(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101914558 B
(45)授权公告日 2013.05.29

(21)申请号 201010239173.0

(22)申请日 2003.10.02

(30)优先权数据
60/415,765 2002.10.04 US
PCT/IB02/05070 2002.12.02 IB

(62)分案原申请数据
200380100941.2 2003.10.02

(73)专利权人 弗门尼舍有限公司
地址 瑞士日内瓦

(72)发明人 米歇尔·沙尔克 安东尼·克拉克

(74)专利代理机构 北京三幸商标专利事务所 11216
代理人 刘激扬

(51)Int.Cl.
C12N 15/60(2006.01)
C12N 9/88(2006.01)
C12N 15/63(2006.01)
C12N 1/21(2006.01)
C12N 1/19(2006.01)
C12N 5/10(2006.01)
C12P 5/00(2006.01)
C12P 7/02(2006.01)

(56)对比文件
Xia,D. 等.《GenBank:Q8RVR2》.《GenBank》.2002,
Xiao YaChen 等.Cloning, Expression, and Characterization of (+)-δ-Cadinene Synthase: A Catalyst for Cotton Phytoalexin Biosynthesis.《Archives of Biochemistry and Biophysics》.1995, 第 324 卷(第 2 期),255-266.
Xiao YaChen 等.Cloning and Heterologous Expression of a Second (+)-δ-Cadinene Synthase from Gossypium arboreum Cloning and Heterologous Expression of a Second (+)-δ-Cadinene Synthas from Gossypium arboreum.《J. Nat. Prod》.1996, 第 59 卷(第 10 期),944-951.

审查员 刘新蕾

权利要求书1页 说明书19页 附图21页

(54)发明名称

倍半萜烯合成酶及其使用方法

(57)摘要

本发明涉及倍半萜烯合成酶及其制备和使用方法。一个实施方案中,本发明提供了含有文中所述核苷酸序列的核酸,该核苷酸序列编码至少一种倍半萜烯合成酶。另一实施方案中,本发明还提供了倍半萜烯合成酶及其制备和使用方法。例如,本发明的倍半萜烯合成酶可用于将法呢基焦磷酸酯转化为各种氧化的和脂肪族倍半萜烯,包括朱栾倍半萜、二环-吉马烯、苹澄茄醇和 δ-杜松烯。

CN 101914558 B

1. 一种分离的核酸，其编码多肽 SEQ ID NO :1。

2. 权利要求 1 所述的分离的核酸，其中该核酸由核苷酸序列 SEQ ID NO :6 构成。

3. 一种分离的多肽，该多肽由氨基酸序列 SEQ ID NO :1 构成。

4. 一种载体，含有权利要求 1 所述的核酸。

5. 权利要求 4 所述的载体，其中所述载体是病毒载体或质粒。

6. 一种用权利要求 1 所述的核酸修饰的微生物。

7. 权利要求 6 所述的微生物，其中所述微生物是原核生物或酵母。

8. 权利要求 7 所述的微生物，其中所述原核生物选自大肠杆菌、枯草杆菌、假单胞属菌种或链霉菌属菌种，并且其中所述酵母是选自酵母属、毕赤酵母属或克鲁维酵母属的酵母。

9. 一种制备重组宿主细胞的方法，包括将权利要求 4 的载体导入宿主细胞。

10. 一种制备莖澄茄醇合成酶的方法，该方法包括 :

在能够产生所述莖澄茄醇合成酶的条件下培养用至少一种核酸序列修饰的宿主，该核酸序列编码多肽 SEQ ID NO :1。

11. 权利要求 10 所述的方法，其中该核酸由核苷酸序列 SEQ IDNO :6 构成。

12. 权利要求 10 所述的方法，其中所述宿主是植物或微生物。

13. 权利要求 10 所述的方法，其中所述宿主选自原核细胞、酵母细胞、植物细胞和动物细胞。

14. 一种制备至少一种萜类化合物的方法，该方法包括 :

A) 让法呢基焦磷酸酯与由氨基酸序列 SEQ ID NO :1 构成的多肽接触 ;并且

B) 分离 A) 中制得的至少一种萜类化合物。

15. 权利要求 14 所述的方法，其中所述至少一种萜类化合物选自倍半萜烯。

16. 权利要求 15 所述的方法，其中所述倍半萜烯是莖澄茄醇和 α - 莖澄茄油烯。

17. 权利要求 15 所述的方法，其中所述倍半萜烯是莖澄茄醇。

18. 权利要求 14 所述的方法，其中接触时的 pH 值为 7。

19. 权利要求 14 所述的方法，其中接触时的 pH 值小于 7。

20. 权利要求 14 所述的方法，其中所述至少一种萜类化合物是通过选自提取和蒸馏的至少一种方法分离的。

21. 权利要求 14 所述的方法，其中所述多肽是通过培养含有核酸的宿主细胞制备的，其中所述宿主细胞选自原核细胞、酵母细胞、植物细胞和动物细胞。

22. 权利要求 14 所述的方法，其中所述多肽是通过培养含有权利要求 1 所述核酸的微生物宿主细胞制备的。

23. 权利要求 14 所述的方法，其中所述多肽是通过培养含有权利要求 1 所述核酸的大肠杆菌宿主细胞制备的。

倍半萜烯合成酶及其使用方法

[0001] 本发明是2003年10月2日申请的发明名称为“倍半萜烯合成酶及其使用方法”的第200380100941.2号发明专利申请的分案申请。

[0002] 本申请要求2002年10月4日提交的申请号为No.60/415,765美国临时申请的优先权，该文献内容在此引入作为参考。本申请还要求2002年12月2日提交的申请号为PCT/IB 02/05070的国际申请的优先权，该文献内容在此引入作为参考。

技术领域

[0003] 本申请涉及倍半萜烯合成酶及其制备和使用方法。一个实施方案中，本发明提供了含有本申请所述核苷酸序列的核酸，该核苷酸序列编码至少一种倍半萜烯合成酶。另一实施方案中，本发明还提供了倍半萜烯合成酶及其制备和使用方法。例如，本发明的倍半萜烯合成酶可用于将法呢基焦磷酸酯转化为各种氧化的和脂肪族倍半萜烯，包括朱栾倍半萜、二环－吉马烯、荜澄茄醇和 δ－杜松烯。

背景技术

[0004] 萜烯化合物是一大类结构差异很大的天然分子。植物界中有最为多样性的单萜和倍半萜烯。通常它们在植物抵抗病虫害及草食动物以及吸引授粉昆虫方面发挥作用。

[0005] 萜烯的生物合成已经在很多种生物体内进行了广泛的研究。萜烯的常见前体为焦磷酸异戊烯酯(IPP)，催化产生IPP步骤的酶大多数已经研究清楚。IPP生物合成的两个不同途径现已清楚（图1）。甲羟戊酸酯途径发现于植物细胞液和酵母中，而非－甲羟戊酸酯途径（或脱氧木酮糖-5-磷酸(DXP)途径发现于植物质体和大肠杆菌中。

[0006] 例如，IPP在IPP异构酶作用下异构化为二甲基烯丙基二磷酸酯，可用异戊烯基转移酶将这两种C 5化合物缩合生成每一类萜烯的脂肪族焦磷酸酯萜烯前体，即对于单萜而言是牻牛儿基焦磷酸酯(GPP)，对于倍半萜烯而言是法呢基焦磷酸酯(FPP)，对于二萜烯为牻牛儿基牻牛儿基焦磷酸酯(GGPP)（图2）。催化脂肪族前体环化步骤的酶命名为萜烯环化酶或萜烯合成酶，在本申请中称之为萜烯合成酶。

[0007] 这些酶还能够催化复合体多步骤环化形成萜烯或倍半萜烯化合物的碳骨架。例如，催化环化的起始步骤可以是二磷酸酯基团的离子化形成烯丙基阳离子。然后，该底物进行异构和重排，这些反应可以通过酶活性位点控制。产物可以是，例如非环、单、二或三－环。然后，质子从碳正离子释放出来或者碳正离子与水分子反应，释放出萜烯烃或醇。一些萜烯合成酶产生单一产物，但是大部分产生多种产物。

[0008] 自然界中发现了很多种的萜烯结构和倍半萜烯合成酶。现已从不同的植物来源克隆和定性了几种编码cDNA或基因的倍半萜烯合成酶，例如，源自烟草(Facchini，P.J.and Chappell，J.(1992)Proc.Natl.Acad.Sci.U.S.A.89,11088-11092.)和辣椒(Back，K.，et al.(1998)Plant Cell Physio.39(9),899-904.)的5-表－马兜铃烯(aristolochene)合成酶，源自Hyoscyamus muticus的vetispiradiene合成酶(Back,K.and Chappell,J.(1995)J.Biol.Chem.270(13),7375-7381.)，源自欧薄荷的(E)-β－金合欢烯合成酶(Crock，J.，

et al.(1997)Proc.Natl.Acad.Sci.U.S.A.94(24),12833-12838.),源自北美冷杉的 δ-芹子烯合成酶和 γ-葎草烯合成酶 (Steele,C.L.,et al.(1998)J.Biol.Chem.273(4),2078-2089.),源自亚洲棉的 δ-杜松烯合成酶 (Chen,X.Y.,et al.(1995)Arch.Biochem.Biophys.324(2),255-266;Chen,X.Y.,et al.(1996)J.Nat Prod.59,944-951.),源自北美冷杉的 E-α-没药烯合成酶 (Bohlmann,J.,et al.(1998)Proc.Natl.Acad.Sci.U.S.A.95(12),6756-6761.),源自番茄的吉马烯 C 合成酶 (Colby,S.M.,et al.(1998)Proc.Natl.Acad.S ci.U.S.A.95(5),2216-2221.),源自黄花蒿的表-雪松醇合成酶和紫穗槐 -4,11- 二烯合成酶 (Mercke,P.,et al.(1999)Arch.Biochem.Biophys.369(2),213-222;Mercke,P.,et al.(2000)Arch.Biochem.Biophys.381(2),173-180.),以及源自莴苣、菊苣和加拿大一枝黄花的吉马烯 A 合成酶 (Bennett,M.H.,et al.(2002)Phytochem.60,255-261;Bouwmeester,H.J.,et al.(2002)Plant Physiol.129(1),134-144;Prosser I,et al.(2002)Phytochem.60,691-702)。

[0009] 很多倍半萜烯化合物被用于香料(例如,广藿香醇 (patchoulol),诺卡酮,檀香醇,韦愒酮,甜橙醛),且很多提取自植物。因此,它们的产量和价格取决于植物来源的波动以及主产国的稳定性。因此,人们希望获得非植物依赖性的倍半萜烯生产系统。由于一些倍半萜烯结构的复杂性,所以其化学合成的造价往往让人难以接受。

发明内容

[0010] 一个实施方案中,本发明涉及编码倍半萜烯合成酶的分离的核酸。本申请中的倍半萜烯合成酶还可以根据该酶与脂肪族焦磷酸酯萜烯前体例如法呢基焦磷酸酯接触时产生的至少一种化合物来命名。例如,能生成二环吉马烯作为其一种产物的倍半萜烯合成酶可称为二环吉马烯合成酶。按照这一惯例,本发明核酸的实例包括编码莙澄茄醇合成酶 (GFTpsC)(SEQ ID NO:1);δ-杜松烯合成酶 (GFTpsE)(SEQ ID NO:2);二环吉马烯合成酶 (GFTpsB)(SEQ IDNO:3);朱栾倍半萜合成酶 (GFTpsD1&GFTpsD2)(SEQ ID NO:4&SEQ ID NO:5) 的 cDNA。

[0011] 在一个实施方案中,本发明提供了一种分离的核酸,该核酸选自:(a) 含有基本上如 SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸;(b) 编码基本上如 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽的核酸;以及 (c) 能够在低度严谨条件下与 (a) 或 (b) 中核酸杂交的核酸,其中由所述核酸编码的多肽是倍半萜烯合成酶。在一个实施方案中,定义的条件为中度严谨条件,另一个实施方案中为高度严谨条件。其他实施方案包括:本发明核酸编码的多肽;含有本发明核酸的宿主细胞;经修饰含有本发明核酸的非人生物体;以及制备多肽的方法,其中包括培养本发明的宿主细胞。

[0012] 在另一个实施方案中,本发明提供了一种分离的多肽,其含有基本如 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ IDNO:4 或 SEQ ID NO:5 所示的氨基酸序列。

[0013] 在另一个实施方案中,本发明提供了一种含有至少一种核酸的载体,该核酸选自:(a) 含有基本上如 SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸;(b) 编码基本上如 SEQ ID NO:1、SEQ ID NO:2、SEQID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽的核酸;以及 (c) 能够在低度严谨条件下与 (a) 或 (b)

中核酸杂交的核酸，其中由所述核酸编码的多肽是倍半萜烯合成酶。其他实施方案包括，制备重组宿主细胞的方法，其中包括将本发明载体导入宿主细胞。

[0014] 在一个实施方案中，本发明提供了一种制备至少一种倍半萜烯合成酶的方法，该方法包括将经修饰以含有至少一种核酸序列的宿主在能够产生所述至少一种倍半萜烯合成酶的条件下培养。在一个实施方案中，所述的至少一种核酸选自：(a) 含有基本上如 SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸；(b) 编码基本上如 SEQID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4 或 SEQID NO:5 所示多肽的核酸；以及 (c) 能够在低度严谨条件下与 (a) 或 (b) 中核酸杂交的核酸，其中由所述核酸编码的多肽是倍半萜烯合成酶。所述宿主可选自，例如，植物、微生物、菌体细胞、酵母细胞、植物细胞和动物细胞。

[0015] 在另一个实施方案中，本发明提供了一种制备至少一种萜类化合物的方法，该方法包括：1) 让至少一种脂肪族焦磷酸酯萜烯前体与由核酸编码的至少一种多肽接触，在一个实施方案中，所述核酸选自：(a) 含有基本上如 SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸；(b) 编码基本上如 SEQ ID NO:1、SEQ ID NO:2、SEQ IDNO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽的核酸；以及 (c) 能够在低度严谨条件下与 (a) 或 (b) 中核酸杂交的核酸，其中由所述核酸编码的多肽是倍半萜烯合成酶，2) 分离 (1) 中制备的至少一种萜类化合物。在一个实施方案中，所述至少一种萜类化合物选自倍半萜烯。在另一个实施方案中，所述至少一种脂肪族焦磷酸酯萜烯前体为法呢基焦磷酸酯。本发明方法制备的倍半萜烯包括，但不限于，二环吉马烯、荜澄茄醇、朱栾倍半萜、α－荜澄茄油烯、吉马烯 D 和 δ－杜松烯 (图 3)。

[0016] 应该理解的是上面的概述和下面的详述都只是例举和解释性的，而并不是对本发明的限制。现在对本发明的范例性实施方案进行详述。

附图说明

[0017] 图 1：焦磷酸异戊烯酯生物合成途径的实例 (甲羟戊酸酯途径 (A) 和脱氧木酮糖途径 (B))。

[0018] 图 2：由焦磷酸异戊烯酯生物合成萜烯的实例。

[0019] 图 3：倍半萜烯化合物实例的结构。

[0020] 图 4：两组倍半萜烯合成酶的氨基酸序列比对的中间部分 (番茄的吉马烯 C δ－杜松烯合成酶 (SEQ ID NO:11)；欧薄荷的 (E)－β－金合欢烯 (SEQ ID NO:12)；北美冷杉的 δ－芹子烯 (SEQ ID NO:13)；香橙 (citrus junos) 的倍半萜烯合成酶 (SEQ ID NO:14)；烟草的 5－表－马兜铃烯 (SEQ ID NO:15)；辣椒的 5－表－马兜铃烯 (SEQ ID NO:16)；马铃薯的 Vetispiradiene(SEQ ID NO:17)；H.muticus 的 Vetispiradiene(SEQ ID NO:18)；亚洲棉的 δ－杜松烯 (SEQ ID NO:19)；黄花蒿的紫穗槐 -4,11－二烯 (SEQ ID NO:20)；黄花蒿的表－雪松醇 (SEQ ID NO:21)；北美冷杉的 δ－葎草烯 (SEQ ID NO:22)) 以及推导的简并引物的序列 (TpsVF1(SEQ ID NO:23)、TpsVF2(SEQ ID NO:24)、TpsVR3(SEQ ID NO:25)；TpsCF1(SEQ ID NO:26)、TpsCF2(SEQ ID NO:27)、TpsCR3(SEQ ID NO:28))。每一比对列下面的箭头表示用于设计简并引物及其方向的比对区域。

[0021] 图 5：柚子总 RNA RT-PCR 扩增产物推导出的氨基酸序列比对 (GFTpsA(SEQ ID NO:

