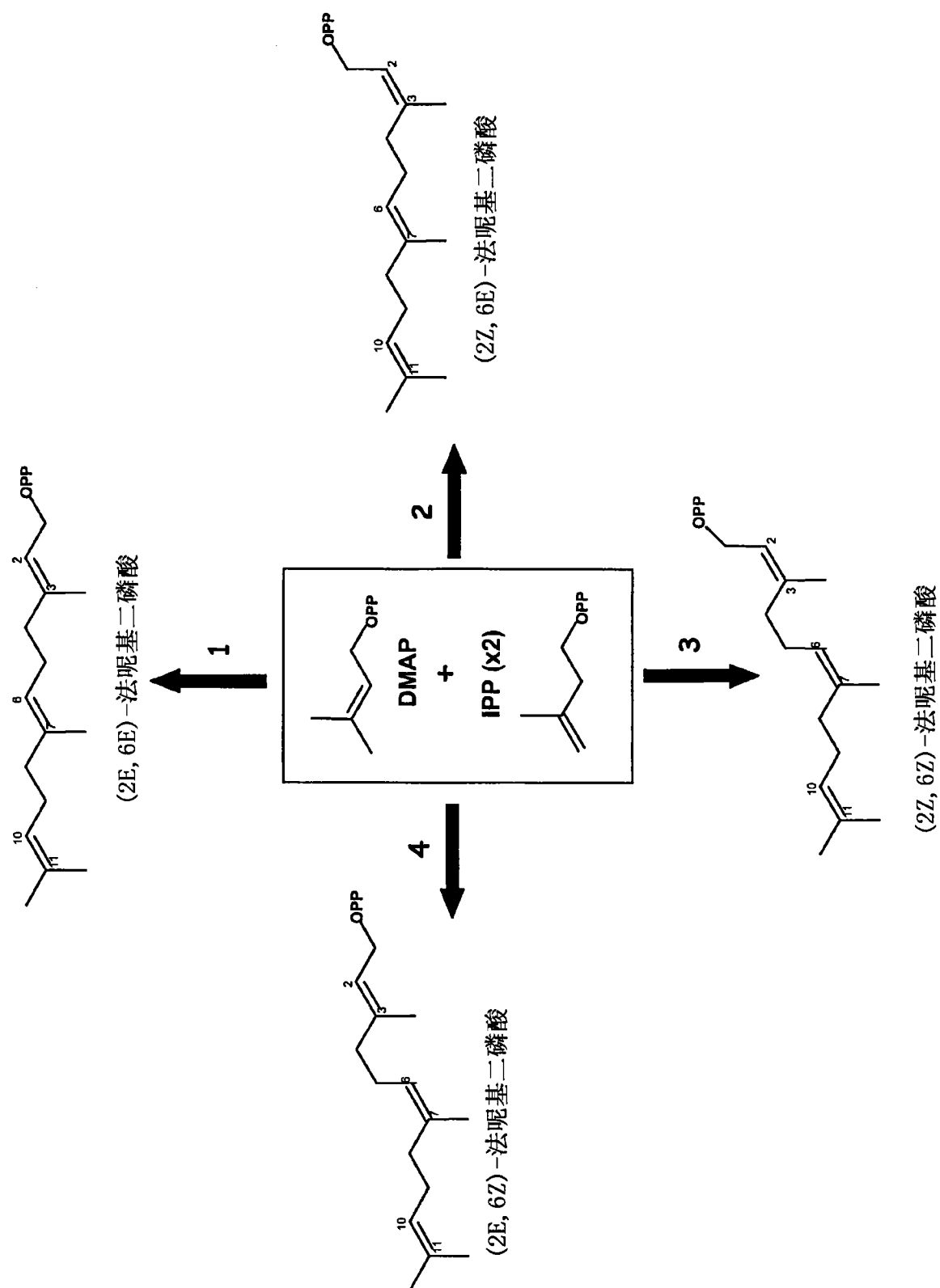


图 1



图 2A

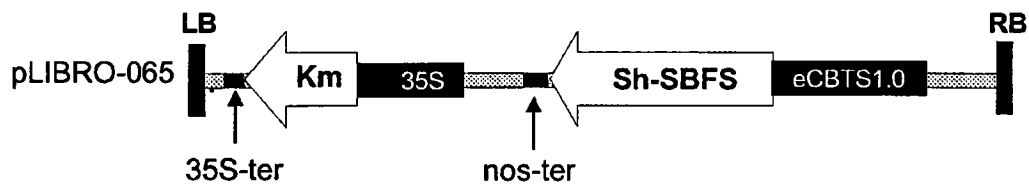


图 2B

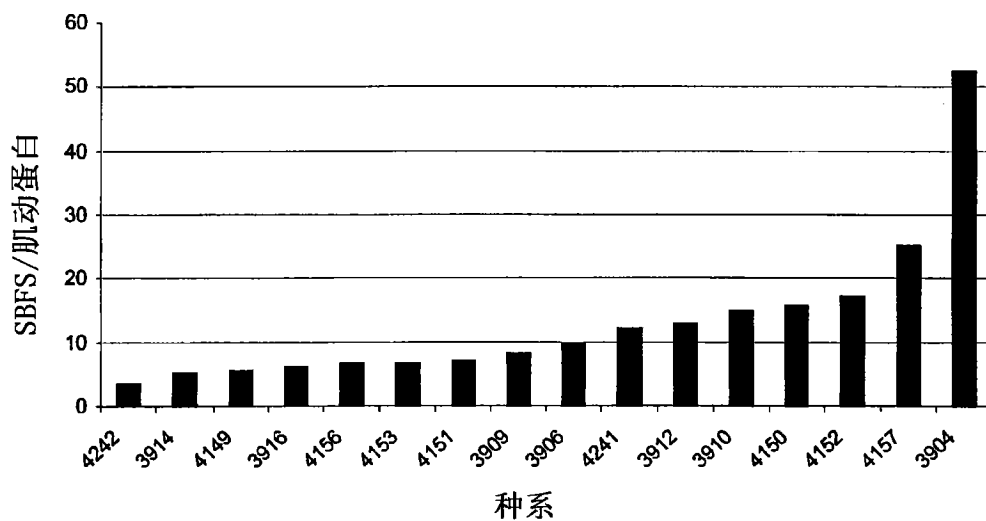


