



(12) 发明专利

(10) 授权公告号 CN 102725410 B

(45) 授权公告日 2016. 03. 16

(21) 申请号 201080038885. 4

(51) Int. Cl.

(22) 申请日 2010. 06. 25

C12N 15/29(2006. 01)

(30) 优先权数据

2009903016 2009. 06. 29 AU

(56) 对比文件

CN 102245778 A, 2011. 11. 16,

JONES, C. G. 等. EU798692.

(85) PCT国际申请进入国家阶段日

2012. 02. 29

《GenBank》. 2008,

(86) PCT国际申请的申请数据

PCT/AU2010/000802 2010. 06. 25

审查员 樊颖

(87) PCT国际申请的公布数据

W02011/000026 EN 2011. 01. 06

(73) 专利权人 西澳大学

地址 澳大利亚西澳洲

专利权人 森林产品委员会

英属哥伦比亚大学

(72) 发明人 K·G·楚拉克 C·琼斯

J·莫尼奥迪斯 J·博尔曼

(74) 专利代理机构 永新专利商标代理有限公司

72002

代理人 左路 林晓红

权利要求书3页 说明书36页

序列表21页 附图22页

(54) 发明名称

来自檀香属的萜烯合酶

(57) 摘要

分离的核酸分子,其编码萜烯合酶并选自:a) 包含 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示核苷酸序列的核酸分子;b) 核酸分子,其为(a)的片段;c) 核酸分子,其包含与(a)或(b)互补的核苷酸序列;和 d) 核酸分子,其编码萜烯合酶并与(a)-(c)的任一核酸分子具有至少或至少约60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%的相同性;其中所述核酸分子编码萜烯合酶。

1. 分离的核酸分子,其编码檀香属 (*Santalum* L.) 物种萜烯合酶,所述萜烯合酶的氨基酸序列是 (i) SEQ ID NO:2 所示的氨基酸残基序列,或 (ii) SEQ ID NO:2 的连续片段,其含有 SEQ ID NO:2 的氨基酸残基 32-426,

其中所述核酸分子编码催化 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯中的一或多种的产生的萜烯合酶。

2. 权利要求 1 的分离的核酸分子,其中所述核酸分子的核苷酸序列示于 SEQ ID NO:1。

3. 分离的核酸分子,其编码檀香属物种萜烯合酶,所述萜烯合酶的氨基酸序列是 (i) SEQ ID NO:2 所示的氨基酸残基序列,或 (ii) SEQ ID NO:2 的连续片段,其含有 SEQ ID NO:2 的氨基酸残基 32-426,其中:

所编码的合酶在 SEQ ID NO:2 的残基 143 处包含替换脯氨酸的丝氨酸;且

所述核酸分子编码催化 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯中的一或多种的产生的萜烯合酶。

4. 分离的核酸分子,其编码催化 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯中的一或多种的产生的萜烯合酶,其中所述核酸分子编码多肽,所述多肽的残基序列是 SEQ ID NO:4 或 6。

5. 权利要求 1-4 中任一项的核酸分子编码的分离的萜烯合酶,其同时催化 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯的产生。

6. 权利要求 1-4 中任一项的核酸分子编码的分离的萜烯合酶。

7. 制备至少一种萜烯的方法,包括:

a) 在产生萜烯的条件下培养细胞,其中:

所述细胞编码萜烯合酶或其催化活性片段,其中:

所述合酶是檀香属萜烯合酶;

所述合酶对所述细胞是异源的;

所述萜烯合酶的氨基酸序列是 (i) SEQ ID NO:2 所示氨基酸残基序列,或 (ii) SEQ ID NO:2 的连续片段,其含有 SEQ ID NO:2 的氨基酸残基 32-426;和

所述细胞表达无环焦磷酸萜烯前体;和

b) 任选地,分离所述萜烯。

8. 用于制备具有萜烯合酶活性的变体多肽的方法,包括:

a) 选择权利要求 1-4 中任一项的核酸分子;

b) 修饰所述核酸分子以产生突变核酸分子;

c) 用所述突变核酸分子转化宿主细胞并表达所编码的萜烯合酶;和

d) 筛选所述萜烯合酶以选择与起始分子相比具有修饰的特性的任何萜烯合酶,其中所述修饰的特性是酶促活性。

9. 权利要求 8 的方法,其中重复步骤 a)-d) 直到产生具有所需特性的合酶。

10. 权利要求 8 的方法,其中所述修饰的核酸分子通过 DNA 改组、随机诱变或定点诱变产生。

11. 用于检测样品中萜烯合酶的方法,包括以下步骤:

a) 使含有权利要求 6 的萜烯合酶的样品与特异性结合所述萜烯合酶的抗体在使得包含所述抗体和萜烯合酶的反应复合物形成的条件下相接触;和

b) 检测样品中包含所述抗体和权利要求 6 的萜烯合酶的反应复合物的形成,其中检测出所述反应复合物的形成表明在样品中存在萜烯合酶。

12. 权利要求 11 的方法,其进一步包括步骤:c) 评估所形成的反应复合物的量,从而确定生物学样品中萜烯合酶的量。

13. 权利要求 11 的方法,其中所述样品来自檀香树。

14. 权利要求 13 的方法,其中:在来自两种不同檀香树的至少两个样品上实施所述方法;测定每个样品中的萜烯合酶或编码萜烯合酶的核酸的量;选择含有最高量的萜烯合酶或编码萜烯合酶的核酸的样品所来源的檀香树;并对所选择的檀香树进行选择育种或收获以获取檀香油。

15. 载体,其包含权利要求 1-4 中任一项的核酸分子。

16. 权利要求 15 的载体,其中所述载体为原核载体、病毒载体或真核载体。

17. 权利要求 15 的载体,其为表达载体。

18. 细胞,其包含权利要求 15 的载体。

19. 权利要求 18 的细胞,其为原核细胞或真核细胞。

20. 权利要求 18 的细胞,其为细菌细胞。

21. 权利要求 19 的细胞,其中所述真核细胞为酵母细胞或植物细胞。

22. 权利要求 21 的细胞,其中所述植物细胞来自茄科 (Solanaceae) 或唇形科 (Lamiaceae) 植物。

23. 用于制备萜烯合酶的方法,其包括步骤:在使所编码的萜烯合酶表达的条件下培养权利要求 18 的细胞。

24. 权利要求 23 的方法,其进一步包括纯化所述萜烯合酶。

25. 通过权利要求 23 的方法产生的萜烯合酶。

26. 制备萜烯的方法,其包括:

a) 使无环焦磷酸萜烯前体与权利要求 5 的萜烯合酶相接触,和,

b) 任选地,分离在步骤 (a) 中产生的萜烯,其中所述萜烯是 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯中的一或多种。

27. 权利要求 26 的方法,其中:所述萜烯合酶在细胞异源表达;所述无环焦磷酸萜烯前体与所述萜烯合酶表达于同一细胞中;和所述与无环焦磷酸前体接触的步骤发生在所述细胞中。

28. 权利要求 7 的方法,其中所述无环焦磷酸萜烯前体选自焦磷酸香叶酯、焦磷酸金合欢酯和焦磷酸香叶基香叶酯。

29. 权利要求 7 的方法,其中所产生的萜烯包括 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯,它们彼此之间的比例使得在进一步将所述萜烯处理为醇时,檀香醇和香柠檬醇的产量会与印度檀香天然油相似。

30. 权利要求 29 的方法,其中所述萜烯合酶由 SEQ ID NO:2 所示的氨基酸序列组成。

31. 权利要求 29 的方法,其中所述萜烯选自 (+)-表- β -檀香烯、(-)- β -檀香烯、(+)- β -檀香烯、(+)- α -檀香烯和 (-)- α -檀香烯。

32. 权利要求 7 的方法,其中产生萜烯的混合物。

33. 权利要求 32 的方法,其中所述萜烯被处理为醇并且所述醇彼此之间的比例为

25-65%的 α -檀香醇、1-20%的 α -反式-香柠檬醇、1-15%的表- β -檀香醇、20-50%的 β -檀香醇。

34. 权利要求 7 的方法,其进一步包括将获得的萜烯处理为醇的步骤。

35. 权利要求 34 的方法,其中所述醇选自 α -檀香醇、 β -檀香醇、 α -反式-香柠檬醇和表- β -檀香醇。

36. 权利要求 34 的方法,其中产生至少 4 种醇。

来自檀香属的萜烯合酶

发明领域

[0001] 本发明涉及新的萜烯合酶。本发明还涉及编码萜烯合酶的核酸、用于制备变体萜烯合酶的方法和表达本发明多肽的宿主生物体。本发明还包括用于制备萜烯合酶的方法和用于制备萜类化合物（如萜烯）的方法。

背景技术

[0002] 以下技术背景的讨论仅为了便于理解本发明。该讨论并非确认或承认，截至本申请的优先权日，其所涉及的任何材料是或曾是普通常识的一部分。

[0003] 檀香（印度檀香 *Santalum album*）（檀香科 *Santalaceae*）是具有巨大的经济价值的小型半寄生热带树木，其被发现生长于印度南部、斯里兰卡、印尼东部和澳大利亚北部。由于其细粒性、高密度和出色的雕刻特性，檀香木材受到高度追捧。檀香木材含有树脂和精油，尤其是檀香醇、檀香烯以及数十种其他微量倍半萜。这些化学物质提供了独特的檀香香味。香木通常被磨碎并蒸汽蒸馏，其挥发油用作许多高端香水的固定剂。

[0004] 几个世纪的过度开采已导致檀香天然林消亡。正在整个澳大利亚北部建立大型种植园，以满足需求并保护剩余的储备。檀香心材中含有高达 6%（干重）的倍半萜油，主要是 α -和 β -檀香醇， α -反式-香柠檬醇和表- β -檀香醇，以及倍半萜烯 α -和 β -檀香烯， α -香柠檬烯和表- β -檀香烯， β -红没药烯， α -、 β -和 γ -姜黄烯。即使是在近乎相同的生长条件下，产于树的心材油的量相差很大。这种产量差异的原因还不是很清楚，但可能是遗传和环境因素两者的结果。

[0005] 关于檀香中倍半萜类化合物（如倍半萜）的生物合成或其精油产生如何调控所知甚少。

[0006] 本发明解决了本领域对用于产生萜烯的方法的需要，所述萜烯与檀香所产生的萜烯类似。

[0007] 发明概述

[0008] 在一个实施方案中，本发明提供了分离的核酸分子，其编码萜烯合酶并且选自：

[0009] a) 包含 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示的核苷酸序列的核酸分子；

[0010] b) 核酸分子，其为 (a) 的片段；

[0011] c) 核酸分子，其包含与 (a) 或 (b) 互补的核苷酸序列；和

[0012] d) 核酸分子，其编码萜烯合酶并与 (a)-(c) 中任一核酸分子具有至少或至少约 60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99% 的相同性；

[0013] 其中所述核酸分子编码萜烯合酶。

[0014] 其它的实施方案包括：本发明的核酸所编码的多肽；包含本发明核酸的宿主细胞；非人生物体，其被修改以含有本发明的核酸；以及产生多肽的方法，其包括培养本发明的宿主细胞。

[0015] 在一个实施方案中，本发明提供了制备至少一种萜烯合酶的方法，其包括在有助

于产生所述至少一种萜烯合酶的条件下培养宿主细胞,所述宿主细胞被修改以含有至少一种核酸序列。

[0016] 本发明还提供分离的萜烯合酶,其中所述萜烯合酶为檀香属 (*Santalum*) 萜烯合酶;并且所述萜烯合酶催化檀香烯的产生。

[0017] 在另一个实施方案中,本发明提供分离的萜烯合酶,其包括:

[0018] a) SEQ ID NO:2、SEQ ID NO:4 或 SEQ ID NO:6 所示的氨基酸序列;

[0019] b) 由权利要求 1-5 中任一项的核酸分子所编码的氨基酸序列;

[0020] c) 氨基酸序列,其与 SEQ ID NO:2 所示的氨基酸序列具有至少或至少约 60%、65%、70%、75%、80%、85%、90%、95%或更高相同性;或

[0021] d) (a)、(b) 或 (c) 的片段;

[0022] 其中所述萜烯合酶催化萜烯产生。

[0023] 本发明还提供萜烯合酶,其中所述萜烯合酶同时催化 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯的产生。

[0024] 本发明还提供用于检测样品中萜烯合酶多肽或核酸的存在的方法。

[0025] 在另外的实施方案中,本发明提供了产生萜烯合酶的方法,所述方法包括以下步骤:

[0026] a) 选择宿主生物体和/或细胞,其不表达具有 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸分子;

[0027] b) 用具有 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸分子转化所述生物体;和

[0028] c) 在有助于所述核酸编码的萜烯合酶产生的条件下培养所述生物体。

[0029] 在另外的实施方案中,本发明提供了产生萜烯合酶的方法,所述方法包括以下步骤:

[0030] a) 选择宿主生物体和/或细胞,其不表达具有 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸分子;

[0031] b) 用 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸分子以较高的量转化所述生物体;和

[0032] c) 在有助于所述核酸编码的萜烯合酶产生的条件下培养所述生物体。

[0033] 本发明还提供了制备萜烯的方法,包括:

[0034] a) 使无环焦磷酸萜烯前体与本发明的萜烯合酶接触,和,

[0035] b) 任选地,分离在步骤 (a) 中产生的萜烯。

[0036] 优选地,所述方法实施于在细胞中异源表达的萜烯合酶上;其中无环焦磷酸萜烯前体与所述萜烯合酶在同一细胞中表达;并且接触无环焦磷酸萜烯前体的步骤发生于所述细胞中。更优选地,所述至少一种萜烯选自 (+)-表- β -檀香烯、(-)- β -檀香烯、(+)- β -檀香烯、(+)- α -檀香烯、(-)- α -檀香烯、顺式- α -香柠檬烯、反式- α -香柠檬烯,反式- β -香柠檬烯和顺式- β -香柠檬烯。

[0037] 本发明的萜烯合酶产生的萜烯可进一步加工为醇,优选为 α -檀香醇, β -檀香醇, α -反式-香柠檬醇和/或表- β -檀香醇。

[0038] 附图简述

[0039] 图1:印度檀香 (*S. album*) 天然檀香油的色谱图。4个主峰是 α -檀香烯、E- α -香柠檬烯、表- β -檀香烯和 β -檀香烯。

[0040] 图2:与FPP温育后的SaSSy产物谱的GC图谱(trace),以及用GC-MS检测的来自与FPP温育后的SaSSy产物谱的 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯的质谱数据。

[0041] 图3:与FPP温育后的SaSSy、SauSSy和SspiSSy产物谱的叠加GC图谱,以及用GC-MS检测的来自与FPP温育后的SaSSy产物谱的 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯、顺式- β -金合欢烯和反式- β -金合欢烯的质谱数据。

[0042] 图4:本发明的SaSSy萜烯合酶的核酸序列。

[0043] 图5:本发明的SaSSy萜烯合酶的氨基酸序列。

[0044] 图6A和6B:萜烯合酶的Clustal比对:FB299123-125-香根草 (*Vetiveria zizanoides*) 萜烯合酶 (WO 2006134523)、AF484125-烟草 (*Nicotiana attenuata*) 5-表-马兜铃烯 (aristolochene) 合酶、AB438045-柠檬香桃 (*Backhousia citriodora*) 芳樟醇合酶、Santalene-本发明的来自檀香的SaSSy檀香烯合酶。

[0045] 图7:图6中比对的蛋白质的系统发生树。

[0046] 图8:28种不同的萜烯合酶基因的UPGMA比对树。参考代码如下:
1 SaSSy[印度檀香 (*Santalum album*)]; 2 SauSSy[新克里多尼亚檀香 (*Santalum austrocaledonicum*)]; 3 SspiSSy[澳洲檀香 (*Santalum spicatum*)]; 4 ACF24767.1 单萜合酶[印度檀香]; 5 SauMonoTPS1[新克里多尼亚檀香]; 6 SspiMonoTPS1[澳洲檀香]; 7 ACF24768.1 SaSesquiTPS1[印度檀香]; 8 SausSesquiTPS1[新克里多尼亚檀香]; 9 SspiSesquiTPS1[澳洲檀香]; 10 AAS79351.1(-)- α -松油醇合酶[葡萄 (*Vitis vinifera*)]; 11 BAG82825.1 芳樟醇合酶[柠檬香桃 (*Backhousia citriodora*)]; 12 AAV63788.1 α -姜烯合酶[罗勒 (*Ocimum basilicum*)]; 13 AAR99061.1(-)-大根香叶烯D合酶[毛果杨 (*Populus trichocarpa*) $\times$ 美洲黑杨 (*Populus deltoides*)]; 14 CAA06614.1 5-表-马兜铃烯[辣椒原变种 (*Capsicum annuum* var. *annuum*)]; 15 CAA77191.1(+)- δ -杜松烯合酶[亚洲棉 (*Gossypium arboreum*)]; 16 AA073863.1(+)-3-萆烯合酶[欧洲云杉 (*Picea abies*)]; 17 AAC05727.1 d-芹子烯合酶[大冷杉 (*Abies grandis*)]; 18 AAF61453.1 β -水芹烯合酶[大冷杉]; 19 AAC05728.1 γ -蛇麻烯合酶[大冷杉]; 20 AAS47691.1 LAS[欧洲云杉]; 21 AAC39443.1 ent-贝壳杉烯合酶[拟南芥 (*Arabidopsis thaliana*)]; 22 ADB55710.1(-)-ent-贝壳杉烯合酶[北美云杉 (*Picea sitchensis*)]; 23 AAM53944.1AF514287_1(+)-柠檬烯合酶1[柠檬 (*Citrus limon*)]; 24AAA86337.1vetispiradiene合酶[纯天仙子 (*Hyoscyamus muticus*)]; 25AAF61439.1 紫穗槐-4,11-二烯合酶[青蒿 (*Artemisia annua*)]; 26BAF02832.1 单萜合酶[蓝桉 (*Eucalyptus globulus*)]; 27 桉叶素合酶[希腊鼠尾草 (*Salvia fruticosa*)]; 28 香桉烯合酶[*Salvia pomifera*]

[0047] 图9:28种不同的萜烯合酶基因的Neighbor joining比对树。参考代码如下:
1SaSSy:SaSSy[印度檀香]; 2SauSSy:SauSSy[新克里多尼亚檀香]; 3SspiSSy:SspiSSy[澳洲檀香]; 4SaMono:ACF24767.1 单萜合酶[印度檀香]; 5SauMono:SauMonoTPS1[新克里多尼亚檀香]; 6SspiMono:SspiMonoTPS1[澳洲檀香]; 7SaSequiTPS1:ACF24768.1

SaSesquiTPS1[印度檀香];8SausSesquiTPS:SausSesquiTPS1[新克里多尼亚檀香];9SspiSesquiTPS:SspiSesquiTPS1[澳洲檀香];10Vvterp:AAS79351.1(-)- α - 松油醇合酶 [葡萄];11BcitLiNS:BAG82825.1 芳樟醇合酶 [柠檬香桃];12obasAZS:AAV63788.1 α - 姜烯合酶 [罗勒];13PtriGDS:AAR99061.1(-)- 大根香叶烯 D 合酶 [毛果杨 $\times$ 美洲黑杨];14CannEAS:CAA06614.15- ϵ - 马兜铃烯 [辣椒原变种];15GarbCDS CAA77191.1(+)- δ - 杜松烯合酶 [亚洲棉];16PabiCRS:AA073863.1(+)-3- 萜烯合酶 [欧洲云杉];17AgradSELS:AAC05727.1d- 芹子烯合酶 [大冷杉];18Argaphel:AAF61453.1 β - 水芹烯合酶 [大冷杉];19AragHUMS:AAC05728.1 γ - 蛇麻烯合酶 [大冷杉];20PabiLAS:AAS47691.1LAS[欧洲云杉];21AthaEKS:AAC39443.1ent- 贝壳杉烯合酶 [拟南芥];22PsitEKS:ADB55710.1(-)-ent- 贝壳杉烯合酶 [北美云杉];23ClimLIS:AAM53944.1AF514287_1(+)- 柠檬烯合酶 1 [柠檬];24HmutVTS:AAA86337.1vetispiradiene 合酶 [纯天仙子];25AannADS:AAF61439.1 紫穗槐 -4, 11- 二烯合酶 [青蒿];26EgloTPS:BAF02832.1 单萜合酶 [蓝桉];27DQ785794.1: 桉叶素合酶 [希腊鼠尾草];28DQ785793: 香桉烯合酶]。

[0048] 图 10a-10g :新克里多尼亚檀香、澳洲檀香和印度檀香的萜烯合酶核酸序列的比对。

[0049] 图 11a-c: 新克里多尼亚檀香、澳洲檀香和印度檀香的萜烯合酶蛋白质氨基酸序列的比对。

[0050] 图 12a-d: 新克里多尼亚檀香、澳洲檀香和印度檀香的萜烯合酶蛋白质氨基酸序列的比对,其与 DQ785793.1- 希腊鼠尾草桉叶素合酶和 DQ785794.1-Salvia pomifera 香桉烯合酶的氨基酸序列相比较。

[0051] 详述

[0052] 发明详述

[0053] 根据本发明,发现了来自印度檀香的新的萜烯合酶基因 SaSSy。还发现了来自其它两个不同系统发生的物种的直系同源基因 (来自澳洲檀香的 SspiSSy 和来自新克里多尼亚檀香的 SauSSy)。新基因由 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示的 DNA 序列表征。

[0054] 在 SEQ ID NO:1 中公开的新基因以下一般称作 SaSSy,在 SEQ ID NO:3 中公开的新基因以下一般称作 SauSSy,而在 SEQ ID NO:5 中公开的新基因 以下一般称作 SspiSSy。SaSSy、SauSSy 和 SspiSSy 的 DNA 和蛋白质序列是非常高度保守的,在 ORF 的氨基酸上具有 94-98% 的相同性。该基因的关键结构域是非常高度保守的 (见图 4 和 5)。

[0055] 新的萜烯合酶催化由 FPP 产生萜类化合物,优选为萜烯,更优选为选自以下的倍半萜: α - 檀香烯、 α - 反式 - 香柠檬烯、 β - 檀香烯和 β - 檀香烯。

[0056] 更具体地, SaSSy、SauSSy 和 SspiSSy 为倍半萜合酶,并且最具体地,为檀香烯合酶。如本文中所用,“萜烯合酶”是催化由底物产生一或多种萜类化合物,或更优选地产生一或多种萜烯的酶;“倍半萜合酶”是催化合成倍半萜类化合物,或更优选地合成倍半萜烯的酶,而“檀香烯合酶”是催化檀香烯合成的酶。由底物形成萜类化合物和 / 或萜烯可使用本领域已知的任何方法评估,包括但不限于,下述实施例 6 和 7 中描述的酶测定和质谱法。

[0057] 萜类化合物在本文定义为通过萜烯合酶活性和可能的其它酶的后续修饰而衍生自异戊二烯基二磷酸底物的化合物。包含在术语萜类化合物中的是非氧化的萜类烯烃 (萜

烯)、氧化的萜烯和其它衍生物如萜醇。

[0058] 如本文中所用,萜烯为基于异戊二烯单元(C_5H_8)的不饱和烃,并且具有 $C_{10}H_{16}$ 的通式。萜烯可以为无环的、单环的或多环的。萜烯包括,但不限于,单萜,其含有10个碳原子;倍半萜,其含有15个碳原子;二萜,其含有20个碳原子;和三萜,其含有30个碳原子。提及萜烯时包括所述萜烯的立体异构体。

[0059] 本发明中提及檀香烯时包括 α -檀香烯和 β -檀香烯,以及其任何立体异构体,包括,例如,(+)-表- β -檀香烯、(-)- β -檀香烯、(+)- β -檀香烯、(+)- α -檀香烯和(-)- α -檀香烯。

[0060] 这些新萜烯合酶基因的分离将使得可以合成与天然油相似的檀香油。迄今为止,使用其它来源的倍半萜合酶合成一些倍半萜是可能的,但是其组分倍半萜的范围和每个组分的终比例与天然檀香油并不相似。

[0061] 优选地,本发明的萜烯合酶基因分离自檀香属的成员。更优选地,其分离自印度檀香(Indian Sandalwood、白檀、Chandana)、澳洲檀香(Australian Sandalwood)或新克里多尼亚檀香。然而,所述基因还可分离自选自以下的植物:密花澳洲檀香(*S. acuminatum*) (Desert Quandong, Sweet Quandong, Native Peach);滨海夏威夷檀香(*S. ellipticum*) (滨海檀香);智利檀香(*S. femandezianum*);垂枝夏威夷檀香(*S. freycinetianum*);榄绿夏威夷檀香(*S. haleakalae*);大花澳洲檀香(*S. lanceolatum*) (北澳檀香);巴布亚檀香(*S. macgregorii*);钩叶澳洲檀香(*S. murrayanum*) (苦 Quandong);伞花澳洲檀香(*S. obtusifolium*);亮叶夏威夷檀香(*S. paniculatum*);柳叶檀香(*S. salicifolium*) (柳叶檀香);或斐济檀香(*S. yasi*)。

[0062] 因此,提供了分离的萜烯合酶,其中所述萜烯合酶为檀香属萜烯合酶;并且所述萜烯合酶催化檀香烯产生。

[0063] 优选地,这样的萜烯合酶包括:

[0064] a) 选自 SEQ ID NO:2、SEQ ID NO:4 或 SEQ ID NO:6 所示序列的氨基酸序列;或

[0065] b) 氨基酸序列,其与 SEQ ID NO:2 所示序列具有至少或至少约 60%、65%、70%、75%、80%、85%、90%、95%或更高的相同性;或

[0066] c) a) 或 b) 的片段;

[0067] 其中所述萜烯合酶催化檀香烯产生。

[0068] 在另一个实施方案中,本发明提供了分离的萜烯合酶,其包括:

[0069] a) SEQ ID NO:2、SEQ ID NO:4 或 SEQ ID NO:6 所示的氨基酸序列;

[0070] b) 由权利要求 1-5 中任一项的核酸分子所编码的氨基酸序列;

[0071] c) 氨基酸序列,其与 SEQ ID NO:2 所示的氨基酸序列具有至少或至少约 60%、65%、70%、75%、80%、85%、90%、95%或更高的相同性;或

[0072] d) (a)、(b) 或 (c) 的片段;

[0073] 其中所述萜烯合酶催化萜烯产生。

[0074] 优选地,这样的萜烯合酶包含选自对应于 SEQ ID NO:2 的 32-42、221-425、321-325、314-315 和 423-426 位氨基酸的氨基酸。

[0075] 更优选地,所述萜烯合酶催化萜烯的产生,所述萜烯选自单环倍半萜、双环倍半萜和三环倍半萜,特别地,其中所述萜烯是从无环焦磷酸萜烯前体如焦磷酸金合欢酯(FPP)

合成。

[0076] SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 的核苷酸序列涵盖单一 ORF。本发明提供了分离的 DNA 核苷酸序列,其对应于 SEQ ID NO :1 所示的萜烯合酶基因 SaSSy 的核苷酸序列、SEQ ID NO :3 所示的萜烯合酶 SauSSy 的核苷酸序列或 SEQ ID NO :5 所示的 SspiSSy 的核苷酸序列,或与 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 基本同源的序列,或它们的片段。本发明还提供了 DNA 序列,其包含以下序列的互补序列:SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5,或与 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 基本上同源的序列,或它们的片段。

[0077] 因此本发明提供了编码萜烯合酶的分离的核酸分子,其选自:

[0078] a) 包含 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 所示的核苷酸序列的核酸分子;

[0079] b) 核酸分子,其为 (a) 的片段;

[0080] c) 核酸分子,其包含与 (a) 或 (b) 互补的核苷酸序列;和

[0081] d) 核酸分子,其编码萜烯合酶并与 (a)-(c) 中任一核酸分子具有至少或至少约 60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99% 的相同性;

[0082] 其中所述核酸分子编码萜烯合酶。

[0083] 分离的核酸分子编码优选地与 (a)-(d) 中任一核酸分子所编码的萜烯合酶具有至少或至少约 61%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99% 相同性的萜烯合酶,其中序列中的差异为氨基酸取代。

[0084] SaSSy、SauSSy 和 SspiSSy 的核苷酸序列是非常高度保守的(见图 4)。

[0085] SEQ ID NO :1 所示的 SaSSy 核苷酸序列编码在第 143 位氨基酸具有脯氨酸的多肽(SEQ ID NO :2)。然而,所编码的多肽在第 143 位还可具有丝氨酸(SEQ ID NO :7 和 SEQ ID NO :8)。两种变体序列为基本同源的,尤其由于这两个残基都是极性的,并且随后的测试已表明,这两种蛋白质在测定时具有基本相同的活性,并以基本相同的比例产生各种化合物。一般地,两种变体蛋白具有相同的活性并以相同的比例产生各种化合物。

[0086] 所述 DNA 序列还可对应于 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 的片段。优选地,所述片段选自 SEQ ID NO :1 的下列位置:961-975 位(通常对应于 DDxxD 基序),94-126 位(通常对应于 R(R/P)X₈W 基序)。不受任何特定理论的约束,相信 DDxxD 基序负责螯合二价金属离子(例如,镁离子),并且如果被去除,可能造成该蛋白质完全无功能。在蛋白质起始处的 R(R/P)X₈W 基序也被相信是萜烯合酶独有的,并且被认为涉及非手性 FPP 或 GPP 分别特定地电离为手性的橙花叔醇二磷酸和橙花基二磷酸。

[0087] 通过 Kampranis 等人(2007)的工作推断,其它对本发明的三种檀香烯合酶的特 定功能来说可能是必不可少的区域是氨基酸位置 314 和 315(核苷酸位置 940-945)。用具有相似极性的较大或较小的残基取代这些残基可能会改变活性中心的大小,并因此改变催化产生的产物。如 Kampranis 等人所证实的,所述檀香烯合酶的第 422-426 位氨基酸残基(核苷酸 1264-1278)还可能负责终产物谱。在 α 19 螺旋起始处的脯氨酸残基的重要性是紧紧把握底物,而去除该脯氨酸可能引起功能性缺失。氨基酸位置 221-426 定义了涵盖许多优选保留的更重要区域的较大的区域。因此,在这些位置的残基优选 为保守的(完整功能),或有稍微变化(改变的功能)。因此, DNA 序列还可能优选对应于 SEQ ID NO :1、SEQ

ID NO :3 或 SEQ ID NO :5 的片段,其选自 SEQ ID NO :1 的如下位置 :940-945、661-1278 或 1264-1278 位。

[0088] 同源核酸分子是指预先确定的相同或同源的核苷酸数目。同源性包括不改变所编码氨基酸的取代(即“沉默取代”)和相同的残基。基本同源的核酸分子通常以中度严格性或高度严格性与感兴趣的核酸分子的全长或至少约 70%、80%或 90%的全长杂交。也考虑这样的核酸分子,其含有替换杂交核酸分子中的密码子的简并密码子。