29)、GFTpsB(SEQ ID NO:30) 和 GFTpsC(SEQ ID NO:31))。

[0022] 图 6:源自葡萄柚的倍半萜烯合成酶 GFTpsA(部分克隆)(SEQ ID NO:32),GFTpsB(SEQ ID NO:3),GFTbsC(SEQ ID NO:1),GFTpsD1(SEQ ID NO:4),GFTpsD2(SEQ ID NO:5) 和 GFTpsE(SEQ ID NO:2) 的氨基酸序列比对。

[0023] 图 7:由重组柚子倍半萜烯合成酶制备的倍半萜烯的 GC 谱。

[0024] 图 8:(a)GFTpsA(SEQ ID NO:32 和 SEQ ID NO:33)、(b)GFTpsB(SEQ ID NO:3 和 SEQ ID NO:8)、(c)GFTbsC(SEQ ID NO:1 和 SEQ ID NO:6)、(d)GFTpsD1(SEQ ID NO:4 和 SEQ ID NO:9)、(e)GFTpsD2(SEQ ID NO:5 和 SEQ ID NO:10) 以及 (f)GFTpsE(SEQ ID NO:2 和 SEQ ID NO:7) 的氨基酸及核苷酸序列。

具体实施方式

[0025] 萜烯是含有异戊二烯单元 (C5H8) 的不饱和烃,可以是脂肪族或环状。萜烯衍生物,包括但不限于樟脑、薄荷醇、萜品醇和茨醇、香叶醇。本申请中的萜烯或萜类化合物包括萜烯和萜烯衍生物,包括经过一个或多个步骤功能化的化合物,例如羟基化、异构化、氧化-还原或酰化作用。在本申请中,倍半萜烯是具有 C15 结构的萜烯,包括倍半萜烯和倍半萜烯衍生物,包括经过一个或多个步骤功能化的化合物,例如羟基化、异构化、氧化-还原或酰化作用。

[0026] 本申请中,衍生物是指获自已知或假定化合物且含有亲体结构基本单元的任一化合物。

[0027] 本申请中,倍半萜烯合成酶是任一种能催化倍半萜烯合成的酶。

[0028] “等同”、“基本上等同”或“基本如所示”这些词是指相关序列与给定序列之间具有至少 70%、75%、80%、85%、90%、92%、95%、96%、97%、98%或 99%的同一性。例如,所述序列可以是等位变体、来自不同物种的序列,或者它们可以是从给定序列通过截短、缺失、氨基酸取代或添加得到的。就多肽而言,比对序列的长度通常为至少 20、30、50、100 或更多个氨基酸。就核酸而言,比对序列的长度通常为至少 50、100、150、300 或更多个核苷酸。两序列之间的同一性百分比可以通过常规对齐算法得到,例如,Altschul et al.(1990)J.Mol.Biol.,215:403-410 中所述的 Basic Local Alignment Tool(BLAST)、Needleman et al.(1970)J.Mol.Biol.,48:444-453 中的算法、或者 Meyers et al.(1988) Comput.Appl.Biosci.,4:11-17 中的算法。

[0029] 因此,在一个实施方案中,本发明提供了一种分离的核酸,该核酸选自:(a) 含有基本上如 SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸;(b) 编码基本上如 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽的核酸;以及 (c) 能够在低度严谨条件下与 (a) 或 (b) 中核酸杂交的核酸,其中由所述核酸编码的多肽是倍半萜烯合成酶。在一个实施方案中,定义的条件为中度严谨条件,另一个实施方案中为高度严谨条件。

[0030] 本申请中,本领域技术人员可以利用本申请实施例中描述的酶特性试验 (enzyme characterization assay) 测定本发明核酸编码的多肽是否是倍半萜烯合成酶。

[0031] 本申请中,杂交或在特定条件下杂交是用来描述杂交条件,在该条件下相互基本等同或同源的核苷酸序列在洗涤后仍然保持彼此结合。所述条件可以是具有至少约 70%,

例如至少约 80%，以及例如至少约 85-90%同一性的序列仍然保持彼此结合。本发明中提供了低度、中度和高度严谨条件的定义。

[0032] 本领域技术人员可以根据 Ausubel et al.(1995)，Current Protocols in Molecular Biology，John Wiley&Sons，一书第 2、4 和 6 章中例举的最小限度试验 (minimal experimentation) 选择合适的杂交条件。另外，还有 Sambrook et al.(1989)Molecular Cloning：A Laboratory Manual，2nd ed.，Cold Spring Harbor Press，第 7、9 和 11 章中描述的严谨条件。本申请中的低度严谨条件如下所述。将含有 DNA 的滤膜 40℃下在含下列物质的溶液中预处理 6 小时：35%甲酰胺，5xSSC，50mMTris-HCI(pH 7.5)，5mM EDTA，0.1% PVP，0.1% Ficoll，1% BSA，和 500μg/ml 变性的鲑鱼精子 DNA。杂交在相同溶液中进行，除下列改动外：0.02% PVP，0.02% Ficoll，0.2% BSA，100μg/ml 鲑鱼精子 DNA，10% (wt/vol) 葡聚糖硫酸酯，使用 5-20x10^6 ^{32}P- 标记探针。将滤膜在杂交混合液中 40℃下孵育 18-20 小时，然后 55℃下用含有 2x SSC，25mM Tris-HCI(pH 7.4)，5mM EDTA 和 0.1% SDS 的溶液洗涤 1.5 小时。用新鲜溶液取代洗涤溶液，60℃下再孵育 1.5 小时。滤膜进行印迹干燥并进行放射自显影。

[0033] 本申请中定义的中度严谨条件如下所述。将含有 DNA 的滤膜 50℃下在含下列物质的溶液中预处理 7 小时：35%甲酰胺，5xSSC，50mM Tris-HCI(pH7.5)，5mM EDTA，0.1% PVP，0.1% Ficoll，1% B SA，和 500μg/ml 变性的鲑鱼精子 DNA。杂交在相同溶液中进行，除下列改动外：0.02% PVP，0.02% Ficoll，0.2% BSA，100μg/ml 鲑鱼精子 DNA，10% (wt/vol) 葡聚糖硫酸酯，使用 5-20x10^6 ^{32}P- 标记探针。将滤膜在杂交混合液中 50℃下孵育 30 小时，然后 55℃下用含有 2xSSC，25mM Tris-HCI(pH 7.4)，5mM EDTA 和 0.1% SDS 的溶液洗涤 1.5 小时。用新鲜溶液取代洗涤溶液，60℃下再孵育 1.5 小时。滤膜进行印迹干燥并进行放射自显影。

[0034] 本申请中定义的高度严谨条件如下所述。将含有 DNA 的滤膜 65℃下在由下列物质组成的缓冲液中预杂交 8 小时至过夜：6xSSC，50mM Tris-HCI(pH7.5)，1mM EDTA，0.02% PVP，0.02% Ficoll，0.02% BSA 和 500μg/ml 变性的鲑鱼精子 DNA。将滤膜在 65℃下在含 100μg/ml 变性鲑鱼精子 DNA 和 5-20x10^6 ^{32}P- 标记探针的预杂交混合液中预杂交 48 小时。将滤膜在 37℃下用含有 2xSSC，0.01% PVP，0.01% Ficoll 和 0.01% BSA 的溶液洗涤 1 小时。然后再用 0.1xSSC 在 50℃下洗涤 45 分钟。

[0035] 如果上面给出的条件不合适（例如，当用于种间杂交时），还可以选用本领域已知的其他低度、中度和高度严谨条件（例如，当用于种间杂交时）。

[0036] 在一个实施方案中，本发明所述核酸和 / 或多肽分离自柑桔类植物，例如柚子或橙子。在一个具体实施方案中，本发明涉及特定的分离的核苷酸序列，包括那些基本上不含内源性污染物质的核苷酸序列。术语“核酸”或“核酸分子”是指单链或双链形式的脱氧核糖核酸或核糖核酸多聚体，除另有说明外，其中还包括与天然生成核苷酸具有类似功能的天然核苷酸的已知类似物。“核苷酸序列”还指以单个片段或作为更大核酸组分形式存在的多核苷酸分子或寡核苷酸分子。所述核苷酸序列或分子还被称为“核苷酸探针”。本发明中的一些核酸分子源自分离的 DNA 或 RNA，这些分离的 DNA 或 RNA 至少一度曾以基本纯化形式存在，其量或浓度足以保证能够用常规生化方法对其组分核苷酸序列进行鉴定、扩增和回收。这些方法的实例，包括本申请中 PCR 方案所使用的方法，参见 Sambrook et al.，Molecular

Cloning:A Laboratory Manual,2nd ed.,Cold Spring Harbor Laboratory,Cold Spring Harbor,NY(1989),Current Protocols in Molecular Biology edited by F.A.Ausubel et al.,John Wiley and Sons,Inc.(1987),和 Innis,M.et al.,eds.,PCR Protocols:A Guide to Methods and Applications,Academic Press(1990)。

[0037] 在本申请中,本发明所述核酸分子包括单链和双链形式的DNA及其RNA互补链。DNA包括,例如,cDNA、基因组DNA、化学合成DNA、PCR扩增的DNA及其组合。基因组DNA包括翻译区、非翻译区及调控区,可以用常规技术来分离,例如使用本发明的任一种cDNA或其合适的片段作为探针对一段基因组DNA进行鉴定,然后可以使用本领域常规方法克隆该基因组DNA。通常,本发明范围内的核酸分子包括:在上述杂交和洗涤条件下能够与本发明所述序列杂交的序列,比本发明所述序列DNA双螺旋熔融温度低5℃、10℃、15℃、20℃、25℃或30℃的序列,包括被这些范围囊括在内的任意范围的条件。

[0038] 在另一个实施方案中,本发明所述核酸包括基本如SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9,或SEQ ID NO:10所示的序列。在一个实施方案中,所述核酸与SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9或SEQ ID NO:10所示核苷酸之间具有至少85%、至少90%或至少95%的同一性。在一个实施方案中,所述核酸包含SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9或SEQ ID NO:10所示的核苷酸序列。在另一个实施方案中,所述核酸编码一种蛋白质,该蛋白质是一种倍半萜烯合成酶,如实施例部分酶测定中所述。本发明还提供了含有不同物种间保守区域的核酸。

[0039] 在另一实施方案中,所述核酸包含由SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9或SEQ ID NO:10中的至少50、100、250、500或750个连续核苷酸组成的连续片段。所述的这些核苷酸的连续片段还包括至少1处突变,只要突变体序列仍然保留原序列的功能,仍然能够与这些核苷酸在低或高度严谨条件下杂交,例如,中度或高度严谨条件。所述片段可以源自,例如,SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9或SEQ ID NO:10中的第200-1600位核苷酸(nt)、第800-1600位nt、第1000-1600位nt、第200-1000位nt、第200-800位nt、第400-1600位nt或第400-1000位nt。

[0040] 如上所述,由本发明所述核酸编码的多肽也在本发明范围之内。本发明所述的分离的核酸可选自编码基本如SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4或SEQ ID NO:5所示多肽的核酸。在一个实施方案中,所述多肽与SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4或SEQ ID NO:5之间具有至少85%、至少90%或至少95%的序列同一性。

[0041] 在一个实施方案中,本发明的多肽含有SEQ ID NO:1、SEQ IDNO:2、SEQ ID NO:3、SEQ ID NO:4或SEQ ID NO:5所示的氨基酸序列。在另一个实施方案中,所述多肽含有基本如SEQ IDNO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4或SEQ IDNO:5所示的氨基酸序列。另一个实施方案中,所述多肽含有与SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4或SEQ ID NO:5所示氨基酸序列间具有至少85%、至少90%或至少95%序列同一性的氨基酸序列。在一个实施方案中,所述多肽是一种倍半萜烯合成酶,如下面实施例部分酶测定中所述。

[0042] 由于遗传密码子的简并性,能够编码同一氨基酸的密码子不止一个,所以多个DNA序列能编码同一多肽。这些变体DNA序列可以获自遗传漂变或人工操作(例如,发生于PCR

扩增过程中或者作为天然序列目的诱变(deliberate mutagenesis)的产物)。因此,本发明包括能够编码源自SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9或SEQ ID NO:10蛋白质的任一核酸或其变体。

[0043] 天然序列的目的诱变可以用本领域已知的许多种技术完成。例如,可以使用寡核苷酸-定点-特异性诱变方法,特别是在期望使基因突变从而通过取代、缺失或插入使预定限制性位点的核苷酸或密码子变化的情况下。进行所述变化的方法的实例见Walder et al.(Gene 42:133,1986);Bauer et al.(Gene 37:73,1985);Craik(BioTechniques, January 12-19,1985);Smith et al.(Genetic Engineering:Principles and Methods, Plenum Press,1981);Kunkel(Proc.Natl.Acad.Sci.USA 82:488,1985);Kunkel et al.(Methods in Enzymol.154:367,1987);和美国专利US4,518,584和US4,737,462。

[0044] 在一个实施方案中,本发明提供了分离的多肽。本申请中的"多肽"是指一类多肽或肽片段,其中含有本申请鉴定出的氨基酸序列,以及更小的片段。替代地,可以根据多肽与本发明核酸序列编码的任一肽之间的抗原性相关程度来对该多肽进行定义。因此,在一个实施方案中,本发明范围内的多肽定义为其中含有与本发明核酸序列编码的任一肽相同线性或三维表位的氨基酸序列。替代地,本发明范围内的多肽是能够被具有下述特性的抗体识别的,即该抗体能够特异性识别本发明核酸序列编码的任意一种肽。如果抗体与本发明多肽结合时的Ka值大于或等于约10^7M^{-1},例如大于或等于10^8M^{-1},那么就将该抗体定义为特异性结合。

[0045] 本申请中的多肽"变体"是指基本上与天然多肽同源的多肽,但是其氨基酸序列因为一处或多处缺失、插入或取代而不同于本发明任一核酸所编码的多肽。

[0046] 变体可以包括保守取代序列,即一特定氨基酸残基被一具有类似理化特性的残基所取代。保守取代的实例包括用一个脂肪族残基取代另一个脂肪族残基,例如,Ile、Val、Leu或Ala之间相互取代,或者用一个极性残基取代另一个极性残基,例如在Lys和Arg;Glu和Asp;或Gln和Asn之间相互取代。参见,Zubay,Biochemistry,Addison-Wesley Pub. Co.,(1983)。这类取代的效果可以用取代打分矩阵(substitution score matrices)来计算,如Altschul,(J.Mol.Biol.219:555-65,1991)中所述的PAM-120、PAM-200和PAM-250。其他这类保守取代,例如具有近似疏水特性的完整区域的取代,已为本领域熟知。

[0047] 天然生成的肽变体也在本发明范围之内。这种变体的实例是本申请所述多肽由于改变mRNA剪接事件或者蛋白水解切割得到的蛋白质。蛋白水解引起的变化包括,例如,在不同类型宿主细胞中表达时,本发明序列编码的多肽由于蛋白水解除去了一个或多个末端氨基酸而导致N-或C-末端的差异。

[0048] 本发明倍半萜烯合成酶的变体可以用于实现酶促活性的增强或减弱,区域化学(regiochemistry)或立体化学的变化,或者底物利用或者产物分配的改变。可以用本领域已知的任一方法得到变体或实现定点诱变。

[0049] 如上所述,本发明提供了重组及非重组、分离且纯化的多肽,例如从柑桔类植物。天然多肽的变体及衍生物可以通过下列方法得到:从其他或同一植物品系或品种中分离天然生成的变体或变体的核苷酸序列,或者对编码天然柑桔多肽的核苷酸序列的突变进行人工设计。改变天然氨基酸序列可以用许多种常规方法中的任一种来完成。可以通过下述方法将突变导入特定位置:合成含有突变序列的寡核苷酸,与限制性酶切位点侧接以能够与

天然序列的片段连接。连接后，得到的重构序列编码含有预期的氨基酸插入、取代或缺失的类似物。替代地，可以使用寡核苷酸－定点－特异性诱变方法提供一种经过改变的基因，其中预定的密码子因取代、缺失或插入而改变。

[0050] 在一个实施方案中，本发明涉及：含有本发明所述核酸的载体。例如，含有至少一种选自下列核酸的载体：(a) 含有基本上如 SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸；(b) 编码基本上如 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽的核酸；以及 (c) 能够在低度严谨条件下与 (a) 或 (b) 中核酸杂交的核酸，其中由所述核酸编码的多肽是倍半萜烯合成酶。

[0051] 本发明使用的载体包括任一种重组载体，包括但不限于病毒载体、噬菌体及质粒。

[0052] 含有本发明所述核酸序列的重组载体可以用熟知的方法制备。在一个实施方案中，所述表达载体中含有编码多肽的 cDNA 序列，该 cDNA 序列可操作地连接有合适的转录或翻译调控核苷酸序列，例如源自哺乳动物、细菌、病毒或昆虫基因的调控序列。调控序列的实例包括转录启动子、操纵子、或增强子，mRNA 核糖体结合位点、以及调控转录及翻译起始和终止的合适序列。当调控序列与本发明的 cDNA 序列功能性相连时那么核苷酸序列就是“可操作地连接”。因此，如果一启动子核苷酸序列能够调控 cDNA 序列的转录，那么该启动子序列就是可操作地连接于该 cDNA 序列。此外还可以向表达载体中引入在预期宿主细胞中的复制能力和选择基因，所述复制能力通常由复制起点赋予，选择基因可以是转化株得以通过其鉴定的那些。