图 3A

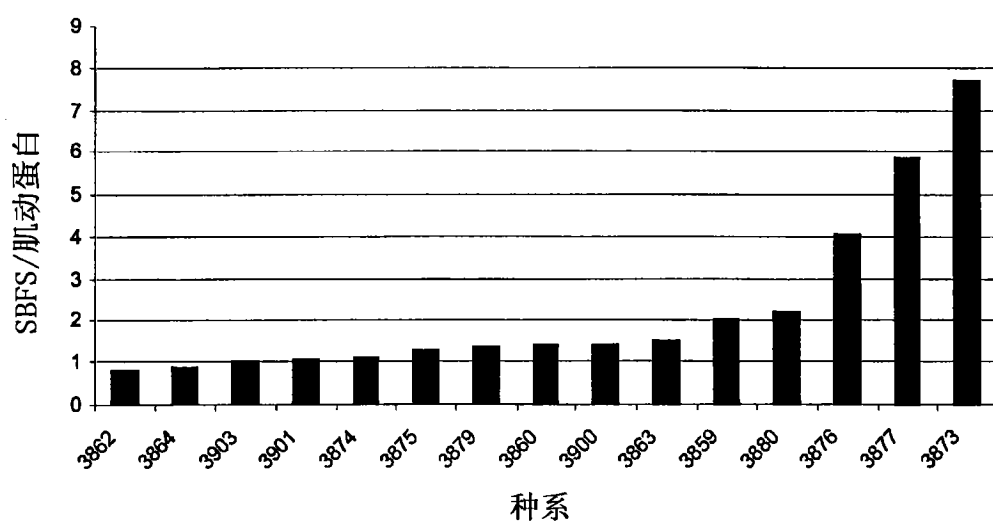


图 3B

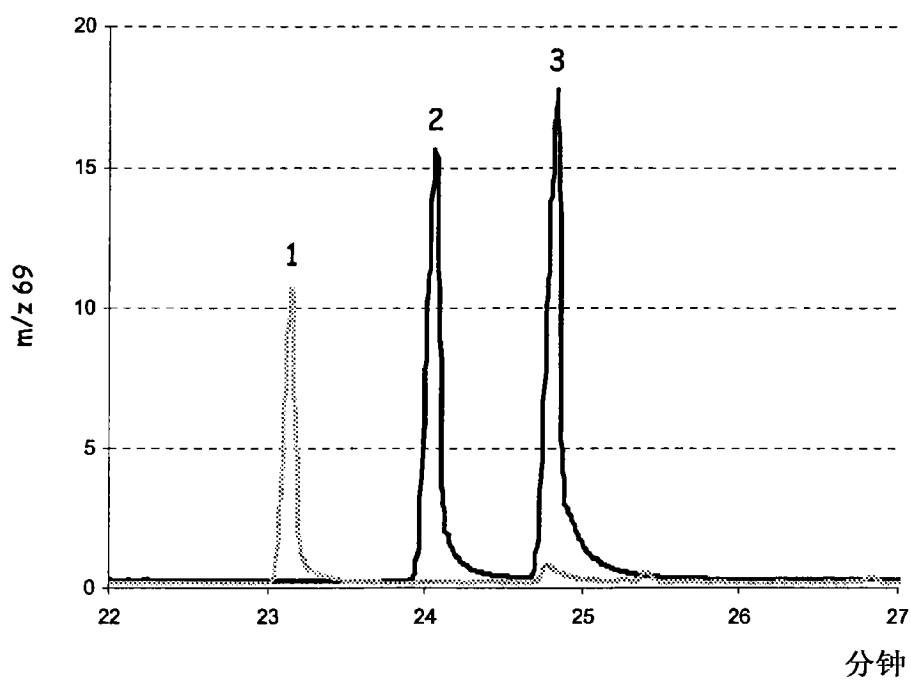


图 4A