[0089] 可用已知的计算机算法来确定任何两个核酸分子是否具有至少 80%、85%、90%、95%、96%、97%、98%或 99%的“相同性”,例如“FAST A”程序,其使用如 Pearson 等人, Proc. Natl. Acad. Sci. USA 85 :2444(1988) 所述的默认参数(其它的程序包括 GCG 程序包(Devereux, J. 等人, Nucleic Acids Research 12(I) :387(1984))、BLASTP、BLASTN、FASTA(Atschul, S. F. 等人, J. Molec. Biol. 215 :403(1990) ;Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego(1994) 和 Carillo 等人, SIAM J Applied Math 48 :1073(1988))。其它的商业或公开可用的程序包括 DNASTar 的“MegAlign”程序(Madison, WI) 和威斯康星大学遗传学计算机组(UWG)的“Gap”程序(Madison WI))。例如,核酸分子同源性或相同性的百分比可通过使用 GAP 计算机程序(例如, Needleman 等人, J. Mol. Biol. 48 :443(1970), 如 Smith 和 Watermand 所修改的(Adv. Appl. Math. 2 :482(1981)) 来比较序列信息而确定。简言之, GAP 程序将相似性定义为相似的比对符号(即核苷酸)的数目除以两条序列中较短那条的符号总数。GAP 程序的默认参数可包括:(1) 一元比较矩阵(对与相同性,值为 1 ;和对于非相同性,值为 0) 和加权比较矩阵, Gribskov 等人, Nucl. Acids Res. 14 :6745(1986), 如 Schwartz 和 Dayhoff, eds., Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, pp. 353-358(1979) 中所描述的;(2) 对每个缺口 3.0 的罚分和每个缺口中每个字符 0.10 的额外罚分;和(3) 对末端缺口不罚分。

[0090] 当本发明的多核苷酸序列或其片段特异性地与另一 SaSSy、SauSSy 或 SspiSSy 多核苷酸(或其互补链)在选择性的杂交条件下杂交时,则存在基本上的同源性或相同性。如本文中所用,“特异性杂交”是指核酸分子(如,寡核苷酸)与靶核酸分子通过互补的碱基配对退火。本领域技术人员熟悉影响 特异性杂交的体外和体内参数(如特定分子的长度和组成)。与体外杂交特别相关的参数进一步包括退火和洗涤温度、缓冲液组成和盐浓度。去除非特异性结合的核酸分子的示范性洗涤条件在高度严格性下为 0.1×SSPE, 0.1% SDS, 65℃,而在中度严格性为 0.2×SSPE, 0.1% SDS, 50℃。等效的严格性条件是本领域已知的。技术人员可轻易地调整这些参数,以在低、中或高度严格性条件下实现核酸分子与靶核酸分子(适合于特定应用的)的特异性杂交。

[0091] 一般地,当至少约 14 个核苷酸的片段存在至少约 55%, 优选至少约 65%, 更优选至少约 75%, 最优选至少约 90%相同性时,选择性杂交将会发生。同源性长度比较,如所描述的,可在更长的片段上进行并且在某些实施方案中经常会在至少约 9 个核苷酸、通常至少约 20 个核苷酸、更通常至少约 24 个核苷酸、一般至少约 28 个核苷酸、更通常至少约 32 个核苷酸并且优选至少约 36 个或更多核苷酸的片段上进行。

[0092] 因此,本发明的多核苷酸序列优选与本文中的序列表所示的序列具有至少 75%、更优选至少 85%、更优选至少 90%的同源性。更优选具有至少 95%、更优选至少 98%的同

源性。核苷酸源性比较可如下描述的用于多肽的比对进行。优选的序列比较程序是 GCG Wisconsin Bestfit 程序。

[0093] 在本发明中, 源性序列包括在核酸水平上与相应的 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 所示的核苷酸序列在至少 20、50、100、200、300、500、1000、1500 或 1710 个核苷酸上具有至少 60、70、80 或 90% 相同性, 优选至少 95% 或 98% 相同性的核苷酸序列。尤其是, 源性一般应考虑编码已知对萜烯合酶基因功能是必不可少的连续氨基酸序列的核酸序列区域, 而不是非必要的邻近序列。例如, 编码氨基酸位置 32-42 (核苷酸 94-126) 和 / 或氨基酸位置 321-325 (核苷酸 961-975) 的核酸序列, 和 / 或可选地编码氨基酸位置 221-426 (核苷酸 661-1278), 位置 314-315 (核苷酸 940-945) 和 / 或位置 422-426 (核苷酸 1264-1278) 的核酸序列。

[0094] 本发明的 SaSSy、SauSSy 或 SspiSSy 多核苷酸序列片段将优选为至少 15 个核苷酸的长度, 更优选至少 20、30、40、50、100 或 200 个核苷酸的长度。一般地, 多核苷酸序列的长度越短, 获得选择性杂交所要求的源性越高。因此, 当本发明的多核苷酸序列由少于约 30 个核苷酸组成时, 与本文中序列表所列出的多核苷酸序列相比, 应优选相同性的百分比高于 75%, 优选高于 90% 或 95%。相反地, 当本发明的多核苷酸序列由例如多于 50 或 100 个核苷酸组成时, 与本文中序列表所列出的多核苷酸序列相比, 相同性的百分比可较低, 例如高于 50%, 优选高于 60% 或 75%。

[0095] 本发明的与 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 所示序列同源的核酸序列可根据本领域已知的任何技术表征并分离, 如通过序列特异性的引物扩增、在较高或较低严格性条件下与序列特异性的探针杂交、血清学筛查方法或通过 LiPA 分型系统。

[0096] SaSSy 的基因组 DNA 序列在 SEQ ID NO :9 中提供。

[0097] 还提供了 SaSSy、SauSSy 和 SspiSSy 基因的 RNA。RNA 序列优选源自以上描述并提供于 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 的 DNA 序列。

[0098] 本发明还提供了可与 SaSSy、SauSSy 或 SspiSSy 基因组 DNA 杂交的 RNA 片段。所述 RNA 或 RNA 片段序列还可源自 SaSSy、SauSSy 或 SspiSSy 或其片段的 cDNA 序列。

[0099] 本发明还提供了包含一些不会实质上改变其与 SaSSy、SauSSy 或 SspiSSy 杂交能力的缺失或突变的核酸序列和片段。这样的变体被认为形成了上述 DNA、RNA 或片段的显而易见的等效物。

[0100] 本发明其它优选的变体核酸序列包括与任何上述给定的本发明的核酸序列相比冗余的序列, 冗余由遗传密码简并性导致。这些变体核酸序列会因此与其所来源的核酸序列一样编码相同的氨基酸序列。优选地, 这些变体的 DNA、RNA 或 cDNA 可与 SaSSy、SauSSy 或 SspiSSy 基因序列相应的部分杂交。

[0101] 本发明还包括 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 的 DNA 序列或相应的 RNA 序列或其片段的序列变体, 其包含一或多个核苷酸的缺失和 / 或插入, 尤其是一或多个密码子的插入或缺失。

[0102] 还包括一些非必须核苷酸由其它核苷酸 (包括修饰的核苷酸和 / 或肌苷) 的取代。

[0103] 本发明特别优选的变体多核苷酸还包括在严格性条件下与任何本发明的核酸序列杂交的序列。因此, 优选与如上描述的本发明的任何序列显示有高精度源性 (相似性)

的序列。尤其优选的是与所述本发明的核酸序列具有至少 80%、85%、90%、95%或更高同源性的序列。优选地,所述序列将 具有所述核酸序列原始核苷酸的少于 20%、15%、10%或 5%的变异。

[0104] 还提供了引物和探针,其可从本发明的任何 DNA 或 RNA 序列或序列片段出发来制备。优选地,这样的探针或引物长约 5 至 50 核苷酸,更优选约 10 至 25 核苷酸。优选地,探针或引物寡核苷酸包含来自萆烯合酶核酸分子的至少或至少约 15、20、25、30、35、40、45、50、60 个或更多的连续的核苷酸。本发明的探针和引物可用于 PCR、测序反应、杂交反应和技术人员已知的其它应用。优选地,探针和 / 或引物会从易于由技术人员确定的高 G 和 C 含量的区域生成。

[0105] 本发明还涉及寡核苷酸引物,其包含 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 的部分,所述引物能够作为特异性扩增 SaSSy、SauSSy 或 SspiSSy 核酸的引物。优选地,所述引物是单链 DNA 寡核苷酸序列,其能够作为引物延伸产物的合成起始点,所述引物延伸产物与待复制的核酸链是互补的。所用引物的具体长度和序列取决于所需的 DNA 或 RNA 靶的复杂度以及引物使用的条件,如温度和离子强度。扩增引物不是必须与对应的模板序列准确匹配以保证正确的扩增的事实文献 (Kwok 等人,1990) 中充分记录。

[0106] 所用扩增方法可为聚合酶链式反应 (PCR ;Saiki 等人,1988)、连接酶链式反应 (LCR ;Langdren 等人,1988 ;Wu 和 Wallace,1989 ;Barany,1991)、基于核酸序列的扩增 (NASBA ;Guatelli 等人,1990 ;Compton,1991)、基于转录的扩增系统 (TAS ;Kwoh 等人,1989)、链置换扩增 (SDA ;Duck,1990 ;Walker 等人,1992) 或通过 Q β 复制酶的扩增 (Lizardi 等人,1988 ;Lomeli 等人,1989) 或任何其它合适的使用引物延伸扩增核酸分子的方法。在扩增期间,扩增产物可通过使用标记的引物或掺入标记的核苷酸被方便地标记。标记可能为同位素的 (32P、35S 等) 或非同位素的 (生物素、地高辛等)。扩增反应重复 20-70 次,最好 25-45 次。

[0107] 本发明还涉及寡核苷酸探针,其包含 SEQ ID NO :1、SEQ ID NO :3 或 SEQ ID NO :5 的部分,所述探针能够作为 SaSSy、SauSSy 或 SspiSSy 基因的杂交探针。优选地,所述探针为单链的序列特异性寡核苷酸序列,其具有与待测定的 SaSSy、SauSSy 或 SspiSSy 基因的靶序列互补的序列。

[0108] 本领域技术人员会认可杂交的严格性会由这样的条件所影响 :盐浓度、温度或有机溶剂,还有碱基组成、互补链长度和杂交核酸间错配核苷酸碱 基的数目。严格性温度条件一般会包括超过 30℃ 的温度,通常超过 37℃,并且优选超过 45℃。严格性盐条件通常会低于 1000mM,一般低于 500mM,并且优选低于 200mM。然而,参数的组合比任何单个的参数更重要。严格性杂交条件的一个实例是 65℃、0.1 $\times$ SSC (1 $\times$ SSC = 0.15M NaCl,0.015M 柠檬酸钠, pH 7.0)。

[0109] 任选地,本发明的探针被标记和 / 或连接至固体基质。固体基质指寡核苷酸探针可偶联的任何基质,条件是所述寡核苷酸探针保持杂交特性并使杂交背景保持低水平。固体基质通常是微量滴定板、膜 (如尼龙膜或硝酸纤维素膜) 或微球 (珠)。在施加至膜或固定之前,可方便地修饰核酸探针以促进固定或改进杂交效率。这样的修饰可涵盖同聚物加尾、偶联不同的活性基团 (如脂肪基团、NH₂ 基团、SH 基团、羧基基团) 或偶联生物素或半抗原。

[0110] 本发明的探针还包括连接至标记或报告分子的分离的多核苷酸并且可用于通过标准方法分离其它具有序列相似性的多核苷酸序列。用于制备和标记探针的技术见,如 Sambrook 和 Russell(2001) 或 Ausubel 等人,(2001)。

[0111] 本发明的用作引物或探针的寡核苷酸还可能含有或由核苷酸类似物(如硫代磷酸酯(Matsukura 等人,1987)、烷基磷酸酯(alkylphosphoriate)(Miller 等人,1979)或肽核酸(Nielsen 等人,1991 ;Nielsen 等人,1993))组成,或可能包含插层剂(Asseline 等人,1984)。为了积极地影响如杂交动力学、杂交物形成的可逆性、寡核苷酸分子的生物学稳定性等等性状,引入这些修饰可能是有益的。

[0112] 本发明还提供了含有 SaSSy、SauSSy 或 SspiSSy 基因 DNA 序列片段的重组 DNA,并且其可能用作,例如,探针。优选地,用于产生重组 DNA 的质粒是在原核或真核细胞中可扩增的并携带所述片段的质粒。例如,使用含有 SaSSy 基因 DNA 片段的克隆的 DNA 作为分子杂交探针,通过放射性核苷酸或荧光试剂标记,例如在檀香树的不同组织中,该基因可被直接检测。

[0113] SaSSy、SauSSy 或 SspiSSy 多核苷酸序列(优选以探针的形式)还可被固定到固相支持物上来检测 SaSSy 基因。在本发明可选的形式中,SaSSy、SauSSy 或 SspiSSy 多核苷酸序列与其他多核苷酸序列(如来自其他萜烯合酶基因)一起可以这样的方式被固定于固相支持物,即其允许确定结合固相支持物的适当萜烯合酶基因和/或任何其他多核苷酸序列在材料(如树样品)中的存在。

[0114] 在本领域中已经描述了用于产生 DNA 分子的固定化文库的技术。一般地,多数现有技术方法描述了单链核酸分子文库的合成,例如使用屏蔽技术在固体基质上的不同的离散位置建立不同的序列排列(permutation)。美国专利号 5,837,832 描述了用于产生固定于硅基质上的 DNA 阵列的基于非常大规模的整合技术的改进方法。特别地,美国专利号 5,837,832 描述了叫做“覆瓦(tiling)”的用于在基质上空间限定的位置合成特定探针组的策略,其可用于产生本发明的固定化 DNA 文库。美国专利号 5,837,832 还提供了关于可使用的早期技术的参考文献。因此多核苷酸序列探针可在基质表面原位合成。

[0115] 另外,单链分子可不在固体基质上合成而将每个预先形成的序列施加至固体基质上的离散位置。例如,多核苷酸序列可用配备了针脚或压电器件的机械设备直接印于基质上。

[0116] 文库序列一般固定于固体基质的离散区域上或离散区域内。基质可以是多孔的(以允许在基质内的固定)或基本上非多孔的(在该情况下文库序列通常固定于基质表面)。固体基质可由多肽可直接地或间接地结合的任何材料制成。适当的固体基质的实例包括平板玻璃、硅片、云母、陶瓷和有机聚合物(如塑料,包括聚苯乙烯和聚甲基丙烯酸酯)。也还可能使用半透膜(如广泛可获得的硝酸纤维素或尼龙膜)。所述半透膜可安装于更坚硬的固体表面如玻璃上。其表面可任选地涂有金属层,如金、铂或其它过渡族金属。合适的固体基质的具体实例是商业可得的 BiaCore™ 芯片(Pharmacia Biosensors)。

[0117] 优选地,固体基质通常为具有刚性或半刚性表面的材料。在优选的实施方案中,所述基质的至少一个表面为基本平的,不过在一些实施方案中,可能需要用例如凸起区域或蚀刻坑道为不同聚合物进行物理上的分区。所述固体基质还优选为适合于在一般为 50 至 100 μm 的离散区域以高密度施加 DNA 序列,给出 10000 至 40000 点/ cm^2 的密度。

[0118] 所述固体基质可方便地分为区段(section)。这可通过例如光蚀刻技术,或通过应用疏水性油墨,例如基于聚四氟乙烯的油墨(Cel-line, USA)来实现。

[0119] 多核苷酸序列与基质的连接可通过共价或非共价的方法。所述多核苷酸序列可通过文库序列所结合的一层分子连接于基质。例如,多核苷酸序列可标记有生物素并且基质用抗生物素蛋白和/或链霉亲和素包被。使用生物素化多核苷酸序列的一个便利的特点是可容易地测定偶联至固体基质的效率。由于多核苷酸序列可能与一些固体基质仅很差地结合,因此通常有必要在固体基质(例如玻璃)与核酸序列之间提供化学界面。适当的化学界面的实例包括六乙二醇。另一个实例是使用聚赖氨酸包被的玻璃,然后使用标准方法化学修饰聚赖氨酸来引入亲和配体。其它的通过使用偶联剂将分子连接到固体基质表面的方法是本领域已知的,见例如 W098/49557。

[0120] 互补多核苷酸序列与被固定的核酸文库的结合可用各种方法测定,如结合的多核苷酸序列的光学特征的变化(即通过使用溴化乙锭)或通过使用标记的核酸如用荧光团标记的多核苷酸。其它不需要使用标记的检测技术包括光学技术如光声学、反射法、椭圆偏光法和表面等离子共振(见 W097/49989)。

[0121] 因此,本发明提供了固体基质,其具有固定于其上的至少一种本发明的多核苷酸,优选两或多种不同的本发明的多核苷酸序列。在优选的实施方案中固体基质还含有源自除 SaSSy、SauSSy 或 SspiSSy 多核苷酸序列以外的基因的多核苷酸序列。

[0122] 优选地, SaSSy、SauSSy 或 SspiSSy 基因的转录被上调或促进(例如在人工系统如宿主细胞或重组植物中),其可导致提高的油产量。这种上调或提高可由极多方法完成,如插入额外的或可选的调控序列至考虑中的基因的上游和/或下游,或产生已存在的调控序列的突变体,所述突变体能够由于例如其调控元件结合的提高而提高转录。

[0123] 本发明还覆盖了由上述 RNA 和 DNA 核苷酸序列及其片段所编码的多肽。本发明还提供了如 SEQ ID NO:2、SEQ ID NO:4 或 SEQ ID NO:6 所示的分离的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列及其片段。更合乎需要地,所述 SaSSy、SauSSy 或 SspiSSy 氨基酸序列以基本上纯化的形式提供。也提供多肽片段,其具有更低分子量并具有与 SEQ ID NO:2、SEQ ID NO:4 或 SEQ ID NO:6 所示那些一样的肽序列或片段。

[0124] 如本文中所用,“氨基酸”为含有氨基基团和羧酸基团的有机化合物。多肽含有两个或多个氨基酸。为了本文中的目的,氨基酸包括 20 种天然存在的氨基酸、非天然氨基酸和氨基酸类似物(即其中的 α 碳具有侧链的氨基酸)。

[0125] 为了符合在 J. Biol. Chem., 243:3552-3559(1969)中描述的和 37 C. F. R. § 1.821-1.822 所采用的标准多肽命名法,氨基酸残基的缩写示于下列的对应表中:

[0126]

符号		
1-字母	3-字母	氨基酸
Y	Tyr	酪氨酸
G	Gly	甘氨酸
F	Phe	苯丙氨酸
M	Met	甲硫氨酸
A	Ala	丙氨酸
S	Ser	丝氨酸
I	Ile	异亮氨酸
L	Leu	亮氨酸
T	Thr	苏氨酸
V	Val	缬氨酸
P	Pro	脯氨酸
K	Lys	赖氨酸
H	His	组氨酸
Q	Gln	谷氨酰胺
E	Glu	谷氨酸
Z	Glx	Glu 和/或 Gln
W	Trp	色氨酸
R	Arg	精氨酸
D	Asp	天冬氨酸
N	Asn	天冬酰胺
B	Asx	Asn 和/或 Asp
C	Cys	半胱氨酸
X	Xaa	未知或其它

[0127] 应当指出,本文中由式表示的所有氨基酸残基序列自左至右为常规的氨基端至羧基端的方向。此外,短语“氨基酸残基”被广泛地定义为包括对应表中列出的氨基酸和修饰的和罕见的氨基酸,如在 37C.F.R. § § 1.821-1.822 中提及并通过引用并入本文的那些。此外,应指出,在氨基酸残基序列开头和结尾处的横线表示与另外的一或多个氨基酸残基序列、氨基端基团如 NH₂ 或羧基端基团如 COOH 相结合的肽键。

[0128] 如本文中所用,“天然存在的氨基酸”是指存在于多肽中存在的 20 种 L-氨基酸,而“非天然氨基酸”是指含有氨基基团和羧基基团的有机化合物,其并非对应表中列出的天然存在的氨基酸。因此,非天然存在的氨基酸包括,例如,20 个天然存在氨基酸之外的氨基酸或氨基酸类似物,包括但不限于,D-立体异构体(isostereomer)氨基酸。非天然存在的氨基酸可并入本文中提供的萜烯合酶及其变体。

[0129] 术语“分离的”用于描述本发明的氨基酸序列,其已从其天然状态所伴随的组分中分离出来。另外,当至少约 60 至 75% 的样品展现出单一 SaSSy 氨基酸序列时,本发明的氨基酸序列为“基本上纯化的”。基本上纯化的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列一般会包含样品的约 60 至 90% W/W,更通常约 95%,并优选将超过约 99% 纯。蛋白质纯度或同质性可通过本领域已知的大量方法表明,如蛋白质样品的聚丙烯酰胺凝胶电泳,随后在染色的胶上观察单一的氨基酸序列条带。出于某些目的,通过使用 HPLC 或其它本领域熟知的方法(其已被用于应用)可提供更高的分辨率。

[0130] 本发明还考虑了 SaSSy、SauSSy 或 SspiSSy 氨基酸序列的片段。根据本发明这方

面的氨基酸序列片段为至少约 5-7 个连续氨基酸的一段氨基酸残基,通常为至少约 7-9 个连续的氨基酸,一般至少约 9-13 个连续的氨基酸,并且最优选地,至少约 20-30 个或更多的连续的氨基酸。

[0131] 在本发明高度优选的形式中,所述片段展现 SaSSy、SauSSy 或 SspiSSy 氨基酸序列的配体结合、免疫学活性和 / 或其它生物学活性性状。更优选地,所述片段具有与天然 SaSSy、SauSSy 和 SspiSSy 氨基酸序列上存在的免疫学表位相一致的免疫学表位。

[0132] 如本文中所用,“表位”指多肽的抗原决定簇。表位可包含表位独特的空间构象的 3 个氨基酸。一般地,表位至少由 5 个氨基酸组成,更通常由至少 8-10 个氨基酸组成。测定这样的氨基酸的空间构象的方法是本领域已知的。

[0133] 本发明优选的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列将具有天然全长氨基酸序列的一或多个生物学特性(如体内、体外或免疫学特性)。另外,SaSSy、SauSSy 或 SspiSSy 全长氨基酸序列的片段可能具有一或多个编码全长氨基酸序列的基因的生物学特性。

[0134] 针对 SaSSy、SauSSy 或 SspiSSy 基因特定区域的抗体可由本领域技术人员确定,例如使用在 G. W. Turner 和 R. Croteau, Organization of monoterpene biosynthesis in *Mentha*. Immunocytochemical localizations of geranyl diphosphate synthase, limonene-6-hydroxylase, isopiperitenol dehydrogenase, and pulegone reductase. Plant Physiology 136 (2004) 4215-4227 中记载的技术。

[0135] SaSSy、SauSSy 或 SspiSSy 基因的氨基酸序列,包括类似物、片段和衍生物,可合成制备(如,使用众所周知的固相或液相肽合成技术)。优选地,采用固相合成技术。另外,本发明的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列可使用众所周知的基因工程技术制备,如下文描述。在另一个实施方案中,氨基酸序列可从生物学材料(如树木心材、树液等)中纯化(如通过免疫亲和纯化)。

[0136] 针对 SaSSy、SauSSy 或 SspiSSy 基因特定区域的氨基酸序列和衍生的肽生物标记物可由本领域技术人员确定,例如,使用 Zulak KG, Lippert DN, Kuzyk M, Domanski D, Chou T, Borchers CH and J Bohlmann (2009) Targeted proteomics using selected reaction monitoring (SRM) reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate treated Norway spruce (*Picea abies*). The Plant Journal 60 :1015-1030 中所记载的方法。

[0137] SaSSy、SauSSy 或 SspiSSy 氨基酸序列类似物优选包括那些其中一或多个氨基酸被另一种氨基酸取代的氨基酸序列,所述取代基本上不改变该分子的生物学活性。

[0138] 本发明的 SaSSy、SauSSy 或 SspiSSy 萜烯合酶的变体可用于实现所需的酶活性提高或降低、修饰的区域选择性化学或立体化学、或改变的底物利用或产物分布。变体或定点突变体可由本领域已知的任何方法制备。天然多肽的变体和衍生物可通过分离其它或相同植物株系或物种中天然存在的变体或变体的核苷酸序列来获取,或通过人工编程编码天然檀香多肽的核苷酸序列的突变来获取。

[0139] 就本发明上下文中,同源性序列包括在至少 20、50、100、200 或 570 个氨基酸与 SEQ ID NO :2、SEQ ID NO :4 或 SEQ ID NO :6 所示氨基酸序列在氨基酸水平上具有至少 60、70、80 或 90% 同源性,优选至少 95% 或 98% 同源性的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列。尤其是,同源性一般应考虑编码已知对萜烯合酶基因功能是必不可少的连续氨基酸序列的

序列区域,而不是非必要的邻近序列。例如,32-42 位氨基酸和 / 或 321-325 位氨基酸,或可选地 221-425 位、314-315 位和 / 或 422-426 位氨基酸。

[0140] 虽然同源性可认为是依据相似性(即具有相似化学特性 / 功能的氨基酸残基),就本发明上下文中,优选依据序列相同性来表示同源性。当涉及 SaSSy 氨基酸序列时,术语“基本上同源”或“基本上相同”表明,考虑中的氨基酸序列展现出和天然存在的 SaSSy、SauSSy 或 SspiSSy 完整氨基酸序列或其部分至少约 70% 的相同性,通常至少约 80% 的相同性并且优选至少约 90% 或 95% 的相同性。

[0141] 如本文中所用,“序列相同性”是指在测试和参考多肽或多核苷酸之间的比较中相同氨基酸(或核苷酸碱基)的数目。同源多肽是指预先确定的相同或同源氨基酸残基数。同源性包括保守氨基酸取代和相同的残基。序列相同性可通过使用由各个供应商建立的默认缺口罚分的标准比对算法程序确定。

[0142] 为了确定蛋白质同源性,可比对保守氨基酸以及相同的氨基酸;在该情况下,相同性百分比和同源性百分比不同。可用已知的计算机算法来确定任何两个氨基酸序列是否具有至少 80%、85%、90%、95%、96%、97%、98% 或 99% 的“相同性”,例如“FASTA”程序,其使用如 Pearson 等人,Proc. Natl. Acad. Sci. USA 85:2444(1988)所述的默认参数(其它的程序包括 GCG 程序包(Devereux, J. 等人, Nucleic Acids Research 12(I):387(1984))、BLASTP、BLASTN、FASTA(Atschul, S. F. 等人, J. Molec. Biol. 215:403(1990); Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego(1994) 和 Carillo 等人, SIAM J Applied Math 48:1073(1988))。例如,国家生物技术信息中心数据库的 BLAST 功能可用于确定相同性。其它的商业或公开可用的程序包括 DNASTar 的“MegAlign”程序(Madison, WI) 和威斯康星大学遗传学计算机组(UWG)的“Gap”程序(Madison WI)。

[0143] 例如,蛋白质和 / 或核酸分子同源性或相同性的百分比可通过使用 GAP 计算机程序(例如, Needleman 等人, J. Mol. Biol. 48:443(1970), 如 Smith 和 Watermand 所修改的(Adv. Appl. Math. 2:482(1981))比较序列信息而测定。简言之, GAP 程序将相似性定义为比对的相似符号(即核苷酸或氨基酸)的数目除以两条序列中较短那条的符号总数。GAP 程序的默认参数可包括:(1)一元比较矩阵(包含相同性的值为 1 和非相同性的值为 0)和加权比较矩阵, Gribskov 等人, Nucl. Acids Res. 14:6745(1986), 如 Schwartz 和 Dayhoff, eds., Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, pp. 353-358(1979)中所描述的;(2)对每个缺口 3.0 的罚分和每个缺口中每个符号 0.10 的额外罚分;和(3)对末端缺口不罚分。

[0144] 因此,如本文中所用,术语“相同性”表示在测试和参考多肽或多核苷酸之间的比较。在一个非限制性的实例中,“至少 90% 相同”指相对于参考多肽,相同性百分比为 90 至 100%。90% 或更高水平的相同性表明,假设(为了例证的目的)长度 100 个氨基酸的测试和参考多核苷酸进行比较,在测试多肽中不超过 10% (即 100 个中的 10 个)的氨基酸与参考多肽中的不同。这样的差异可由在整个氨基酸序列长度上随机分布的点突变所代表,或它们可聚集于不同长度上(多至所允许的最大值)的一或多个位点,如 10/100 氨基酸差异(大约 90% 相同性)。差异定义为氨基酸取代、插入或缺失。在超过约 85-90% 同源性或相同性的水平上,结果应不依赖于程序和缺口参数设定;这样高水平的相同性可容易地被评估,通常不依靠于软件。

[0145] 在本发明高度优选的形式中, SaSSy、SauSSy 或 SspiSSy 氨基酸序列类似物与 SEQ ID NO :2、SEQ ID NO :4 或 SEQ ID NO :6 中列出的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列将具有 80% 或更高的氨基酸序列相同性。在本发明范围内这样的氨基酸序列类似物的实例包括 SEQ ID NO :2、SEQ ID NO :4 或 SEQ ID NO :6 的氨基酸序列, 其中: (a) 一或多个天冬氨酸残基由谷氨酸取代; (b) 一或多个异亮氨酸残基由亮氨酸取代; (c) 一或多个甘氨酸或缬氨酸残基由丙氨酸取代; (d) 一或多个精氨酸残基由组氨酸取代; (e) 一或多个酪氨酸或苯丙氨酸残基由色氨酸取代; 或 (f) 一或多个脯氨酸残基由丝氨酸残基取代。

[0146] 本发明还涵盖天然存在的肽变体。这样的变体的实例为由 mRNA 可变剪切事件或本文中描述的多肽的蛋白水解裂解得到的蛋白质。例如, 归因于蛋白质水解的变异包括例如在不同类型宿主细胞中表达的 N- 或 C- 末端差异, 其由于蛋白质水解从由本发明的序列所编码的多肽上去除一或多个末端氨基酸。

[0147] 本发明还提供了 SaSSy、SauSSy 或 SspiSSy 氨基酸序列衍生物, 其包括在一级结构上为基本上同源但包含化学和 / 或生物化学修饰或非常见氨基酸的氨基酸序列、类似物或其片段。例如, 这样的修饰包括乙酰化、羧基化、磷酸化、糖基化、羟基化、硫酸化、泛素化、标记 (如, 用放射性核苷酸) 和多种酶修饰, 其将容易地由本领域熟练技术人员鉴别。