[0053] 此外，还可以向表达载体引入信号肽编码序列，天然状态下该信号肽与本发明多肽并不相连。例如，可以将编码信号肽（分泌前导肽）的 DNA 序列与本发明的核苷酸序列融合在同一阅读框内从而使得本发明的多肽最初翻译为含有信号肽的融合蛋白。在预期宿主细胞中发挥作用的信号肽可以增进所表达多肽的胞外分泌。可在分泌离开细胞时将信号肽从多肽上切除。

[0054] 可以在本发明多肽的氨基及羧基端融合其他的肽序列，用以提高多肽的表达或者促进蛋白质的纯化。

[0055] 在一个实施方案中，本发明包括含有本发明所述核酸的宿主细胞。本发明的另一实施方案是一种制备重组宿主细胞的方法，该方法包括将本发明载体导入宿主细胞。在另一个实施方案中，提供了制备多肽的方法，该方法包括将本发明宿主细胞培养于能够产生多肽的条件下。一个实施方案中，对所述多肽进行回收。本发明所述方法包括制备至少一种本发明所述倍半萜烯合成酶的方法，该方法包括：培养含有本发明所述核酸的宿主细胞，然后回收积聚的倍半萜烯合成酶。

[0056] 适于表达本发明多肽的宿主细胞包括原核细胞、酵母或高等真核细胞。适用于细菌、真菌、酵母及哺乳动物细胞宿主的克隆和表达载体的描述参见，例如，Pouwels et al., Cloning Vectors:A Laboratory Manual, Elsevier, New York, (1985)。另外还可以采用无细胞翻译系统，用源自本发明所述 DNA 构建体的 RNA 制备公开的多肽。

[0057] 原核细胞包括革兰氏阴性或革兰氏阳性生物体，例如，大肠杆菌或杆菌。适于转化的原核细胞包括，例如，大肠杆菌、枯草杆菌、鼠伤寒沙门氏菌和假单胞属的其他菌种、链霉菌、以及葡萄球菌。原核宿主细胞中，例如大肠杆菌中，所述多肽可以含有有助于重组多肽

在原核细胞中表达的N-末端甲硫氨酸残基。可以将N-末端甲硫氨酸残基从表达的重组多肽上切除。

[0058] 适用于原核宿主细胞的表达载体的实例包括从可商购的质粒如克隆载体pET质粒(Novagen，Madison，WI，USA)或pBR322(ATCC37017)得到的载体。pBR322中含有氨苄青霉素及四环素抗性基因，因而提供了鉴定转化细胞的便捷方法。就使用pBR322构建表达载体而言，可以将合适的启动子和编码一种或多种本发明所述多肽的DNA序列插入pBR322载体。其他可商购的载体包括，例如，pKK223-3(Pharmacia Fine Chemicals，Uppsala，Sweden)和pGEM-1(Promega Biotec，Madison，WI，USA)。其他可商购的载体包括为蛋白表达专门设计的载体；这些载体包括用于与麦芽糖结合蛋白融合的蛋白质表达的pMAL-p2和pMAL-c2载体(New England Biolabs，Beverly，MA，USA)。

[0059] 重组原核宿主细胞表达载体常用的启动子序列包括细菌噬菌体T7启动子(Studier F.W.and Moffatt B.A.，J.Mol.Biol.189：113，1986)、β-内酰胺酶(青霉素酶)、乳糖启动子系统(Chang et al.，Nature 275：615，1978；和Goeddel et al.，Nature 281：544，1979)、色氨酸(trp)启动子系统(Goeddel et al.，Nucl.Acids Res.8：4057，1980；和EP-A-36776)、以及tac启动子(Maniatis，MoLecular Cloning：A Laboratory Manual，Cold Spring Harbor Laboratory，p.412，1982)。一种特别有用的原核宿主细胞表达系统使用λ噬菌体PL启动子和c1857ts热敏感抑制序列(thermolabile repressor sequence)。可从美国典型培养物保藏中心(″ATCC″)获得的其中引入PL启动子衍生物的质粒载体，包括质粒pHUB2(存在于大肠杆菌JMB9菌株(ATCC 37092))和pPLc28(存在于大肠杆菌RR1菌株(ATCC 53082))。

[0060] 另外还可以将本发明所述多肽表达于酵母宿主细胞、优选来自酵母属(例如，酿酒酵母)。其他酵母属也可使用，例如毕赤酵母属或克鲁维酵母属(例如，乳酸克鲁维酵母)。酵母载体通常含有来自2μ酵母质粒复制序列起点、自主复制序列(ARS)、启动子区域、聚腺苷酸化用序列、转录终止序列和可筛选标志基因。其中适于酵母载体的合适启动子包括，例如，金属硫蛋白、3-磷酸甘油酸激酶(Hitzeman et al.，J.Biol.Chem.255：2073，1980)的启动子，或其它糖酵解酶的启动子(Hess et al.，J.Adv.Enzyme Reg.7：149，1968；和Holland et al.，Biochem.17：4900，1978)，例如烯醇化酶，甘油酸-3-磷酸脱氢酶、己糖激酶、丙酮酸脱羧酶、果糖磷酸激酶、葡糖-6-磷酸异构酶、3-磷酸甘油酸变位酶、丙酮酸激酶、丙糖磷酸异构酶、葡糖磷酸异构酶和葡糖激酶。其他适于酵母表达的载体和启动子的描述参见Hitzeman，EPA-73，657或Fleer et.al.，Gene，107：285-195(1991)；和van den Berg et.al.，Bio/Technology，8：135-139(1990)。另一种替代方案是葡萄糖-抑制ADH2启动子，其描述见Russell et al.(J.Biol.Chem.258：2674，1982)和Beier et al.(Nature 300：724，1982)。通过将来自pBR322中在大肠杆菌中负责筛选和复制的DNA(Ampr基因和复制起始位点)插入上述酵母载体，可以构建出在大肠杆菌和酵母中均可复制的穿梭载体。

[0061] 本发明的一个实施方案是经过修饰以含有本发明所述核酸的非人生物体。可以通过本领域已知的任一种基因转移方法对所属非人生物体和/或宿主细胞进行修饰，包括，例如，使用递送载体，如脂质及病毒载体、裸露DNA、电穿孔及微粒介导基因转移。在一个实施方案中，所述非人生物体是植物、昆虫或微生物。

[0062] 例如，在一个实施方案中，本发明提供了一种制备至少一种倍半萜烯合成酶的方

法，该方法包括将经修饰以含有至少一种核酸的宿主培养在利于所述至少一种倍半萜烯合成酶产生的条件下，其中所述的至少一种核酸选自：(a) 含有基本上如SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸；(b) 编码基本上如 SEQ ID NO:1、SEQID NO:2、SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽的核酸；以及 (c) 能够在低度严谨条件下与 (a) 或 (b) 中核酸杂交的核酸，其中由所述核酸编码的多肽是倍半萜烯合成酶。

[0063] 在另一个实施方案中，所述宿主是一种植物（如烟草）、动物或微生物，还包括但不限于菌体细胞、酵母细胞、植物细胞和动物细胞。在本申请中，植物细胞和动物细胞包括用植物和动物作为宿主。例如，在本方明的一些实施方案中，表达是在经过基因修饰后的非人生物体中进行。

[0064] 在一个实施方案中，使用哺乳动物或昆虫宿主细胞培养系统表达本发明的重组多肽。用于在昆虫细胞中制备异源蛋白的杆状病毒系统的综述见 Luckow and Summers,Bio/Technology6 :47(1988)。另外还可以使用从哺乳动物来源建立的细胞系。合适的宿主细胞系的实例包括 COS-7 猴肾细胞系 (ATCC CRL1651)(Gluzman et al.,Cell 23 :175,1981),L 细胞、C127 细胞、3T3 细胞 (ATCC CCL163)、中华仓鼠卵巢 (CHO) 细胞、HeLa 细胞和 BHK(ATCC CRL10) 细胞系，以及源自非洲绿猴肾细胞系 CVI(ATCC CCL70) 的细胞系 CV-1/EBNA-1(ATCC CRL10478)，描述见 McMahan et al.(EMBO J.10 :2821,1991)。

[0065] 为将 DNA 导入哺乳动物细胞而建立的方法现已公开 (Kaufman，R.J.，Large Scale Mammalian Cell Culture,1990，pp.15-69)。其他使用可商购的试剂例如 Lipofcctamine(Gibco/BRL) 或 Lipofectamine-Plus 的方法，也可用于转染细胞 (Felgner et al.，Proc.Natl.Acad.Sci.USA 84 :7413-7417,1987)。此外，可以使用常规方法通过电穿孔转染哺乳动物细胞，例如 Sambrook et al.Molecular Cloning :A Laboratory Manual,2ed.Vol.1-3,Cold Spring Harbor Laboratory Press,1989) 中所描述的方法。可以用细胞毒性物质抗性作为筛选方法筛选得到稳定的转化体。Kaufman et al.，Meth.In Enzymology 185 :487-511,1990 中描述了几种筛选方案，例如二氢叶酸还原酶 (DHFR) 抗性。适于 DHFR 筛选的宿主细胞株有 CHO 细胞株 DX-B11，该细胞株是 DHFR 缺陷型的 (Urlaub and Chasin,Proc.Natl.Acad.Sci.USA77 :4216-4220,1980)。可以将能表达 DHFR cDNA 的质粒导入细胞株 DX-B11，这时只有那些含有该质粒的细胞才能在合适的筛选培养基中生长。

[0066] 哺乳动物细胞表达载体用的转录及翻译调控序列可以从病毒基因组切取。常用的启动子序列和增强子序列源自多瘤病毒、腺病毒 2、猴病毒 40(SV40) 和人巨细胞病毒。源自 SV40 病毒基因组的 DNA 序列，例如 SV40 起点、早期和晚期启动子、增强子、剪接（位点）以及聚腺苷酸化位点，可用于提供哺乳动物宿主细胞表达用的其他基因元件。病毒早期和晚期启动子特别有用，因为它们容易以片段形式从病毒基因组获得，片段中还含有病毒复制起点 (Fiers et al.，Nature 273 :113,1978 ;Kaufman，Meth.in Enzymology,1990)。

[0067] 现已有数种本领域已知的用于制备转基因植物的方法。这些方法包括，但不限于：植物原生质体电穿孔、脂质体介导的转化、聚乙二醇介导的转化、植物细胞的微注射以及使用病毒进行的转化。在一个实施方案中，使用的是粒子轰击定向基因转移。

[0068] 粒子轰击定向基因转移为转化植物组织提供了一个实例。该技术中，包被了 DNA 的粒子或微抛射体被射出穿越细胞的物理屏障。粒子轰击可用于将 DNA 导入 DNA 包被粒子

可穿透的任一靶组织，但是为了稳定转化，必须使用可再生的细胞。通常，粒子由金或钨制成。可以使用 $CaCl_2$ 或乙醇沉淀的方法将 DNA 包被在粒子上，这些方法都是本领域已知的常规方法。

[0069] DNA 包被粒子用粒子枪射出。合适的粒子枪可从 Bio-Rad Laboratories(Hercules,CA) 购买。粒子的穿透由改变参数来控制，例如爆裂强度、粒子大小或者粒子到达靶组织必须行经的距离。

[0070] 包被粒子用的 DNA 包括适于驱动目的基因表达的表达框，其中含有与目的基因可操作连接的启动子。

[0071] 粒子轰击定向基因转移方法的描述参见 Tomes 等人的美国专利 US5,990,387。

[0072] 在一个实施方案中，本发明的 cDNA 可用产生正义或反义 RNA 的方法表达。反义 RNA 就是与基因编码 mRNA（正义 RNA）反向互补的 RNA。驱动反义 RNA 表达的载体是这样一种载体，cDNA 以相对于启动子而言“相反方向”放入其中从而使得非编码链（而不是编码链）被转录。反义 RNA 的表达被用于下调该反义 RNA 互补 mRNA 编码的蛋白质的表达量。产生反义 RNA 的载体可用于制备转基因植物，如上所述。

[0073] 在一个实施方案中，转染后的 DNA 整合入非人生物体的染色体从而得到稳定的重组系统。本领域已知的任一染色体整合方法都可用于实施本发明，包括但不限于，重组酶介导的表达框置换 (recombinase-mediated cassette exchange)(RMCE)、病毒位点特异性染色体插入、腺病毒以及原核注射。

[0074] 本发明的另一实施方案是制备萜类化合物和倍半萜烯化合物的方法，例如，使用本发明所述核苷酸及多肽。实施例包括用于制备至少一种萜类化合物的方法，该方法包括：让至少一种脂肪族焦磷酸酯萜烯前体与核酸编码的至少一种多肽接触，该核酸选自 (a) 含有基本上如 SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9 或 SEQ ID NO:10 所示核苷酸序列的核酸；(b) 编码基本上如 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽的核酸；以及 (c) 能够在低度严谨条件下与 (a) 或 (b) 中核酸杂交的核酸，其中由所述核酸编码的多肽是倍半萜烯合成酶，然后将制备得到的至少一种萜类化合物分离出来。另一实施例是制备至少一种萜类化合物的方法，该方法包括：让至少一种脂肪族焦磷酸酯萜烯前体与至少一种基本如 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4 或 SEQ ID NO:5 所示多肽接触，然后将制备得到的至少一种萜类化合物分离出来。

[0075] 本申请中使用的脂肪族焦磷酸酯萜烯前体是任意的脂肪族焦磷酸酯化合物，其是用于制备至少一种萜烯的前体，包括但不限于，牻牛儿基焦磷酸酯 (GPP)，法呢基焦磷酸酯 (FPP) 和牻牛儿基牻牛儿基焦磷酸酯 (GGPP)。

[0076] 在一个实施方案中，所述的至少一种萜类化合物选自倍半萜烯。在一个实施方案中，所述的脂肪族焦磷酸酯萜烯前体是法呢基焦磷酸酯。在另一个实施方案中，所述的至少一种倍半萜烯选自二环吉马烯 ((E,E)-3,7,11,11- 四甲基 - 二环 [8.1.0] 十一 -2,6- 二烯）、荜澄茄醇 ((1R,4S,5R,6R,7S,10R)-7- 异丙基 -4,10- 二甲基 - 三环 [4.4.0.0(1,5)] 癸 -4- 醇，朱栾倍半萜 ((+)-(1R)-1,2,3,5,6,7,8,8A- 八氢 -7- 异丙烯基 -1@,8A@- 二甲基萘）、α - 荜澄茄油烯、吉马烯 D 和 δ - 杜松烯 (rel-(1R,8AS)-1,2,3,5,6,8A- 六氢 -1- 异丙基 -4,7- 二甲基萘）（图 2）。本发明的萜类化合物可以用本领域使用的任一方法分离，包

括但不限于层析、提取及蒸馏。

[0077] 在一个实施方案中，可以通过改变合成酶与脂肪族焦磷酸酯萜烯前体（例如法呢基焦磷酸酯）接触时的pH值来调整产物分配或形成的实际产物。在一个实施方案中，pH为7。另一实施方案中，pH小于7，例如，6、5、4和3。

[0078] 另外，本发明还提供了一种生物体（例如，微生物或植物），该生物体用于构建一种平台，用于以高水平生产倍半萜烯合成酶底物（例如，FPP）以及将本发明核酸导入该生物体。例如，将至少一种编码倍半萜烯合成酶的本发明所述核酸引入制备FPP的非人生物体，从而完成FPP向倍半萜烯的转化以及随后的倍半萜烯的代谢产生。在一个实施方案中，这样就得到一种用于高水平生产倍半萜烯的平台。

[0079] 在一个实施方案中，本发明所述核酸可用于制备编码倍半萜烯合成酶的其他核酸。例如，本发明提供了一种鉴定倍半萜烯合成酶的方法，该方法包括：用本发明所述核酸构建DNA文库，从该文库中筛选出编码至少一种倍半萜烯合成酶的核酸。本发明所述核酸的文库可以用本领域已知的任一方法构建，其中用本发明核酸作为起始位点制备DNA序列，包括但不限于DNA重构（DNA shuffling）。在该方法中，用功能试验发现编码倍半萜烯合成酶的目标核酸从而从所述文库中筛选倍半萜烯合成酶。倍半萜烯合成酶的活性可以用例如本申请所述方法测定。在一个实施方案中，使用高通量筛选分析编码多肽的活性。

[0080] 本申请中的“核苷酸探针”是指能够通过形成一种或多种化学键、通过互补碱基配对或者通过形成氢键，与互补序列的目标核酸结合的寡核苷酸或多核苷酸。如上所述，所述寡核苷酸探针可以含有天然（即A，G，C或T）或者修饰碱基（7-脱氮鸟嘌呤、次黄嘌呤核苷等）。此外，核苷酸探针中的碱基可以通过磷酸二酯键以外的键连接，只要其不妨碍杂交即可。因此，寡核苷酸探针中可以含有通过磷酸二酯键以外的肽键连接的连续碱基。