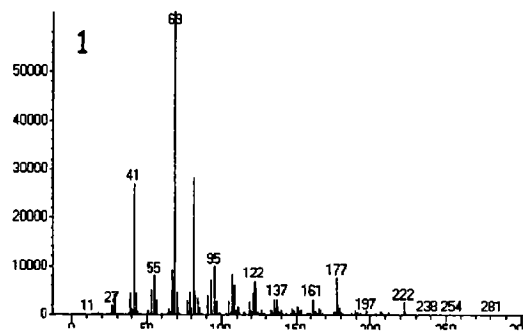


图 4B

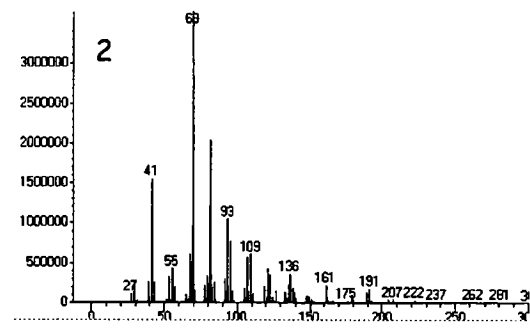


图 4C

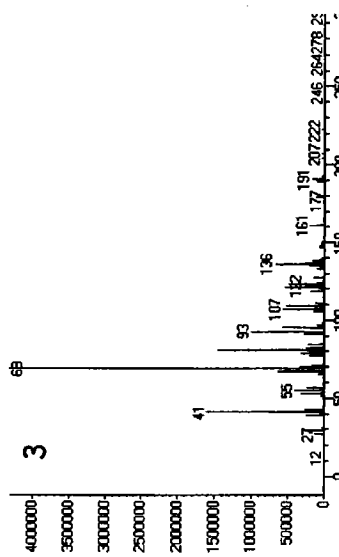


图 4D