[0148] 本发明的萆烯合酶还可能以“嵌合蛋白”或“融合蛋白”的形式提供。这样的蛋白质为可操作地连接于不同多肽的多肽。本文中所提供的嵌合或融合蛋白可包括一或多种萆烯合酶多肽或其部分, 以及用于以下一或多种的一 或多种其它多肽: 转录 / 翻译控制信号、信号序列、用于定位的标签、用于纯化的标签、免疫球蛋白 G 的结构域的一部分, 和 / 或靶向剂。这些嵌合或融合蛋白包括通过重组方法产生的融合蛋白质、通过化学方法如化学偶联 (例如, 偶联于巯基基团) 产生的, 和通过任何其它方法产生的, 据此至少有一种多肽 (如萆烯合酶) 或其部分通过接头直接或间接地连接于另一多肽上。

[0149] SaSSy、SauSSy 或 SspiSSy 氨基酸序列以标记的形式提供时, 多种用于标记氨基酸序列的方法为本领域众所周知的, 其包括放射性同位素如 ^3H 或 ^{14}C 、结合于标记的抗配体 (如抗体) 的配体、荧光团、化学发光剂、酶和可作为标记的配体的特异性结合对成员的抗配体。标记的选择取决于所需的灵敏度、稳定性要求和可用的仪器。标记氨基酸序列的方法是本领域众所周知的 [见, 如 Sambrook 等人, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989); 和 Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K. *Current protocols in molecular biology*. Greene Publishing Associates/Wiley Intersciences, New York (2001)]。

[0150] 本发明的 SaSSy、SauSSy 和 SspiSSy 氨基酸序列, 如果可溶, 可偶联于固相支持物, 如, 硝酸纤维素、尼龙、柱填充材料 (如, Sepharose 珠)、磁珠、玻璃棉、塑料、金属、聚合物凝胶、细胞或其它基质。这样的支持物可采取例如珠、孔、试纸或膜的形式。

[0151] 本发明还提供融合多肽, 其包含 SaSSy、SauSSy 或 SspiSSy 氨基酸序列和片段。因此, SaSSy、SauSSy 或 SspiSSy 氨基酸序列可能为两或多个本发明的氨基酸序列间的融合物或为来自 SaSSy、SauSSy 或 SspiSSy 和相关蛋白质的氨基酸序列之间的融合物。同样, 可构造异源融合物, 其将展现出来源蛋白质的性质或活性的组合。例如, 配体结合或其它结构域可能在不同的融合多肽或片段间“交换”。例如, 这样的同源或异源融合多肽可能表现出

改变的结合强度或特异性。融合配对物 (partner) 包括免疫球蛋白、细菌 β -半乳糖苷酶、trpE、蛋白 A、 β -内酰胺酶、 α -淀粉酶、乙醇脱氢酶和酵母 α -交配因子。

[0152] 可使用常规技术合成或通过修饰的多核苷酸序列编码并使用重组核酸方法产生修饰的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列。修饰的多核苷酸序列还可通过常规技术制备。融合蛋白一般会通过重组核酸方法制备或可化学合成。

[0153] 本发明的萆烯合酶的变体可用于实现所需的酶活性提高或降低、修饰的区域选择性化学或立体化学、或改变的底物利用或产物分布。此外,可制备至少具有一种修饰的特性的变体,例如对底物提高的亲合力、用于产生一或多种所需化合物的改进的特异性、不同的产物分布、不同的酶活性、提高的酶反应速度、在特定环境 (pH、温度、溶剂等) 中更高的活性或稳定性或在所需的表达系统中改进的表达水平。变体或定点突变体可由本领域已知的任何方法制备。

[0154] 如上所述,本发明提供重组和非重组的、分离的和纯化的多肽。天然多肽的变体和衍生物可通过分离其它或相同植物株系或物种中天然存在的变体或变体的核苷酸序列来获取,或通过编码天然檀香多肽的核苷酸序列的人工编程突变来获取。天然氨基酸序列的改变可通过许多常规方法中的任何一个完成。

[0155] 由本发明多肽的氨基和羧基末端上额外的肽序列的融合导致的多肽变体可用于在所需的环境或表达系统中提高该多肽的表达、帮助纯化该蛋白质或改进该酶活性。例如,这种额外的多肽序列可以是信号肽。因此,本发明涵盖本发明多肽的变体,如通过与其它寡肽或多肽融合所获取的变体和 / 或连接于信号肽的多肽。

[0156] 因此,在一个实施方案中,本发明提供用于制备具有所需萆烯合酶活性的变体多肽的方法,所述方法包括步骤:

[0157] a) 选择由如上描述的 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5,其片段或变体组成的组中的任一核酸;

[0158] b) 修饰所选的核酸以获取至少一种突变核酸;

[0159] c) 用所述突变核酸序列转化宿主细胞以表达由所述突变核酸序列编码的多肽;

[0160] d) 在所述多肽中筛选具有至少一种修饰的特性的功能多肽。

[0161] e) 任选地,如果所述多肽不具有所需的变体萆烯合酶活性,重复处理步骤 (a) 至 (d) 直至获取具有所需变体萆烯合酶活性的多肽 (即 DNA 改组)。

[0162] 在步骤 (b) 中,可创建大量的突变核酸序列,例如通过随机诱变、定点突变或 DNA 改组。基因改组的详细步骤可在 Stemmer, W.P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci U S A. 91(22):10747-1075 中发现。简言之, DNA 改组是指体外随机重组已知序列的过程,其涉及至少 2 种选择用于重组的核酸。例如,突变可通过合成包含突变序列的寡核苷酸 (两侧为限制性位点以使其能连接至天然序列的片段) 引入至特定的位点。连接后,所得的重建序列编码具有所需氨基酸插入、取代或缺失的类似物。另外,可采用寡核苷酸定向的定点突变步骤以提供改变的基因,其中预先确定的密码子可通过取代、缺失或插入而改变。

[0163] 因此,SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5,其片段和变体的任何一个可能与选自 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5,其片段和变体的不同序列和 / 或与其

它的（例如，分离自除印度檀香、新克里多尼亚檀香或澳洲檀香之外的生物的）萜烯合酶编码核酸重组。因此，可能获取并分离突变核酸，其可用于根据标准程序转化宿主细胞。

[0164] 在步骤 (d) 中，对获取于步骤 (e) 的多肽筛选修饰的特性（例如，所需的修饰的酶活性）。可能对所表达的多肽筛选的所需酶活性的实例包括提高或降低的酶活性（通过 K_M 或 V_{max} 值衡量），例如，修饰的区域选择性化学或空间化学、改变的底物利用或产物分布。酶活性的筛选可根据技术人员熟悉的步骤进行。测定动力学数据的方法和分析萜烯产物的方法，例如，在实施例 13 和 14 中给出。

[0165] 步骤 (e) 提供处理步骤 (a)–(d) 的重复，优选地，其可平行地实施。因此，通过创建大量的突变核酸，许多宿主细胞可同时用不同的突变核酸转化，使得可以随后筛选更大量的多肽。因此根据技术人员的判断，获取所需变体多肽的机会可被提高。

[0166] 在一个实施方案中，本发明提供了用于制备编码具有萜烯合酶活性的变体多肽的核酸的方法，所述方法包括上述公开的步骤 (a)–(e) 并且进一步包括步骤：(f) 如果鉴定了具有所需变异萜烯活性的多肽，获得步骤 (c) 中获取的突变核酸，其用于在步骤 (c) 和 (d) 之后转化宿主细胞以表达变体萜烯合酶。

[0167] 优选地，本发明的萜烯合酶多肽催化单、双和 / 或三环倍半萜产生。优选地，所述萜烯合酶由无环焦磷酸萜烯前体底物产生或合成倍半萜。无环焦磷酸萜烯前体是任何无环的焦磷酸化合物，其是至少一种萜烯的产生的前体，包括但不限于，焦磷酸香叶酯 (GPP)、焦磷酸金合欢酯 (FPP) 和焦磷酸 香叶基香叶酯 (GGPP)。一般来说，GPP 是单萜的前体、FPP 是倍半萜的前体并且 GGPP 是二萜的前体。优选地，所述前体是 FPP。

[0168] 如本文中所用，“催化萜烯产生”或“催化檀香烯产生”是指萜烯合酶从底物（如无环萜烯前体）分别产生萜烯或其混合物，或具体地檀香烯或其混合物的能力。萜烯（如檀香烯）通过萜烯合酶由底物的形成可使用本领域已知的任何方法评估，包括但不限于，下述实施例 6 和 7 中描述的酶测定和质谱方法。

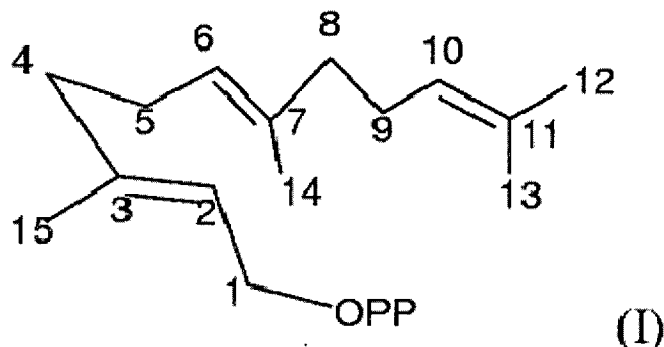
[0169] 通常，“催化萜烯产生”的多肽从特定的底物以总萜烯的至少或至少约 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90% 或更多的量（按重量计算）产生萜烯或其混合物，其使用体外测定（如在实施例 6、7、13 和 14 中描述的测定）来测量，其中所述萜烯合酶（如檀香烯合酶）与底物（如 FPP）在 Mg^{2+} 存在下混合。由所述合酶产生的檀香烯可包括 α -檀香烯和 / 或 β -檀香烯，包括 (+)- β -檀香烯、(-)- β -檀香烯、(+)- β -檀香烯、(+)- α -檀香烯、和 (-)- α -檀香烯。例如，在本文中提供的萜烯合酶包括檀香属物种的萜烯合酶，其催化由 FPP 产生 α -檀香烯，其中由 FPP 所产生的 α -檀香烯的量至少为或至少约为产生自 FPP 的总萜烯（按重量计算）的 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90% 或更多。在另外的实施例中，萜烯合酶催化由 FPP 产生 β -檀香烯，其中由 FPP 所产生的 β -檀香烯的量至少为或至少约为产生自 FPP 的总萜烯（按重量计算）的 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90% 或更多。

[0170] 优选地，本发明分离的多肽能够由 FPP 形成红没药阳离子 (bisabolyl cation) 并且能够进一步在 FPP 的 C_3 和 C_7 碳原子之间形成键以产生包含 C_3 – C_7 键的双环或三环倍半萜。相似地，本发明的多肽能够由 FPP 形成红没药阳离子并且能够进一步在 FPP 的 C_2 和 C_7

碳原子之间形成键以产生包含 C_2-C_7 键的双环或三环倍半萜。

[0171] 术语“能够合成”化合物（如特定的倍半萜）和术语“萜烯合酶活性”，优选“倍半萜合酶活性”，是指本发明的多肽和编码这些多肽的核酸，其能够由至少一种起始化合物（其优选为无环焦磷酸萜烯前体）合成萜烯，优选倍半萜，更优选檀香烯和最优选本文中提及的倍半萜和 / 或檀香烯化合物。优选地，所述无环萜烯前体为 FPP，其以倍半萜碳骨架的标准编号在下述的式 (I) 中给出。OPP 是指焦磷酸。

[0172]



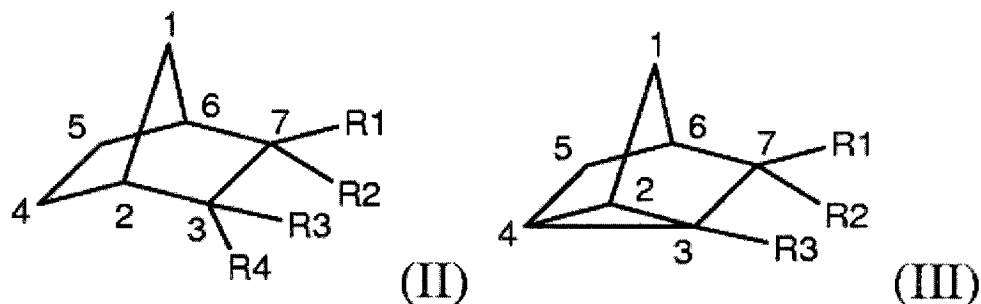
[0173] 优选地，所述分离的萜烯合酶多肽能够合成至少一种倍半萜，更优选至少一种具有檀香烯或香柠檬烯的碳骨架的倍半萜。在优选的实施方案中，所述多肽能够由 FPP 形成红没药阳离子并且能够进一步在 FPP 的 C_3 或 C_2 碳原子与 C_7 碳原子之间形成化学键以产生一或若干种双环和 / 或三环倍半萜。

[0174] 术语“键”是指单一的共价键。

[0175] 本发明涉及编码能够合成至少一种包含 C_3-C_7 键的双环和 / 或三环倍半萜的多肽的核酸和所述多肽本身。优选地，包含 C_3-C_7 键的倍半萜构成了所述萜烯合酶所合成的倍半萜产物总量至少约 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、80% 或更多，其为具有 C_3-C_7 键的倍半萜。为了本发明的目的，萜烯合酶的定量倍半萜产物分布可以通过采用实施例 7 中详述的步骤或通过使用本领域技术人员熟知的技术测定组分百分比来确定。

[0176] 因此，本发明涉及分离的多肽，其能够形成具有下式 (II) 和 / 或 (III) 的具有 C_3-C_7 键的化合物

[0177]

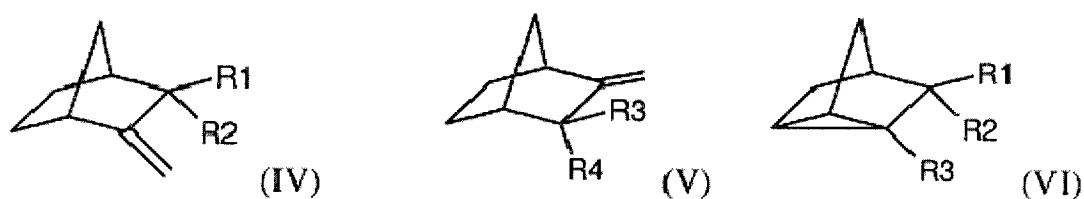


[0178] 其中 R_1 、 R_2 、 R_3 、 R_4 相互独立地是线性或分支的 C_1-C_{20} 的烷基或烯基基团，并且其中 R_1 及 R_2 和 / 或 R_3 及 R_4 可形成双键而不是两个单独的单键。

[0179] 优选地， R_1 、 R_2 、 R_3 、 R_4 相互独立地是线性或分支的 C_1-C_{15} ，更优选 C_1-C_{10} ，最优选 C_1-C_8 的烷基或烯基基团。

[0180] 特别地,本发明的多肽能够形成下式 (IV)、(V) 和 / 或 (VI) 的化合物

[0181]



[0182] 其中 R_1 、 R_2 、 R_3 、 R_4 如上述定义。

[0183] 优选地,在式 (IV) 和 / 或 (VI) 中, R_1 或 R_2 一个为 C_1 - C_5 烷基而另一个是 C_2 - C_8 烯基。此外,式 (VI) 中的 R_3 优选为 C_1 - C_5 烷基,更优选为 C_1 - C_3 烷基。优选地,在式 (V) 中, R_3 和 R_4 如上 R_1 和 R_2 在式 (IV) 中的定义。最优选地,化合物 VI 的 R_3 是甲基基团并且 R_1 和 R_2 可选地为甲基和 C_6 线性烷基基团。

[0184] 优选地,所述倍半萜为檀香烯,更优选为 α -檀香烯或 β -檀香烯。所述檀香烯可为选自 (+)- β -檀香烯、(-)- β -檀香烯、(+)- β -檀香烯、(+)- α -檀香烯和 (-)- α -檀香烯的立体异构体。

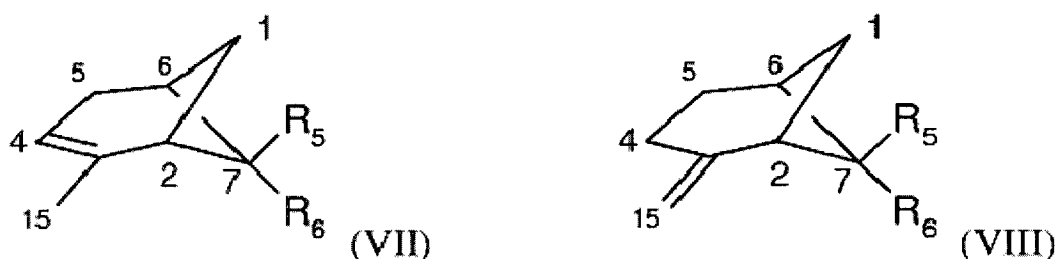
[0185] 优选地,所述萜烯合酶催化 α -檀香烯的产生,并且所产生的 α -檀香烯的量至少或至少约为所产生的萜烯总量的 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、80% 或更多。

[0186] 本发明涉及编码能够形成至少一种包括 C_2 - C_7 键的倍半萜的多肽的核酸和所述多肽本身。根据优选的实施方案,包括 C_2 - C_7 键的倍半萜构成了所述萜烯合酶所合成的倍半萜产物总量的至少约 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、80% 或更多,其为具有 C_2 - C_7 键的倍半萜。

[0187] 所述倍半萜可为香柠檬烯、檀香烯和 / 或这些化合物的异构体的一种,优选立体异构体。优选地,所述产物主要为一种立体异构体,不过其还可包含较少量的其它立体异构体或对映异构体。

[0188] 根据一个实施方案,本发明涉及分离的多肽,其能够形成具有下式 (VII) 和 / 或 (VIII) 的含 C_2 - C_7 键的化合物

[0189]



[0190] 其中 R_5 和 R_6 如上 R_1 和 R_2 的述定义。优选地, R_5 为甲基并且 R_6 为 C_2 - C_{10} 烯基,或相反。

[0191] 优选地,可能存在于在上述 R_1 、 R_2 、 R_3 、 R_4 、 R_5 或 R_6 残基之一中的至少一个烯基为 4-甲基-3-戊烯基,而另一个连接于同一碳原子的残基为甲基。

[0192] 能够合成上式 (II)、(III)、(IV)、(V)、(VI)、(VII) 和 / 或 (VIII) 中化合物的多肽优选为具有 SEQ ID NO :2、SEQ ID NO :4 或 SEQ ID NO :6 的氨基酸序列的多肽,或其多肽

变体。

[0193] 优选地,具有 C₂-C₇键的倍半萜为香柠檬烯,包括其立体异构体,尤其是,例如顺式- α -香柠檬烯、反式- α -香柠檬烯、反式- β -香柠檬烯和顺式- β -香柠檬烯。

[0194] 优选地,所述萜烯合酶催化香柠檬烯产生,并且所产生的香柠檬烯的量至少为或至少约为所产生的萜烯总量的 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、80%或更多。

[0195] 另一方面,本发明提供了分离的多肽,其能够合成檀香烯和/或香柠檬烯。在优选的实施方案中,本发明提供了分离的多肽,其能够合成以下的一或多种: α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯、 β -檀香烯和反式-反式-金合欢醇。优选地,分离的多肽至少能够合成 β -檀香烯。

[0196] 最优选地,本发明的 SaSSy、SauSSy 和 SspiSSy 萜烯合酶能够合成至少两种或多种萜烯,优选选自 (+)-表- β -檀香烯、(-)- β -檀香烯、(+)- β -檀香烯、(+)- α -檀香烯和 (-)- α -檀香烯、E- α -香柠檬烯、 γ -姜黄烯、 β -红没药烯、 β -姜黄烯、 α -姜黄烯、反式-反式-金合欢醇、 α -蒎烯、蒎烯、柠檬烯和 α -异松油烯、更优选 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯中的萜烯。最优选地,所述萜烯合酶可同时合成两种或多种上述化合物。

[0197] 所述酶可优选为能在同一反应混合物中、相同的反应条件下合成至少 4 种上述列出的化合物。因此,所述酶优选为能在单一的一组条件下将 FFP 转换为至少 4 种所列出的萜烯;不需要另外加入不同的起始化合物、不同的附加因子或改变反应条件来获取一或多种萜烯。

[0198] 另一方面,本发明提供包含本发明核酸的载体。

[0199] 如本文中所用的“载体”包括原核载体、病毒载体、或真核载体和/或任何重组载体,包括但不限于,噬菌体和质粒。所述载体还可为表达载体。表达载体包括能够表达可操作地连接于调控序列的 DNA 的载体,所述调控序列,如启动子区域,能够影响这种 DNA 片段的表达。这样的附加片段可包括启动子和终止子序列,并且任选地可包括一或多个复制原点、一或多个可选择标记、增强子、多腺苷酸信号等等。表达载体一般源自质粒或病毒 DNA,或可包含两者的元件。因此,表达载体是指重组 DNA 或 RNA 构建体,如质粒、噬菌体、重组病毒或其它载体,其在引入至适当的宿主细胞后可导致克隆的 DNA 表达。适当的表达载体对本领域技术人员是众所周知的,包括那些在真核细胞和/或原核细胞中为可复制的以及那些保持游离的或那些整合入宿主细胞基因组的表达载体。病毒载体为可操作地连接于外源基因以将外源基因转运(如运载或穿梭)入细胞中的工程化病毒。

[0200] 技术人员能够根据表达系统选择适当的载体。在一个实施方案中,表达载体包含可操作地连接于调控序列的编码多肽的 cDNA 序列,所述调控序列如转录启动子、操纵子或增强子、mRNA 核糖体结合位点和控制转录和翻译(例如起始及终止)的适当的序列。

[0201] 当调控序列与本发明的 cDNA 序列有功能地相关时,核苷酸序列即为“可操作的连接”。可操作的或可操作地连接 DNA 区段意指排列所述区段从而使其按预期的目的发挥功能,例如转录在启动子起始并通过编码区段继续进行至终止子。

[0202] 本发明的载体可用于制备遗传学修饰的宿主生物体和/或细胞的方法中,用于含有本发明的核酸的宿主生物体和/或细胞中,并且可用于产生或制备萜烯合酶的方法中,

其如下述进一步展示。因此,本发明提供了包含本发明载体的细胞。

[0203] 一方面,本发明提供了用于制备萜烯合酶的方法,其包括步骤:在使所述编码的萜烯合酶表达的条件下培养被修饰以含有至少一种核酸序列的宿主生物体和/或细胞,其中所述至少一种核酸为本发明的核酸。然后可任选地纯化所述萜烯合酶。这种产生方法的实例在例如实施例 4、5 和 12 中提供。

[0204] 例如,产生萜烯合酶的方法可包括以下步骤:

[0205] a) 选择宿主生物体和/或细胞,其不表达具有 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸;

[0206] b) 用具有 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸分子转化所述生物体;和

[0207] c) 在有助于所述核酸编码的萜烯合酶产生的条件下培养所述生物体。

[0208] 本发明还提供了产生萜烯合酶的方法,所述方法包括以下步骤:

[0209] a) 选择宿主生物体和/或细胞,其表达具有 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸分子;

[0210] b) 用具有 SEQ ID NO:1、SEQ ID NO:3 或 SEQ ID NO:5 所示序列的核酸分子以较高的量转化所述生物体;和

[0211] c) 在有助于所述核酸编码的萜烯合酶产生的条件下培养所述生物体。

[0212] 还提供了制备萜烯的方法,包括使无环焦磷酸萜烯前体与本发明的萜烯合酶相接触,其中所述萜烯合酶在细胞中异源表达;所述无环焦磷酸萜烯前体与所述萜烯合酶表达于同一细胞中;与无环焦磷酸前体接触的步骤发生在所述细胞中。

[0213] 还提供了制备至少一种萜烯的方法,包括:在有助于萜烯产生的条件下培养细胞,其中所述细胞异源表达本发明的萜烯合酶;以及所述细胞表达无环焦磷酸萜烯前体。

[0214] 如此产生的萜烯合酶可任选地被进一步分离。

[0215] 无环焦磷酸萜烯前体可选自焦磷酸香叶酯 (GPP)、焦磷酸金合欢酯 (FPP) 和焦磷酸香叶基香叶酯 (GGPP)。

[0216] 另外,所述萜烯优选选自倍半萜、二萜和单萜,更优选选自二环或三环倍半萜,并且最优选地,所述萜烯选自 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯、 β -檀香烯、 γ -姜黄烯、 β -红没药烯、 β -姜黄烯、 α -姜黄烯、反式-反式-金合欢醇、 α -蒎烯、蒎烯、柠檬烯和 α -异松油烯。

[0217] 所述萜烯可优选地选自 (+)-表- β -檀香烯、(-)- β -檀香烯、(+)- β -檀香烯、(+)- α -檀香烯和 (-)- α -檀香烯。

[0218] 另一方面,可产生两种或多种萜烯。

[0219] 当所述萜烯为 α -檀香烯时,所产生的 α -檀香烯的量优选至少或至少约为所产生的萜烯总量的 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、80%或更多。

[0220] 当所述萜烯为 β -檀香烯时,所产生的 β -檀香烯的量优选至少或至少约为所产生的萜烯总量的 1%、2%、3%、4%、5%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、80%或更多。

[0221] 最优选地,至少产生 4 种萜烯;这四种萜烯包括 α -檀香烯、 α -反式-香柠檬

烯、表- β -檀香烯和 β -檀香烯；并且 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯以下述彼此之间的比例产生： α -檀香烯(38.0%)、 α -反式-香柠檬烯(12.1%)、表- β -檀香烯(4.7%)和 β -檀香烯(45.2%)。

[0222] 如此产生的萜烯可转换为醇，例如选自 α -檀香醇、 β -檀香醇、 α -反式-香柠檬醇和表- β -檀香醇的醇。在进一步处理后，这样的醇可具有下述彼此之间的比例： α -檀香醇(25-65%)、 α -反式-香柠檬醇(1-20%)、表- β -檀香醇(1-15%)和 β -檀香醇(20-50%)；或更优选地， α -檀香醇(59.28%)、 α -反式-香柠檬醇(7.32%)、表- β -檀香醇(3.45%)和 β -檀香醇(29.0%)。

[0223] 另一方面，本发明提供了重组宿主生物体和/或细胞，其被转化以含有本发明的核酸。所述宿主生物体可为单细胞或多细胞（但非人的）生物体。例如，宿主可为多细胞生物体的细胞。优选地，所述宿主生物体异源地包含本发明的核酸。因此，本发明的宿主细胞可为原核细胞、细菌细胞或真核细胞（例如酵母细胞或植物细胞）。

[0224] 如本文中所用，异源核酸为表达其的细胞正常不体内产生的核酸，或正常地产生于所述细胞但在不同基因座或差异地表达的核酸，或介导或编码改变内源核酸（如DNA）表达（通过影响转录、翻译或其他可调控的生化过程）的中间体（mediator）的核酸。异源核酸对其所引入的细胞通常为非内源的，但已从另一细胞获取或合成制备。异源核酸还可以是内源的表达于不同基因座或其表达被改变的核酸。一般地，虽然并非必须的，这样的核酸编码了通常不由所述细胞产生，或不以与其所表达的细胞中相同的方式而表达的RNA和蛋白质。异源核酸，如DNA，还可以指外来核酸，如DNA。因此，异源核酸或外来核酸包括不像发现于基因组的对应核酸分子（如DNA）一样存在于确切的方向或位置的核酸分子。它还可指来自另一生物体或物种的核酸分子（即外源的）。

[0225] 异源表达是指在宿主细胞中表达由异源核酸编码的多肽，所述异源核酸通过如转化、电转化、转导或任何其他方法引入至所述宿主细胞中。

[0226] 优选地，所述宿主生物体为细菌，例如大肠杆菌。

[0227] 其它优选的宿主生物体包括真菌，优选酵母，最优选酿酒酵母（*Sacharomyces cerevisiae*）。其它合适的宿主细胞包括在US6,531,303中描述的BTR酵母株系。

[0228] 用于表达本发明多肽的适当的宿主生物体可另外为高等真核细胞，优选为植物或植物细胞。优选地，所述植物为属于茄科（*Solaniaceae*）或唇形科（*Lamiaceae*），更优选为烟草（*Nicotiana*）属的物种；或另外为长春花（*Catharanthus*）属的成员，如长春花（*Catharanthus vinca*）。所述宿主生物体可另外为檀香属的物种，如印度檀香或澳洲檀香。

[0229] 一方面，本发明提供了表达本发明多肽的重组宿主生物体或细胞。优选地，所述宿主生物体被转化以比在未如此转化的相同的生物体中更高量地表达所述多肽。

[0230] 术语“转化的”是指所述宿主被遗传改造以包含本发明的任何核酸的一、二或多个拷贝。优选地，所述宿主异源地表达本发明的核酸和/或多肽。

[0231] 因此，在一个实施方案中，本发明提供了转化的生物体，其中本发明的多肽以比在未如此转化的相同生物体中更高的量表达。

[0232] 本领域已知若干创建转基因、重组宿主生物体或细胞（如植物、酵母、细菌或高等真核生物体细胞培养物）的方法。例如，用于细菌、真菌、酵母和哺乳动物细胞宿主的适当的克隆和表达载体描述于，例如，Pouwels等人，*Cloning Vectors :A Laboratory Manual*,

Elsevier, New York, (1985) 和上述引用的 Sambrook 等人。技术人员尤其可获得用于高等植物和 / 或植物细胞的克隆和表达载体, 例如见 Schardl 等人, (1987) Gene 61 :1-11。

[0233] 用于转化宿主生物体, 例如, 产生转基因植物, 修饰宿主生物体或细胞以含有转基因核酸 (如本发明的那些) 的方法是技术人员所熟悉的。例如, 对制备转基因植物来说, 目前的方法包括: 植物原生质体电穿孔技术、脂质体介导的转化、农杆菌介导的转化、聚乙二醇介导的转化、颗粒轰击、植物细胞显微注射和使用病毒的转化。

[0234] 在一个实施方案中, 转化的 DNA 整合入非人宿主生物体和 / 或细胞的染色体中, 使得获得稳定的重组系统。本发明的实践中可以使用本领域已知的任何染色体整合方法, 包括但不限于, 重组酶介导的盒交换 (RMCE)、病毒位点特异性染色体插入、腺病毒、通过宿主介导加工的同源重组和前核注射。

[0235] 本发明的宿主细胞优选可天然产生萜烯合酶的 FPP 底物。另外, 如果宿主细胞不能天然产生所述化合物, 可用例如异源 FPP 基因转化改造来产生 FPP。另外, 宿主细胞可被改造以产生更多 (其天然会产生的) FPP, 从而提供大量的为本发明的萜烯合酶所用的底物 (见, 例如 US6, 531, 303 或 W02009109597)。

[0236] 可预计, 本发明的酶的倍半萜产物可进一步被处理, 其通过在 C₁₂ 位置羟基化以形成檀香醇和香柠檬醇。例如, α -檀香烯会进一步被处理以产生 α -檀香醇, α -反式-香柠檬烯会产生 α -反式-香柠檬醇, 表- β -檀香烯会产生表- β -檀香醇并且 β -檀香烯会产生 β -檀香醇。本发明的 SaSSy、SauSSy 或 SspiSSy 萜烯合酶优选能够合成 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香, 它们彼此之间的比例使得在进一步将化合物处理为等量醇时, 檀香醇和香柠檬醇的产量会与例如印度檀香天然油的产量相似。