[0081] 本申请中的“目标核酸”是指能够与所述核苷酸探针或分子特异性杂交的核酸。可以设计探针测定目标核酸的存在与否以及目标核酸的含量。目标核酸的序列与将相应探针中导向目标的核酸序列互补。正如本领域技术人员理解的那样，所述探针还可以含有与目标不能特异性杂交的其他核酸或其他部分，例如标记物。术语目标核酸可以是探针指向的较大核酸中的特定的核苷酸序列也可以是全序列（例如，基因或mRNA）。本领域技术人员可以认识到各种条件下的所有用途。

[0082] 除操作实施例或者另有说明之外，所有表示组分含量、反应条件的数值以及说明书和权利要求中使用的数值应理解为全部用“约”进行了限定。因此，除有相反说明外，说明书和权利要求中的所有数值参数都是近似值，可以根据本发明期望获得的理想特性来变动。绝对无意限制等同原则在权利要求范围上的应用，所以每一个数字参数都应该是根据有效数字的数目和常规舍入法得到的。

[0083] 尽管表示本发明较宽范围时使用的数值范围和参数是近似值，但是特定实施例中列出的数值却是尽可能精确的。但是，任一数值中都必然含有因各自测量存在的标准差所不可避免带来的一定误差。下列实施例是为了阐述本发明但决不对本发明的范围构成限制。百分比是根据重量计算出的。

[0084] 下列实施例是为了阐述本发明但决不对本发明的范围构成限制。

[0085] 实施例

[0086] 材料

[0087] 用柚子（葡萄柚）皮作为本申请所述试验的起始原材料。皮（果皮外部有颜色的部分）内含有油腺，油腺是萜烯生物合成的部位。用新鲜采摘的成熟果实在 Simone Gatto(Sicily) 制得新鲜的柚子皮。将皮 (3-4mm 厚）剥离后，立即冷冻，而且在运输过程和随后的所有步骤中都保持冷冻以防降解。

[0088] 实施例 1 ：用 RT-PCR 分离倍半萜烯合成酶 cDNA。

[0089] 将植物倍半萜烯合成酶的推导氨基酸序列进行比对，鉴定保守区并设计植物倍半萜烯合成酶－特异性寡核苷酸。为了获得较高的序列同源性，将序列分成两组（图 4)。第一组包括下列酶的序列 ：源自蕃茄 (Lycopersicon esculentum)cv.VFNT cherry(Colby et al,1998) 的吉马烯 C 合成酶、源自欧薄荷 (Mentha x piperita)(Crock et al,1997) 的 (E)－β－金合欢烯合成酶、源自北美冷杉 (Steele et al,1998) 的 δ－芹子烯合成酶、源自香橙 (GenBank 登记号 AF288465) 的倍半萜烯合成酶、源自烟草 (Facchini 和 Chappell，1992) 和辣椒 (Back et al,1998) 的 5- 表－马兜铃烯合成酶、源自马铃薯和 Hyoscyamus muticus(Back and Chappel,1995) 的 vetispiradiene 合成酶。第二组包括下列酶的序列 ：源自亚洲棉 (Chen et al.1995) 的 (+)－δ－杜松烯合成酶，源自黄花蒿的紫穗槐 -4,11- 二烯合成酶 (Mercke et al,2000) 和表－雪松醇合成酶 (Merck et al,1999) 以及源自北美冷杉 (Steele et al,1998) 的 γ－葎草烯合成酶。最高序列同源性发现于序列的中间部分。选择三个含有足够保守氨基酸的区域，设计这些区域的特异性简并寡核苷酸（即，从每一比对组得到两个正向和一个反向引物）（图 4)。

[0090] 用从 Hot B orate 技术制备柚子皮得到的总 RNA 以及正向和反向简并引物的不同组合，进行 RT-PCR。提取总 RNA 用的 Hot Borate 方法是由 Wan and Wilkins(Wan et al.(1994)Anal.Biochem.223,7-12.）改进得到的。在液氮中将这些组织用研钵和杵研磨成细粉。将得到的细粉加入 5ml 提取缓冲液 (200mM 硼酸盐、30mM EGTA,10mM DTT,1％ SDS,1％脱氧胆酸钠、2％ PVP,0.5％乙基苯基聚乙二醇 (NP40))80℃预热。将上述混合物用 UltraturaxTM 匀浆器均质化，然后用两层 MiraclothTM 滤膜 (Calbiochem) 过滤。将混合物在 0.5mg/ml 蛋白酶 K（用于蛋白质 / 核糖核酸酶消化）存在条件下 42℃孵育 90 分钟。然后加入 KCl 至 1M 终浓度，置于冰上孵育 1 小时，将混合物置于 4℃离心机中 10000g 离心 10 分钟。回收上清，加入 1/3 体积的 8M LiCl。将混合物 4℃孵育过夜，然后 4℃下 10000g 离心 20 分钟。弃上清，用 2M LiCl 洗涤沉淀，用上述条件再次离心。然后，将沉淀重悬于无 RNase 的蒸馏水，然后加入 0.15 倍体积的 2M 乙酸钾溶液和 2.5 倍体积的无水乙醇。-20℃下孵育 2 小时沉淀 RNA，然后照前述条件离心。用 80％乙醇洗涤 RNA 沉淀，然后重悬于无 RNase 的蒸馏水。

[0091] RNA 浓度由 260nm 处的 OD 值估测。用琼脂糖凝胶通过查证核糖体 RNA 条带的整齐度测定 RNA 纯度。

[0092] 用 Qiagen OneStep RT-PCR 试剂盒和 Eppendorf Mastercycler 梯度热循环仪进行 RT-PCR。常用反应混合物 50μl 终体积中含有 ：10μl 5XQiagen OneStep RT-PCR 缓冲液、400μM 每一 dNTP、400nM 每一引物、2μl Qiagen OneStep RT-PCR Enzyme Mix、1μl RNasin® 核糖核酸酶抑制剂 (Promega Co.）和 1μg 总 RNA。热循环仪条件为 ：50℃、30 分钟（反转录）；95℃、15 分钟 (DNA 聚合酶活化）；94℃、45 秒，42-45℃、10 秒（根据使用引物而定），72℃、45-90 秒（根据待扩增 DNA 片段长度而定）共 40 个循环 ；然后 72℃、

10 分钟。

[0093] 用 1％琼脂糖凝胶测算 PCR 产物的大小。将出现在预期大小对应位置上的条带从凝胶上切割下来，用QIAquick®凝胶提取试剂盒 (Qiagen) 纯化，然后用 TOPO TA 克隆试剂盒 (Invitrogen) 克隆入pCR®2.1-TOPO载体。然后对插入的cDNA进行DNA测序，将该序列用BLASTX算法 (Altschul et al1990) 与GenBank非冗余蛋白质数据库 (non redundant protein database)(NCBI) 进行比对。

[0094] 对预期长度 (80-240bp 大小)PCR 产物进行 DNA 测序和 Blast 搜索表明有三种倍半萜烯合成酶片段，命名为 GFTpsA、GFtpsB 和 GFTpsC。180-bp 的 GFTpsA 片段用引物 TpsVF1 和 TpsVR3 扩增，115-bp 的 GFtpsB 片段用 TpsVF2 和 TpsVR3 扩增，115-bp 的 GFTpsA 片段用 TpsVF1 和 TpsVR3 扩增。推导的氨基酸序列表明这三种片段之间差异显著 (图 5)。

[0095] 实施例 2 :用 5′ /3′ -RACE 分离倍半萜烯合成酶 cDNA

[0096] 为了分离倍半萜烯合成酶的全长序列，首先使用 5′ /3′ -RACE 方法，用 Marathon™cDNA 扩增试剂盒 (Clontech)，从纯化自 Hot Borate 技术制备总 RNA 的 1μg mRNA 起始。合成 cDNA 的质量较差，基本上都是很小的 cDNA(平均大小为 0.5Kb)。但是，用该 cDNA，通过基因特异性引物得到了 GFTpsB 和 GFTpsC 的 3’末端，GFTpsB 使用 GFTpsBRF1 和 GFTpsBRF2，GFTpsC 使用 GFTpsCRF1 和 GFTpsCRF2(参见表 1)。用硫氰酸胍 / 苯酚提取方法制备的 mRNA 得到了较高质量的 cDNA，平均大小为 2Kb，从而可以用基因特异性引物 GFTpsCRR1 和 GFTpsRR2(参见表 1) 分离 GFTpsC 的 5′ - 末端序列，从而完成该克隆的全长序列。GFTpsB 的 5′ - 末端和 GFTpsA 的 5′ - 末端及 3′ - 末端用该方法不能获得。

[0097] 硫氰酸胍 / 苯酚提取法采用 Chomczynski 和 Sacchi 所述技术，使用 ThermoHybaid 的 RNAClean™ 溶 液 (Chomczynski, P., and Sacchi, N., (1987)Anal.Biochem.162, 156-159.)。简而言之，在液氮中将 2g 冷冻组织用研钵和杵研磨成细粉。将得到的细粉转移到 20mlRNAClean™ 溶液，将悬液用 UltraturaxTM 匀浆器均质化，然后用 MiraclothTM 滤膜 (Calbiochem) 过滤。加入 0.1 倍体积的氯仿后，将试管置于冰上，然后 4℃下 12000g 离心 20 分钟。回收上层水相，加入 1 倍体积的异丙醇。-20℃孵育 20 分钟后，将样品在 4℃下 12000g 离心 20 分钟。将得到的白色沉淀大块用 70％乙醇洗涤，室温干燥。将这种含有总 RNA 的沉淀立即用FastTrack® 2.0mRNA 分离试剂盒 (Invitrogen) 通过 oligodT- 纤维素亲和层析进行 mRNA 纯化。按照厂商推荐方法进行，除下述改动外 :在将总 RNA 重悬于裂解缓冲液后，将样品由 65℃改为 100℃下加热。

[0098] cDNA3′ 和 5′ 末端快速扩增 (RACE) 用 Marathon™ cDNA 扩增试剂盒 (Clontech) 完成。该方法从合成第一链 cDNA 开始，用 oligo(dT) 引物从 mRNA 起始。第二链合成后，将特异性衔接子连接于双链 cDNA(ds cDNA) 末端。该方法提供了一种衔接子连接的 ds cDNA 的非克隆文库。用纯化的柚子 mRNA(1μg) 作为起始材料。用琼脂糖凝胶测算 cDNA 的质和量。

[0099] 使用Advantage® 2 聚合酶 Mix，用基因特异性和衔接子特异性寡核苷酸的组合扩增 cDNA 的 3’- 或 5’- 末端。常规 RACE 反应混合物 50μl 终体积中含有 :5μl 10XcDNAPCR 反应缓冲液 (clontech)、200nM 每一 dNTP、1μl Advantage® 2 聚合酶 Mix、200μM 衔接子特异性引物 (Clontech)、200μM 基因特异性引物 (参见表 1) 和 5μl 经 50-250 倍稀释后的衔接子连接的 cDNA。扩增在 Eppendorf Mastercycler 梯度热循环仪中进行。热循环仪

条件为：94℃、1 分钟；94℃、30 秒，72℃、2-4 分钟，共 5 个循环；94℃、30 秒，70℃、2-4 分钟，共 5 个循环；94℃、30 秒，68℃、2-4 分钟，共 20 个循环。必要时用巢式衔接子特异性引物(Clontech)和巢式基因特异性引物进行第二轮扩增。

[0100] 按照上文 RT-PCR 产物使用的方法，对扩增产物进行评定、亚克隆及序列分析。

[0101] 表 1：3′ /5′ -RACE 试验中使用的引物

[0102]

名称	定义	序列(从 5'到 3'端)
GFTpsARF1 (SEQ ID NO:33)	3'-RACE 正向引物	CTGGGAGTGTCCTATGAGCT CAATTTGG
GFTpsARF2 (SEQ ID NO:34)	GFTpsA 3'-RACE 正向巢式引物	GTAGAATTTTTGCATCCAAAG TGGTGTGC
GFTpsARR1 (SEQ ID NO:35)	GFTpsA 5'-RACE 反向引物	CACACCACTTTGGATGCAAA AATTCATC
GFTpsARR2 (SEQ ID NO:36)	GFTpsA 5'-RACE 反向巢式引物	CCAAATTGGGCTCATAGGAC ACTCCCAG
GFTpsBRF1 (SEQ ID NO:37)	GFTpsB 3'-RACE 正向引物	TAGGGACGTATTTTGAACCA AAGTAC
GFTpsBRF2 (SEQ ID NO:38)	GFTpsB 3'-RACE 正向巢式引物	AAATAATGACCAAAACAATTT ACACGG
GFTpsBRR1 (SEQ ID NO:39)	GFTpsB 5'-RACE 反向引物	GCACTTTCATGTATTCTGGAA G
GFTpsBRR2 (SEQ ID NO:40)	GFTpsB 5'-RACE 反向巢式引物	GTTTGAGCTCTTCAAAGAAA CC
GFTpsCRF1 (SEQ ID NO:41)	GFTpsC 3'-RACE 正向引物	AATGGGAGTGTATTTTGAGC CTCGATACTCC
GFTpsCRF2 (SEQ ID NO:42)	GFTpsC 3'-RACE 正向巢式引物	GATATTTTCCAAAGTAATTGC AATGGCATCC
GFTpsCRR1 (SEQ ID NO:43)	GFTpsC 5'-RACE 反向引物	GTTGCTAATATCCCACCTTTT GATAGC
GFTpsCRR2 (SEQ ID NO:44)	GFTpsC 5'-RACE 反向巢式引物	AAGTGTGCCATAGGCGTCGT AGG
GFTpsDRR1 (SEQ ID NO:45)	GFTpsD 5'-RACE 反向引物	CTGTTCCGCAAGCTTAGGGG TTACATG
GFTpsDRR2 (SEQ ID NO:46)	GFTpsD 5'-RACE 反向巢式引物	CTGAGCTACCAATGACTTCA GGTGAGTGG
GFTpsERR1 (SEQ ID NO:47)	GFTpsE 5'-RACE 反向引物	CAATTTTGCCATACACATCAT AGATATCATC
GFTpsERR2 (SEQ ID NO:48)	GFTpsE 5'-RACE 反向巢式引物	AACAGAAGTCATGGAGATCA CTTTCGTC
GFTpsERR3 (SEQ ID NO:49)	GFTpsE 5'-RACE 反向引物	CGCAAGAGATGTTTTAAAGTT CCCATCC
GFTpsERR4 (SEQ ID NO:50)	GFTpsE 5'-RACE 反向巢式引物	TGAACATCAGCGGAAATTTTA TAGCC

[0103] 发现用 3′ -RACE 得到的部分 GFTpsB cDNA 与公共数据库(GeneBank 登记号 AF288465)中的及分离自香橙的推定萜烯合成酶 cDNA 之间具有高度序列同一性。这种萜烯

合成酶的特异性引物设计为：根据香橙 (C. junos) 萜烯合成酶非编码区设计一个正向和反向引物对 (junosF1 和 junosR1)，根据含有起始和终止密码子在内的编码区设计一正向和反向引物对 (junosF2 和 junosR2)。用引物对 junosF1 和 junosR1 对柚子 Marathon™ cDNA 文库进行 PCR 扩增没有得到任何扩增子。用引物对 junosF2 和 junosR2 扩增出了一 1.6Kb 的片段。DNA 测序证实该 PCR 产物是 GFTpsB 克隆的 1669-bp 全长倍半萜烯合成酶 (图 6)。

[0104] 实施例 3 ：cDNA 文库筛选和 EST 测序

[0105] 为了得到 PCR 方法获得的部分克隆的全长 cDNA，构建了从柚子皮 mRNA 制备的 cDNA 文库。

[0106] 用 Uni-ZAP® XR 文库构建试剂盒 (Stratagene)，按照厂商推荐方法，从 7.5μg 柚子皮 mRNA(用硫氰酸胍 / 苯酚法制得) 起始进行 cDNA 合成和文库构建。文库的初始滴度为 $2x10^7$PFU(噬斑形成单位)，插入片段的平均长度为 1.1Kb。

[0107] 使用两种方法从上述 cDNA 文库分离编码 cDNA 的倍半萜烯合成酶 ：EST 测序以及用 DNA 探针筛选。就 EST(已表达序列标志) 测序而言，使用文库的一部分，用 Stratagene 的大量切除法 (mass excision protocol) 切除 Uni-ZAP XR 载体中的 pBluescript 噬菌粒。随机挑取得到的转化菌落 (576)，纯化质粒。对于每一克隆都用 T3 引物进行一测序反应。首先对序列进行编辑除去载体序列，然后用 BLASTX 算法 (Altschul et al1990) 与 GenBank 非冗余蛋白质数据库 (NCBI) 进行比对。