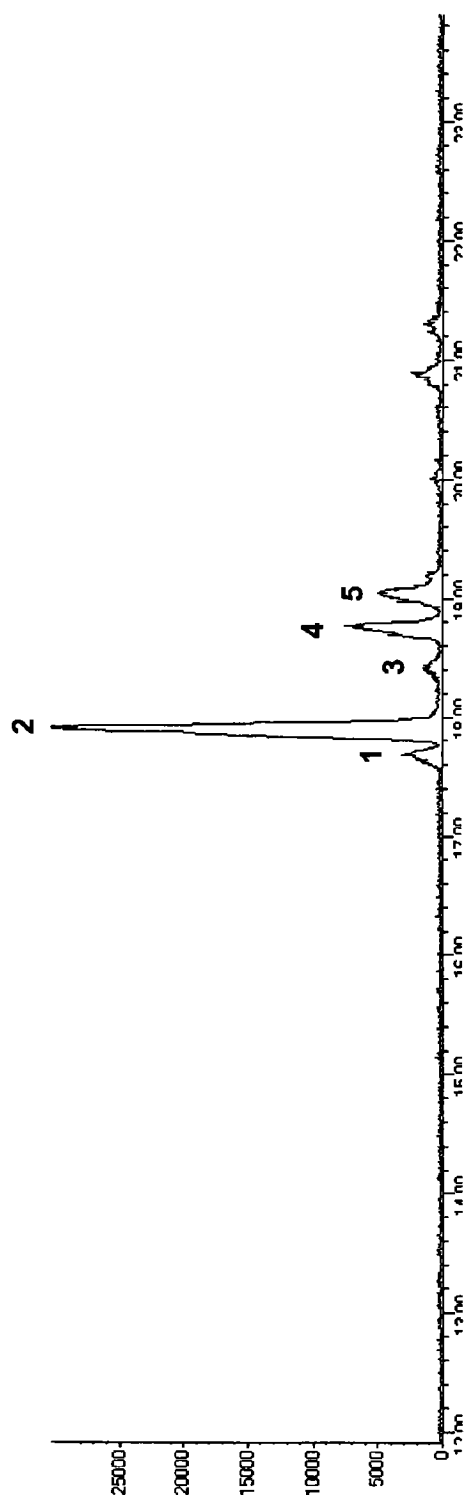


图 5A

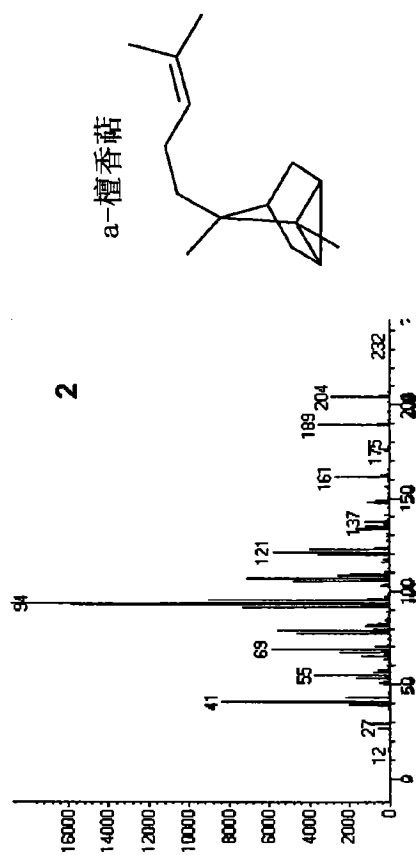
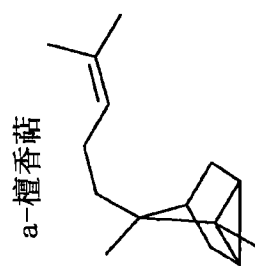