[0237] 一般地, 来自印度檀香的檀香油以给出的比例包含表 1 中所示的化合物。

[0238] 表 1. 天然檀香油 (来自印度檀香) 的组分示例

[0239]

化合物	通过 GC 的组分%
α -檀香烯	1.26
β -檀香烯	0.22
E- α -香柠檬烯	1.28
表- β -檀香烯	1.81
g-姜黄烯	0.049
b-红没药烯	0.029
b-姜黄烯	0.226
a-姜黄烯	0.180
a-红没药醇	0.01
Z-a-檀香醇	51.48
Z-a-反式-香柠檬醇	6.36

[0240]

Z-表- β -檀香醇	3.80
Z-b-檀香醇	25.0
E-b-檀香醇	0.211
Z-澳白檀醇	1.76
Z-坚果醇	1.30

[0241] 由上述表 1 可确定, α -檀香醇、 α -反式-香柠檬醇、表- β -檀香醇和 β -檀香醇彼此之间 (除去其它的油组分) 的比例为: α -檀香醇 (59.28%)、 α -反式-香柠檬醇

(7.32%)、表- β -檀香醇(4.35%)和 β -檀香醇(29.0%)。

[0242] 本发明的萜烯合酶优选合成 α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯和 β -檀香烯,使得在进一步处理后,所述化合物以下述彼此之间的比例产生各种醇: α -檀香醇(25-65%)、 α -反式-香柠檬醇(1-20%)、表- β -檀香醇(1-15%)和 β -檀香醇(20-50%)。

[0243] 最优选地,本发明的萜烯合酶合成倍半萜化合物,使得进一步处理时,所述化合物以下述彼此之间的比例产生各种醇: α -檀香醇(34.7%)、 α -反式-香柠檬醇(1.1%)、表- β -檀香醇(4.3%)和 β -檀香醇(41.3%)。剩余的8.6%的化合物一般由一系列其它少量化合物组成。如果将其余少量化合物的比例消除,剩余的4种主要化合物,进一步处理以产生各自的醇时,优选以下述彼此之间的比例生成: α -檀香烯(38.0%)、 α -反式-香柠檬烯(12.1%)、表- β -檀香烯(4.7%)和 β -檀香烯(45.2%)。

[0244] 用分离的SaSSy萜烯合酶产生的油中4种主要组分以及其转化为各自的醇的相对比例由GC-MS和GC-FID测定。已知GC-MS往往由于易于电离而偏向一些化合物。然而,两种方法都表明,本发明萜烯合酶产物的相对组成反映了檀香油中各自的醇的组成(表2)。

[0245] 表2

[0246]

化合物	天然油中相对%组成 (GC-MS)	SaSSy 生成的油中相对%组成 (GC-MS)	SaSSy 生成的油中相对%组成 (GC-FID)	SauSSy 生成的油中相对%组成 (GC-FID)	SspiSSy 生成的油中相对%组成 (GC-FID)
α -檀香醇	59.28	38.0	45.3	51.3	47.3
β -檀香醇	29.0	45.2	31.5	27.4	26.3
<i>E</i> - α -香柠檬醇	7.32	12.1	16.0	14.5	19.2
表- β -檀香醇	4.35	4.7	4.7	4.8	4.2

[0247] 可通过技术人员已知的任何方法进一步的处理以产生正确的顺式醇,如利用合适的细胞色素P450酶,或化学反应如碱金属化、硼化(borylation)和氧化。细胞色素P450技术是为其它倍半萜(最著名为青蒿素)的产生而建立的。

[0248] 本发明的核酸探针可用于选择涉及印度檀香或檀香属其它成员的合适的树,用于育种方案或株系改进方案。已知并非所有新克里多尼亚檀香、澳洲檀香和印度檀香树都以相同的量产生檀香油,并且油中萜烯的相对比例可能是树之间差异的主体,特别是涉及SaSSy、SauSSy或SspiSSy基因核酸的控制、表达转录和/或翻译。

[0249] 因此,本发明的核酸探针可用于使用本文中描述的和/或技术人员熟知的方法确定给定的一组树中哪些高水平地表达SaSSy、SauSSy或SspiSSy。另外,可测定SaSSy、SauSSy或SspiSSy氨基酸表达的水平以评估哪些树产生大量的所需萜烯。优选地,探针可用于检测表达萜烯合酶基因的树,所述萜烯合酶生成一或多种下列萜烯: α -檀香烯、 α -反式-香柠檬烯、表- β -檀香烯、 β -檀香烯和反式-反式-金合欢醇。

[0250] 另外,本发明的多肽或多肽片段可用于产生针对SaSSy、SauSSy或SspiSSy基因特定区域的抗体。这些抗体然后可用于筛选来自树的样品以确定哪些树表达所述合酶基因,

以及表达的水平。同样,这样的检测方法可用于选择大量产生所需萜烯或产生所需比例的不同萜烯的树。

[0251] 因此,本发明提供用于检测样品中萜烯合酶存在的方法,包括以下步骤:

[0252] a) 使令可能含有本发明的萜烯合酶的样品与特异性结合所述萜烯合酶的抗体在使得包含所述抗体和萜烯合酶的反应复合物形成的条件下相接触;和

[0253] b) 检测样品中包含所述抗体和萜烯合酶的反应复合物的形成,其中检测到所述反应复合物的形成表明样品中存在萜烯合酶氨基酸序列。

[0254] 所述方法可包含可选的评估所形成的反应复合物的量的步骤,从而确定所述生物学样品中萜烯合酶的量。

[0255] 优选地,这些方法中所使用的抗体源自亲和纯化的多克隆抗体,并且更优选为 mAb。另外,本文中所用的抗体分子优选 Fab、Fab'、F(ab')₂ 或 F(v) 部分的形式或完整抗体分子。

[0256] 特别优选的用于检测本发明萜烯合酶基因(根据以上方法)的方法包括酶联免疫吸附测定、放射性免疫测定、免疫放射测定和免疫酶测定,包括使用单克隆和/或多克隆抗体的竞争和夹心测定。

[0257] 在每个例子中,本发明的氨基酸序列形成含有一或多个抗体或结合配对物的复合物,并且复合物的一个成员标记有可检测的标记。复合物的形成和(如果需要的话)其量可由已知的可应用于检测标记的方法确定。

[0258] 最通常被这类研究所采用的标记为放射性元素、酶、暴露于紫外线时发荧光的化学物和其它。

[0259] 许多荧光材料是已知的并可利用作标记。这些材料包括,例如,荧光素、罗丹明和金胺。特定的检测材料是在山羊中制备的抗兔抗体,其通过异硫氰酸酯缀合于荧光素。

[0260] SaSSy、SauSSy 或 SspiSSy 氨基酸序列或其它结合配对物还可标记有放射性元素或酶。放射性标记可用任何现有可用的计数方法检测。

[0261] 酶标记也是同样有用的,并可用任何目前利用的比色法、分光光度法、荧光分光光度法、安培计或气体定量技术检测。所述酶通过与桥接分子如碳二亚胺、二异氰酸酯、戊二醛等等反应而缀合于所选择的颗粒。已知有许多可用于这些方法的酶可被利用。优选的酶是过氧化物酶、 β -葡萄糖醛酸酶、 β -D-葡萄糖苷酶、 β -D-半乳糖苷酶、脲酶、葡萄糖氧化酶加过氧化物酶和碱性磷酸酶。美国专利号 3,654,090、3,850,752 和 4,016,043 通过例举其公开的可选的标记材料和方法而引用。

[0262] 根据以上方法,其它优选的用于检测本发明萜烯合酶基因和蛋白质的方法包括选择性反应监测或多重反应监测,如在 Zulak KG, Lippert DN, Kuzyk M, Domanski D, Chou T, Borchers CH and J Bohlmann(2009) Targeted proteomics using selected reaction monitoring(SRM) reveals the induction of specific terpene synthases in a multi-level study of methyl jasmonate treated Norway spruce(*Picea abies*). The Plant Journal 60:1015-1030 中所讨论的。

[0263] 本发明还提供了检测生物学样品中萜烯合酶核酸分子存在的方法,其包括如下步骤:

[0264] a) 在合适的杂交条件下使生物学样品与包含本发明萜烯合酶多核苷酸的多核苷

酸探针或引物相接触 ;和

[0265] b) 检测所述探针或引物与样品中的核酸序列间形成的任何双链体。

[0266] 根据本发明的一个实施方案, SaSSy、SauSSy 或 SspiSSy 基因的检测可通过使用已知技术从生物学样品直接扩增萆烯合酶多核苷酸序列, 然后检测 SaSSy、SauSSy 或 SspiSSy 多核苷酸序列的存在来完成。

[0267] 因此, 本发明还涉及用于检测生物学样品中的萆烯合酶的方法, 包括:

[0268] a) 用至少一种上述定义的引物或寡核苷酸扩增所述核酸分子, b) 检测所述扩增的核酸分子。

[0269] 优选地, 在扩增前提取和 / 或纯化 (如从组织样品中) 所述核酸。

[0270] 本发明还涉及检测存在于生物学样品中的 SaSSy、SauSSy 或 SspiSSy 核酸的方法, 包括:

[0271] a) 在合适的条件下, 使所述生物学样品的核酸与如上述定义的一或多个探针杂交,

[0272] b) 在合适的条件下洗涤, 和

[0273] c) 检测形成的杂交体。

[0274] 优选地, 杂交条件为变性条件。

[0275] 优选地, 在杂交前提取和 / 或纯化 (如从组织样品中) 所述核酸。更优选地, 在提取后, 或至少在杂交前, 用至少一种如上述定义的引物扩增核酸样品。优选地, 所述探针连接于固体基质, 或在液相中用光度法或荧光检测或其它可视化方法如琼脂糖凝胶电泳来检测。

[0276] 本发明还涉及如上述定义的方法, 其中在扩增期间或之后标记所述核酸。

[0277] 为了本发明的目的, 用于检测在寡核苷酸探针和样品的核酸序列之间形成的杂交体的合适的测定方法可包括本领域已知的任何测定形式, 如常规的点印迹形式、夹心杂交或反向杂交。例如, 检测可使用点印迹形式完成, 未标记的扩增样品结合于膜, 在适当的杂交和洗涤条件下, 向所述膜加入至少一种标记的探针, 并且监测存在的结合探针。

[0278] 可选的和优选的方法是“反向”点印迹形式, 其中扩增的序列包含标记。在此形式中, 未标记的寡核苷酸探针结合于固体支持物并在适当的严格性杂交和随后的洗涤条件下暴露于标记的样品。应该理解, 可使用任何其它基于样品核酸与本发明的寡核苷酸探针间杂交体的形成的测定方法。

[0279] 本发明的一种形式中, 靶核酸序列通过 PCR 扩增, 然后用任何上述提及的特定方法检测。其它有用的用于检测 SaSSy、SauSSy 或 SspiSSy 多核苷酸序列存在的诊断技术包括但不限于: 1) 等位基因特异性 PCR; 2) 单链构象分析; 3) 变性梯度凝胶电泳; 4) RNA 酶保护测定; 5) 使用识别核酸错配的蛋白质, 如大肠杆菌 mutS 蛋白; 6) 等位基因特异性寡核苷酸; 和 7) 荧光原位杂交。

[0280] 除上述方法之外, 还可使用常规探针技术检测 SaSSy、SauSSy 或 SspiSSy 多核苷酸序列。当探针用于检测 SaSSy、SauSSy 或 SspiSSy 多核苷酸序列存在时, 如果需要的话, 可处理待分析的生物学样品 (如木质组织, 特别是心材组织) 以提取核酸。所述样品多核苷酸序列可能以多种便于检测靶序列的方法制备; 如变性、限制性酶消化、电泳或点印迹。样品多核苷酸序列的被靶向区域通常必须至少部分为单链的, 以与探针的靶向序列形成杂交

体。靶序列的变性将可能是必须的并且可通过多种本领域已知的技术实现。

[0281] 样品多核苷酸序列和探针在可促进探针中靶序列与样品中假定的 SaSSy、SauSSy 或 SspiSSy 多核苷酸序列形成稳定杂交体的条件下温育。优选地,使用高严格性的条件以防止假阳性。

[0282] 得到的杂交体(如果有的话)的检测通常通过使用标记的探针来完成。另外,探针可以是未标记的,但可通过与标记的配体(直接地或间接地)结合而被检测。用于标记探针和配体的合适的标记和方法是本领域已知的,并包括,例如,可通过已知方法(如缺口平移、随机引物或激酶激活)并入的放射性标记、生物素、荧光基团、化学发光基团(如二恶二酮,尤其是触发的二恶二酮)、酶、抗体等等。所述基本方案的变化是本领域已知的,并包括那些便于将待检测的杂交体从无关材料中分离的和/或那些放大标记基团的信号的变化。

[0283] 优选地,探针是标记的。更优选地,探针是放射性标记的或荧光-或酶-标记的。

[0284] 一旦确定合适的树,它们可用于选择性育种方案,或仅确定为可收获用于产生檀香油的树。

[0285] 在另一方面,本发明提供用于制备萜类化合物的工艺和/或方法。

[0286] 因此,本发明提供了制备至少一种萜类化合物的方法,包括:

[0287] a) 使至少一种无环焦磷酸萜烯前体与至少一种本发明的或由本发明的任何核酸编码的多肽相接触,和,

[0288] b) 任选地,分离步骤(a)中产生的至少一种萜类化合物。

[0289] 此外,本发明提供了制备至少一种萜类化合物的方法,包括:

[0290] a) 在有助于萜类化合物产生的条件下,培养被转化以表达或提高表达本发明多肽的非人生物体,和,

[0291] b) 任选地,从所述非人生物体分离至少一种萜类化合物。

[0292] 根据优选的实施方案,所述方法还包括如下步骤:在步骤(a)之前,用重组核酸转化非人生物体以表达或提高表达本发明的多肽。

[0293] 优选地,所述至少一种萜类化合物选自: α -檀香烯, α -反式-香柠檬烯,表- β -檀香烯和 β -檀香烯。

[0294] 制备至少一种萜类化合物的方法包括使至少一种无环焦磷酸萜烯前体与至少一种本发明的多肽相接触的步骤。例如,可使用在用于产生萜烯合酶的上述方法中获取的多肽。这样的多肽可从表达本发明核酸的宿主生物体中根据标准蛋白质或酶提取技术来提取。如果宿主生物体是可将本发明的多肽释放入培养基中的单细胞生物体或细胞,所述多肽可从培养基中收集,例如,通过离心,任选地随后通过洗涤步骤和重悬于适当的缓冲液溶液中。

[0295] 如果所述宿主生物体是将本发明的多肽积累于细胞中的植物或单细胞生物体或细胞,所述多肽可通过破坏或裂解细胞并从细胞裂解物中提取多肽而获取。

[0296] 分离的多肽然后可在最优的 pH 和温度下悬浮于缓冲溶液中。如果需要,可加入盐、BSA 和其它种类的酶辅助因子以优化酶活性。

[0297] 可将萜烯前体加入至多肽悬浮液或溶液中,随后在最优温度如 30°C 温育。温育后,优选在从溶液中除去多肽后,可用标准分离步骤(如溶剂萃取和蒸馏)将萜类化合物从温育溶液中分离。

[0298] 在制备至少一种萜类化合物的方法的步骤中,在有助于产生萜类化合物的条件下培养宿主生物体或细胞。因此,如果宿主是转基因植物,提供最优的生长条件,例如最优的光照、水分和营养条件。如果宿主是单细胞生物体,有助于产生所述萜类化合物的条件可能包括向宿主培养基中添加适当的辅助因子。另外,可选择已证明能最大化萜类化合物合成的培养基。外部因子,如最优的 pH 和温度通常也可在给定的表达系统中有助于萜类化合物产生。

[0299] 其它因子如激发子也可用于上调某些基因的转录,尤其是萜烯,其可涉及植物防御机制。因此在檀香或其它宿主生物体(包括宿主树)中,激发子可用于上调本发明的萜烯合酶基因。一系列的激发子对本领域的技术人员是众所周知的,并且易于购买。优选地,所选的激发子是在所提供的特定环境中能上调本发明的萜烯合酶基因表达的激发子。适当的激发子的实例包括茉莉酮酸甲酯和水杨酸。所述萜烯合酶基因还可提供评估在用激发子处理如檀香树后是否存在转录上调的方法。这样的评估还可包括,在所述树中测量油的水平以确定施用的激发子是否具有上调的效果。

[0300] 在本发明另外的实施方案中,可制备试剂盒以确定在檀香树中 SaSSy、SauSSy 或 SspiSSy 基因存在或不存在和/或确定所述基因的活性。

[0301] 根据上文讨论的测试技术,用于检测 SaSSy、SauSSy 或 SspiSSy 蛋白质的一类所述试剂盒将至少包含标记的 SaSSy、SauSSy 或 SspiSSy 氨基酸序列结合配对物(例如特异性的抗体)和基于所选方法(如“竞争性”、“夹心”、“DASP”等等)的说明书。所述试剂盒还可包含周边试剂如缓冲液、稳定剂等。

[0302] 因此,可制备用于证明 SaSSy、SauSSy 或 SspiSSy 酶存在的测试试剂盒,其包括:

[0303] a) 预定量的至少一种标记的免疫化学反应性组分,其通过 SaSSy、SauSSy 和 SspiSSy 氨基酸序列特异性结合配对物直接或间接连接至可检测的标记而获得;

[0304] b) 其它试剂;和

[0305] c) 使用所述试剂盒的说明书。

[0306] 所述标记的结合配对物(如抗体)通常可结合于固相。

[0307] 本发明还提供用于检测 SaSSy、SauSSy 或 SspiSSy 核酸序列的试剂盒。

[0308] 因此,本发明还提供用于证明 SaSSy、SauSSy 或 SspiSSy 核酸序列存在的试剂盒,其包括:

[0309] a) 预定量的至少一种标记的核酸序列,其来源于 SaSSy、SauSSy 或 SspiSSy 基因序列;

[0310] b) 其它试剂;和

[0311] c) 使用所述试剂盒的说明书。

[0312] 例如,所述多核苷酸序列可以是一或多种如上文例举的引物,并且使用说明书可为对来自对象的组织样品中提取的 RNA 或 DNA 实施 PCR 的说明。

[0313] 另一方面,本发明提供了用于证明萜烯合酶存在的试剂盒,包括:

[0314] a) 预定量的至少一种结合本发明萜烯合酶的配体,其中所述配体包含可检测的标记;

[0315] b) 其它试剂;和

[0316] c) 使用所述试剂盒的说明书。

[0317] 另一方面,提供了用于证明编码萜烯合酶的核酸分子存在的试剂盒,包括:

[0318] a) 预定量的至少一种标记的本发明的核酸分子或引物;

[0319] b) 其它试剂;和

[0320] c) 使用所述试剂盒的说明书。

[0321] 上述试剂盒可使用来自檀香树的样品。可在来自两种不同檀香树的至少两个样品上实施所述方法;并且确定每个样品中的萜烯合酶或编码萜烯合酶的核酸的量;然后选择含有最高量萜烯合酶或编码萜烯合酶的核酸的样品所来源的檀香树;对所选择的檀香树进行选择育种或进行收获来获取檀香油。

[0322] 一般原则

[0323] 本领域技术人员应明白,本文中描述的发明除了那些明确描述的之外的变化和修改是容许的。应理解,本发明包括所有的这些变化和修改。本发明还包括所有在说明书中单独地或共同地提及或表明的步骤、特征、组合物和化合物,以及任意的和所有的组合,或者任何两或多个所述步骤或特征。

[0324] 本发明不被本文中描述的具体实施方案(仅为了示范性目的)限定范围。功能上等同的产品、组合物和方法明确地在本文中所描述的本发明的范围内。

[0325] 所有出版物(包括专利、专利申请、期刊文章、实验室手册、书籍或其它文档)的完整公开通过引用并入本文。不承认任何所述文献构成现有技术或为本发明相关领域的工作人员普通常识知识的一部分。

[0326] 如本文中所用,术语“来源”和“来源于”应视为表明了特定的对象可能获取自特定来源,虽然并非必须直接来自所述来源。

[0327] 如本文中所用,单数形式“a”、“an”和“the”也包括复数指代,除非上下文另有明确定义。因此,例如,催化萜烯形成的萜烯合酶的指代包括催化一或多种萜烯合成的合酶。

[0328] 在本说明书中,除非上下文另有要求,单词“包括”,或变体如“包括”或“包括”将理解为暗示了包括对象或一组对象的表述,但不排除任何其它的对象或对象组。

[0329] 除实施例或另有说明外,在本说明书和权利要求书中所使用的所有表达分量、反应条件等的数字在所有情况下应被理解为被术语“大约”所修饰。因此,除非有相反表示,说明书和权利要求书中表述的数字参数均为近似值,其可能有所不同,取决于本发明试图获取的所需特性。因此“约 80%”意指“约 80%”和“80%”。至少,每一数字参数应考虑到有效位数字与普通取整方法。

[0330] 虽然,表述了本发明广泛范围的数字范围和参数是近似值,在具体实例中,仍尽可能准确地报告所表述的数字值。但是,任何数字值都内在包含某些必然的误差,其由在其各自检验分析中发现的标准差所引起。

[0331] 本文中所用其它所选术语的定义可在发明详述发现并可应用于全文。除非另有定义,本文中所用的所有其它科学和技术术语具有如本发明所属领域的普通技术人员通常所理解的同样的含义。

[0332] 下述的实施例用于更完整地说明使用以上描述的发明的方式,和表述设想的实施本发明各方面的最好的模式。应理解这些方法不以任何方式用于限定本发明的真正范围,而只是用于说明的目的。

实施例

[0333] 所有的试剂、溶剂、抗生素和前体化学品从商业来源购买。二磷酸金合欢酯和二磷酸香叶酯来自 Sigma 公司 (St. Louis, MO, USA)。限制性内切酶和 T4DNA 连接酶来自 New England Biolabs (Ipswich, MA)。

[0334] 实施例 1

[0335] 植物材料收集和 RNA 提取

[0336] 在种植于 FPC Kununurra 植物园, Kununurra, WA 的成熟印度檀香树上钻若干个洞。收集木质部过渡区的刨花并立即冷冻于液氮中运至 UWA, Perth。

[0337] 使用修改的方案提取 RNA [Kolosova N. 等人, Isolation of high-quality RNA from gymnosperm and angiosperm trees. *BioTechniques* 36 (2004) 821-824]。刨花 (50g) 放置于液氮中并加入 RNA 提取缓冲液 (200mM Tris-HCl, pH 值 8.5, 1.5% 十二烷基磺酸锂, 300mM LiCl, 10mM EDTA, 1% w/v 脱氧胆酸钠, 1% w/v Tergitol Nonidet P-40)。使用前加入 5mM 硫脲、1mM 金黄三羧酸、10mM 二硫苏糖醇和 2% (w/v) 聚乙烯吡咯烷酮 (PVPP)。所有溶液由 DEPC 处理的和 / 或高压灭菌的水制备。在 TE/NaCl 重悬步骤可合并若干管以浓缩 RNA 样品。沉淀后, RNA 储存于 -80℃ 直至运至 UBC Vancouver, 加拿大, 建立 cDNA 文库。

[0338] 实施例 2

[0339] cDNA 合成和文库构建

[0340] 1.4 μg 的印度檀香木质部总 RNA 用 SuperScript III 反转录酶 (Invitrogen) 在 42℃ 反转录 1 小时, 其使用 Clontech 公司的寡 dT 引物和 SMART 5' oligo。cDNA 用试剂盒中提供的 M1 引物扩增。

[0341] 扩增的 cDNA 然后用限制性内切酶 SfiI (New England Biolabs) 消化。获得的混合物通过 Chromaspin 尺寸排阻色谱, 首先洗脱最大的片段。cDNA 以 200ml 等分收集并且合并具有所需高分子量范围的 cDNA 并在 Qiagen MinElute 离心柱洗脱。

[0342] 消化后的 cDNA 片段然后被克隆进预先切割的 pDNR-LIB 载体并用电穿孔转化入 25 μl 噬菌体耐受的大肠杆菌电感受态细胞中。在与甘油混合并储存于 -80℃ 前, 所述细胞在 SOC 培养基中于 37℃ 振荡 1 小时。分装物涂布于含 30mg ml⁻¹ 氯霉素的琼脂上。cDNA 文库滴定至每毫升大约 3×10⁶ 的菌落形成单位。该文库被送往 Genome Science Centre, Vancouver 加拿大, 进行 5' 和 3' 测序。

[0343] 实施例 3

[0344] 从印度檀香木质部 EST 文库鉴定萜烯合酶基因

[0345] 使用 CAP3 程序的默认设置组装短片段 (read)。用 BLAST 比较 EST 文库检索与 NCBI 数据库之前发表的萜烯合酶 (TPS) 和细胞色素 P450 基因序列同源的基因。鉴定了若干候选基因, 包括一个与蓖麻 (*Ricinus communis*) 柠檬烯合酶和柠檬香桃 (*Backhousia citriodora*) 芳樟醇合酶具有中度相同性的全长基因。还发现了约 12 个假定的细胞色素 P450 氧化酶, 其中有些是全长的。

[0346] 实施例 4

[0347] 细菌表达

[0348] 将产生已知萜烯合酶 (TPS) 基因的阳性命中的菌落纯化质粒以产生 pDNR-LIB 结合的全长 cDNA。选择一个标记为 SaSSy 的基因。SaSSy 推导出的氨基酸序列与其它被子

植物基因的比对示于图 6, 并且推导出的 SaSSy、SauSSy 和 SspiSSy 氨基酸序列与其它被子植物萜烯合酶基因的比对示于图 12。

[0349] 定位了两个版本的 SaSSy 基因, 其仅在位置 143 不同, 克隆 P143 在 143 位具有脯氨酸残基而克隆 S143 在所述位置具有丝氨酸残基。产生于这些克隆的两种蛋白质之后被发现具有相同的活性, 暗示了在 143 位将一个极性残基替换为另一个对催化活性几乎没有影响。

[0350] SaSSy 的单萜合酶基因 TPS-b 亚家族 [J. Bohlmann, J. 等人, *Plant terpenoid synthases: Molecular biology and phylogenetic analysis*. Proc. Natl. Acad. Sci. USA 95(1998)4126-4133] 典型特征基序包括天冬氨酸富集的 DDxxD 金属离子结合位点 (位置 321-325) 和 RRX8W 基序 (位置 32-42), 所述基序涉及二磷酸基团迁移 [Williams, D. C. 等人, *Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair*. Biochemistry 37(1998)12213-12220]。

[0351] 上述实施例 4 中定位的两个全长 SaSSy 基因序列 (P143 和 S143) 被克隆入带有 C 末端组氨酸标记的 pET28b(+) 表达载体 (Novagen, San Diego CA)。用 NcoI 和 XhoI 切开所述载体以产生适合于连接进相同酶切 cDNA 的突出末端。含有与镍亲和标签 (His₆) 符合读框的倍半萜合酶基因的环状质粒首先被转化入化学感受态细胞并在含卡那霉素 (50 μ g ml⁻¹) 的 LB 平板上生长。

[0352] 转化子在选择性 LB 平板 (50 μ g ml⁻¹ 卡那霉素) 上生长, 并且通过质粒制备 (Invitrogen) 提取 DNA。包含插入的 pET28b(+) 载体被测序验证并且转化入化学感受态 C41 大肠杆菌细胞中 (Avidis, Saint-Beauzire, France), 其包含分离自 Rosetta 2 感受态细胞 (Novagen) 的 pRARE 2 质粒。菌落生长于含卡那霉素和氯霉素 (50 μ g ml⁻¹) 的 LB 平板上。

[0353] 挑取独立的菌落并在 5ml 含同样抗生素的 LB 中于 37°C 在振荡器中过夜生长。2ml 该培养物加入到 4 批 100ml 含卡那霉素和氯霉素的 TB 中。细胞悬液在 37°C 振荡生长 4 小时, 直至 OD₆₀₀ = 0.8。加入异丙基- β -D 型-硫代半乳糖苷 (IPTG) 至终浓度 0.2mM 并且该混合物于 16°C 振荡过夜。通过 4°C 离心收获细胞悬液并且将细胞沉淀冷冻于 -80°C 以备将来使用。

[0354] 细胞沉淀 (~7g) 重悬于 5ml 裂解液中 (25mg 溶菌酶, 0.25mg DNA 酶, 0.25mg RNA 酶, 19.3mg DTT, 5 μ l 鸡尾酒蛋白质酶抑制剂 [PMSF 和硫脲] 和 250 μ l 100mM MgCl₂)。放置于冰上时, 混合物用玻璃棒充分搅拌 30 分钟。随后裂解物用 Bronson 超声探头冰上超声至半透明。混合物于 4°C 2000 $\times$ g 离心 30 分钟。倒出上清并冰上储存。

[0355] 实施例 5

[0356] 蛋白质纯化

[0357] 澄清的裂解物上清液 (~3ml) 依次装上 HisTrap Spin 柱 (GE Healthcare) 并于 4°C 1000rpm 旋转 2 分钟。用新鲜的结合缓冲液洗涤后, 蛋白质用 300 μ l 洗脱缓冲液 (20mM HEPES pH 7.5, 150mM NaCl 和 350mM 咪唑) 洗脱两次, 至终体积 600 μ l。

[0358] 两种 SaSSy 萜烯合酶变体 (P143 和 S143) 中每种蛋白的洗脱液在 2ml 葡聚糖凝胶脱盐柱 (BioRad, Hercules CA, USA) 上脱盐入无咪唑的洗脱缓冲液中。

[0359] 洗脱液用分光光度法分析以确定近似的蛋白质浓度。终浓度为 $\sim 8\text{mgml}^{-1}$ 。这些溶液用于测定。纯化后的蛋白质分装与 80% 甘油混合 (1 : 1) 并于液氮中冷冻。其被储存于 -80°C 直至使用。用 SDS-PAGE、免疫印迹和免疫标记用 His₆ 抗体 (Sigma) 分析蛋白质。抗体通过 NBT/BCIP 法可视化 (Roche, USA)。P143 和 S143 各自的重组蛋白证实为 $\sim 66\text{kDa}$ 的预期大小。

[0360] 实施例 6

[0361] 酶测定

[0362] 重组同工酶蛋白质 P143 和 S143 的酶活性测定一式两份地使用 GC 小瓶完成, 如 O' Maille P.E. 等人, [A single-vial analytical and quantitative gas chromatography-mass spectrometry assay for terpene synthases. Anal. Biochem. 335 (2004) 210-217]。对每一测定, $20\mu\text{l}$ 蛋白质 (8mg ml^{-1}) 加入至 $560\mu\text{l}$ 含 25mM HEPES, 10% 甘油, 5mM DTT 和 10mM Mg^{2+} 的反应缓冲液中。加入 $20\mu\text{l}$ 底物 FPP 和 GPP ($\sim 1\text{mg ml}^{-1}$) 并且轻轻振荡混合。酶终浓度为 0.3mgml^{-1} 。小瓶由 $500\mu\text{l}$ 正戊烷覆盖以捕捉挥发性产物并且在 30°C 温育 2 小时。