[0108] 用地高辛 (DIG) 标记的 DNA 探针对文库进行筛选。该探针是用 PCR DIG(地高辛) 探针合成试剂盒 (Roche Diagnostics) 通过 PCR 扩增将 D1G-dUTP 掺入 DNA 片段而合成的。从一等份文库中扩增出 149-bp 的 GFTpsA- 来源 DNA 探针，使用的引物为正向引物 GFTpsAproF(SEQ ID NO ：51)(5′ -ACGATTTAGGCTTCCCTAAAAAGG-3′) 和反向引物 GFTpsAproR(SEQ ID NO ：52)(5′ -TATTATGGAATATTATGCACACCAC-3′)。从 pET-GFTpsB2-2 质粒中扩增出 1036-bp 的 GFTpsB- 来源探针，使用的引物为正向引物 junosF2(SEQ ID NO ：53)(5′ -AAATGTCCGCTCAAGTTCTAGCAA CGG-3′) 和反向引物 GFTpsBRR2(SEQ ID NO ：54)(5′ -GTTTGAGCTCTTCAAAGAAACC-3′)。从 pET-GFTpsC 质粒中扩增出 1008-bp 的 GFTpsC- 来源 DNA 探针，使用的是正向引物 GFTpsCpetF1(SEQ ID NO ：55)(5′ -ATGGCACTTCAAGATTCAGA AGTTCC-3′) 和反向引物 GFTpsCRR1(SEQ ID NO ：56)(5′ -GTTGCTAATATCCCACCTTTTGATAGC -3′)。

[0109] 探针在 40℃下在 Dig Easy Hyb 杂交溶液 (Roche Diagnostics) 中杂交出噬斑挑取物 (plaques lifts)，噬斑挑取物制备自含有 5,000-25,000PFU 的平板。通过化学发光检测膜上的探针 - 目标杂合体，使用抗 - 地高辛 - 碱性磷酸酶 (Roche Diagnostics) 和 CDP-Star 碱性磷酸酶底物 (Roche Diagnostics)，用 VersaDoc Imaging System(BioRad) 显色。

[0110] 从琼脂糖平板上挖髓取出阳性信号的噬菌体，进行第二轮筛选，如果必要在进行第三轮筛选，直至获得单一噬斑阳性信号。体内切割阳性分离物，按照上述方法对插入片段测序。

[0111] 用从 GFTpsA、GFTpsB 和 GFTpsC 制备的 DNA 探针对文库进行筛选。该方法生产出三种不同倍半萜烯合成酶 cDNA。其中两种此前已有克隆 ：一种克隆名为 GF2-30-1，是 GFTpsC cDNA 的 5’- 末端 300-bp 截短形式 ；第二种名为 9-13-6，是 GFtpsA 的部分克隆，与其他倍半

萜烯合成酶相比 5’- 末端截短了大约 168-bp。与上述通过 RT-PCR 得到的部分 GFTpsA 克隆相比，9-13-6 克隆提供了 618-bp5’- 末端附加序列的信息。在其 3’- 末端，9-13-6 克隆也是不完整的。大约丢失了 680 个核苷酸，取而代之的是一功能未知的 600-bp 片段。筛选获得的第三种编码 cDNA 的倍半萜烯合成酶，GF2-5-11，其编码出一种新的名为 GFTpsD 的倍半萜烯合成酶。该克隆 5’- 末端截短，但是用表 1 中的引物通过 5′ -RACE 回收到了所丢失的 414-bp。然后，从 Marathon™ cDNA 文库扩增出全长 GFTpsD，并将其克隆入上述细菌表达载体。几种全长 GFTpsDcDNA 的 DNA 序列分析发现 ：有两种倍半萜烯合成酶密切相关，只有极小的序列差异，将其命名为 GFTpsD1 和 GFTpsD2（图 6）。这两种变体的推导氨基酸序列彼此间相差 4 个碱基。

[0112] 就 EST 测序方法而言，对 576 个克隆进行了测序。其中只有 3 个克隆编码推定萜烯合成酶 - 样蛋白，更确切地说，是两种不同的倍半萜烯合成酶 - 样蛋白和一种单萜合成酶 - 样蛋白。两个倍半萜烯合成酶 - 样克隆中的一个（克隆 GF002-G3）此前被鉴定为克隆 GFTpsD1。第二个克隆（克隆 GF006-G7）是一种截短的 cDNA，其编码一种新的名为 GFTpsE 的倍半萜烯合成酶。该克隆在 5’- 末端也是截短的（大约 800bp），但是该丢失的片段可以按照下述两个步骤通过 5′ -RACE 扩增得到。第一个 5′ -RACE 使用引物 GFTpsERR1 和 GFTpsERR2（参见表 1）提供了 500 个附加的 5’- 末端核苷酸，第二个 5′ RACE 使用引物 GFTpsERR3 和 GFTpsERR4（参见表 1）提供了丢失的 5’- 末端 DNA 序列，重构得到了全长 GFTpsE cDNA（图 6）。

[0113] 实施例 4 ：质粒构建和酶表达

[0114] 表达质粒的构建

[0115] 将 cDNA 亚克隆入 pET11a 表达质粒 (Novagen)，进行倍半萜烯合成酶的功能表达。就 GFTpsB，GFtpsD 和 GFTpsE 而言，通过 PCR 扩增全长 cDNA，以向 5’- 末端导入其中含有起始密码子的 Nde1 位点，在 3’- 末端紧接终止密码子之后导入 BamH1 位点。就 GFTpsB 而言，使用正向引物 GFTpsBNde1(SEQ ID NO ：57)5′ -GCATGTTC*CATATG* TCCGCTCAAGT TCTAGCAACGGTTTCC-3′ (Nde1 位点用斜体表示，起始密码子用下划线标出）和反向引物 GFTpsBBam(SEQ ID NO ：58)5′ -CGC*GGATCC* TCAGATGGTAACAGGGTCTCTGAGCACTGC-3′ (BamH1 位点用斜体表示，起始密码子用下划线标出）。同样，正向引物 GFTpsDNde1(SEQ ID NO ：59)5′ -GCATGTTC *CATATG* TCGTCTGGAGAAACATTTCGTCC-3′ 和反向引物 GFTpsDBam(SEQ ID NO ：60)5′ -CGC *GGATCC* TCAAAATGGAACGTGGTCTCCTAG-3′ 用于 GFTpsD ；以及正向引物 GFTpsENde1(SEQ ID NO ：61)5′ -GCATGTTC *CATATG* TCTTTGGAAGTTTCAGCCTCTCCTG-3′ 和反向引物 GFTpsEBam(SEQ ID NO ：62)5′ -CGC *GGATCC* TCATATCGGCACAGGATTAATAAACA AAGAAGC-3′ 用于 GFTpsE。

[0116] 扩增反应是使用 Pfu DNA 聚合酶 (Promega) 在含下列物质的 50μl 终体积中进行的 ：5μl Pfu DNA 聚合酶 10X 缓冲液、每一种 dNTP 各 200μM、正向和反向引物各 0.4μM、2.9 单位 Pfu DNA 聚合酶以及 5μl 经 100 倍稀释的 cDNA（按照上述方法用 Marathon™cDNA 扩增试剂盒 (Clontech) 制备）。热循环条件为 ：95℃、2 分钟 ；95℃、30 秒，55℃、30 秒，72℃、4 分钟，共 25 个循环 ；然后 72℃、10 分钟。PCR 产物在琼脂糖凝胶上纯化，用 QIAquick® 凝胶提取试剂盒 (Qiagen) 洗脱，接着用 Nde1 和 BamH1 消化，然后将其连接入经同样消化后的 pET11a 质粒。

[0117] 就 GFTpsC-pET11a 表达载体的构建而言，分别进行两次 PCR 扩增，使插入片段与 NdeI 和 BamHI 粘性末端侧接。第一次 PCR 在同上所述条件下进行，使用正向引物 GFTpsCpETF1(SEQ ID NO:63)5′ -TAATGGCACTTCAAGATTCAGAAGTTCCTC-3′ 和反向引物 GFTpsCpETR1(SEQ ID NO:64)5′ -AAAAGGGAACAGGCTTCTCAAGCAATG-3′，第二次 PCR 使用正向引物 GFTpsCpETF2(与 GFTpsCpETF1 相比 5'-末端少了 AT)(SEQ ID NO:65)5′ -ATGGCACTTCAAGATTCAGAAGTTCCTC-3′ 和反向引物 GFTpsCpETR2(与 GFTpsCpETR1 相比 5'-末端多出了 GATC)(SEQ ID NO:66)5′ -GATCAAAAGGGAACAGGCTTCTCAAGCAATG-3′ 。将两次 PCR 产物用上述方法纯化后合并，通过下述操作变性：煮沸 5 分钟然后冰上冷却 5 分钟。将得到的 cDNA 直接连接入 pET11a 质粒。

[0118] 起初将连接产物转化入 JM109 大肠杆菌细胞，通过限制性酶切消化和 DNA 测序验证构建体。

[0119] 倍半萜烯合成酶表达

[0120] 就蛋白表达而言，将含倍半萜烯合成酶 cDNA 的 pET11a 质粒和 pET11a 空质粒转化入 BL21(DE3) 大肠杆菌细胞 (Novagen)。将得到的单菌落接种 5ml LB 培养基。37℃孵育 5-6 小时后，将培养物转移到 20℃孵箱，放置 1 小时平衡。加入 0.5mM IPTG 诱导蛋白质的表达，然后将培养物在 20℃孵育过夜。

[0121] 第二天，离心收集细胞，重悬于 0.5ml 提取缓冲液 (50mMMOPSO pH7，5mM EDTA，5mM EDTA，10％甘油)，然后超声处理 3 次，每次 30 秒。18,000g 离心 30 分钟沉淀细胞碎片，回收含可溶蛋白的上清液。通过下述操作对倍半萜烯合成酶的表达进行评估：将蛋白提取物用 SDS-PAGE(SDS-聚丙烯酰胺凝胶电泳）分离，然后用考马氏亮蓝染色，然后与用空质粒转化细胞得到的蛋白提取物进行比较。

[0122] 所有构建体都观察到了一具有预期计算分子量的明显条带，但是用空质粒转化的大肠杆菌得到的可溶性蛋白中没有出现该条带。

[0123] 实施例 4 酶功能测定

[0124] 酶促试验在密封的玻璃试管中进行，使用终体积中含 50-100μl 蛋白提取物的提取缓冲液，该提取缓冲液补加了 15mM MgC12 和 100-250μM FPP(Sigma)。培养基上层用 1ml 戊烷覆盖，试管置于 30℃孵育过夜。回收含有倍半萜烯的戊烷相，用第二体积的戊烷提取培养基。将合并后的戊烷级分在氮中浓缩，用 Hewlett-Packard6890Series GC 系统的气相色谱分析，使用内径 0.25mm 长 30m 的 SPB-1(Supelco) 毛细管层析柱。载气为恒定流速 1.6ml/min 的 He 气。注射按 2 ：1 的比例分开进行，注射器温度设定为 200℃，烤箱 (oven) 程序设定为：从 80℃（保持 0 分钟）开始以 7.5℃／分钟的速率升至 200℃（保持 0 分钟），接着以 20℃／分钟的速率升至 280℃（保持 2 分钟）。检测使用火焰离子化检测器。在权威标准能够获得的情况下，化合物识别是根据与权威标准具有相同保留时间进行的。为了证实产物身份 (identities)，使用 Hewlett-Packard 6890GC-四极质谱检测器，用安装有内径 0.25mm 长 30m 的 SPB-1(Supelco) 毛细层析柱的组合毛细 (combined capillary)GC-MS 对样本进行分析。烤箱程序设定为：从 80℃（保持 0 分钟）以 7.5℃速率升至 200℃，He 气流的恒定流速为 1.5ml/min。用 2200V 电压的电子倍增器记录 70eV 时的光谱。

[0125] 用该试验测定不同重组酶的酶活性，使用可溶性蛋白级分，以法呢基-二磷酸酯作为底物。从所测的所有克隆获得倍半萜烯合成酶活性，生成的产物用保留时间以及气相

色谱－质谱联用技术 (GC-MS)(图 7) 来定性。GFTpsB cDNA 编码出的倍半萜烯合成酶，产生主产物二环－吉马烯 (图 3) 和至少 15 种次要的倍半萜烯链烯烃或氧化的倍半萜烯例如 δ－杜松烯 (图 3)(由二环－吉马烯热重排得到的二环－榄香烯 (图 3)，也出现在了 GC 图中)。GFTpsC cDNA 编码出一种生成多产物的倍半萜烯合成酶，主产物鉴定为荜澄茄醇 (图 3)。该酶还生产出了占相当大比例的 3 种其他倍半萜烯 -(-)-α－荜澄茄油烯和 2 种氧化后的倍半萜烯，以及少量的至少 11 种倍半萜烯链烯烃或氧化后的倍半萜烯。GFtpsD 编码出的倍半萜烯合成酶，可以生产出主产物朱栾倍半萜。10 个其他的小峰也被鉴定为倍半萜烯链烯烃或醇。其中，β－榄香烯是吉马烯 A 的热降解产物。GFTpsE 编码出的倍半萜烯合成酶，生产出的是倍半萜烯化合物与 δ－杜松烯 (图 3) 的复合混合物，其中 δ－杜松烯以微弱优势成为含量最丰富的组分。另外，从 GFTpsE 生成的产物中还鉴定出了荜澄茄醇和吉马烯 D(图 3)。

[0126] 上述试验中，分离的倍半萜烯合成酶表现出生成多产物的特性。例如，荜澄茄醇合成酶和 δ－杜松烯可以生成占相对较大比例的 3 或 5 种产物。二环吉马烯合成酶产生了主产物倍半萜烯和几种微量的次产物。

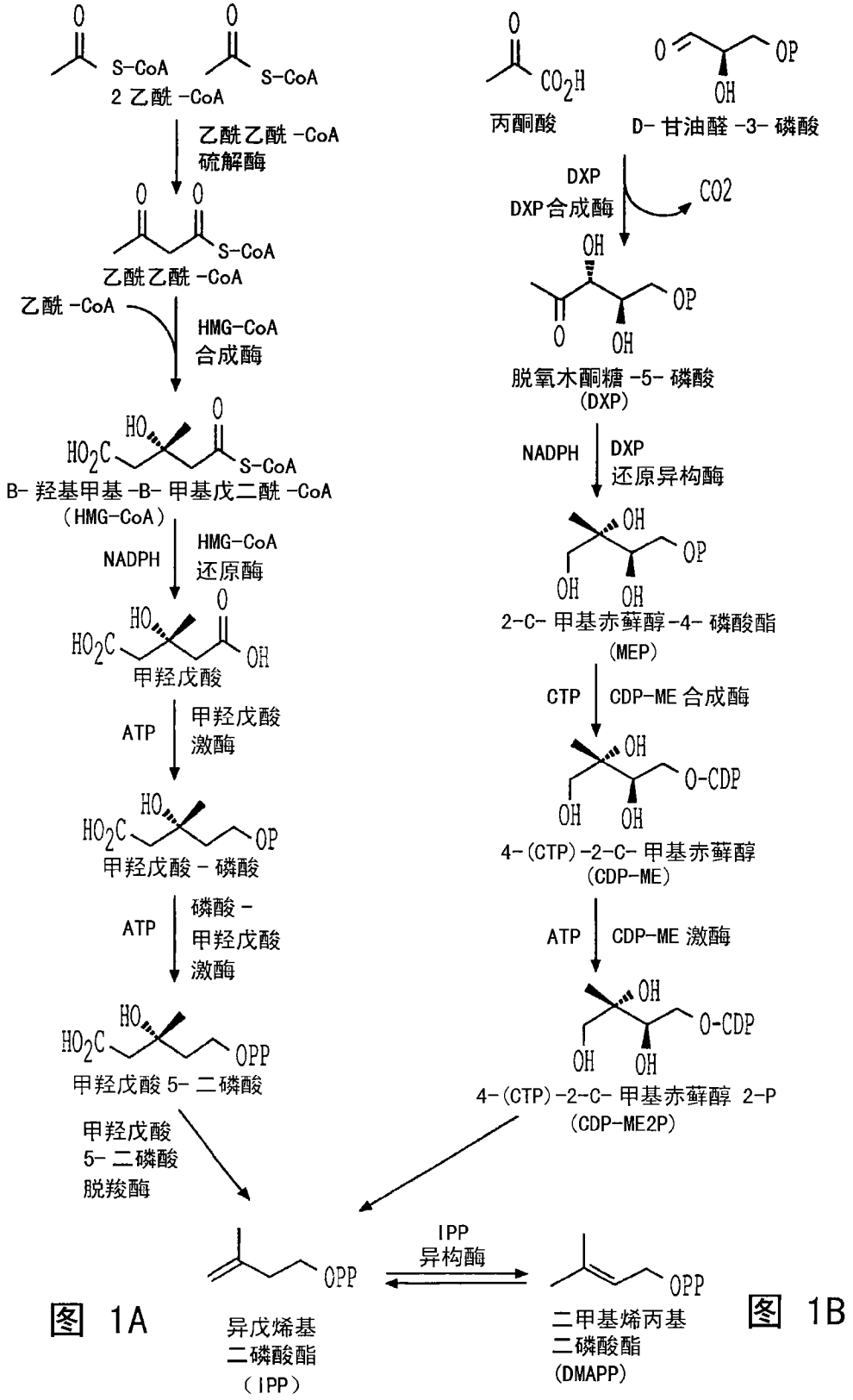

2 乙酰 -CoA
乙酰乙酰 -CoA 硫解酶
乙酰乙酰 -CoA
乙酰 -CoA
HMG-CoA 合成酶
B- 羟基甲基 -B- 甲基戊二酰 -CoA
（HMG-CoA）
NADPH
HMG-CoA 还原酶
甲羟戊酸
ATP
甲羟戊酸激酶
甲羟戊酸 - 磷酸
ATP
磷酸 - 甲羟戊酸激酶
甲羟戊酸 5- 二磷酸
甲羟戊酸 5- 二磷酸脱羧酶
异戊烯基二磷酸酯（IPP）
IPP 异构酶
二甲基烯丙基二磷酸酯 (DMAPP)
丙酮酸
D- 甘油醛 -3- 磷酸
DXP
DXP 合成酶
CO2
脱氧木酮糖 -5- 磷酸 (DXP)
NADPH
DXP 还原异构酶
2-C- 甲基赤藓醇 -4- 磷酸酯 (MEP)
CTP
CDP-ME 合成酶
4-(CTP)-2-C- 甲基赤藓醇 (CDP-ME)
ATP
CDP-ME 激酶
4-(CTP)-2-C- 甲基赤藓醇 2-P (CDP-ME2P)