图 5B



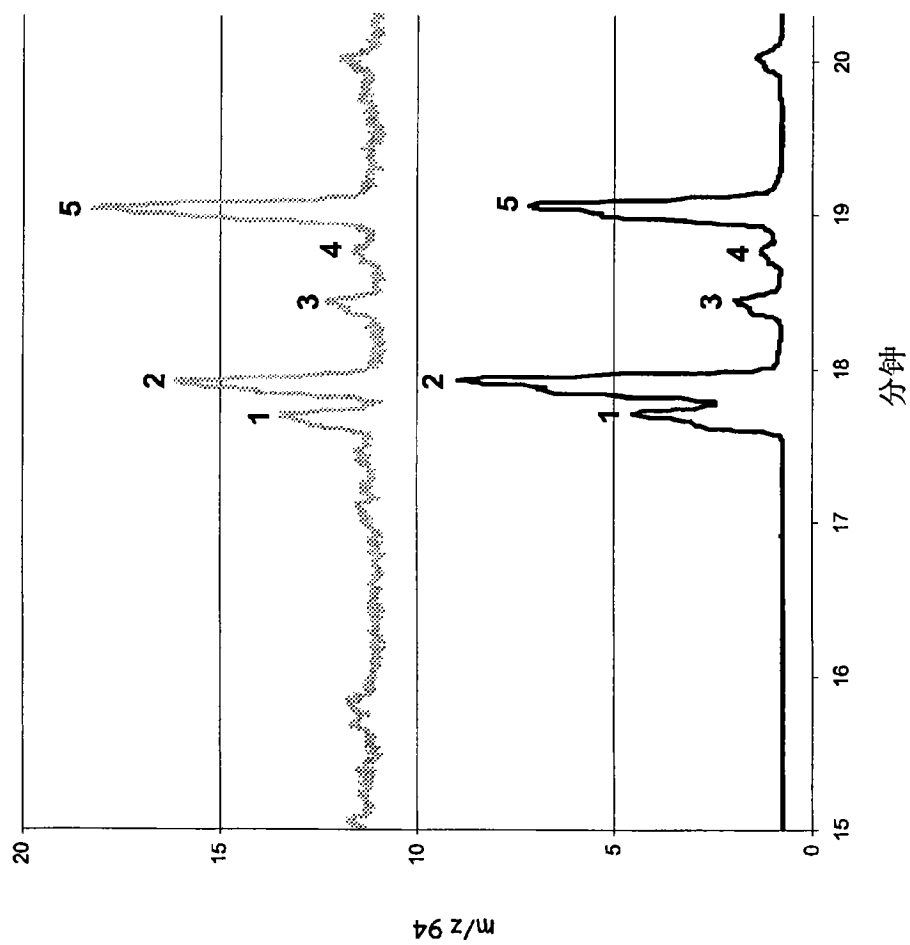


图 9

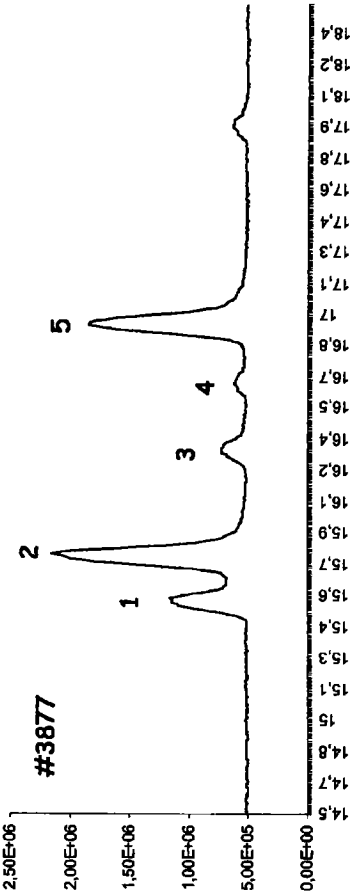


图 7A

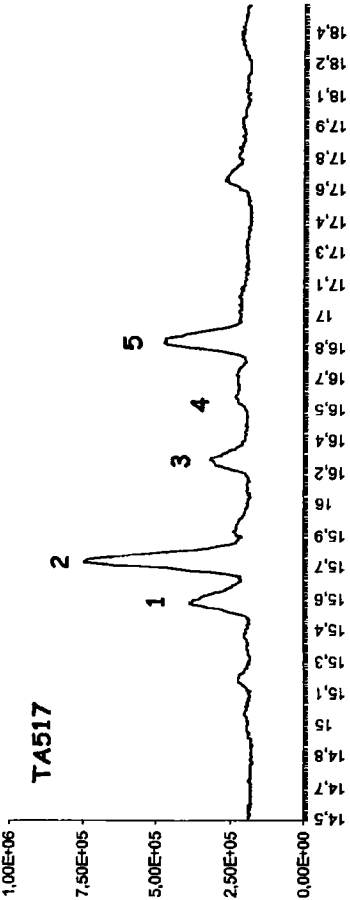


图 7B