[0363] 实施例 7

[0364] GC-MS 分析和产物鉴定

[0365] SaSSy 酶的产物混合物在带有使用氦气做气体载体的 5937MSD 的 Agilent 6890GC 上分析。使用 HP5-MS 毛细管柱 (30m, 0.25mm ID, 0.25mm 薄膜厚度) 完成峰鉴定。进样器温度为 240°C , 检测器温度设定在 250°C , 进样体积为 $1\mu\text{l}$ 并且柱流速为 1ml min^{-1} 。柱箱温度程序如下: 在 40°C 保持 1 分钟, 然后以 7.5°C 每分钟升至 250°C 并保持。所有的质谱在 70eV, 多重离子扫描模式下检测。

[0366] SaSSy 温育产物的身份通过与可靠的檀香油的质谱保留时间 (NIST 2005 库, Adams, R. P. 出版 [Identification of essential oil components by gas chromatography/mass spectrometry, Allured Publishing Corporation, Carol Stream, Illinois, 1995] 和 MassFinder (www.massfinder.com)) 比较而确认。

[0367] 使用重组 SaSSy 酶以 GPP 作为底物在 Mn^{2+} 存在下的酶活测定产生了 α -蒎烯、蒎烯、柠檬烯和 α -异松油烯的混合物, 以及其它由 GC-MS 确定的单萜产物。使用 SaSSy 以 FPP 作为底物在包含 Mg^{2+} 离子的缓冲液中的酶活测定产生了 α -檀香烯 (34.7%)、 α -反式-香柠檬烯 (11.1%)、 β -檀香烯 (4.3%) 和 β -檀香烯 (41.3%), 以及少量的由 GC-MS 确定的若干其它化合物 (图 2)。SaSSy 酶产生的 4 种化合物的质谱 (示于图 2) 与产于檀香树 (见用于比较谱的上述引用的 Adams (1995)) 中的天然 β -檀香烯是相同的。

[0368] FPP 与热变性的酶温育得不到可检测的产物, 暗示了所述酶对产物的转换是不可避免的。

[0369] 实施例 8

[0370] 核酸序列比对

[0371] 萜烯合酶的 Clustal 多重序列比对 [FB299123-125- 香根草 (Vetiveria zizanoides) 萜烯合酶 (WO 2006134523), AF484125- 烟草 (Nicotiana attenuata) 5-表-aristolochene 合酶, AB438045- 柠檬香桃 (Backhousia citriodora) 芳樟醇合酶, Santalene- 本发明的来自印度檀香的 SaSSy 檀香烯合酶] 示于图 6。柠檬香桃

芳樟醇合酶是最同源的蛋白质,然而,在檀香烯和其它 TPS 基因间存在普遍的相似性缺失。顶部的单线表明近似的区域(由 Back 和 Chappell(1996) [Proc. Natl. Acad. Sci. USA 93] 指出)其对特异性产物的形成是重要的,而双线表明 TEAS 的比率特异性区域。

[0372] 在氨基酸水平, SaSSy 檀香烯合酶与之前发现的单萜烯合酶 SamonoTPS1 (EU798692.1) 具有 56% 相同性,与假定的蓖麻柠檬烯合酶 (EQ973796.1) 具有 48% 相同性,并且与柠檬香桃芳樟醇合酶 (BAG82825.1) 具有 45% 相同性。

[0373] 实施例 9

[0374] 植物材料收集和 RNA 提取

[0375] 在种植于由 Forest Products Commission of WA 管理的种植园中的成熟印度檀香、新克里多尼亚檀香和澳洲檀香树上钻若干 25mm 的洞。心材-边材过渡区的刨花被收集并立即冷冻于液氮中。样品被运至实验室,在那里使用修改的方案提取 RNA [N. Kolosova, B. Miller, S. Ralph, B. Ellis, C. Douglas, K. Ritland, J. Bohlmann, BioTechniques 36(2004)821-824]。刨花 (10g) 放置于液氮中并且加入 RNA 提取缓冲液 (200mM Tris-HCl, pH 值 8.5, 1.5% 十二烷基磺酸锂, 300mM LiCl, 10mM EDTA, 1% w/v 脱氧胆酸钠, 1% w/v Tergitol Nonidet P-40)。使用前向缓冲液中加入 RNA 酶抑制剂 (5mM 硫脲、1mM 金黄三羧酸、10mM 二硫苏糖醇和 2% 聚乙烯吡咯烷酮)。所有溶液由 DEPC 处理的、高压灭菌的水制备。如果可能,来自同一树的样品在 TE/NaCl 重悬步骤可合并以浓缩 RNA 样品。LiCl 沉淀后, RNA 储存于 -80℃ 直至运至 UBC Vancouver, 加拿大, 建立 cDNA 文库。

[0376] 实施例 10

[0377] 印度檀香 cDNA 文库构建

[0378] 1.4 μg 印度檀香木质部总 RNA 用于创建 cDNA 文库, 其使用 Clontech SMART-Creator 文库构建试剂盒。RNA 用 SuperScript III 反转录酶 (Invitrogen) 反转录。cDNA 用 Phusion high-fidelity DNA 聚合酶和通用引物 (Clontech) 来扩增。用 SfiI 限制性内切酶消化并将其克隆进预先切开的 pDNR-LIB 载体。混合物被电穿孔转化入 25 μl 噬菌体耐受的电感受态大肠杆菌细胞。确定该文库的滴度并将其送往 Genome Science Centre, Vancouver, 加拿大, 进行双向 Sanger 法 DNA 测序。使用 CAP3 程序的默认设置组装短片段。序列 (6000 独特的短片段) 与 GenBank 数据库进行比对以寻找关键的特定代谢基因, 尤其是 TPS 基因和细胞色素 P450。

[0379] 实施例 11

[0380] 通过 RACE 发现直系同源 TPS 基因

[0381] 以前述相同的方式产生新克里多尼亚檀香和澳洲檀香的 cDNA, 只不过所述 cDNA 是直接用作 PCR 模板。基于每一基因 ORF 的引物用于扩增 (表 3)。凡不能扩增产物, 5' - 和 3' - RACE 分别用于获得各自的 UTR 以设计更特异的引物。使用巢式引物方法两轮扩增 SesquiTPS1 基因的直系同源基因。

[0382] 表 3. 引物

[0383] 开放读码框引物

[0384]

基因	正向引物	反向引物
<i>SaSSy</i>	<i>ATGGATTCTTCCACCGCCACCGCC</i>	<i>CGAGCTTACTACTCCTCGCCGAGAGG</i>
<i>SauSSy</i>	<i>ATGGATTCTTCCACCGCCACCGCC</i>	<i>CGAGCTTACTACTCCTCGCCGAGAGG</i>
<i>SspiSSy</i>	<i>ATGGATTCTTCCACCGCCACCGCC</i>	<i>CGAGCTTACTACTCCTCGCCGAGAGG</i>

[0385] pET28b(+) 克隆引物下划线的为并入的限制性位点

[0386]

基因	正向引物	反向引物
<i>SaSSy</i>	<i>ATCCATGGATTCTTCCACCGCC</i>	<i>ATCTCGAGCTCCTCGCCGAGAGG</i>
<i>SauSSy</i>	<i>ATCCATGGATTCTTCCACCGCC</i>	<i>ATCTCGAGCTCCTCGCCGAGAGG</i>
<i>SspiSSy</i>	<i>ACGGATCCAATGGATTCTTCCACCGCCAC</i>	<i>TACTCGAGTTACTACTCCTCGCCGAG</i>

[0387] 所有的产物首先克隆入高拷贝存储载体 (TOP0 Zero Blunt, Invitrogen) 中以在克隆入表达载体前进行测序。从印度檀香和澳洲檀香的若干基因型中扩增的 TPS 基因被克隆并测序以检验 ORF 中潜在的核苷酸多态性。所有三个檀香种的三种 TPS 基因的基因组序列也被克隆并测序。将与成功的 cDNA 扩增所用的相同的 ORF 引物用于基因组 DNA 上, 其提取于与 RNA 提取所实施的同样个体。这些较大的 gDNA 片段 (3-4kb) 被克隆并送去测序 (Macrogen, Korea)。

[0388] 实施例 12

[0389] 细菌表达和蛋白质分离

[0390] 所有的 TPS 基因符合读码地克隆入带有多聚组氨酸标签的 pET28b(+) 表达载体 (Novagen, San Diego CA)。根据可用的限制性酶切位点, His₆ 标签为 N 末端的或 C 末端的。带有适当限制性位点的引物用于扩增每个基因, 其然后被消化, 凝胶纯化并克隆入 pET28b(+) 载体 (表 3)。所有 pET28b(+) 构建体都在进行重组蛋白质表达之前进行测序验证。

[0391] 包含 TPS 基因的表达载体被转化入化学感受态 C41 大肠杆菌细胞 (Avidis, Saint-Beauzire, France), 其包含分离自 Rosetta 2 感受态细胞 (Novagen) 的 pRARE 2 质粒。菌落生长于含卡那霉素和氯霉素 (50 μ g ml⁻¹) 的 LB 平板上。挑取三个独立的菌落并在 5ml 含同样抗生素的 LB 中于 37°C 在振荡器中过夜生长。培养物用于接种至 400ml 选择性 Terrific 肉汤中。细胞悬液在 37°C 振荡生长 4 小时, 直至 OD₆₀₀ = 0.8。加入异丙基- β -D 型-硫代半乳糖苷 (IPTG) 至终浓度 0.2mM 并且混合物于 16°C 振荡过夜。通过 4°C 离心收获细胞悬液并且细胞沉淀 (~1g) 冷冻于 -80°C 以备将来使用。

[0392] 细胞沉淀重悬于 5ml 含 1mg ml⁻¹ 溶菌酶, 1mM MgCl₂, 5mM DTT, 0.01mg ml⁻¹ DNA 酶 I 和 RNA 酶 I, 100 μ l 鸡尾酒蛋白质酶抑制剂 (Sigma) 的裂解缓冲液中并且置于 His-trap 结合缓冲液 (20mM Na₂HPO₄ pH 值 7.4, 500mM NaCl, 30mM 咪唑, pH 值 7.4) 中。在冰上, 细胞重悬液用玻璃棒充分搅拌 30 分钟。随后裂解物用高压细胞粉碎仪均质化, 直至混合物为半透明, 并进一步用 5ml 裂解缓冲液漂洗。在转出前, 裂解液于 4°C 12000 $\times$ g 离心 1.25 小时。澄清的裂解液 (~12ml) 用 Ni²⁺ 亲和层析 (GE 医疗) 法纯化并且用 600 μ l 洗脱缓冲液洗脱 (20mM Na₂HPO₄ pH 值 7.4, 500mM NaCl, 500 mM 咪唑, pH 值 7.4)。洗脱的蛋白质在 PD-10 脱盐柱 (GE 医疗) 上用 25mM HEPES pH 值 7.4, 10% 甘油和 100mM KCl 脱盐。收集 3.5ml 洗

脱液组分,其中间部为最浓的。用 NanoDrop 分光光度计以通过氨基酸组成计算出的消光系数测定蛋白质浓度。用 SDS-PAGE 可视化纯化的蛋白质。

[0393] 实施例 13

[0394] 酶功能表征和动力学测定

[0395] 对所有重组蛋白质的酶测定一式三份地用 GC 小瓶方法完成,其由 O' Maille 等人 [Anal. Biochem. 335 (2004) 210-2175] 描述。对仅需要产物鉴定的酶测定,10 μ g 蛋白质用于终体积 500 μ l 的反应缓冲液 (25mM HEPES, 10% 甘油, 5mM DTT 和 10mM Mg^{2+} 或 Mn^{2+})。加入底物 (FPP 和 GPP) 至最终反应浓度 100 μ M。小瓶由 500 μ l 正己烷覆盖以捕捉挥发性产物并且在 30°C 温育 2 小时。混合物涡旋混合 1 分钟以萃取所有挥发物,并且离心小瓶以分离有机层。

[0396] 为了确定稳态酶动力学常数,条件如之前所述,除了酶浓度保持在 10nM。底物浓度范围为 1 μ M 至 100 μ M,并且反应在 30°C 温育整整 5 分钟。加入 500 μ l pH 8.0 的 0.5M EDTA,并且如上述涡旋混匀使反应淬灭。

[0397] 当 FPP 用作底物时,三个檀香烯合酶的动力学特性非常相似,每个酶都具有约 1.65 μ M 的 K_m 。使用 FPP 的测定的 V_{max} 范围为, SaSSy 0.42 μ M min^{-1} 至 SauSSy 0.54 μ M min^{-1} 。Kcat 为 SaSSy 0.67 min^{-1} 和 SauSSy 1.66 min^{-1} ,与发表的其它合酶的 Kcat 相似。

[0398] 实施例 14

[0399] GC-MS 分析和产物鉴定

[0400] 产物混合物通过扫描模式下的 GC-MS 分析进行产物鉴定。包含三种檀香烯和 α -反式-香柠檬烯的标准样通过将 2ml 印度檀香纯油在硅柱上快速层析并用正己烷洗脱来制备。25mg 的最终产物重悬于 EtOAc,并且用 GC-FID 按以下描述的条件确定纯度。所有的质谱与 NIST 2005 库和文献相比对。使用正烷烃标准样确定了所有化合物的保留指数,并且与文献 [R. P. Adams, Identification of essential oil components by gas chromatography/mass spectrometry, Allured Publishing Corporation, Carol Stream, Illinois, 19956] 进行比较。

[0401] GC-FID 在 Shimadzu GC2010 上用 30m 0.25mm 的 ID、0.25 μ m 膜 DB-WAX 柱,以 He 作为气体载体进行。所有分析使用不分流进样 (2 μ l)。条件如下:进样器 200°C,检测器 250°C,柱流速为 1ml min^{-1} 。温度程序:40°C 3 分钟,然后每分钟 8°C 至 180°C,保持 5 分钟,然后每分钟 10°C 至 220°C,保持 10 分钟。针高度调至仅吸取所有样品小瓶中上部的有机层。GC-MS 在 Shimadzu GC2010 上进行,使用同样的 DB-WAX 柱,并以 He 作为气体载体。条件如下:进样器 200°C,MS 界面 240°C,离子源 200°C。柱箱程序:40°C 3 分钟,然后每分钟 8°C 至 180°C,保持 5 分钟,然后每分钟 10°C 至 220°C,保持 10 分钟。溶剂扣除时间设定为 5 分钟。对产物鉴定,使用总离子监测,从 m/z 45 扫描至 m/z 300。对动力学测定,使用倍半萜基本离子 m/z 91、93 和 94 的单离子监测。同样,监测了单萜烯基本离子 (m/z 69、71 和 93)。内部标准样 (异丁基苯, 30 μ M) 被加入用于覆盖每个反应的正己烷中。检测器响应因子根据更早制备的萜烯标准物来计算,并且用于在动力学分析中确定产物浓度。

[0402] 表 4. 镁或锰存在下萜烯合酶由 FPP 得到的反应产物的 GC-FID 测定。

[0403]

		SaSSy 用 FPP Mg				SaSSy 用 FPP Mn			
	保留时间	1	2	3	平均值	1	2	3	平均值
a-檀香烯	15.101	46.4	43.8	45.6	45.3	9.5	9.9	10.1	9.8
a-E-香柠檬烯	15.288	16.3	15.3	16.4	16.0	75.7	73.4	74.9	74.7
表-b-檀香烯	16.132	4.6	5.1	4.3	4.7	0.9	0.8	0.5	0.7
b-檀香烯	16.362	30.4	33.1	31.2	31.5	6.0	7.3	6.4	6.6
Z-b-金合欢烯	16.468	0.7	1.0	0.7	0.8	0.6	0.5	0.5	0.5
E-b-金合欢烯	16.927	1.7	1.7	1.7	1.7	7.4	8.0	7.5	7.6
		SauSSy 用 FPP Mg				SauSSy 用 FPP Mn			
		1	2	3	平均值	1	2	3	平均值
a-檀香烯	15.093	51.7	51.5	50.7	51.3	10.3	10.5	9.7	10.2
a-E-香柠檬烯	15.278	13.8	14.6	15.2	14.5	72.5	73.8	74.4	73.6
表-b-檀香烯	16.123	4.7	4.6	5.1	4.8	1.4	0.8	1.2	1.2
b-檀香烯	16.352	27.8	27.5	27.0	27.4	5.3	5.5	5.1	5.3

[0404]

Z-b-金合欢烯	16.458	0.6	0.6	0.5	0.6	0.9	0.9	0.9	0.9
E-b-金合欢烯	16.917	1.5	1.3	1.5	1.4	8.3	8.5	8.6	8.5
		SspiSSy 用 FPP Mg				SspiSSy 用 FPP Mn			
		1	2	3	平均值	1	2	3	平均值
a-檀香烯	15.093	46.3	47.4	48.1	47.3	7.2	7.4	7.3	7.3
a-E-香柠檬烯	15.28	19.6	19.2	18.9	19.2	79.0	78.6	79.1	78.9
表-b-檀香烯	16.125	4.6	4.2	4.0	4.2	1.4	1.7	0.7	1.3
b-檀香烯	16.354	26.4	26.4	26.3	26.3	3.3	3.4	3.5	3.4
Z-b-金合欢烯	16.46	0.7	0.6	0.8	0.7	1.0	1.0	1.0	1.0
E-b-金合欢烯	16.919	2.5	2.2	2.0	2.2	8.1	8.0	8.4	8.2

[0405] 根据上述涉及所公开的发明的教导,以上描述的本发明多种实施方案实施模式的修改对本领域技术人员是显而易见的。上述本发明的实施方案仅仅是示范性的并且不应以任何方式被解释为限制性的。

[0001]

序列表

<110> 西澳大学

<120> 新基因和产物

<211> Length : 1713

SequenceName : SEQ ID NO:1

SequenceDescription : SaSSy DNA

<212> DNA

<213> Santalum album

<400> 1

```
atggattcctt ccaccgccac cgccatgaca gctccattca ttgatcctac tgatcatgtg 60
aatctcaaaa ctgatacggg tgcctcagag aatcgaagga tgggaaatta taaaccacgc 120
atttggaatt atgatttttt acaatcactt gcaactcacc acaatattgt ggaagagagg 180
catctaaagc tagctgagaa gctgaagggc caagtgaagt ttatgtttgg ggcaccaatg 240
gagccgttag caaagctgga gcttgtggat gtggttcaaa ggcttgggct aaaccaccta 300
tttgagacag agatcaagga agcgcgtgtt agtatttaca aggatgggag caatggatgg 360
tgglttggcc accttcacgc gacatctctc cgatttaggc tgctacgaca gtgtgggctt 420
tttattcccc aagatgtgtt taaaacgttc caaaacaaga ctggggaatt tgatatgaaa 480
ctttgtgaca acgtaaaagg gctgctgagc ttatatgaag cttcatactt ggcatggaag 540
ggtgaaaaca tcctagatga agccaaggcc ttcaccacca agtgcttgaa aagtgcattg 600
gaaaatatat ccgaaaagtg gttagccaaa agagtgaagc atgcattggc tttgcctttg 660
cattggagag tccctcgaat cgaagctaga tggttcattg aggcataatg gcaagaagcg 720
aatatgaacc caacactact caaactcgca aaattagact ttaatatggt gcaatcaatt 780
catcagaaag agattgggga attagcaagg tgggtgggtga ctactggctt ggataagtta 840
gcctttgccg ggaataattt actgcagagc tatatgtgga gctgcgcgat tgcttccgac 900
ccgaagtcca aacttgctag agaaactatt gtcgaaatcg gaagtgtact cacagtgtt 960
gacgatggat atgacgtcta tggttcaatc gacgaacttg atctctacac aagctccgtt 1020
gaaagggtga gctgtgtgga aattgacaag ttgccaaaca cgttaaaatt aatttttatg 1080
tctatgttca acaagaccaa tgaggttggc cttcgagtcc agcatgagcg aggtacaaat 1140
agcatcccta cttttatcaa agcgtggggt gaacagtgtg aatcatacca gaaagaagca 1200
```

[0002]

```

agatggttcc acgggggaca cacgcctcca ttggaagaat atagcttgaa tggacttgtt 1260
tccataggat tccctctctt gttaatcacg ggctacgtgg caatcgctga gaacgaggct 1320
gcactggata aagtgcaccc ccttcctgat cttctgcact actcctccct ccttagtcgc 1380
cicatcaatg atataggaac gtctccggat gagatggcaa gaggcgataa tctgaagtca 1440
atccattgtt acatgaacga aactggggct lccgaggaag ttgctcgtga gcacataaag 1500
ggagtaatcg aggagaattg gaaaatactg aatcagtgtt gctttgatca atctcagttt 1560
caggagcctt ttataacctt caatttgaac tctgttcgag ggtctcattt cttctatgaa 1620
tttggggatg gctttggggg gacggatagc tggacaaagg ttgatatgaa gtccgttttg 1680
atcgacccta ttcctctcgg cgaggagtag taa 1713

```

<110> University of Western Australia

<120> Novel gene and product

Length : 569

SequenceName : SEQ ID NO:2

SequenceDescription : SaSSy Protein

<212> PRT

<213> Santalum album

<400> 1

```

Met Asp Ser Ser Thr Ala Thr Ala Met Thr Ala Pro Phe Ile Asp Pro
1           5           10           15

```

```

Thr Asp His Val Asn Leu Lys Thr Asp Thr Asp Ala Ser Glu Asn Arg
          20           25           30

```

```

Arg Met Gly Asn Tyr Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln
          35           40           45

```

```

Ser Leu Ala Thr His His Asn Ile Val Glu Glu Arg His Leu Lys Leu
          50           55           60

```

```

Ala Glu Lys Leu Lys Gly Gln Val Lys Phe Met Phe Gly Ala Pro Met
65           70           75           80

```

[0003]

Glu Pro Leu Ala Lys Leu Glu Leu Val Asp Val Val Gln Arg Leu Gly		
85	90	95
Leu Asn His Leu Phe Glu Thr Glu Ile Lys Glu Ala Leu Phe Ser Ile		
100	105	110
Tyr Lys Asp Gly Ser Asn Gly Trp Trp Phe Gly His Leu His Ala Thr		
115	120	125
Ser Leu Arg Phe Arg Leu Leu Arg Gln Cys Gly Leu Phe Ile Pro Gln		
130	135	140
Asp Val Phe Lys Thr Phe Gln Asn Lys Thr Gly Glu Phe Asp Met Lys		
145	150	155
Leu Cys Asp Asn Val Lys Gly Leu Leu Ser Leu Tyr Glu Ala Ser Tyr		
165	170	175
Leu Gly Trp Lys Gly Glu Asn Ile Leu Asp Glu Ala Lys Ala Phe Thr		
180	185	190
Thr Lys Cys Leu Lys Ser Ala Trp Glu Asn Ile Ser Glu Lys Trp Leu		
195	200	205
Ala Lys Arg Val Lys His Ala Leu Ala Leu Pro Leu His Trp Arg Val		
210	215	220
Pro Arg Ile Glu Ala Arg Trp Phe Ile Glu Ala Tyr Glu Gln Glu Ala		
225	230	235
Asn Met Asn Pro Thr Leu Leu Lys Leu Ala Lys Leu Asp Phe Asn Met		
245	250	255
Val Gln Ser Ile His Gln Lys Glu Ile Gly Glu Leu Ala Arg Trp Trp		
260	265	270
Val Thr Thr Gly Leu Asp Lys Leu Ala Phe Ala Arg Asn Asn Leu Leu		
275	280	285

[0004]

Gln Ser Tyr Met Trp Ser Cys Ala Ile Ala Ser Asp Pro Lys Phe Lys
290 295 300
Leu Ala Arg Glu Thr Ile Val Glu Ile Gly Ser Val Leu Thr Val Val
305 310 315 320
Asp Asp Gly Tyr Asp Val Tyr Gly Ser Ile Asp Glu Leu Asp Leu Tyr
325 330 335
Thr Ser Ser Val Glu Arg Trp Ser Cys Val Glu Ile Asp Lys Leu Pro
340 345 350
Asn Thr Leu Lys Leu Ile Phe Met Ser Met Phe Asn Lys Thr Asn Glu
355 360 365
Val Gly Leu Arg Val Gln His Glu Arg Gly Tyr Asn Ser Ile Pro Thr
370 375 380
Phe Ile Lys Ala Trp Val Glu Gln Cys Lys Ser Tyr Gln Lys Glu Ala
385 390 395 400
Arg Trp Phe His Gly Gly His Thr Pro Pro Leu Glu Glu Tyr Ser Leu
405 410 415
Asn Gly Leu Val Ser Ile Gly Phe Pro Leu Leu Leu Ile Thr Gly Tyr
420 425 430
Val Ala Ile Ala Glu Asn Glu Ala Ala Leu Asp Lys Val His Pro Leu
435 440 445
Pro Asp Leu Leu His Tyr Ser Ser Leu Leu Ser Arg Leu Ile Asn Asp
450 455 460
Ile Gly Thr Ser Pro Asp Glu Met Ala Arg Gly Asp Asn Leu Lys Ser
465 470 475 480
Ile His Cys Tyr Met Asn Glu Thr Gly Ala Ser Glu Glu Val Ala Arg
485 490 495
Glu His Ile Lys Gly Val Ile Glu Glu Asn Trp Lys Ile Leu Asn Gln

[0005]

	500	505	510
Cys Cys Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Ile Thr Phe Asn	515	520	525
Leu Asn Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly	530	535	540
Phe Gly Val Thr Asp Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu	545	550	555
			560
Ile Asp Pro Ile Pro Leu Gly Glu Glu	565		

<110> University of Western Australia

<120> Novel gene and product

<211> Length : 1713

SequenceName : SEQ ID NO:3

SequenceDescription : SauSSy DNA

<212> DNA

<213> Santalum austrocaledonicum

<400> 1

```

atggattctt ccaccgccac cgccatgaca gtcattca ttgacctac tgatcatgtg 60
aatctcaaaa ctgatactga tgcctcagag aatcgaagga tggggaatta taaaccagc 120
atttgaatt atgatttttt acaatcactt gcaactcatc acaatattgt ggaagagagg 180
catctaaagc tagctgagaa gctgaagggc caagtgaagt ttatgtttgg ggcaccaatg 240
gagccgttag caaagctgga gcttgtggat gtggttcaaa ggctcgggct aaaccaccga 300
tttgagacag agatcaagga agcgtatatt agtatttata aggacgagag caatggatgg 360
tggtttggcc accttcatgc gacatctctc cgatttaggc tgctacgaca gtgtgggctt 420
tttatcccc aggatgtgtt taaaacgttc caaaacaaaa ctggtgaatt tgatatgaaa 480
ctgtgtgaca acgtaaaagg gctgctgagc ttatatgaag cttcatactt gggatggaag 540
ggtgaaaaca tcctagatga agccaaggcc ttcgccacca agtacttgaa aagtgcattg 600
gaaaatatat ctgaaaagtg gttagccaaa agagtgaagc atgcattggc ttacctttg 660

```

[0006]

cattggagag tccctcgaat cgaagctaga tggttcattg aggcataatga gcaagaagcg 720
 aatatgaacc caacactact caaactcgca aaattagact ttaatatggt gcaatcaatt 780
 catcagaaag agattgggga attagcaagg tggtaggtga ctactggctt ggataagtta 840
 gcctttgcta ggaataatct actgcaaagc tatatgtgga gctgcgcgat tgcctccgac 900
 ccgaagtcca aacttgctag agaaactatt gtcgaaatcg gaagtgtact cacagttggt 960
 gatgatgcat atgacgtcta tggttcaatg gacgaacttg atctctacac aagctccgtt 1020
 gaaagggtga gctgtgtaga aattgacaag ttgccaaaca cgtaaaatt gatTTTTatg 1080
 tctatgttta ataagaccaa tgaggttggc cttcgagtcc agcatgagcg aggctacaat 1140
 agcatcccta cttttatcaa agcgtgggtt caacagtgtg aatcatacca gaaagaagca 1200
 agatggttcc acgggggaca cacgcctccg ttggaagaat atagcttgaa tggacttggt 1260
 tccataggat tccctctctt gtgatcacc ggctacgtgg caatcgctga gaacgaggct 1320
 gcactggata aagtgcaccc ctttctgat cttctgcact actcctccct ccttagtcgc 1380
 ctcalcaatg atataggaac gtcctcggat gagatggcaa gaggcgataa tctgaagta 1440
 atccattggt acatgaacgg aactggggct tccgaggaag ttgctcgtga gcacataaag 1500
 ggagtaatcg aggagaattg gaaaatactg aatcagtgtc gctttgatca atctcagttt 1560
 caggagcctt ttataacctt caatttgaac tctgttcgag ggtctcattt cttctatgaa 1620
 tttggggatg gctttggggt gacggatagc tggacaaagg ttgatatgaa gtccgttttg 1680
 attgacccta ttctctcgg cgaggagtag taa 1713

<110> University of Western Australia

<120> Novel gene and product

<211> Length : 569

SequenceName : SEQ ID NO:4

SequenceDescription : SauSSy Protein

<212> PRT

<213> Santalum austrocaledonicum

<400> 1

Met Asp Ser Ser Thr Ala Thr Ala Met Thr Ala Pro Phe Ile Asp Pro

[0007]

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

[0008]

Pro Arg Ile Glu Ala Arg Trp Phe Ile Glu Ala Tyr Glu Gln Glu Ala			
225	230	235	240
Asn Met Asn Pro Thr Leu Leu Lys Leu Ala Lys Leu Asp Phe Asn Met			
	245	250	255
Val Gln Ser Ile His Gln Lys Glu Ile Gly Glu Leu Ala Arg Trp Trp			
	260	265	270
Val Thr Thr Gly Leu Asp Lys Leu Ala Phe Ala Arg Asn Asn Leu Leu			
	275	280	285
Gln Ser Tyr Met Trp Ser Cys Ala Ile Ala Ser Asp Pro Lys Phe Lys			
	290	295	300
Leu Ala Arg Glu Thr Ile Val Glu Ile Gly Ser Val Leu Thr Val Val			
305	310	315	320
Asp Asp Ala Tyr Asp Val Tyr Gly Ser Met Asp Glu Leu Asp Leu Tyr			
	325	330	335
Thr Ser Ser Val Glu Arg Trp Ser Cys Val Glu Ile Asp Lys Leu Pro			
	340	345	350
Asn Thr Leu Lys Leu Ile Phe Met Ser Met Phe Asn Lys Thr Asn Glu			
	355	360	365
Val Gly Leu Arg Val Gln His Glu Arg Gly Tyr Asn Ser Ile Pro Thr			
	370	375	380
Phe Ile Lys Ala Trp Val Gln Gln Cys Lys Ser Tyr Gln Lys Glu Ala			
385	390	395	400
Arg Trp Phe His Gly Gly His Thr Pro Pro Leu Glu Glu Tyr Ser Leu			
	405	410	415
Asn Gly Leu Val Ser Ile Gly Phe Pro Leu Leu Leu Ile Thr Gly Tyr			
	420	425	430