图 1A 图 1B

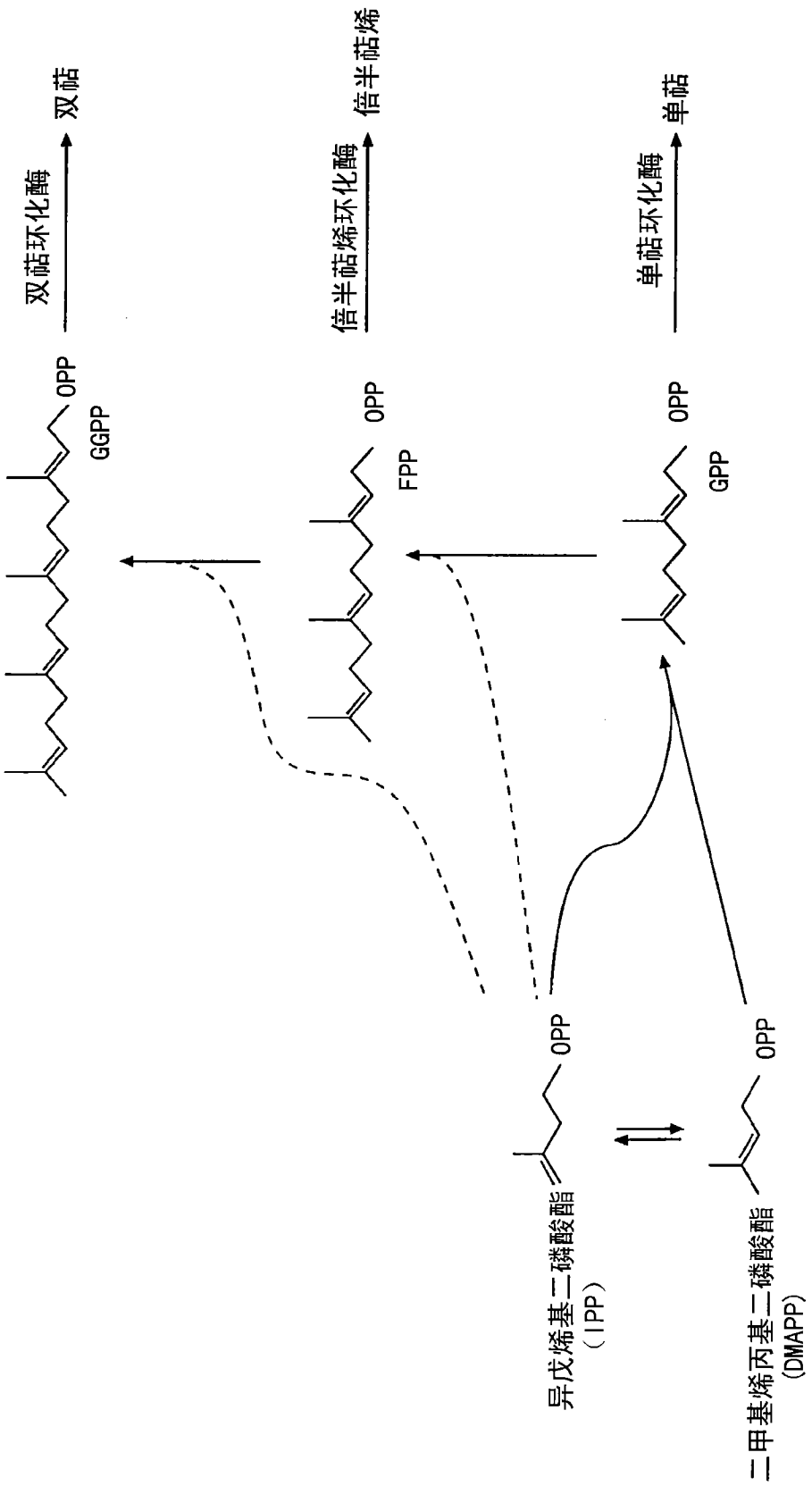

双萜环化酶
双萜
OPP
GGPP
倍半萜烯环化酶
倍半萜烯
OPP
FPP
单萜环化酶
单萜
OPP
GPP
OPP
异戊烯基二磷酸酯
(IPP)
OPP
二甲基烯丙基二磷酸酯
(DMAPP)

图 2

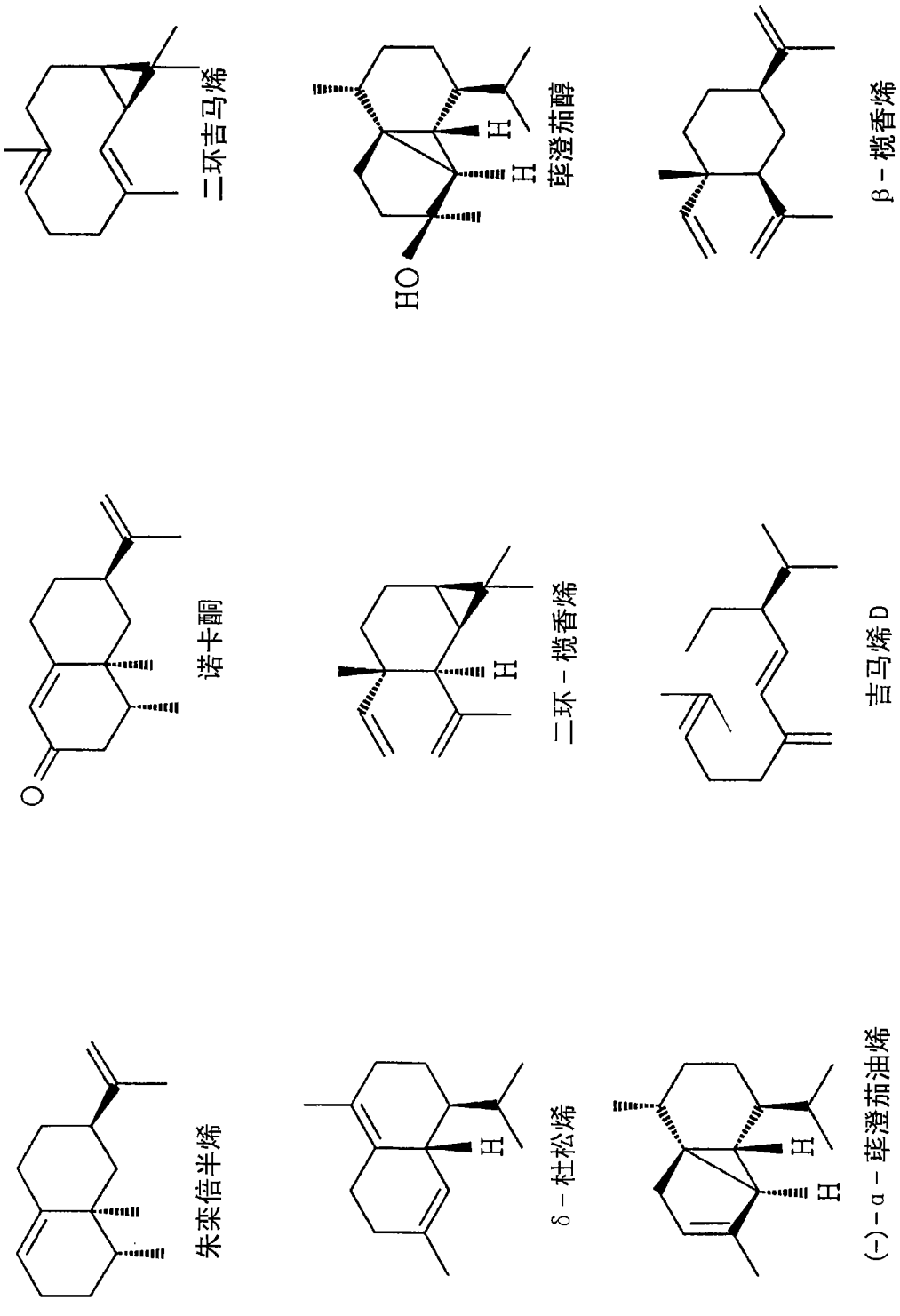


图 3

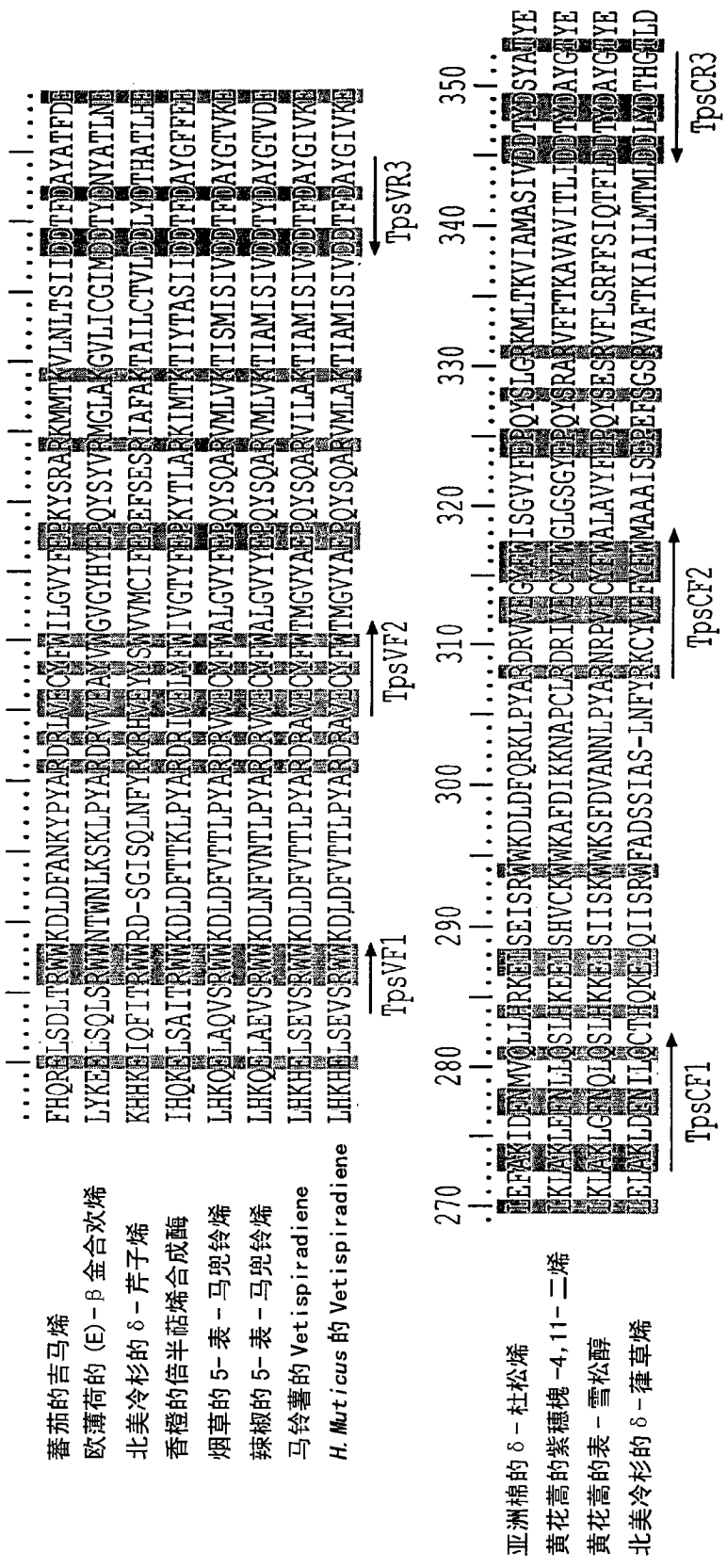

番茄的吉马烯
欧薄荷的 (E)-β 金合欢烯
北美冷杉的 δ- 芹子烯
香橙的倍半萜烯合成酶
烟草的 5- 表 - 马兜铃烯
辣椒的 5- 表 - 马兜铃烯
马铃薯的 Vetispiradiene
H. Mutticus 的 Vetispiradiene
亚洲棉的 δ- 杜松烯
黄花蒿的紫穗槐 -4,11- 二烯
黄花蒿的表 - 雪松醇
北美冷杉的 δ- 葎草烯
TpsVF1
TpsVF2
TpsVR3
TpsCF1
TpsCF2
TpsCR3
TpsVF1: 5′-(A/T/C) (A/T/C) (A/C/G)T(A/T/C)(A/T)C(A/C) (A/C)GGTGGTGGA-3′
TpsVF2: 5′-GT(G/A/T/C/)GA(G/A) (G/T) (G/T/C) (G/T/C)TAT(T/G)TTGG-5′
TpsVR3: 5′-C(A/G)T(A/G) (A/G) (G/T)T(A/G)TCG(T/A)A(A/T/G)G(T/G)GTC(A/G)TC-3′
TpsCF1: 5′-GC(T/G)AAG(A/T)T(A/C/G)GG(T/G)TTCAAT(A/C/T) (A/T) (T/G) (C/T/G)T(A/T/G)C-3′
TpsCF2: 5′-GGGA(A/T) (A/T)G(A/T) (A/T/G/C) (A/T) (C/T/G)GTTGAA(T/G) (T/G)TTATTTTTGG-3′
TpsCR3: 5′-GT(A/T) (C/G)CGT(A/T/G/C/)G(A/C/T)GTCGTA(A/C/T)GTCATC-3′

图 4

10
20
30
40
GFTpsA DLGFPKKVPY ARDRVVETYI WMLLGVSYEP NLAEGRIFAS KVVCIISII
GFTpsB ---------- ---------- --IVGTYEEP KYTLARKIMT KTIYTASII
GFTpsC ---------- ---------- ---MGVYEEP RYSFARKILS KVIAMASIL

图 5

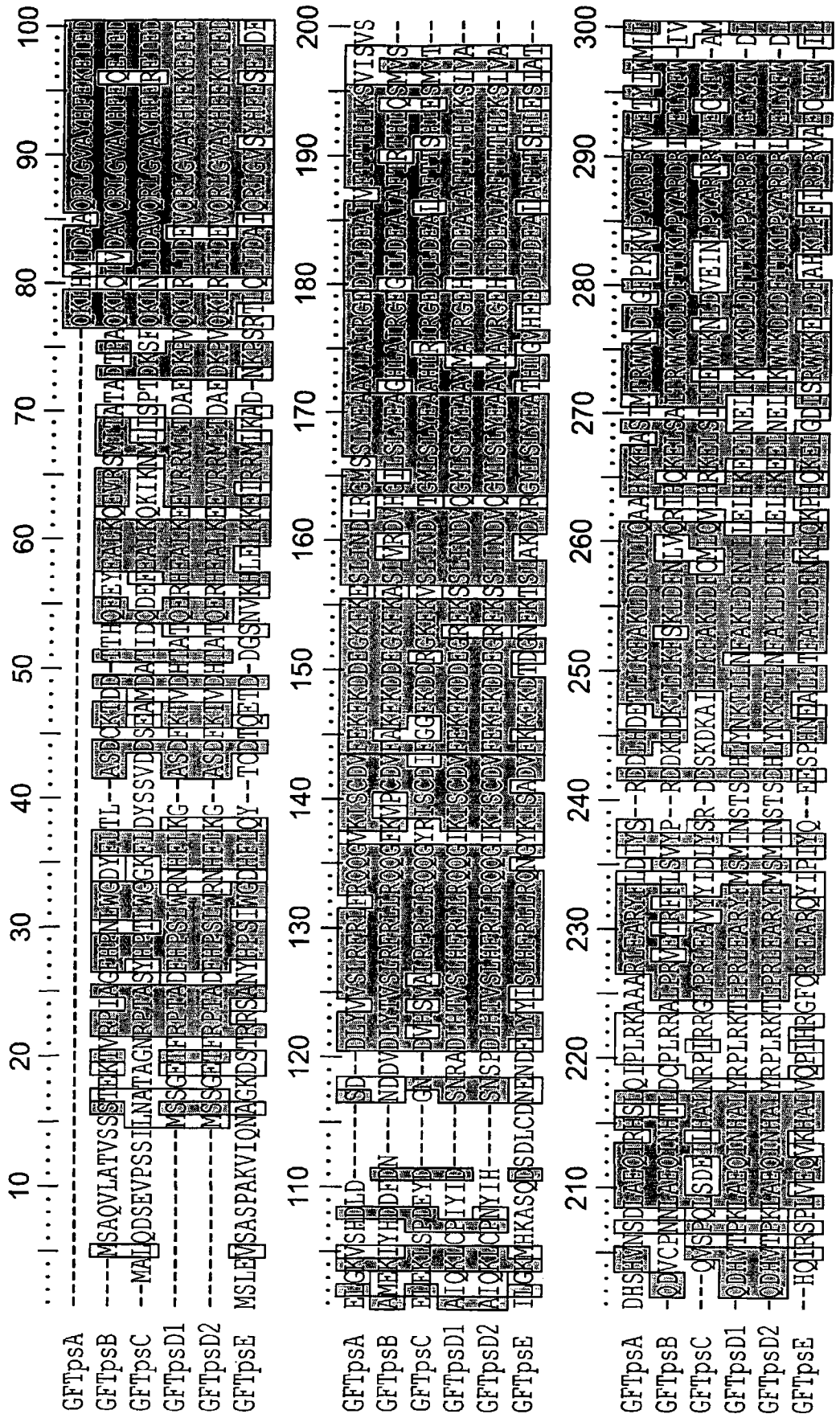


图 6A

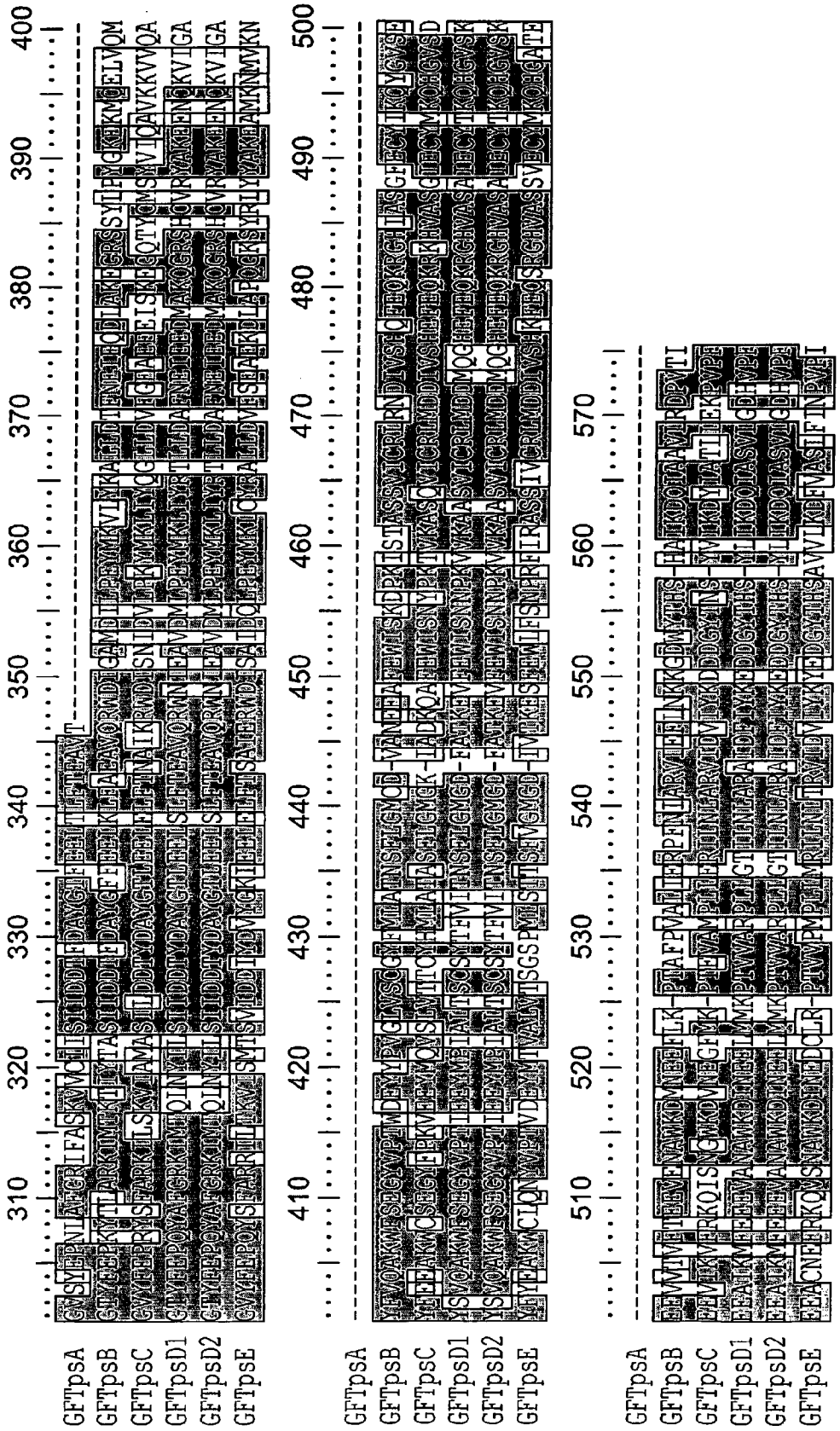


图 6B

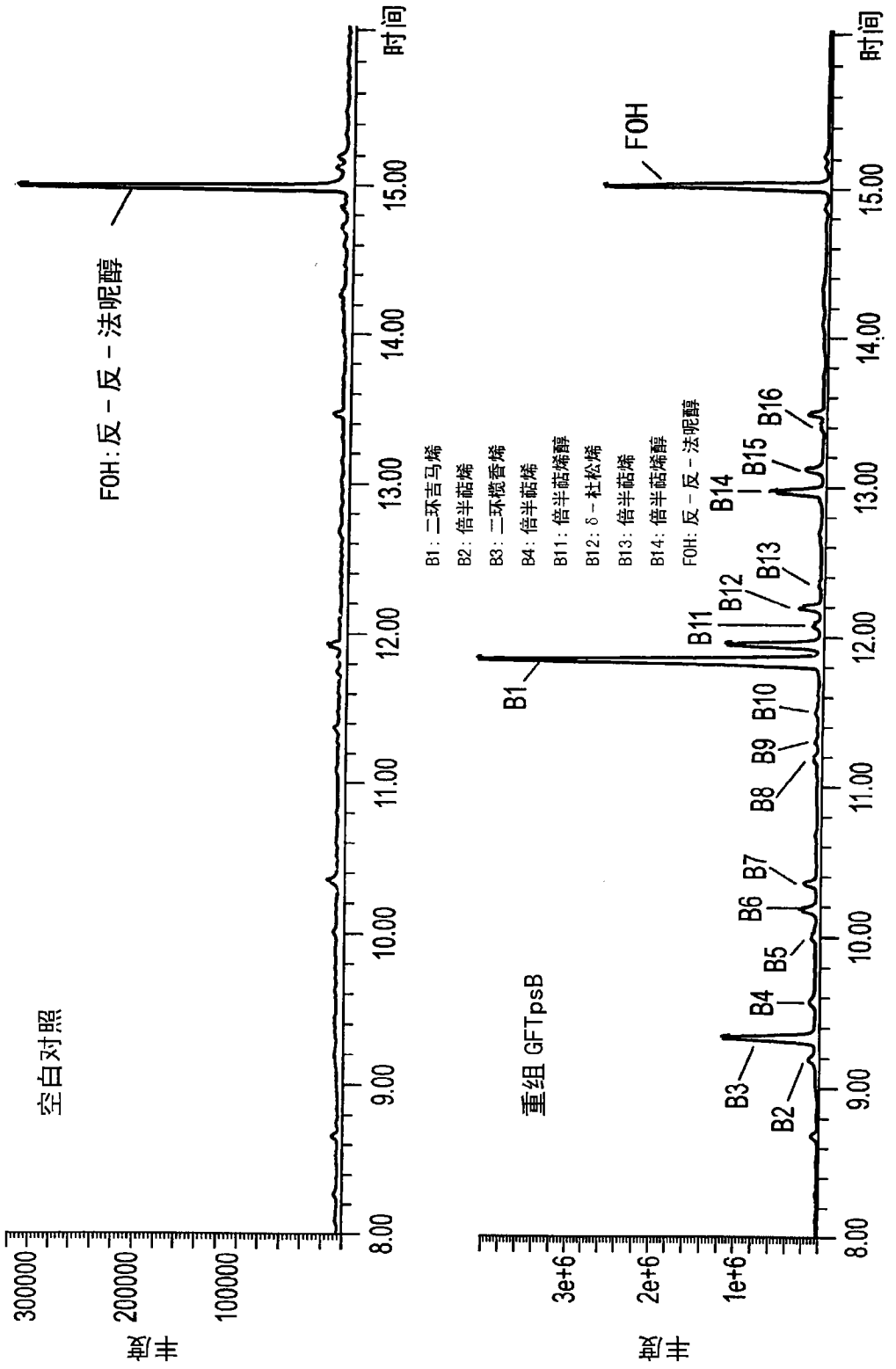

空白对照
FOH:反－反－法呢醇
丰度
时间
300000
200000
100000
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
重组 GFTpsB
B1: 二环吉马烯
B2: 倍半萜烯
B3: 二环橙香烯
B4: 倍半萜烯
B11: 倍半萜烯醇
B12: δ－杜松烯
B13: 倍半萜烯
B14: 倍半萜烯醇
FOH: 反－反－法呢醇
B1
B2
B3
B4
B5
B6
B7
B8
B9
B10
B11
B12
B13
B14
B15
B16
FOH
3e+6
2e+6
1e+6

图 7A

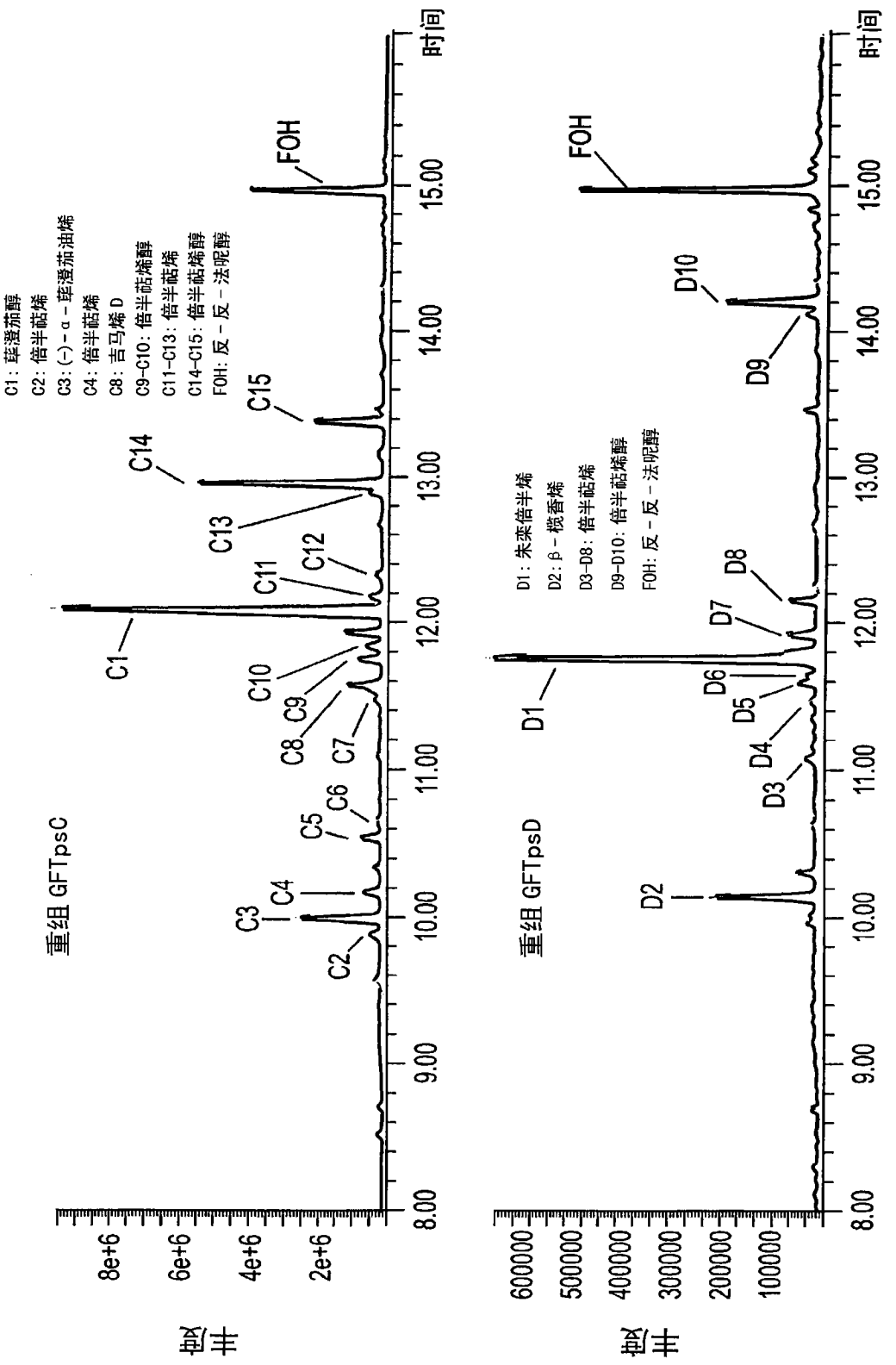

C1: 莙澄茄醇
C2: 倍半萜烯
C3: (-) - α - 莙澄茄油烯
C4: 倍半萜烯
C8: 吉马烯 D
C9-C10: 倍半萜烯醇
C11-C13: 倍半萜烯
C14-C15: 倍半萜烯醇
FOH: 反 - 反 - 法呢醇
重组 GFTpsC
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
FOH
丰度
时间
8e+6
6e+6
4e+6
2e+6
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
D1: 朱栾倍半烯
D2: β - 榄香烯
D3-D8: 倍半萜烯
D9-D10: 倍半萜烯醇
FOH: 反 - 反 - 法呢醇
重组 GFTpsD
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
FOH
丰度
时间
600000
500000
400000
300000
200000
100000
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00

图 7B

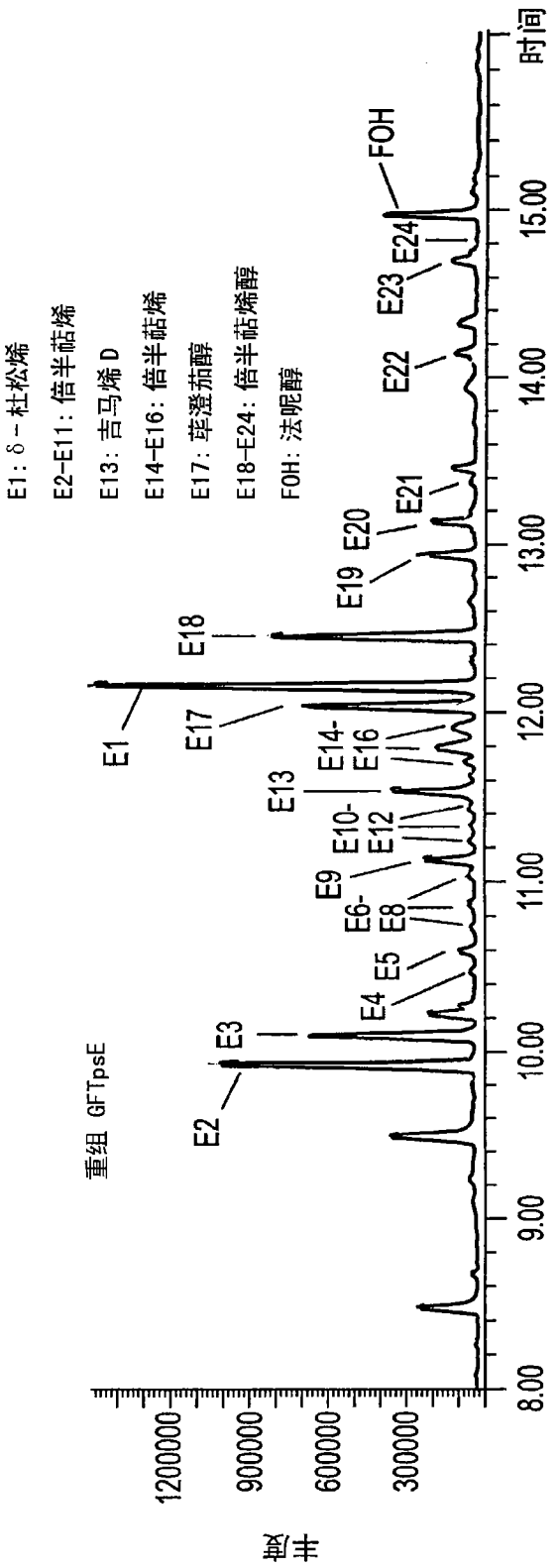

E1: δ- 杜松烯
E2-E11: 倍半萜烯
E13: 吉马烯 D
E14-E16: 倍半萜烯
E17: 萃澄茄醇
E18-E24: 倍半萜烯醇
FOH: 法呢醇
重组 GFTpsE
丰度
时间
1200000
900000
600000
300000
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00
E1
E2
E3
E4
E5
E6-E8
E9
E10-E12
E13
E14-E16
E17
E18
E19
E20
E21
E22
E23
E24
FOH