[0009]

Val Ala Ile Ala Glu Asn Glu Ala Ala Leu Asp Lys Val His Pro Leu
435 440 445

Pro Asp Leu Leu His Tyr Ser Ser Leu Leu Ser Arg Leu Ile Asn Asp
450 455 460

Ile Gly Thr Ser Pro Asp Glu Met Ala Arg Gly Asp Asn Leu Lys Ser
465 470 475 480

Ile His Cys Tyr Met Asn Gly Thr Gly Ala Ser Glu Glu Val Ala Arg
485 490 495

Glu His Ile Lys Gly Val Ile Glu Glu Asn Trp Lys Ile Leu Asn Gln
500 505 510

Cys Cys Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Ile Thr Phe Asn
515 520 525

Leu Asn Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly
530 535 540

Phe Gly Val Thr Asp Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu
545 550 555 560

Ile Asp Pro Ile Pro Leu Gly Glu Glu
565

<110> University of Western Australia

<120> Novel gene and product

<211> Length : 1713

SequenceName : SEQ ID NO:5

SequenceDescription : SspiSSy DNA

<212> DNA

<213> Santalum spicatum

<400> 1

atggattctt ccaccgccac cgccacgaca gtcatttta ttgatcatac tgatcatgtg 60

aatcttaaaa ttgataatga ttctccgag agtcgaagga tgggaatta taaaccagat 120

[0010]

atttgaatt atgattttct gcaatcactt gcaatccatc acaatatgtt ggaagagaag 180
catctaaagc tagctgagaa gctgaagggc caagtgatgt ctatgtttgg ggcaccaatg 240
gagccgttag caaagctgga gcttgtggat gtggttcaaa ggcttgggct aaaccaccaa 300
tttgagacag agatcaagga agccctatit agtgtttaca aggatgggag caatggatgg 360
tggtttggcc accttcatgc aacatctctt cgatttaggc tactacgaca gtgtgggctt 420
tttatcccc aggatgtgtt taaaacgttc cagagcaaaa ctgatgaatt tgatatgaaa 480
ctgtgtgaca acataaaagg gttgttgagc ttgtatgaag cttcattcct ggggtggaag 540
ggtgaaaaca tcctagatga agccaaggcc ttgccacca agtacttgaa aaatgcatgg 600
gaaaatatat cccaaaagt gctagccaaa agagtgaagc atgactggc tttgcctctg 660
cactggagag tccctcgaat cgaggctaga tggttcattg aggcataatga gcaagaagag 720
aacatgaacc caacactact caaacttgca aaattagact ttaacatggt gcaatcaatt 780
catcagaaag agattgggga attagcaagg tgggtgggtga ctactggctt ggataagtta 840
gcctttgcta ggaataattt actgcaaagc tatatgtgga gctgcgcgat tgcttccgac 900
ccaaagttca aacttgctag agaaactatt gtcgaaatcg gaagtgtact cacagttgtg 960
gacgatgcat atgatgtcta tggttcaatg gatgaactg atcactacac atactccgtt 1020
gaaaggtgga gctgtgtaga aattgacaag ctgccaaaca cgttaaaatt gatttttatg 1080
tctatgttca acaagaccaa tgaggttggc cttcgagtcc agcatgagcg aggcataaac 1140
ggcatcccta cttttatcaa agcatgggtt gaacagtga aagcatacca gaaagaggca 1200
agatggtacc atgggggaca cacgcctcca ttggaggaat atagcttgaa tggacttggt 1260
tccataggat tccctctctt gttgatcacc ggctacatcg caatcgctga gaacgagget 1320
gcactggata aagtgcaccc ctttctgat cttctgcact actcctccct ccttagtcgc 1380
ctcatcaatg acatgggaac gtctccggac gagatggcaa gaggtgacaa tctgaagtca 1440
atccactgtt acatgaacga aactggggct tctgaggaag ttgctcgtga gcacataaaa 1500
ggaataatcg aggagaattg gaaaatactg aatcagtgtc gctttgatca atctcagttt 1560
caggagcctt ttataacctt caatttgaac tctgttcgag ggtctcattt cttctatgaa 1620
tttggggaig gctttggggt gacagatagc tggacaaagg ttgatatgaa gtctgttttg 1680

[0011]

atcgacccta ttcctctcgg cgaggagtag taa

1713

<110> University of Western Australia

<120> Novel gene and product

<211> Length : 569

SequenceName : SEQ ID NO:6

SequenceDescription : SspiSSy Protein

<212> PRT

<213> Santalum spicatum

<400> 1

Met Asp Ser Ser Thr Ala Thr Ala Thr Thr Ala Pro Phe Ile Asp His
1 5 10 15

Thr Asp His Val Asn Leu Lys Ile Asp Asn Asp Ser Ser Glu Ser Arg
20 25 30

Arg Met Gly Asn Tyr Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln
35 40 45

Ser Leu Ala Ile His His Asn Ile Val Glu Glu Lys His Leu Lys Leu
50 55 60

Ala Glu Lys Leu Lys Gly Gln Val Met Ser Met Phe Gly Ala Pro Met
65 70 75 80

Glu Pro Leu Ala Lys Leu Glu Leu Val Asp Val Val Gln Arg Leu Gly
85 90 95

Leu Asn His Gln Phe Glu Thr Glu Ile Lys Glu Ala Leu Phe Ser Val
100 105 110

Tyr Lys Asp Gly Ser Asn Gly Trp Trp Phe Gly His Leu His Ala Thr
115 120 125

Ser Leu Arg Phe Arg Leu Leu Arg Gln Cys Gly Leu Phe Ile Pro Gln
130 135 140

[0012]

Asp Val Phe Lys Thr Phe Gln Ser Lys Thr Asp Glu Phe Asp Met Lys
145 150 155 160

Leu Cys Asp Asn Ile Lys Gly Leu Leu Ser Leu Tyr Glu Ala Ser Phe
165 170 175

Leu Gly Trp Lys Gly Glu Asn Ile Leu Asp Glu Ala Lys Ala Phe Ala
180 185 190

Thr Lys Tyr Leu Lys Asn Ala Trp Glu Asn Ile Ser Gln Lys Trp Leu
195 200 205

Ala Lys Arg Val Lys His Ala Leu Ala Leu Pro Leu His Trp Arg Val
210 215 220

Pro Arg Ile Glu Ala Arg Trp Phe Ile Glu Ala Tyr Glu Gln Glu Glu
225 230 235 240

Asn Met Asn Pro Thr Leu Leu Lys Leu Ala Lys Leu Asp Phe Asn Met
245 250 255

Val Gln Ser Ile His Gln Lys Glu Ile Gly Glu Leu Ala Arg Trp Trp
260 265 270

Val Thr Thr Gly Leu Asp Lys Leu Ala Phe Ala Arg Asn Asn Leu Leu
275 280 285

Gln Ser Tyr Met Trp Ser Cys Ala Ile Ala Ser Asp Pro Lys Phe Lys
290 295 300

Leu Ala Arg Glu Thr Ile Val Glu Ile Gly Ser Val Leu Thr Val Val
305 310 315 320

Asp Asp Ala Tyr Asp Val Tyr Gly Ser Met Asp Glu Leu Asp His Tyr
325 330 335

Thr Tyr Ser Val Glu Arg Trp Ser Cys Val Glu Ile Asp Lys Leu Pro
340 345 350

Asn Thr Leu Lys Leu Ile Phe Met Ser Met Phe Asn Lys Thr Asn Glu

[0013]

355	360	365
Val Gly Leu Arg Val Gln His Glu Arg Gly Tyr Asn Gly Ile Pro Thr 370	375	380
Phe Ile Lys Ala Trp Val Glu Gln Cys Lys Ala Tyr Gln Lys Glu Ala 385	390	395 400
Arg Trp Tyr His Gly Gly His Thr Pro Pro Leu Glu Glu Tyr Ser Leu 405	410	415
Asn Gly Leu Val Ser Ile Gly Phe Pro Leu Leu Leu Ile Thr Gly Tyr 420	425	430
Ile Ala Ile Ala Glu Asn Glu Ala Ala Leu Asp Lys Val His Pro Leu 435	440	445
Pro Asp Leu Leu His Tyr Ser Ser Leu Leu Ser Arg Leu Ile Asn Asp 450	455	460
Met Gly Thr Ser Pro Asp Glu Met Ala Arg Gly Asp Asn Leu Lys Ser 465	470	475 480
Ile His Cys Tyr Met Asn Glu Thr Gly Ala Ser Glu Glu Val Ala Arg 485	490	495
Glu His Ile Lys Gly Ile Ile Glu Glu Asn Trp Lys Ile Leu Asn Gln 500	505	510
Cys Cys Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Ile Thr Phe Asn 515	520	525
Leu Asn Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly 530	535	540
Phe Gly Val Thr Asp Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu 545	550	555 560
Ile Asp Pro Ile Pro Leu Gly Glu Glu 565		

[0014]

<110> University of Western Australia

<120> Novel gene and product

<211> Length : 1713

SequenceName : SEQ ID NO:7

SequenceDescription : SaSSy DNA

<212> DNA

<213> Santalum album

<400> 1

```
atggattctt ccaccgccac cgccatgaca gctccattca ttgatcctac tgatcatgtg 60
aatctcaaaa ctgatacgga tgccicagag aatcgaagga tgggaaatta taaaccacgc 120
atttgggaatt atgatttttt acaatcactt gcaactcatc acaatattgt ggaagagagg 180
catctaaagc tagctgagaa gctgaagggc caagtgaagt ttatgtttgg ggcaccaatg 240
gagccgtagt caaagctgga gcttgtggat gtggttcaaa ggcttgggct aaaccaccta 300
tttgagacag agatcaagga agcgcgtgtt agtatttaca aggatgggag caatggatgg 360
tggtttggcc accitcatgc gacatctctc cgatttaggc tgctacgaca gtgtgggctt 420
tttatttccc aagatgtgtt taaaacgttc caaaacaaga ctggggaatt tgatatgaaa 480
ctttgtgaca acgtaaaagg gctgctgagc ttatatgaag cttcatactt gggatggaag 540
ggtgaaaaca tcctagatga agccaaggcc ttcaccacca agtgcttgaa aagtgcattg 600
gaaaatatat ccgaaaagtg gttagccaaa agagtgaagc atgcattggc ttgcctttg 660
cattggagag tcctcgaat cgaagctaga tggttcattg aggcatatga gcaagaagcg 720
aatatgaacc caacactact caaactcgca aaattagact ttaatatggt gcaatcaatt 780
catcagaaag agattgggga attagcaagg tgggtgggtga ctactggcct ggataagtta 840
gcctttgcca ggaataatth actgcagagc tatatgtgga gctgcgcgat tgcttccgac 900
ccgaagtcca aacttgctag agaaactatt gtcgaaatcg gaagtgtact cacagttggt 960
gacgatggat atgacgtcta tggttcaatc gacgaacttg atctctacac aagctccggt 1020
gaaagggtga gctgtgtgga aattgacaag ttgccaaaca cgttaaaatt aatttttatg 1080
tctatgttca acaagaccaa tgaggttggc cttcgagtcc agcatgagcg aggctacaat 1140
agcatcccta cttttatcaa agcgtgggtt gaacagtgtg aatcatacca gaaagaagca 1200
```

[0015]

agatggttcc acgggggaca cacgcctcca ttggaagaat atagcttgaa tggacttggt 1260
tccataggat tccctctctt gttaatcacg ggctacgtgg caatcgctga gaacgaggt 1320
gcactggata aagtgcaccc ccttcctgat cttctgcaact actcctccct ccttagtcgc 1380
ctcatcaatg atalaggaac gtctccggat gagatggcaa gaggcgataa tctgaagtca 1440
atccattggt acatgaacga aactggggct tccgaggaag ttgctcgtga gcacataaag 1500
ggagtaatcg aggagaattg gaaaatactg aatcagtgtt gctttgatca atctcagttt 1560
caggagcctt ttataacctt caatttgaac tctgttcgag ggtctcattt cttctatgaa 1620
tttggggatg gctttggggt gacggatagc tggacaaagg ttgatatgaa gtccgttttg 1680
atcgacccta ttccctcgg cgaggagtag taa 1713

<110> University of Western Australia

<120> Novel gene and product

Length : 569

SequenceName : SEQ ID NO:8

SequenceDescription : SaSSy Protein

<212> PRT

<213> Santalum album

<400> 1

Met Asp Ser Ser Thr Ala Thr Ala Met Thr Ala Pro Phe Ile Asp Pro
1 5 10 15

Thr Asp His Val Asn Leu Lys Thr Asp Thr Asp Ala Ser Glu Asn Arg
20 25 30

Arg Met Gly Asn Tyr Lys Pro Ser Ile Trp Asn Tyr Asp Phe Leu Gln
35 40 45

Ser Leu Ala Thr His His Asn Ile Val Glu Glu Arg His Leu Lys Leu
50 55 60

Ala Glu Lys Leu Lys Gly Gln Val Lys Phe Met Phe Gly Ala Pro Met
65 70 75 80

[0016]

Glu	Pro	Leu	Ala	Lys	Leu	Glu	Leu	Val	Asp	Val	Val	Gln	Arg	Leu	Gly
				85					90					95	
Leu	Asn	His	Leu	Phe	Glu	Thr	Glu	Ile	Lys	Glu	Ala	Leu	Phe	Ser	Ile
			100					105					110		
Tyr	Lys	Asp	Gly	Ser	Asn	Gly	Trp	Trp	Phe	Gly	His	Leu	His	Ala	Thr
		115					120						125		
Ser	Leu	Arg	Phe	Arg	Leu	Leu	Arg	Gln	Cys	Gly	Leu	Phe	Ile	Ser	Gln
		130					135					140			
Asp	Val	Phe	Lys	Thr	Phe	Gln	Asn	Lys	Thr	Gly	Glu	Phe	Asp	Met	Lys
145					150					155					160
Leu	Cys	Asp	Asn	Val	Lys	Gly	Leu	Leu	Ser	Leu	Tyr	Glu	Ala	Ser	Tyr
				165					170					175	
Leu	Gly	Trp	Lys	Gly	Glu	Asn	Ile	Leu	Asp	Glu	Ala	Lys	Ala	Phe	Thr
			180					185						190	
Thr	Lys	Cys	Leu	Lys	Ser	Ala	Trp	Glu	Asn	Ile	Ser	Glu	Lys	Trp	Leu
		195					200						205		
Ala	Lys	Arg	Val	Lys	His	Ala	Leu	Ala	Leu	Pro	Leu	His	Trp	Arg	Val
		210				215						220			
Pro	Arg	Ile	Glu	Ala	Arg	Trp	Phe	Ile	Glu	Ala	Tyr	Glu	Gln	Glu	Ala
225					230					235					240
Asn	Met	Asn	Pro	Thr	Leu	Leu	Lys	Leu	Ala	Lys	Leu	Asp	Phe	Asn	Met
				245					250					255	
Val	Gln	Ser	Ile	His	Gln	Lys	Glu	Ile	Gly	Glu	Leu	Ala	Arg	Trp	Trp
			260					265					270		
Val	Thr	Thr	Gly	Leu	Asp	Lys	Leu	Ala	Phe	Ala	Arg	Asn	Asn	Leu	Leu
		275					280						285		
Gln	Ser	Tyr	Met	Trp	Ser	Cys	Ala	Ile	Ala	Ser	Asp	Pro	Lys	Phe	Lys

[0017]

290	295	300	
Leu Ala Arg Glu Thr Ile Val Glu Ile Gly Ser Val Leu Thr Val Val			
305	310	315	320
Asp Asp Gly Tyr Asp Val Tyr Gly Ser Ile Asp Glu Leu Asp Leu Tyr			
	325	330	335
Thr Ser Ser Val Glu Arg Trp Ser Cys Val Glu Ile Asp Lys Leu Pro			
	340	345	350
Asn Thr Leu Lys Leu Ile Phe Met Ser Met Phe Asn Lys Thr Asn Glu			
	355	360	365
Val Gly Leu Arg Val Gln His Glu Arg Gly Tyr Asn Ser Ile Pro Thr			
	370	375	380
Phe Ile Lys Ala Trp Val Glu Gln Cys Lys Ser Tyr Gln Lys Glu Ala			
385	390	395	400
Arg Trp Phe His Gly Gly His Thr Pro Pro Leu Glu Glu Tyr Ser Leu			
	405	410	415
Asn Gly Leu Val Ser Ile Gly Phe Pro Leu Leu Leu Ile Thr Gly Tyr			
	420	425	430
Val Ala Ile Ala Glu Asn Glu Ala Ala Leu Asp Lys Val His Pro Leu			
	435	440	445
Pro Asp Leu Leu His Tyr Ser Ser Leu Leu Ser Arg Leu Ile Asn Asp			
	450	455	460
Ile Gly Thr Ser Pro Asp Glu Met Ala Arg Gly Asp Asn Leu Lys Ser			
465	470	475	480
Ile His Cys Tyr Met Asn Glu Thr Gly Ala Ser Glu Glu Val Ala Arg			
	485	490	495
Glu His Ile Lys Gly Val Ile Glu Glu Asn Trp Lys Ile Leu Asn Gln			
	500	505	510

[0018]

Cys Cys Phe Asp Gln Ser Gln Phe Gln Glu Pro Phe Ile Thr Phe Asn
515 520 525

Leu Asn Ser Val Arg Gly Ser His Phe Phe Tyr Glu Phe Gly Asp Gly
530 535 540

Phe Gly Val Thr Asp Ser Trp Thr Lys Val Asp Met Lys Ser Val Leu
545 550 555 560

Ile Asp Pro Ile Pro Leu Gly Glu Glu
565

<110> University of Western Australia

<120> Novel gene and product

Length : 4088

SequenceName : SEQ ID NO:9

SequenceDescription : SaSSy genomic DNA

<212> DNA

<213> Santalum album

<400> 1

atggattcct ccaccgccac cgccatgaca gctccattca ttgatcctac tgatcatgtg 60
aatctcaaaa ctgatacgga tgcctcagag aatcgaagga tgggaaatta taaaccacgc 120
atttgggaatt atgatttttt acaatcactt gcaactcatc acaatattgt ggtacatata 180
tatatatata gtagcctata atatttaatt ttgtttgtgt cagagtttct tttcggtgat 240
aaaaatataa caagtcatac gccaccctag ggtgactcct cctggcgccc agcaaccact 300
tctcttttagt attttttttt cattttatct tttatgtttt tatgcttttt tttttttttt 360
gagaatgtga gtgtcaaact tattaacagg tccactaatt ccttgaggga gttgcgctag 420
tttttagcag cagtcttcgg ataaatgctc gccacatatg cataatatgt cgagtatatc 480
tttcagactc aaatctatca tcttaggggt gtggaccttc tcttctcttc actatttttg 540
ccgaacaata aacacaattg cttagagccc ttgtggacta gcccagagatt tgttttggett 600
aatgtctgcc acacaagctt catattatct aatataactc attcatctgt tttgctctat 660
tategtagcc tgtggatcac cttcaatatg gtgctagtaa atttttgcct ccttaaaatt 720

[0019]

gagtcacata tctctattgt gattttgtcc tacgtattat ttgttagcta atgggaggaa 780
atacacagga agagaggcat ctaaagctag ctgagaagct gaagggccaa gtgaagtta 840
tgtttggggc accaatggag ccgttagcaa agctggagct tgtggatgtg gttcaaagc 900
ttgggctaaa ccacctatit gagacagaga tcaaggaagc gctgtttagt atttacaagg 960
atgggagcaa tggatggtg tttggccacc ttcacgcgac atctctccga tttaggctgc 1020
tacgacagtg tgggcttttt attccccaag gtacgtctag cttaaattct aattaggccc 1080
atgttttaat tgttggaata taataaacgt ttccttateg ttccttagtt ataatacagc 1140
gaccatccaa tcaaaaacca attgggtaac tcttgatgta aatattaaca gagtcacgac 1200
caacatcatt aatataatit agagggtatg actgagcatt aactacgaca gtatttcttg 1260
taagtggagg tcagatatit ttttctaga tcttgtaata gaatatit tagcctctgt 1320
tatacccgtc tgtatgtg cagcacatc ctagaatit gtattcctit ctattaatit 1380
ttgtgttttc tatttgaaa ctcataact tgcataatit tgtgttaatt tctgtgata 1440
cagaacgttc caaaacaaga ctggggaatt tgatatgaaa ctttgtgaca acgtaaaagg 1500
gtgctgagc ttatatgaag cttcatactt gggatggaag ggtgaaaaca tcctagatga 1560
agccaaggcc ttcaccacca agtgcttgaa aagtgcattg gaaaatatat ccgaaaagt 1620
gttagccaaa agagtgaagc atgcattggc tttgcctttg cattggagag tccctcgaat 1680
cgaagctaga tggttcattg aggcatatga gcaagaagc aatatgaacc caacactact 1740
caaactcgca aaattagact ttaatatggt gcaatcaatt catcagaaag agattgggga 1800
attagcaagg tattatcttc atattaaatc cgtcaagtca aatgatctaa tacatggtgc 1860
acttagagt acaggggata caatattcatt atttcaatg agcaaggcta caccttttaa 1920
aggatgctat tttgagaaa gattttaaaa agatagcctt ttttatagt gcattctctt 1980
tgcatgcata catctttgtc tacttataaa caatggtatc cctaaaaag tcaaaaaaa 2040
attccctcca aaaaccttaa ttttttgaga atgggtcgat atcaaatcta ataattttct 2100
tttgccaaa aaaatctaatt acttttctcc agaaaacaga aaacctgaat agtcaaagag 2160
ctttgacccg acaagactgt gcctcgctt tcccataaa gctctacaca cacatgcact 2220
agcaatttat ctttgactt tcttctctt tcttttttt cctttcttt ttcatatit 2280
tgaigattat cattttccct aattattacc aggattatca tttcttggg taatattgtt 2340

[0020]

attattagca ttttccttgt tgaacaaatt tgatttccag gtggtgggtg actactggct 2400
tggataagtt agcctttgcc aggaataatt tactgcagag ctatatgttg agctgcgcga 2460
ttgcttccga cccgaagtgc aaacttgcta gagaaactat tgtcgaaatc ggaagtgtac 2520
tcacagttgt tgacgatgga tatgacgtct atggttcaat cgacgaactt gatctctaca 2580
caagctccgt tgaaggtta tttgtatact ctttcgatgt tccataatat atttgaatc 2640
alatcgatgg tgaagtagt aallattgaa aaatclatit glggatgaal tttcllatcg 2700
tgttacttat tcaggtggag ctgtgtggaa attgacaagt tgccaaacac gttaaaatta 2760
atttttatgt ctatgttcaa caagaccaat gaggttggcc ttcgagtcca gcatgagcga 2820
ggctacaata gcatccctac ttttatcaaa gcggtaaata attaccataa ttactatttt 2880
ctctaagata aaacaagggt taatattcca tttaatgagt attattttgt gaaaaaata 2940
ttatttcagt gggttgaaca gtgtaaatac taccagaaag aagcaagatg gttccacggg 3000
ggacacacgc ctccattgga agaatatagc ttgaatggac ttgtttccat aggattccct 3060
ctcttggtta tcacgggcta cgtggcaatc gctgagaacg aggctgcact ggataaagtg 3120
cacccecttc ctgatcttct gcaactactc tccctectta gtcgcctcat caatgatata 3180
ggaacgtctc cggtaaagtc gtaaactttt ttataagtct ttttataaac atatagtcaa 3240
atattctttt gagtccagat tcataaagtc cagaattttt atttttaaac caattggtaa 3300
gagtcataga tttttgatgt ggaactatag cctcaatata tatggcaaca cgtgggcaaa 3360
ttttacttag ggtgacatga agcatgctta gtagtagaag ttaaagaaga gtaaattgcc 3420
cgaatctcat actatgataa agtcactgtt gggcaatggt ctcgacaagc gttcacaagt 3480
ggcaacaacc tgctttgcta ccatgcaatc gttcgtggag tttcacctta aaatcaatcg 3540
gaattaggtt aaataaacca aaatttatit aattgcatat gtgggacttt aattgctccc 3600
aatattctta attcacttgg agccaccaa aagaactctt gccataatct aatgaagctc 3660
aaattgagtt ggtagtacia tattggaaat gtaaagtatt aaaggcataa gtcttgactt 3720
tgggacatcg cccaacagtc aatatcaagt gaataatatt atttggtggt tttgataatc 3780
aggatgagat ggcaagaggc gataatctga agtcaatcca ttgttacatg aacgaaactg 3840
gggcttccga ggaagttgct cgtgagcaca taaaggaggt aatcgaggag aattggaaaa 3900

[0021]

tactgaatca gtgctgcttt gatcaatctc agtttcagga gccttttata accttcaatt 3960
tgaactctgt tcgagggctct catttcttct atgaatttgg ggatggcttt ggggtgacgg 4020
atagctggac aaaggttgat atgaagtccg ttttgatcga ccctattcct ctcggcgagg 4080
agtagtaa 4088

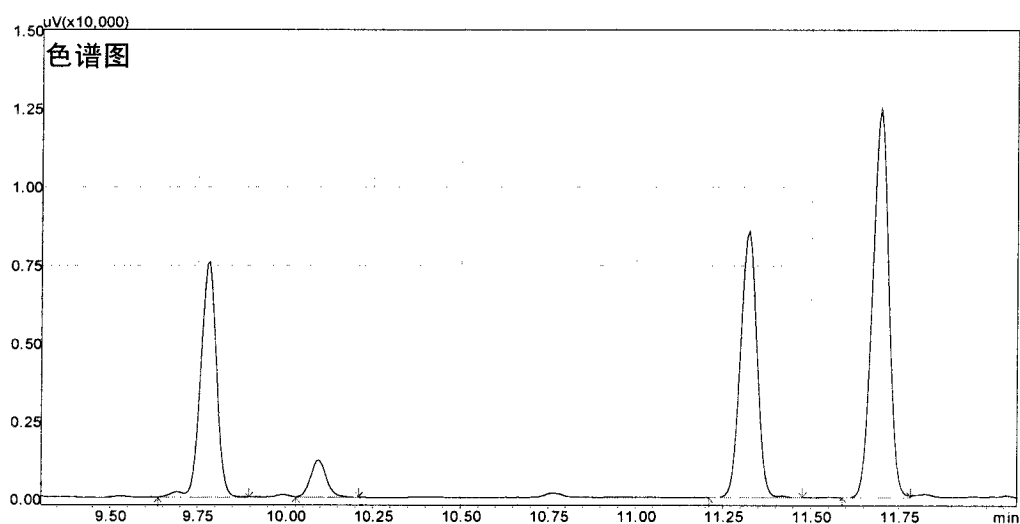


图 1

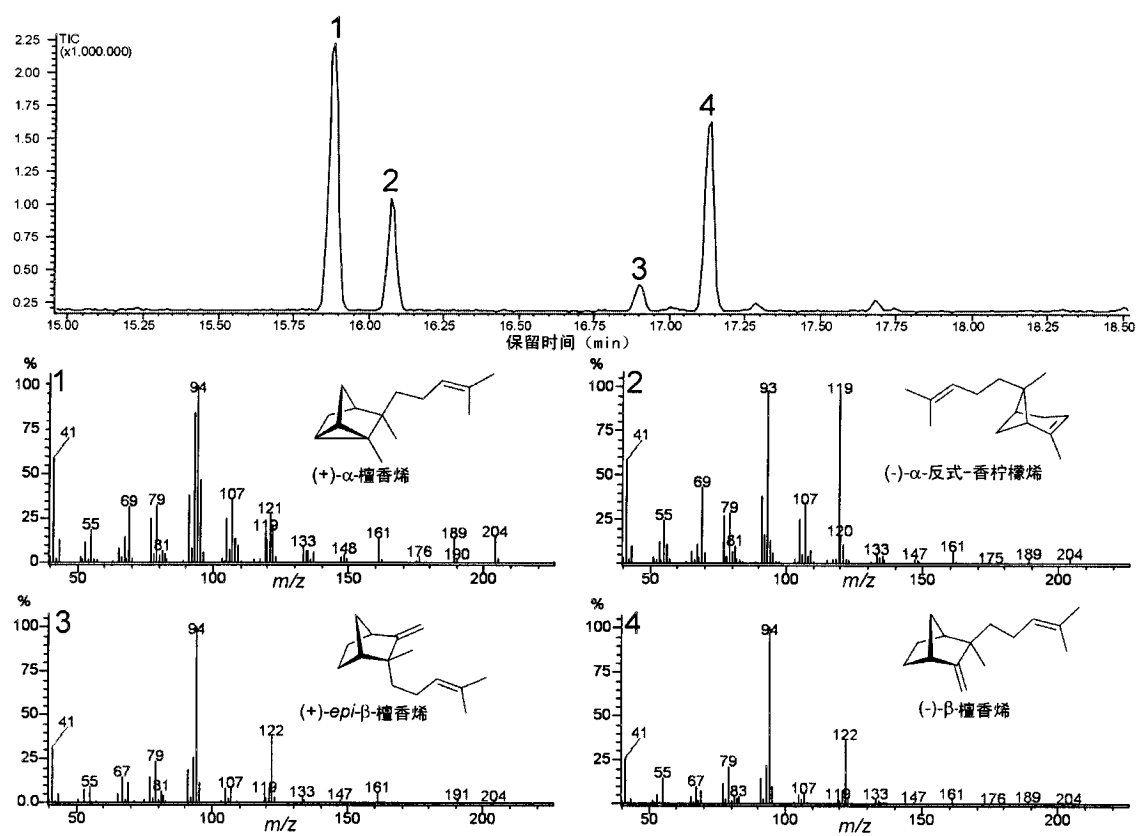


图 2

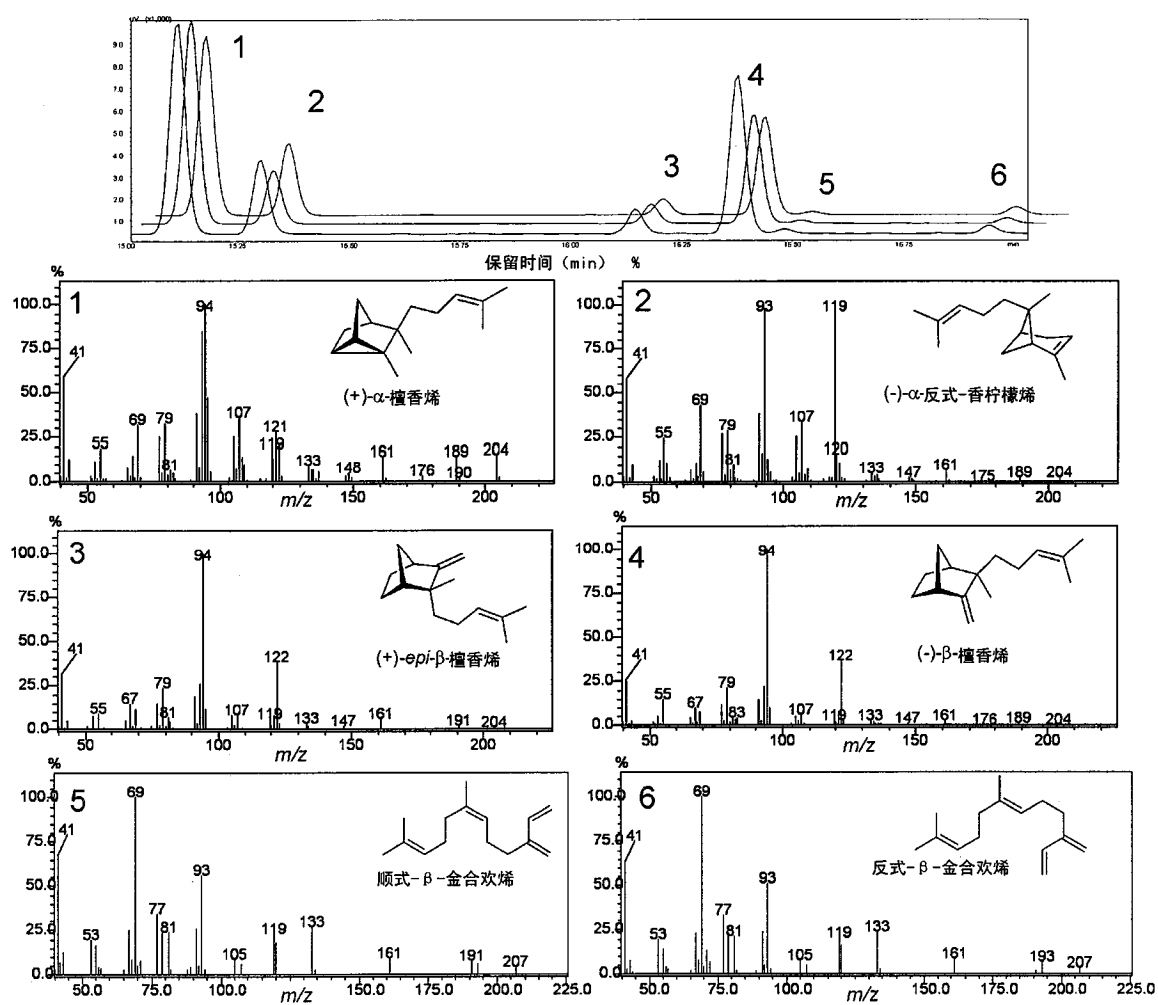


图 3

印度檀香檀香烯合酶.txt

>Contig2333 檀香烯合酶 334 (FL)

```
GGCCACAAAATAAGTCTCTTGTTCTGTTCTTTTGATCTCGTTTTATACTCAGCTCTCTCACTAATGGATTCTTCACCGCCACC
GCCATGACAGCTCCATTCATTGATCCTACTGATCATGTGAATCTCAAACTGATACGGATGCCTCAGAGAATCGAAGGATGGGA
AATTATAAACCCAGCATTTTGGAAATTATGATTTTTTACAATCACTTGCAACTCATCACAAATATTGTGGAAGAGAGGCATCTAAAG
CTAGCTGAGAAGCTGAAGGGCCAAGTGAAGTTTATGTTTGGGGCACCATGGAGCCGTTAGCAAAGCTGGAGCTTGTGGATGTG
GTTCAAAGGCTTGGGCTAAACCACCTATTTGAGACAGAGATCAAGGAAGCGCTGTTTAGTATTTACAAGGATGGGAGCAATGGA
TGGTGGTTTGGCCACCTTCATGCGACATCTCTCCGATTTAGGCTGCTACGACAGTGTGGGCTTTTATCCCAAGATGTGTTT
AAAACGTTCCAAAACAAGACTGGGGAATTTGATATGAAACTTTGTGACAACGTAAAAGGGCTGCTGAGCTTATATGAAGCTTCA
TACTTGGGATGGAAGGGTGAAAAATCCTAGATGAAGCCAAGGCCTTCAACCACCAAGTGCTTGAAAAGTGCATGGGAAAAATATA
TCCGAAAAGTGGTTAGCCAAAAGAGTGAAGCATGCATTGGCTTTGCCTTTGCATTGGAGAGTCCCTCGAATCGAAGCTAGATGG
TTCATTGAGGCATATGAGCAAGAAGCGAATATGAACCCAACACTCAAACTCGCAAAATTAGACTTTAATATGGTGCAATCA
ATTCATCAGAAAGAGATTGGGGAATTAGCAAGGTGGTGGGTGACTACTGGCTTGGATAAGTTAGCCTTTGCCAGGAATAATTTA
CTGCAGAGCTATATGTGGAGCTGCGCGATTGCTTCCGACCCGAAGTTCAAACTTGCTAGAGAAACTATTGTCGAAATCGGAAGT
GTACTCACAGTTGTTGACGATGGATATGACGCTATGGTTCAATCGACGAACTTGATCTCTACACAAGCTCCGTTGAAAGGTGG
AGCTGTGTGGAAATTGACAAGTTGCCAAACACGTTAAAATTAATTTTTATGCTATGTTCAACAAGACCAATGAGGTTGGCCTT
CGAGTCCAGCATGAGCGAGGCTACAATAGCATCCCTACTTTTATCAAAGCGTGGGTTGAACAGTGTAAATCATACCAGAAAGAA
GCAAGATGGTTCCACGGGGGACACACGCCCTCCATTGGAAGAATATAGCTTGAATGGACTTGTTTCCATAGGATTCCCTCTCTTG
TTAATCACGGGCTACGTGGCAATCGCTGAGAACGAGGCTGCACTGGATAAAGTGCACCCCTTCTGATCTTCTGCACTACTCC
TCCCTCCTTAGTCGCCTCATCAATGATATAGGAACGTTCCGGATGAGATGGCAAGAGGCGATAATCTGAAGTCAATCCATTGT
TACATGAACGAACTGGGGCTTCCGAGGAAGTTGCTCGTGAGCACATAAAGGGAGTAATCGAGGAGAATTGGAAAATACTGAAT
CAGTGCTGCTTTGATCAATCTCAGTTTCAGGAGCCTTTTATAACCTTCAATTTGAACTCTGTTGAGGGTCTCATTCTCTAT
GAATTTGGGGATGGCTTTGGGGTGACGGATAGCTGGACAAAGGTTGATATGAAGTCCGTTTTGATCGACCTATTCTCTCGGC
GAGGAGTAGTAAGCTCGAAGCTTATGTTTGGTTTATGGTGATAGTGGTTGACAAAAATATAACATCGACAGGAGTATCTATT
TTTTATGATCGTGAGTTAAATAGGTAAGAAATATCTATTTGCTCATGATTCATAGAAATATTATTCCTTACACGTTATATTC
TATCAAATAACTATTTTATTCCTGTAAGCCAACTTTAACTTTACTCTTAAGGGTCACTAGTGGTTTTACTATATAAATAGTCAA
CTAGCTTGAAGTAATAGAATAAAGGTTTGTGTTTAAAGAGGTTATGCACTACGGTTTCGATTGCTTTTACTAATAAGTTTCTT
AACGAAACTCAGATGCTTTTTGGGATCATTCCCAAGGGATTTGTGCTTGATTATCATCGTCTCTCTTATTAATATATTCCTT
TTATGGCTGGTGTGAAAAAAAAAAAAAAAAAAAAAAAAA
```

图 4

MDSSTATAMTAPFIDPTDHVNLKTDTDASENRRMGNYKPSIWNYDFLQSLATHHNIVEERHLKLAEKLGQV
KFMFGAPMEPLAKLELVDVVQRLGLNHLFETEIKEALFSIYKDGSNGWWFGHLHATSLRFRLLRQCGLFIPQ
DVFKTFQNKTEGFDMKLCDNVKGLLSLYEASYLGWKGENILDEAKAFTTKCLKSAWENISEKWLAKRVKHAL
ALPLHWRVPRIEARWFIEAYEQEANMNPTLLKLAKLDFNMVQSIHQKEIGELARWWVTGLDKLAFARNLL
QSYMWSCAIASDPKFKLARETIVEIGSVLTVVDDGYDVYGSIDELDLYTSSVERWSCVEIDKLPNTLKLIFM
SMFNKTNEVGLRVQHERGYNSIPTFIKAWVEQCKSYQKEARWFHGGHTPPLEEYSLNGLVLSIGFPLLLITGY
VAIAENEAALDKVHPLPDLLHYSSLLSRLINDIGTSPDEMARGDNLKSIHCYMNETHGASEEVAREHIKGVIE
ENWKILNQCCFDQSQFQEPFITFNLNSVRGSHFFYEFGDGFVTDTSWTKVDMKSVLIDPIPLGEE*

MetAspSerSerThrAspThrAspMetThrAspProPheIleAspProThrAspHisValAsnLeuLysThr
AspThrAspAspSerGluAsnArgArgMetglyAsnTyrLysProSerIleTrpAsnTyrAspPheLeuGln
SerLeuAspThrHisHisAsnIleValGluGluArgHisLeuLysLeuAspGluLysLeuLysglyGlnVal
LysPheMetPheglyAspProMetGluProLeuAspLysLeuGluLeuValAspValValGlnArgLeugly
LeuAsnHisLeuPheGluThrGluIleLysGluAspLeuPheSerIleTyrLysAspglySerAsnglyTrp
TrpPheglyHisLeuHisAspThrSerLeuArgPheArgLeuLeuArgGlnCysglyLeuPheIleProGln
AspValPheLysThrPheGlnAsnLysThrglyGluPheAspMetLysLeuCysAspAsnValLysglyLeu
LeuSerLeuTyrGluAspSerTyrLeuglyTrpLysglyGluAsnIleLeuAspGluAspLysAspPheThr
ThrLysCysLeuLysSerAspTrpGluAsnIleSerGluLysTrpLeuAspLysArgValLysHisAspLeu
AspLeuProLeuHisTrpArgValProArgIleGluAspArgTrpPheIleGluAspTyrGluGlnGluAsp
AsnMetAsnProThrLeuLeuLysLeuAspLysLeuAspPheAsnMetValGlnSerIleHisGlnLysGlu
IleglyGluLeuAspArgTrpTrpValThrThrglyLeuAspLysLeuAspPheAspArgAsnAsnLeuLeu
GlnSerTyrMetTrpSerCysAspIleAspSerAspProLysPheLysLeuAspArgGluThrIleValGlu
IleglySerValLeuThrValValAspAspglyTyrAspValTyrglySerIleAspGluLeuAspLeuTyr
ThrSerSerValGluArgTrpSerCysValGluIleAspLysLeuProAsnThrLeuLysLeuIlePheMet
SerMetPheAsnLysThrAsnGluValglyLeuArgValGlnHisGluArgglyTyrAsnSerIleProThr
PheIleLysAspTrpValGluGlnCysLysSerTyrGlnLysGluAspArgTrpPheHisglyglyHisThr
ProProLeuGluGluTyrSerLeuAsnglyLeuValSerIleglyPheProLeuLeuLeuIleThrglyTyr
ValAspIleAspGluAsnGluAspAspLeuAspLysValHisProLeuProAspLeuLeuHisTyrSerSer
LeuLeuSerArgLeuIleAsnAspIleglyThrSerProAspGluMetAspArgglyAspAsnLeuLysSer
IleHisCysTyrMetAsnGluThrglyAspSerGluGluValAspArgGluHisIleLysglyValIleGlu
GluAsnTrpLysIleLeuAsnGlnCysCysPheAspGlnSerGlnPheGlnGluProPheIleThrPheAsn
LeuAsnSerValArgglySerHisPhePheTyrGluPheglyAspglyPheglyValThrAspSerTrpThr
LysValAspMetLysSerValLeuIleAspProIleProLeuglyGluGlu

图 5

```

FB299124.1 | -----MALPVAHRYSSEAEELRE-- 18
FB299123.1 | -----MASSSPVPLVTQVQRK-- 16
FB299125.1 | -----MWNCSTISATAASPPLRQWPGG-----ISWRRPSRLQCSAATTRHDD 43
AF484125.1 | -----MASAAVANYEEEEIVRP-- 16
AB438045.1 | MALPALFGSSSLPSSIRHNQPSLFSFRHPRFCSSSSSASFSSQFILCASKTGDQEI VRR-- 58
檀香烯 | -----MDSSTATAMTAP-----FIDP-----TDHVNLTDT--DASENRR-- 33
*

FB299124.1 | -----ATTFHPSLWGDFFLT Y-----QPPTAAQQAYMEERA EVLREDV 56
FB299123.1 | -----PQFTFPSPWGDFFLHH-----VPCTPSQFLSMKERAQRKKEEV 54
FB299125.1 | DLVLDNKGDNR LRENTGADI FQPSIWGDI FLGNSNPAAAASSQQQIQMEERADKLREEV 103
AF484125.1 | -----VADFSPSLWGDQFLSFS-----IDNQIAEKYAQEIEALKEQT 53
AB438045.1 | -----SANWQPSVWDYDYVQS-----LTVDY--TEDKYTKQVQRLKEEV 95
檀香烯 | -----MGNYKPSIWNYDFLQS-----LATHHNIVEERHLKLA EKLKGQV 72
: ** * : : : : :

FB299124.1 | RKILRD--STQLPETLNLILTLQRLGLDYYYENEIDKLLHRIYN--SDYNDKDLNLVS 110
FB299123.1 | RQIILENFASSNLVRKLELVDTLQRIGVDYHYKEEIDNLLHSIFDD--KDGGSDNLYITS 112
FB299125.1 | AKMIAS---STTTASRLQLIDALERLCLDHLFEEEEIGAALAQIET--ADVSDYDLETVA 157
AF484125.1 | RSMLLAT--ARKLADTLNLIDTIERLGIAYHFEKEIDEILDQIYN--QNSTFDDLCTSA 108
AB438045.1 | KGLFDR---EMKQVAKLEFIDVVQRLGLGYHFKEI KIALSSIHNNTEDAQLSNDLYAAS 152
檀香烯 | KFMFGA---PMEPLAKLELVDVVQRLGLNLHFLFETI KEALFSIYKDGSGNGWWFGHLHATS 129
: : * : : : : : * * * :

FB299124.1 | LRFYLLRKNGYDVSSDVFLSFKTDEGGFAY---GDTISLLSLYNAAYLRRHGEKVLDEAI 167
FB299123.1 | LRFYLLRKHGYGVSSDVFEKFRDEQGNISS---DDISCLLMLYDAAHLRTHGEEILDNI I 169
FB299125.1 | LWFCLLRKHYMVSPDVFRFKDEDDGGFLV---NSPKDLLNLYNAAHMRTHGEI ILEEAV 214
AF484125.1 | LQFRLLRQHGFNISPQIFSKFQDENGKFKESLASDVLGLLNLYEASHVRTHTDNI LEDAL 168
AB438045.1 | LRFRLLRQYGCNVQQDVFQRFMKTGTFKESLNKDVKGILGLYEASFGHMEGETVLDEAW 212
檀香烯 | LRFRLLRQCGLFIPQDVFKTFQNKTFGEFDMKLCNVKGLLSLYEASYLGWKGENILDEAK 189
* * * : : : * * : : : * * * : : : : :

FB299124.1 | SFTRRLQD-ILELPASPFAKEVSASLHTPLFRRVGILEARNYIP-IYEKDATVNEAILE 225
FB299123.1 | TFNKSHLQSLLENLEPELREEVQCTLETFRFRVVRVEARRYIS-VYEKNTTRDATILE 228
FB299125.1 | LFSQRHLET-MVPYMEGLSLAREIKSALDIPRRPRIYEEKYIIS-MYEKDG MVDEKVLQ 272
AF484125.1 | AFSTVHLES-AAPYMN SPLREQVTHALEQCLHKGVPRVETRFFISSIYEKEESKNDMLLR 227
AB438045.1 | NFASKHLKDLNLDEVPTNLASNVSHALDMP IHWRPNRLEARWFMD-MYEKQD LIPSLLR 271
檀香烯 | AFTTKCLKS AWENISEKWLAKRVKHALALPLHWRVPRIEARWFIE-AYEQEANMNPTLLK 248
* * : : : * : : : * : : :

FB299124.1 | LAKLNFNLQQLVFCEELKHCTMWWKEFLAKSKMTFVRDRIVEVYFWMNGACYHPPYSHSR 285
FB299123.1 | FAKLDYNILQAIYCEDELKELTVWVKDFQSQTDL SFARDRMVELHFWM LGVVYEPYPYSR 288
FB299125.1 | LAKLNSNIMQLHHQHELGI VSRWNDINIESRLPHVRDLRVECYLWILGVYEPYCSYR 332
AF484125.1 | FAKLDFNLLQMLHKQELAEVSRWWKDLNFVTTLPYARDRVVECYFWALGVYFEPQYSQAR 287
AB438045.1 | LAKLDFNIVQSIHRKEVSNLARWWELGAN-KMTFFRDLRVESYFWSCILVFEPQYTD FR 330
檀香烯 | LAKLDFNMVQSIHQKEIGELARWWVTGLD-KLAFARNLLQSYMWSCAIASDPKFKLAR 307
: * : * : * : : * : : : : * : *

=====
FB299124.1 | IIQTKITSFVTI IDDMFDTYGTTEECMKFVEAIGRWDES AVPLLP EYMKGFYLFLLDTPH 345
FB299123.1 | IMMTKFIVLASLLDDLYDSYSTTEESNAFIAAMQRWDDRTTEHLPA CLKALFINIVKTTN 348
FB299125.1 | IILTMTTAMVTLLDDTYDSYATPEEC ELF TKCIESWDSMGAQDLPERMKYGLEKIFDSCE 392
AF484125.1 | VMLVKTISMISIVDDTFDAYGTVKELEAYTDAIQRWDINEIDRLPDYMKISYKAILDLYK 347
AB438045.1 | ELNTRIACMATLIDDVYDIYGTPEE LELLTDFILRWDITDIDKL PPTIRNGFMALYNTTN 390
檀香烯 | ETIVEIGSVLTVVDDGYDVYGISDEL DLYTSSVERWSCVEIDKL PNTLKLIFMSMFNKTN 367
. . : : * * * : * : * : * : :

=====
FB299124.1 | SFEDELGPQ-KSYRVLYLKHAMERLVQQYNEIKWRDEDYVPKTMSEHLQVSMESIACIP 404
FB299123.1 | EIEEELKLM-KNKHADLIKRLVIDTAKFYHAEVEWRDQHYIPTDIEEHLQISTRSSVCMQ 407
FB299125.1 | I IENMLHQE-EKYRIWYLRQSIKDLVISYSVEVKMLQEGYI PKSVEEHLKLSLITVGYPI 451
AF484125.1 | DYEKELSSAGRSHIVCHAIERMKEVVRNYNVESTWFI EGYKPP-VSEYLSNALATTTYYY 406
AB438045.1 | KVGYRTMTKRGINPIPYLRKLWGECKADMKEVHWFNNGIKPT-LKEYMDVA VDSIGGLI 449
檀香烯 | EVGLRVQHERGYNSIPTFIKAWVEQCKSYQKEARWFHGGHTFP-L EEYSLNGLVSIGFP L 426
. * * : * : :

```

图 6A

FB299124.1	ITCAAFVGMGDIITKETLEWILSFPQFLMSFGIYVRLSNDVASTMREQTKDHSASTVHCY	464
FB299123.1	ITNLALISLGEVTTTRKDVWALTFPKIIRAACIVGRVGNDIVSHEREQTSEHVGSTVQTC	467
FB299125.1	LACVSFVGMHDIATKDCLDWVSSI PKMVEALS VILRLVDDLESYEREQLVPHVASTIDSY	511
AF484125.1	LATTSYLG MKSVAEQD-FEWLSKNPKILEASV IICRVIDDTATYEVEKSRGQIATGIECC	465
AB438045.1	LLLSYFLTTDYLTEEGLNYVSKI PSVMHSSAQIFRFNDDLSTSSHELARGDNSKALECY	509
檀香烯	LLITGYVAIAEN--EAALDKVHPLPDL LHYSSLLSRLINDIGTSPDEMARGDNLKSIHCY	484
	: : * . . . * . : * . . . :	
FB299124.1	MKEHGTMTNDACEKIKELAEDKWKDMLEQCLALT--ELPKVIPRTVDFARTIVNMYKND	522
FB299123.1	MKQYGVTVVEEANEKLRV IEEAWMDIVEECLEQ--KRPMALLETAVNVARTMDFMYKRE	524
FB299125.1	MKEHNSIEVAREQIYILKEESWKDFNNEWLNPDNNVYPKQLLERMFNLARTAQFLYNKE	571
AF484125.1	MRDYGVTKEAMD K FQKMAETAWKDLNEGLLRPT--PISAEFLTPI LNLARIVEVTYIHN	523
AB438045.1	MNETGASEEIAREH IRLVRETWK MNKEVFEDYPFSGFGPFLSACLNLARASHCFY EYG	569
檀香烯	MNETGASEEVAREHIKGVIEENWKILNQCCFDQSQFQ--EPFITFNLNSVRGSHFFYEFG	542
	* : * : : . * : : : : . : . * *	
FB299124.1	HDGFTS-SEALKEMIELLFVKVPVN---	546
FB299123.1	-DAYTL-SFSLKDVIASMYVNSVRAC--	548
FB299125.1	-EKFTN-SHYLKDTVHSLLLAEPFKIPI	597
AF484125.1	LDGYTHPEKVLKPHIIDLLVESIQIA--	549
AB438045.1	-DGYGLPDHQTRDHLASTIFESVSLD--	594
檀香烯	-DGFGVTDSWTKVDMKSVLIDPIPLGEE	569

图 6B

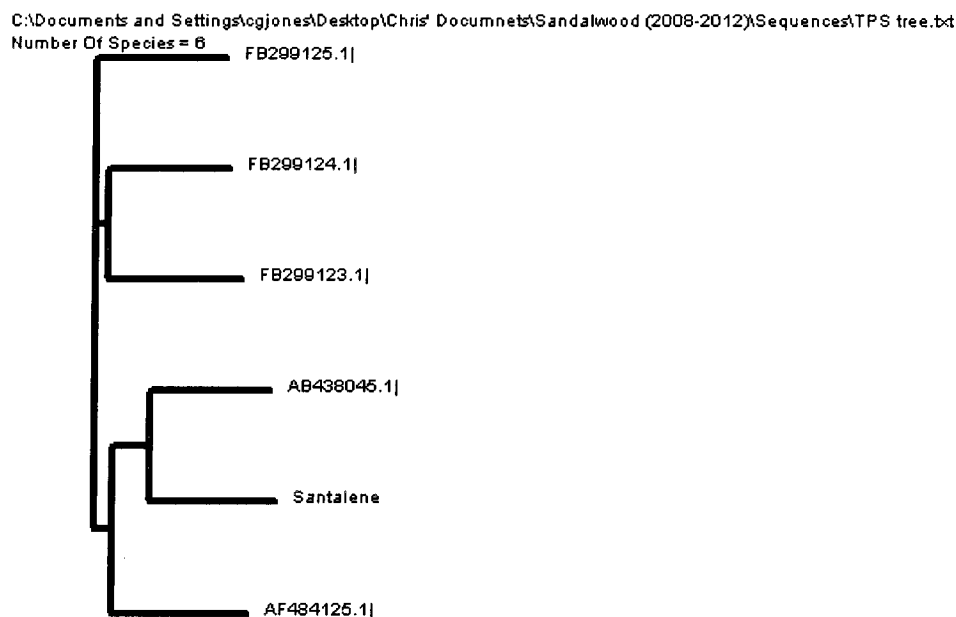


图 7

C:\Documents and Settings\jones\Desktop\Chid\Documents\Santalwood (2008-2012)\Sequences\Santalum TPS genes\multi TPS align\UPMA.ph
Number Of Species = 28

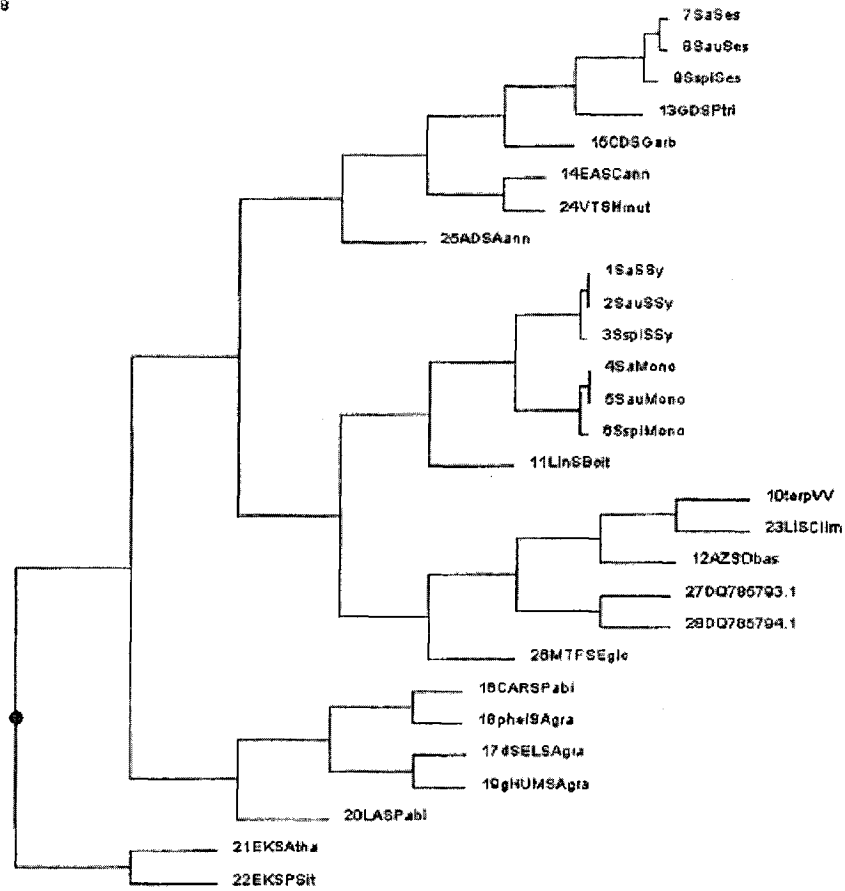


图 8

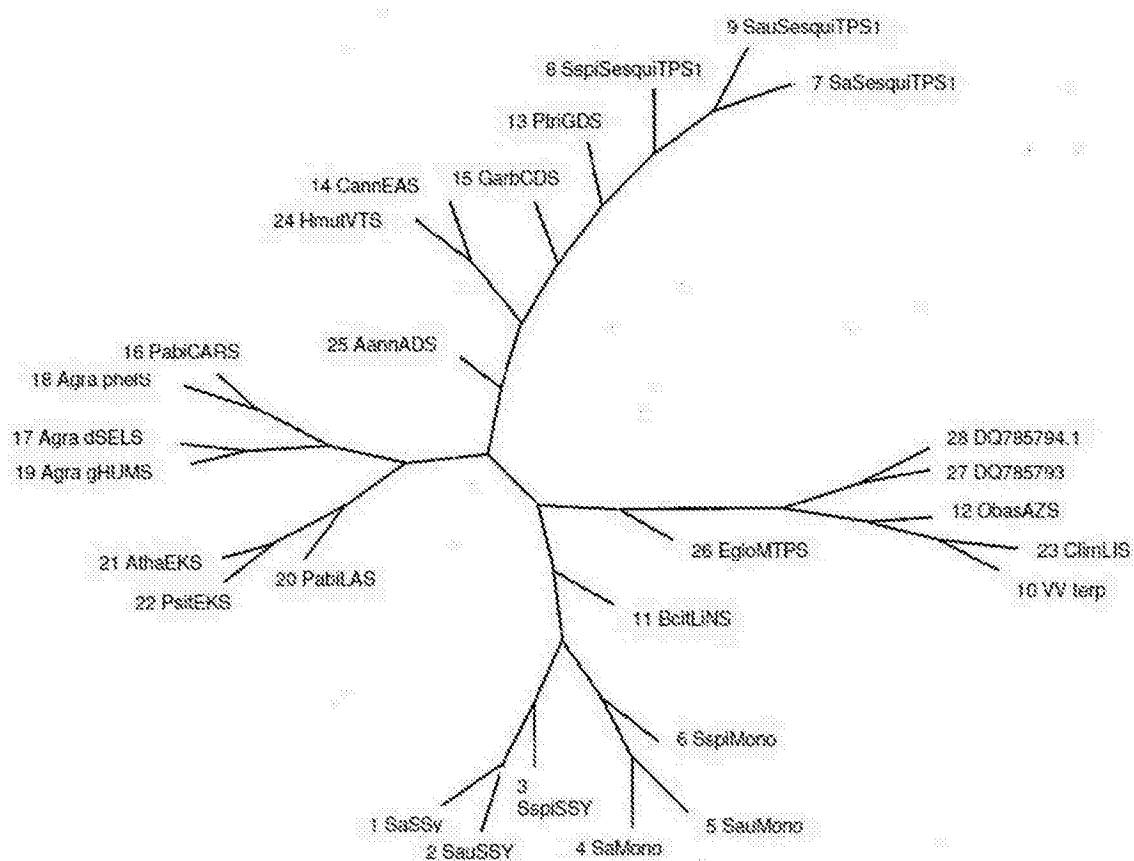


图 9

来自新克里多尼亚檀香、澳洲檀香和印度檀香的檀香烯合酶

						
	10	20	30	40	50	60
SauSSy ORF	ATGGATTCTT	CCACCGCCAC	CGCCATGACA	GCTCCATTCA	TTGATCCTAC	TGATCATGTG
SspiSSy OR	ATGGATTCTT	CCACCGCCAC	CGCCACGACA	GCTCCATTTA	TTGATCATAC	TGATCATGTG
SaSSy ORF	ATGGATTCTT	CCACCGCCAC	CGCCATGACA	GCTCCATTCA	TTGATCCTAC	TGATCATGTG
						
	70	80	90	100	110	120
SauSSy ORF	AATCTCAAAA	CTGATACTGA	TGCCTCAGAG	AATCGAAGGA	TGGGGAATTA	TAAACCCAGC
SspiSSy OR	AATCTTAAAA	TTGATAATGA	TTCTCCGAG	AGTCGAAGGA	TGGGCAATTA	TAAACCCAGT
SaSSy ORF	AATCTCAAAA	CTGATACGGA	TGCCTCAGAG	AATCGAAGGA	TGGGAAATTA	TAAACCCAGC
						
	130	140	150	160	170	180
SauSSy ORF	ATTTGGAATT	ATGATTTTTT	ACAATCACTT	GCAACTCATC	ACAATATTGT	GGAAGAGAGG
SspiSSy OR	ATTTGGAATT	ATGATTTTCT	GCAATCACTT	GCAATCCATC	ACAATATTGT	GGAAGAGAAG
SaSSy ORF	ATTTGGAATT	ATGATTTTTT	ACAATCACTT	GCAACTCATC	ACAATATTGT	GGAAGAGAGG
						
	190	200	210	220	230	240
SauSSy ORF	CATCTAAAGC	TAGCTGAGAA	GCTGAAGGGC	CAAGTGAAGT	TTATGTTTGG	GGCACCAATG
SspiSSy OR	CATCTAAAGC	TAGCTGAGAA	GCTGAAGGGC	CAAGTGATGT	CTATGTTTGG	GGCACCAATG
SaSSy ORF	CATCTAAAGC	TAGCTGAGAA	GCTGAAGGGC	CAAGTGAAGT	TTATGTTTGG	GGCACCAATG

图 10a

```

...|...| ...|...| ...|...| ...|...| ...|...| ...|...|
      250      260      270      280      290      300

SauSSy ORF  GAGCCGTTAG CAAAGCTGGA GCTTGTGGAT GTGGTTCAAA GGCTCGGGCT AAACCACCGA

SspiSSy OR  GAGCCGTTAG CAAAGCTGGA GCTTGTGGAT GTGGTTCAAA GGCTTGGGCT AAACCACCAA

SaSSy ORF   GAGCCGTTAG CAAAGCTGGA GCTTGTGGAT GTGGTTCAAA GGCTTGGGCT AAACCACCTA

...|...| ...|...| ...|...| ...|...| ...|...| ...|...|
      310      320      330      340      350      360

SauSSy ORF  TTTGAGACAG AGATCAAGGA AGCGCTATTT AGTATTTATA AGGACGAGAG CAATGGATGG

SspiSSy OR  TTTGAGACAG AGATCAAGGA AGCCCTATTT AGTGTTTACA AGGATGGGAG CAATGGATGG

SaSSy ORF   TTTGAGACAG AGATCAAGGA AGCGCTGTTT AGTATTTACA AGGATGGGAG CAATGGATGG

...|...| ...|...| ...|...| ...|...| ...|...| ...|...|
      370      380      390      400      410      420

SauSSy ORF  TGGTTTGGCC ACCTTCATGC GACATCTCTC CGATTTAGGC TGCTACGACA GTGTGGGCTT

SspiSSy OR  TGGTTTGGCC ACCTTCATGC AACATCTCTT CGATTTAGGC TACTACGACA GTGTGGGCTT

SaSSy ORF   TGGTTTGGCC ACCTTCATGC GACATCTCTC CGATTTAGGC TGCTACGACA GTGTGGGCTT

...|...| ...|...| ...|...| ...|...| ...|...| ...|...|
      430      440      450      460      470      480

SauSSy ORF  TTTATCCCCC AGGATGTGTT TAAAACGTTT CAAAACAAAA CTGGTGAATT TGATATGAAA

SspiSSy OR  TTTATCCCCC AGGATGTGTT TAAAACGTTT CAGAGCAAAA CTGATGAATT TGATATGAAA

SaSSy ORF   TTTATTCCCC AAGATGTGTT TAAAACGTTT CAAAACAAGA CTGGGGAATT TGATATGAAA

...|...| ...|...| ...|...| ...|...| ...|...| ...|...|
      490      500      510      520      530      540

SauSSy ORF  CTGTGTGACA ACGTAAAAGG GCTGCTGAGC TTATATGAAG CTCATACTT GGGATGGAAG

```

图 10b

SspiSSy OR CTGTGTGACA ACATAAAAGG GTTGTGAGC TTGTATGAAG CTTTCATTCCT GGGGTGGAAG

SaSSy ORF CTTTGTGACA ACGTAAAAGG GCTGCTGAGC TTATATGAAG CTTTCATACTT GGGATGGAAG

.....|.....||.....||.....||.....||.....||.....|

550 560 570 580 590 600

SauSSy ORF GGTGAAAACA TCCTAGATGA AGCCAAGGCC TTCGCCACCA AGTACTTGAA AAGTGCATGG

SspiSSy OR GGTGAAAACA TCCTAGATGA AGCCAAGGCC TTCGCCACCA AGTACTTGAA AAATGCATGG

SaSSy ORF GGTGAAAACA TCCTAGATGA AGCCAAGGCC TTCACCACCA AGTGCTTGAA AAGTGCATGG

.....|.....||.....||.....||.....||.....||.....|

610 620 630 640 650 660

SauSSy ORF GAAAATATAT CTGAAAAGTG GTTAGCCAAA AGAGTGAAGC ATGCATTGGC TTTACCTTTG

SspiSSy OR GAAAATATAT CCCAAAAGTG GCTAGCCAAA AGAGTGAAGC ATGCACTGGC TTTGCCTCTG

SaSSy ORF GAAAATATAT CCGAAAAGTG GTTAGCCAAA AGAGTGAAGC ATGCATTGGC TTTGCCTTTG

.....|.....||.....||.....||.....||.....||.....|

670 680 690 700 710 720

SauSSy ORF CATTGGAGAG TCCCTCGAAT CGAAGCTAGA TGGTTCATTG AGGCATATGA GCAAGAAGCG

SspiSSy OR CACTGGAGAG TCCCTCGAAT CGAGGCTAGA TGGTTCATTG AGGCATATGA GCAAGAAGAG

SaSSy ORF CATTGGAGAG TCCCTCGAAT CGAAGCTAGA TGGTTCATTG AGGCATATGA GCAAGAAGCG

.....|.....||.....||.....||.....||.....||.....|

730 740 750 760 770 780

SauSSy ORF AATATGAACC CAACACTACT CAAACTCGCA AAATTAGACT TTAATATGGT GCAATCAATT

SspiSSy OR AACATGAACC CAACACTACT CAAACTTGCA AAATTAGACT TTAACATGGT GCAATCAATT

SaSSy ORF AATATGAACC CAACACTACT CAAACTCGCA AAATTAGACT TTAATATGGT GCAATCAATT

图 10c


```

      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
      790      800      810      820      830      840

SauSSy ORF  CATCAGAAAG AGATTGGGGA ATTAGCAAGG TGGTGGGTGA CTACTGGCTT GGATAAGTTA

SspiSSy OR  CATCAGAAAG AGATTGGGGA ATTAGCAAGG TGGTGGGTGA CTACTGGCTT GGATAAGTTA

SaSSy ORF   CATCAGAAAG AGATTGGGGA ATTAGCAAGG TGGTGGGTGA CTACTGGCTT GGATAAGTTA

      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
      850      860      870      880      890      900

SauSSy ORF  GCCTTTGCTA GGAATAATTT ACTGCAAAGC TATATGTGGA GCTGCGCGAT TGCTTCCGAC

SspiSSy OR  GCCTTTGCTA GGAATAATTT ACTGCAAAGC TATATGTGGA GCTGCGCGAT TGCTTCCGAC

SaSSy ORF   GCCTTTGCCA GGAATAATTT ACTGCAGAGC TATATGTGGA GCTGCGCGAT TGCTTCCGAC

      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
      910      920      930      940      950      960

SauSSy ORF  CCGAAGTTCA AACTTGCTAG AGAAACTATT GTCGAAATCG GAAGTGTACT CACAGTTGTT

SspiSSy OR  CCAAAGTTCA AACTTGCTAG AGAAACTATT GTCGAAATCG GAAGTGTACT CACAGTTGTG

SaSSy ORF   CCGAAGTTCA AACTTGCTAG AGAAACTATT GTCGAAATCG GAAGTGTACT CACAGTTGTT

      ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
      970      980      990      1000      1010      1020

SauSSy ORF  GATGATGCAT ATGACGTCTA TGGTTCAATG GACGAACTTG ATCTCTACAC AAGCTCCGTT

SspiSSy OR  GACGATGCAT ATGATGTCTA TGGTTCAATG GATGAACTTG ATCACTACAC ATACTCCGTT

SaSSy ORF   GACGATGGAT ATGACGTCTA TGGTTCAATC GACGAACTTG ATCTCTACAC AAGCTCCGTT

```

图 10d

....|....||....||....||....||....||....|

1030 1040 1050 1060 1070 1080

SauSSy ORF GAAAGGTGGA GCTGTGTAGA AATTGACAAG TTGCCAAACA CGTTAAAATT GATTTTTATG

SspiSSy OR GAAAGGTGGA GCTGTGTAGA AATTGACAAG CTGCCAAACA CGTTAAAATT GATTTTTATG

SaSSy ORF GAAAGGTGGA GCTGTGTGGA AATTGACAAG TTGCCAAACA CGTTAAAATT AATTTTTATG

....|....||....||....||....||....||....||....|

1090 1100 1110 1120 1130 1140

SauSSy ORF TCTATGTTTA ATAAGACCAA TGAGGTTGGC CTTGAGTCC AGCATGAGCG AGGCTACAAT

SspiSSy OR TCTATGTTCA ACAAGACCAA TGAGGTTGGC CTTGAGTCC AGCATGAGCG AGGCTACAAC

SaSSy ORF TCTATGTTCA ACAAGACCAA TGAGGTTGGC CTTGAGTCC AGCATGAGCG AGGCTACAAT

....|....||....||....||....||....||....||....|

1150 1160 1170 1180 1190 1200

SauSSy ORF AGCATCCCTA CTTTATCAA AGCGTGGGT CAACAGTGTA AATCATACCA GAAAGAAGCA

SspiSSy OR GGCATCCCTA CTTTATCAA AGCATGGGT GAACAGTGTA AAGCATACCA GAAAGAGGCA

SaSSy ORF AGCATCCCTA CTTTATCAA AGCGTGGGT GAACAGTGTA AATCATACCA GAAAGAAGCA

....|....||....||....||....||....||....||....|

1210 1220 1230 1240 1250 1260

SauSSy ORF AGATGGTTCC ACGGGGACA CACGCCTCCG TTGGAAGAAT ATAGCTTGAA TGGACTTGTT

SspiSSy OR AGATGGTACC ATGGGGGACA CACGCCTCCA TTGGAGGAAT ATAGCTTGAA TGGACTTGTT

SaSSy ORF AGATGGTTCC ACGGGGACA CACGCCTCCA TTGGAAGAAT ATAGCTTGAA TGGACTTGTT

图 10e

.....|.....||.....||.....||.....||.....||.....|

1270 1280 1290 1300 1310 1320

SauSSy ORF TCCATAGGAT TCCCTCTCTT GTTGATCACC GGCTACGTGG CAATCGCTGA GAACGAGGCT

SspiSSy OR TCCATAGGAT TCCCTCTCTT GTTGATCACC GGCTACATCG CAATCGCTGA GAACGAGGCT

SaSSy ORF TCCATAGGAT TCCCTCTCTT GTTAATCACG GGCTACGTGG CAATCGCTGA GAACGAGGCT

.....|.....||.....||.....||.....||.....||.....|

1330 1340 1350 1360 1370 1380

SauSSy ORF GCACTGGATA AAGTGCACCC CCTTCCTGAT CTTCTGCACT ACTCCTCCCT CCTTAGTCGC

SspiSSy OR GCACTGGATA AAGTGCACCC CCTTCCTGAT CTTCTGCACT ACTCCTCCCT CCTTAGTCGC

SaSSy ORF GCACTGGATA AAGTGCACCC CCTTCCTGAT CTTCTGCACT ACTCCTCCCT CCTTAGTCGC

.....|.....||.....||.....||.....||.....||.....|

1390 1400 1410 1420 1430 1440

SauSSy ORF CTCATCAATG ATATAGGAAC GTCTCCGGAT GAGATGGCAA GAGGCGATAA TCTGAAGTCA

SspiSSy OR CTCATCAATG ACATGGGAAC GTCTCCGGAC GAGATGGCAA GAGGTGACAA TCTGAAGTCA

SaSSy ORF CTCATCAATG ATATAGGAAC GTCTCCGGAT GAGATGGCAA GAGGCGATAA TCTGAAGTCA

.....|.....||.....||.....||.....||.....||.....|

1450 1460 1470 1480 1490 1500

SauSSy ORF ATCCATTGTT ACATGAACGG AACTGGGGCT TCCGAGGAAG TTGCTCGTGA GCACATAAAG

SspiSSy OR ATCCACTGTT ACATGAACGA AACTGGGGCT TCTGAGGAAG TTGCTCGTGA GCACATAAAA

SaSSy ORF ATCCATTGTT ACATGAACGA AACTGGGGCT TCCGAGGAAG TTGCTCGTGA GCACATAAAG

图 10f

....|....||....||....||....||....||....|

1510 1520 1530 1540 1550 1560

SauSSy ORF GGAGTAATCG AGGAGAATTG GAAAATACTG AATCAGTGCT GCTTTGATCA ATCTCAGTTT

SspiSSy OR GGAATAATCG AGGAGAATTG GAAAATACTG AATCAGTGCT GCTTTGATCA ATCTCAGTTT

SaSSy ORF GGAGTAATCG AGGAGAATTG GAAAATACTG AATCAGTGCT GCTTTGATCA ATCTCAGTTT

....|....||....||....||....||....||....|

1570 1580 1590 1600 1610 1620

SauSSy ORF CAGGAGCCTT TTATAACCTT CAATTGAAC TCTGTTCGAG GGTCTCATTT CTTCTATGAA

SspiSSy OR CAGGAGCCTT TTATAACCTT CAATTGAAC TCTGTTCGAG GGTCTCATTT CTTCTATGAA

SaSSy ORF CAGGAGCCTT TTATAACCTT CAATTGAAC TCTGTTCGAG GGTCTCATTT CTTCTATGAA

....|....||....||....||....||....||....|

1630 1640 1650 1660 1670 1680

SauSSy ORF TTTGGGGATG GCTTTGGGGT GACGGATAGC TGGACAAAGG TTGATATGAA GTCCGTTTTG

SspiSSy OR TTTGGGGATG GCTTTGGGGT GACAGATAGC TGGACAAAGG TTGATATGAA GTCTGTTTTG

SaSSy ORF TTTGGGGATG GCTTTGGGGT GACGGATAGC TGGACAAAGG TTGATATGAA GTCCGTTTTG

....|....||....||....| ...

1690 1700 1710

SauSSy ORF ATTGACCCTA TTCCTCTCGG CGAGGAGTAG TAA

SspiSSy OR ATCGACCCTA TTCCTCTCGG CGAGGAGTAG TAA

SaSSy ORF ATCGACCCTA TTCCTCTCGG CGAGGAGTAG TAA

图 10g

新克里多尼亚檀香、澳洲檀香和印度檀香的檀香烯合酶的氨基酸序列

	...	...	...	...	...	...	...	...	...	...	...
	10	20	30	40	50	60					
SauSSy ORF	MDSSTATAMT	APFIDPTDHV	NLKTDTDASE	NRRMGNYKPS	IWNYDFLQSL	ATHHNIVEER					
SspiSSy OR	MDSSTATATT	APFIDHTDHV	NLKIDNDSSE	SRRMGNYKPS	IWNYDFLQSL	AIHHNIVEEK					
SaSSy ORF	MDSSTATAMT	APFIDPTDHV	NLKTDTDASE	NRRMGNYKPS	IWNYDFLQSL	ATHHNIVEER					
	...	...	...	...	...	...	...	...	...	...	...
	70	80	90	100	110	120					
SauSSy ORF	HLKLAEKLKG	QVKFMFGAPM	EPLAKLELVD	VVQRLGLNHR	FETEIKEALF	SIYKDESNGW					
SspiSSy OR	HLKLAEKLKG	QVMSMFGAPM	EPLAKLELVD	VVQRLGLNHQ	FETEIKEALF	SVYKDGSNGW					
SaSSy ORF	HLKLAEKLKG	QVKFMFGAPM	EPLAKLELVD	VVQRLGLNHL	FETEIKEALF	SIYKDGSNGW					
	...	...	...	...	...	...	...	...	...	...	...
	130	140	150	160	170	180					
SauSSy ORF	WFGHLHATSL	RFRLLRQCGL	FIPQDVFKTF	QNKTGEFDMK	LCDNVKGLLS	LYEASYLGWK					
SspiSSy OR	WFGHLHATSL	RFRLLRQCGL	FIPQDVFKTF	QSKTDEFDMK	LCDNIKGLLS	LYEASFLGWK					
SaSSy ORF	WFGHLHATSL	RFRLLRQCGL	FIPQDVFKTF	QNKTGEFDMK	LCDNVKGLLS	LYEASYLGWK					
	...	...	...	...	...	...	...	...	...	...	...
	190	200	210	220	230	240					
SauSSy ORF	GENILDEAKA	FATKYLKSAW	ENISEKWLAKE	RVKHALALPL	HWRVPRIEAR	WFIEAYEQEA					
SspiSSy OR	GENILDEAKA	FATKYLKNAW	ENISQKWLAKE	RVKHALALPL	HWRVPRIEAR	WFIEAYEQEE					
SaSSy ORF	GENILDEAKA	FTTKCLKSAW	ENISEKWLAKE	RVKHALALPL	HWRVPRIEAR	WFIEAYEQEA					

图 11a

						
	250	260	270	280	290	300
SauSSy ORF	NMNPTLLKLA	KLDFNMVQSI	HQKEIGELAR	WVTTGLDKL	AFARNNLLQS	YMWSCAIASD
SspiSSy OR	NMNPTLLKLA	KLDFNMVQSI	HQKEIGELAR	WVTTGLDKL	AFARNNLLQS	YMWSCAIASD
SaSSy ORF	NMNPTLLKLA	KLDFNMVQSI	HQKEIGELAR	WVTTGLDKL	AFARNNLLQS	YMWSCAIASD
						
	310	320	330	340	350	360
SauSSy ORF	PKFKLARETI	VEIGSVLTVV	DDAYDVYGS	DELDTYTSSV	ERWSCVEIDK	LPNTLKLIFM
SspiSSy OR	PKFKLARETI	VEIGSVLTVV	DDAYDVYGS	DELDTYTSSV	ERWSCVEIDK	LPNTLKLIFM
SaSSy ORF	PKFKLARETI	VEIGSVLTVV	DDAYDVYGS	DELDTYTSSV	ERWSCVEIDK	LPNTLKLIFM
						
	370	380	390	400	410	420
SauSSy ORF	SMFNKTNEVG	LRVQHERGYN	SIPTFIKAWV	EQCKSYQKEA	RWFHGGHTPP	LEEYSLNGLV
SspiSSy OR	SMFNKTNEVG	LRVQHERGYN	SIPTFIKAWV	EQCKSYQKEA	RWFHGGHTPP	LEEYSLNGLV
SaSSy ORF	SMFNKTNEVG	LRVQHERGYN	SIPTFIKAWV	EQCKSYQKEA	RWFHGGHTPP	LEEYSLNGLV
						
	430	440	450	460	470	480
SauSSy ORF	SIGFPLLLIT	GYVAIAENEA	ALDKVHPLPD	LLHYSSLLSR	LINDIGTSPD	EMARGDNLKS
SspiSSy OR	SIGFPLLLIT	GYVAIAENEA	ALDKVHPLPD	LLHYSSLLSR	LINDIGTSPD	EMARGDNLKS
SaSSy ORF	SIGFPLLLIT	GYVAIAENEA	ALDKVHPLPD	LLHYSSLLSR	LINDIGTSPD	EMARGDNLKS

图 11b

....|....||....||....||....||....||....|

490 500 510 520 530 540

SauSSy ORF IHCYMNGTGA SEEVAREHIK GVIEENWKIL NQCCFDQSQF QEPFITFNLN SVRGSHFFYE

SspiSSy OR IHCYMNETGA SEEVAREHIK GIIEENWKIL NQCCFDQSQF QEPFITFNLN SVRGSHFFYE

SaSSy ORF IHCYMNETGA SEEVAREHIK GVIEENWKIL NQCCFDQSQF QEPFITFNLN SVRGSHFFYE

....|....||....||....| .

550 560 570

SauSSy ORF FGDGFGVTDS WTKVDMKSVL IDPIPLGEE* *

SspiSSy OR FGDGFGVTDS WTKVDMKSVL IDPIPLGEE* *

SaSSy ORF FGDGFGVTDS WTKVDMKSVL IDPIPLGEE* *

图 11c

CLUSTAL 2.0.12 多重序列比对

```

1SaSSy      -----MDSSTATAMTAPF--IDPTDHVNLKTDTDASENRRMG 35
2SauSSy      -----MDSSTATAMTAPF--IDPTDHVNLKTDTDASENRRMG 35
3SspiSSy      -----MDSSTATATTAPF--IDHTDHVNLKIDNDSSESRRMG 35
4SaMono      -----MDAFATSPTSALIKAVNCIAHVTPMAGEDSSENRRAS 37
27DQ785793.1 MSSLIMQVVIPKPAKFFHNNLFSLSKRHRFSTTTTTRGGRWARCSLQTGNEIQTERRTG 60
28DQ785794.1 -----MPLNSLHN-LERKPSK--AWSTSC TAPAARLQASFSLQQE EPRQIRRSG 46
                :      ::      :      **

1SaSSy      NYKPSIWNYDFLQSLATHHNIVEERHLKLAEKLGQVKFMFGAPMEPLAKLELVDVVQRL 95
2SauSSy      NYKPSIWNYDFLQSLATHHNIVEERHLKLAEKLGQVKFMFGAPMEPLAKLELVDVVQRL 95
3SspiSSy      NYKPSIWNYDFLQSLAIHHNIVEEKHLKLAEKLGQVMSMFGAPMEPLAKLELVDVVQRL 95
4SaMono      NYKPSTWDYEF LQSLATSHNTVQE KHKMAEKLKEEVKSMIKGQMEPVAKLELINILQRL 97
27DQ785793.1 GYQPTLWDFSTIQSFDSEYK--EEKHLMRAAGMIDQVKMMLQEEVDSIRRLLELIDDLRL 118
28DQ785794.1 DYQPSLWDFNYIQSLNTPYK--EQRYVNRQAE LIMQVRMLLKVKMEAIQQLELIDDLQYL 104
      .***: *::: **:  :: ::::  :  :*  ::  ::: :***: :: *

1SaSSy      GLNHLFETEIKEALFSIYKDGSNGWWFGHLHATSLRFRLLRQCGLFIPQDVFKTFQ-NKT 154
2SauSSy      GLNHRFETEIKEALFSIYKDESNGWWFGHLHATSLRFRLLRQCGLFIPQDVFKTFQ-NKT 154
3SspiSSy      GLNHQFETEIKEALFSVYKDGSNGWWFGHLHATSLRFRLLRQCGLFIPQDVFKTFQ-SKT 154
4SaMono      GLKYRFESEIKEELFSLYKDGTDAWWVDNLHATALRFRLLRRENGIFVPQDVFETLK-DKS 156
27DQ785793.1 GISCHFEREIVEILNSKYTTN-EIDERDLYSTALRFRLLRQYDFSVSQEVFDCFKNAKG 177
28DQ785794.1 GLSYFFPDEIKQILSSIHNEHR-YFHNNDLYLTALGFRILRQHGFNVSEDFDCFKTEKC 163
      *:  *  ** : * * :      ,*: *:* ***: **:  ::: **  :: *

```

图 12a

1SaSSy	GEFDMKLCNDNVKGLLSLYEASYLGWKGENILDEAKAFTTKCLKSAWENISEK---WLAKR 211
2SauSSy	GEFDMKLCNDNVKGLLSLYEASYLGWKGENILDEAKAFATKYLKSAWENISEK---WLAKR 211
3SspiSSy	DEFDMKLCDNKGLLSLYEASFLGWKGENILDEAKAFATKYLKNAWENISQK---WLAKR 211
4SaMono	GKFKSQLCKDVRGLLSLYEASYLGWEGEDLLDEAKKFSTTNLNNVKESISSN---TLGRL 213
27DQ785793.1	TDFKPSLVDDTRGLLQLYEASFLSAQGEETLRLARDFATKFLQKRVLVDKD---INLLSS 234
28DQ785794.1	SDFNANLAQDTKGMLQLYEASFLLRREGEDTLELARRFSTRSLREKLDEDGDEIDEDLSSW 223
	. * . * . : : * : * . * * * * : * : * : * * : * : * . . . *
1SaSSy	VKHALALPLHWRVPRIEARWFIEAYEQEANMNPTLLKLAKLDFNMVQSIHQKEIGELARW 271
2SauSSy	VKHALALPLHWRVPRIEARWFIEAYEQEANMNPTLLKLAKLDFNMVQSIHQKEIGELARW 271
3SspiSSy	VKHALALPLHWRVPRIEARWFIEAYEQEANMNPTLLKLAKLDFNMVQSIHQKEIGELARW 271
4SaMono	VKHALNLPLHWSAARYEARWFIDEYEKEENVNPNLLKYAKLDFNIVQSIHQQELGNLARW 273
27DQ785793.1	IERALELPTHWRVQMPNARSFIDAYKRRPDMNPTVLELAKLDFNMVQAQFQQELKEASRW 294
28DQ785794.1	IRHSLDLPLHWRIQGLEARWFLDAYARRPDMNPLIFKLAKLNFNIVQATYQEELKDVSrw 283
	: . : : * * * * : * * * : : * : . : : * * : : : * * : * : * : : * : : : * *
	—
1SaSSy	WVTTGL-DKLAFARNNLLQSYMWSCAIASDPKFKLARETIVEIGSVLTVVDDGYDVYGSI 330
2SauSSy	WVTTGL-DKLAFARNNLLQSYMWSCAIASDPKFKLARETIVEIGSVLTVVDDAYDVYGSM 330
3SspiSSy	WVTTGL-DKLAFARNNLLQSYMWSCAIASDPKFKLARETIVEIGSVLTVVDDAYDVYGSM 330
4SaMono	WVETGL-DKLSFVRNTLMQNFMWGCAMVFEPQYGVKVRDAAVKQASLIAMVDDVDYDVYGSL 332
27DQ785793.1	WNSTGLVHELPPFVRDRIVECYWTTGVVVERRQHGYERIMLTKINALVTTIDDVFDIYGTL 354
28DQ785794.1	WNSSCLAELPPFVRDRIVECFFWAIGAFEPHQYSYQRKMAAIIITFVTIIDDVDYDVYGTL 343
	* : * . : * . * . * : : : : * . : . * . : : : * * : * : * : :

图 12b

1SaSSy	DELDLYTSSVERWSCVEIDKLPNTLKLIFMSMFNKTNEVGLRVQHERGYNSIPTFIKAWV	390
2SauSSy	DELDLYTSSVERWSCVEIDKLPNTLKLIFMSMFNKTNEVGLRVQHERGYNSIPTFIKAWV	390
3SspiSSy	DELDHYTYSVERWSCVEIDKLPNTLKLIFMSMFNKTNEVGLRVQHERGYNGIPTFIKAWV	390
4SaMono	EELEIFTDIVDRWDITGIDKLPRNISMILLTMFNTANQIGYDLLRDRGFNGIPHIAQAWA	392
27DQ785793.1	EELQLFTTAIQRWDIESMKQLPPYMQICYLALFNFVNEMAYDTLRDKGFDSTPYLRKVWV	414
28DQ785794.1	EELELFTDMIRRWDNISISQLPYMQVCYLALYNFVSERAYDILKDQHFNSIPYLQRSWV	403

: ** : * : ** . : : ** : : : : * . . : : : : : * : : *

1SaSSy	EQCKSYQKEARWFHGGHTPPLEEYSLNGLVLSIGFPLLLITGYVAIAENE----	AALDKVH 446
2SauSSy	QQCKSYQKEARWFHGGHTPPLEEYSLNGLVLSIGFPLLLITGYVAIAENE----	AALDKVH 446
3SspiSSy	EQCKAYQKEARWYHGGHTPPLEEYSLNGLVLSIGFPLLLITGYIAIAENE----	AALDKVH 446
4SaMono	TLCKKYLKEAKWYHSGYKPTLEEYLENGLVLSISFVLSLVTAYLQTEILENLTYESAAYVN	452
27DQ785793.1	GLIESYLIEAKWYYKGHKPSLEEYMKNSWISIGGIPILSHLFFRLTDSI--	EEEEAESMH 472
28DQ785794.1	SLVEGYLKEAYWYYNGYKPSLEEYLNNAKISISAPTIISQLYFTLANST--	DETVIESLY 461

: * ** * : : * : . * . * * * * * . : * . : : : : :

1SaSSy	PLPDLLHYSSLLSRLINDIGTSPDEMARGDNLKSIHCYMNETHGASEEVAREHIKGVIEEN	506
2SauSSy	PLPDLLHYSSLLSRLINDIGTSPDEMARGDNLKSIHCYMNETHGASEEVAREHIKGVIEEN	506
3SspiSSy	PLPDLLHYSSLLSRLINDIGTSPDEMARGDNLKSIHCYMNETHGASEEVAREHIKGVIEEN	506
4SaMono	SVPPLVRYSGLLNRLYNDLGTSSAEIARGDTLKSICQYMTQTGATEEAAREHIKGLVHEA	512
27DQ785793.1	KYHDIVRASCTILRLADDMGTSLEVERGDVPSVQCYMNEKNASEEEAREHVRSLIDQT	532
28DQ785794.1	EYHNILYLSGTILRLADDLGTSHLELARGDVPKAIQCYMKDTNASEREAVEHVKFLIRET	521

: : * : ** : * : * * * * * : * * * * * : : : * * * : : : :

图 12c

1SaSSy	WKILNQCCFDQS-QFQEPFITFNLNSVRGSHFFYEF-GDGFGVTDSWTKVDMKSVLIDPI 564
2SauSSy	WKILNQCCFDQS-QFQEPFITFNLNSVRGSHFFYEF-GDGFGVTDSWTKVDMKSVLIDPI 564
3SspiSSy	WKILNQCCFDQS-QFQEPFITFNLNSVRGSHFFYEF-GDGFGVTDSWTKVDMKSVLIDPI 564
4SaMono	WKG MNKCLFEQT-PFAEPFVG FNVN TVRGSQFFYQH-GDGYAVTESWTKDLSLSVLIHPI 570
27DQ785793.1	WKMMNKEMMTS--SFSKYFVEVSANLARMAQWIYQHESDGFGMQHSLVNKMLRDLLFHRY 590
28DQ785794.1	WKEMNTVTTASDCPFTDDLVA VATNLARAAQFIYLD-GDGHGVQHSEIHQQMGGLLFQPY 580

** : * , * . : : . * . * : : : * . * * . : : * : . : * : .

1SaSSy	PLGEE- 569
2SauSSy	PLGEE- 569
3SspiSSy	PLGEE- 569
4SaMono	PLNEED 576
27DQ785793.1	E----- 591
28DQ785794.1	V----- 581

图 12d