图 7C

GFTpsA

```
1     CAA AAA CTA CAC ATG ATT GAT GCA GCA CAA CGA TTA GGT GTC GCT   45
1      Q   K   L   H   M   I   D   A   A   Q   R   L   G   V   A   15

46    TAT CAT TTT GAA AAA GAG ATT GAA GAT GAA TTG GGA AAG GTA TCT   90
16     Y   H   F   E   K   E   I   E   D   E   L   G   K   V   S   30

91    CAT GAT CTT GAC AGT GAT GAT CTA TAC GTT GTT TCT CTT CGT TTT  135
31     H   D   L   D   S   D   D   L   Y   V   V   S   L   R   F   45

136   CGA CTT TTT AGA CAG CAA GGA GTT AAG ATT TCA TGT GAT GTG TTT  180
46     R   L   F   R   Q   Q   G   V   K   I   S   C   D   V   F   60

181   GAG AAG TTC AAA GAT GAC GAA GGT AAA TTC AAG GAA TCA TTG ATC  225
61     E   K   F   K   D   D   E   G   K   F   K   E   S   L   I   75

226   AAC GAT ATA CGA GGC ATG TCG AGT TTG TAC GAG GCA GCA TAC CTA  270
76     N   D   I   R   G   M   S   S   L   Y   E   A   A   Y   L   90

271   GCA ATT CGG GGG GAA GAC ATT TTA GAT GAA GCC ATT GTT TTC ACT  315
91     A   I   R   G   E   D   I   L   D   E   A   I   V   F   T  105

316   ACC ACT CAC CTT AAG TCA GTA ATA TCT GTA TCT GAT CAT TCT CAT  360
106    T   T   H   L   K   S   V   I   S   V   S   D   H   S   H  120

361   GTA AAC TCT GAT CTT GCT GAA CAA ATA CGT CAT TCT CTG CAA ATT  405
121    V   N   S   D   L   A   E   Q   I   R   H   S   L   Q   I  135

406   CCT CTC CGT AAA GCC GCA GCA AGG TTA GAG GCA AGG TAT TTT TTG  450
136    P   L   R   K   A   A   A   R   L   E   A   R   Y   F   L  150

451   GAT ATC TAT TCA AGG GAT GAT TTG CAT GAT GAA ACT TTG CTC AAG  495
151    D   I   Y   S   R   D   D   L   H   D   E   T   L   L   K  165

496   TTT GCA AAG TTA GAC TTT AAT ATA TTA CAA GCA GCA CAC AAG AAG  540
166    F   A   K   L   D   F   N   I   L   Q   A   A   H   K   K  180

541   GAA GCA AGT ATC ATG ACC AGG TGG TGG AAC GAT TTA GGC TTC CCT  585
181    E   A   S   I   M   T   R   W   W   N   D   L   G   F   P  195

586   AAA AAG GTG CCT TAT GCA AGA GAT AGA GTA GTA GAG ACA TAT ATT  630
196    K   K   V   P   Y   A   R   D   R   V   V   E   T   Y   I  210

631   TGG ATG TTG CTG GGA GTG TCC TAT GAG CCC AAT TTG GCA TTT GGT  675
211    W   M   L   L   G   V   S   Y   E   P   N   L   A   F   G  225

676   AGA ATT TTT GCA TCC AAA GTG GTG TGC ATA ATA TCC ATA ATA GAC  720
226    R   I   F   A   S   K   V   V   C   I   I   S   I   I   D  240

721   GAC ACA TTT GAT GCT TAC GGT ACT TTT GAA GAG CTC ACA CTT TTT  765
241    D   T   F   D   A   Y   G   T   F   E   E   L   T   L   F  255

766   ACT GAA GCA GTC ACA  780
256    T   E   A   V   T
```

图 8(a)

GFTpsB

```
1       ATG TCC GCT CAA GTT CTA GCA ACG GTT TCC AGT TCG ACA GAA AAA   45
1        M   S   A   Q   V   L   A   T   V   S   S   S   T   E   K   15

46      ACT GTT CGT CCC ATT GCT GGT TTC CAT CCT AAC TTA TGG GGA GAC   90
16       T   V   R   P   I   A   G   F   H   P   N   L   W   G   D   30

91      TAT TTC CTG ACC CTC GCT TCT GAT TGC AAG ACA GAT GAT ACT ACG   135
31       Y   F   L   T   L   A   S   D   C   K   T   D   D   T   T   45

136     CAC CAA GAG GAA TAC GAA GCG CTG AAG CAA GAA GTC AGA AGC ATG   180
46       H   Q   E   E   Y   E   A   L   K   Q   E   V   R   S   M   60

181     ATA ACG GCT ACG GCA GAT ACA CCT GCC CAG AAG TTG CAA TTG GTT   225
61       I   T   A   T   A   D   T   P   A   Q   K   L   Q   L   V   75

226     GAT GCA GTC CAA CGA TTG GGT GTG GCC TAT CAC TTC GAA CAG GAG   270
76       D   A   V   Q   R   L   G   V   A   Y   H   F   E   Q   E   90

271     ATA GAA GAT GCA ATG GAA AAG ATT TAT CAC GAT GAC TTT GAT AAT   315
91       I   E   D   A   M   E   K   I   Y   H   D   D   F   D   N   105

316     AAC GAT GAT GTC GAT CTC TAC ACT GTT TCT CTT CGT TTT CGA CTG   360
106      N   D   D   V   D   L   Y   T   V   S   L   R   F   R   L   120

361     CTT AGG CAG CAA GGA TTT AAG GTT CCG TGT GAT GTG TTC GCG AAG   405
121      L   R   Q   Q   G   F   K   V   P   C   D   V   F   A   K   135

406     TTC AAA GAT GAT GAA GGT AAA TTC AAG GCA TCA TTG GTG CGG GAT   450
136      F   K   D   D   E   G   K   F   K   A   S   L   V   R   D   150

451     GTT CAT GGC ATT CTA AGT TTG TAT GAG GCA GGA CAC TTG GCC ATT   495
151      V   H   G   I   L   S   L   Y   E   A   G   H   L   A   I   165

496     CGC GGA GAA GGG ATA TTA GAT GAA GCC ATT GCT TTC ACT AGA ACT   540
166      R   G   E   G   I   L   D   E   A   I   A   F   T   R   T   180

541     CAC CTT CAG TCA ATG GTA TCT CAG GAT GTA TGC CCT AAT AAT CTT   585
181      H   L   Q   S   M   V   S   Q   D   V   C   P   N   N   L   195

586     GCT GAA CAA ATT AAT CAT ACT CTC GAC TGT CCT CTC CGC AGA GCC   630
196      A   E   Q   I   N   H   T   L   D   C   P   L   R   R   A   210

631     CTT CCA AGA CTC GAG ACA AGA TTT TTC TTG TCG GTC TAT CCA AGA   675
211      L   P   R   V   E   T   R   F   F   L   S   V   Y   P   R   225

676     GAT GAT AAA CAC GAT AAA ACT TTG TTA AAG TTT TCA AAG TTA GAC   720
226      D   D   K   H   D   K   T   L   L   K   F   S   K   L   D   240

721     TTT AAC CTT GTG CAA AGA ATA CAT CAG AAG GAA TTA AGT GCC ATC   765
241      F   N   L   V   Q   R   I   H   Q   K   E   L   S   A   I   255

766     ACA CGG TGG TGG AAA GAT TTA GAC TTC ACT ACA AAG CTA CCT TAT   810
256      T   R   W   W   K   D   L   D   F   T   T   K   L   P   Y   270
```

图 8(b)-1

811	GCA	AGA	GAC	AGA	ATC	GTA	GAG	TTG	TAT	TTT	TGG	ATT	GTA	GGG	ACG	855
271	A	R	D	R	I	V	E	L	Y	F	W	I	V	G	T	285
856	TAT	TTT	GAA	CCA	AAG	TAC	ACT	TTA	GCA	AGA	AAA	ATA	ATG	ACC	AAA	900
286	Y	F	E	P	K	Y	T	L	A	R	K	I	M	T	K	300
901	ACA	ATT	TAC	ACG	GCA	TCT	ATC	ATA	GAT	GAC	ACT	TTC	GAC	GCT	TAT	945
301	T	I	Y	T	A	S	I	I	D	D	T	F	D	A	Y	315
946	GGT	TTC	TTT	GAA	GAG	CTC	AAA	CTC	TTT	GCA	GAA	GCA	GTC	CAG	AGG	990
316	G	F	F	E	E	L	K	L	F	A	E	A	V	Q	R	330
991	TGG	GAC	ATT	GGA	GCC	ATG	GAT	ATA	CTT	CCA	GAA	TAC	ATG	AAA	GTG	1035
331	W	D	I	G	A	M	D	I	L	P	E	Y	M	K	V	345
1036	CTT	TAT	AAG	GCC	CTT	TTA	GAT	ACT	TTC	AAT	GAA	ATT	GAG	CAA	GAC	1080
346	L	Y	K	A	L	L	D	T	F	N	E	I	E	Q	D	360
1081	TTG	GCC	AAG	GAA	GGA	AGA	TCG	TCC	TAC	TTA	CCT	TAT	GGC	AAA	GAA	1125
361	L	A	K	E	G	R	S	S	Y	L	P	Y	G	K	E	375
1126	AAG	ATG	CAA	GAG	CTT	GTT	CAA	ATG	TAC	TTT	GTT	CAA	GCC	AAG	TGG	1170
376	K	M	Q	E	L	V	Q	M	Y	F	V	Q	A	K	W	390
1171	TTC	CGT	GAA	GGT	TAT	GTT	CCG	ACA	TGG	GAC	GAA	TAT	TAT	CCG	GTT	1215
391	F	S	E	G	Y	V	P	T	W	D	E	Y	Y	P	V	405
1216	GGA	CTT	GTA	AGT	TGC	GGC	TAC	TTC	ATG	CTT	GCG	ACA	AAC	TCC	TTC	1260
406	G	L	V	S	C	G	Y	F	M	L	A	T	N	S	F	420
1261	CTT	GGC	ATG	TGT	GAT	GTT	GCA	AAC	GAG	GAA	GCT	TTT	GAA	TGG	ATA	1305
421	L	G	M	C	D	V	A	N	E	E	A	F	E	W	I	435
1306	TCC	AAG	GAC	CCT	AAG	ATT	TCA	ACA	GCG	TCA	TCA	GTT	ATC	TGC	AGA	1350
436	S	K	D	P	K	I	S	T	A	S	S	V	I	C	R	450
1351	CTT	AGG	AAT	GAC	ATT	GTT	TCC	CAC	CAG	TTT	GAA	CAG	AAG	AGA	GGA	1395
451	L	R	N	D	I	V	S	H	Q	F	E	Q	K	R	G	465
1396	CAT	ATT	GCC	TCA	GGA	TTT	GAA	TGC	TAC	ATT	AAG	CAG	TAT	GGT	GTT	1440
466	H	I	A	S	G	F	E	C	Y	I	K	Q	Y	G	V	480
1441	TCA	GAA	GAA	GAG	GTA	GTT	ACA	GTT	TTT	ACT	GAA	GAA	GTT	GAG	AAT	1485
481	S	E	E	E	V	V	T	V	F	T	E	E	V	E	N	495
1486	GCA	TGG	AAA	GAT	ATG	AAT	GAG	GAA	TTC	CTG	AAA	CCA	ACT	GCT	TTT	1530
496	A	W	K	D	M	N	E	E	F	L	K	P	T	A	F	510
1531	CCT	GTG	GCT	TTG	ATT	GAG	AGA	CCT	TTC	AAT	ATC	GCA	CGT	GTG	ATT	1575
511	P	V	A	L	I	E	R	P	F	N	I	A	R	V	I	525
1576	GAA	TTT	CTA	AAC	AAG	AAG	GGT	GAT	TGG	TAC	ACT	CAT	TCT	CAT	GCG	1620
526	E	F	L	N	K	K	G	D	W	Y	T	H	S	H	A	540
1621	ATT	AAA	GAC	CAG	ATT	GCC	GCA	GTG	CTC	AGA	GAC	CCT	GTT	ACC	ATC	1665
541	I	K	D	Q	I	A	A	V	L	R	D	P	V	T	I	555

图 8(b)-2

GFTpsC

```
1       ATG GCA CTT CAA GAT TCA GAA GTT CCT TCT TCC ATT CTG AAT GCT    45
1        M   A   L   Q   D   S   E   V   P   S   S   I   L   N   A    15

46      ACA GCC GGC AAC CGT CCC ACA GCT AGT TAT CAT CCC ACC CTC TGG    90
16       T   A   G   N   R   P   T   A   S   Y   H   P   T   L   W    30

91      GGG GGA AAA TTC CTT GAC TAT TCT TCT GTT GAC GAC TCT GAG GCA    135
31       G   G   K   F   L   D   Y   S   S   V   D   D   S   E   A    45

136     ATG GAT GCC ACA ATT GAT CAA GAC GAA TTT GAA GCA CTT AAG CAA    180
46       M   D   A   T   I   D   Q   D   E   F   E   A   L   K   Q    60

181     AAA ATA AAG AAC ATG TTA ATC TCA CCA ACC GAT AAG TCT TTT CAA    225
61       K   I   K   N   M   L   I   S   P   T   D   K   S   F   Q    75

226     AAA TTG AAC TTG ATT GAT GCC GTC CAA CGC TTA GGA GTG GCT TAC    270
76       K   L   N   L   I   D   A   V   Q   R   L   G   V   A   Y    90

271     CAT TTT GAG AGG GAG ATA GAA GAT GAA CTA GAA AAA CTA TCT CCT    315
91       H   F   E   R   E   I   E   D   E   L   E   K   L   S   P   105

316     GAT GAG TAT GAT GGC AAC GAT GTA CAC TCC GTT GCT CTT CGA TTT    360
106      D   E   Y   D   G   N   D   V   H   S   V   A   L   R   F   120

361     CGG TTA CTC AGA CAA CAA GGA TAT CGC ATA TCA TGC GAT ATT TTT    405
121      R   L   L   R   Q   Q   G   Y   R   I   S   C   D   I   F   135

406     GGC GGT TTC AAA GAT GAT CGA GGA AAG TTC AAG GTA TCC TTA ATT    450
136      G   G   F   K   D   D   R   G   K   F   K   V   S   L   I   150

451     AAT GAT GTG ACC GGC ATG CTA AGT TTG TAT GAG GCT GCA CAT CTT    495
151      N   D   V   T   G   M   L   S   L   Y   E   A   A   H   L   165

496     CGC ATT CGC GGG GAA GAT ATC CTG GAT GAA GCC CTA GCT TTC ACT    540
166      R   I   R   G   E   D   I   L   D   E   A   L   A   F   T   180

541     ACT TCT CAC CTG GAA TCA ATG GTT ACT CAA GTA AGC CCT CAG CTT    585
181      T   S   H   L   E   S   M   V   T   Q   V   S   P   Q   L   195

586     TCT GAT GAA ATA CTT CAT GCC TTG AAT AGG CCA ATC CGC AGA GGC    630
196      S   D   E   I   L   H   A   L   N   R   P   I   R   R   G   210

631     TTA CCA AGG CTG GAG GCA GTC TAT TAC ATC GAT CTC TAC TCA CGA    675
211      L   P   R   L   E   A   V   Y   Y   I   D   L   Y   S   R   225

676     GAT GAT TCA AAG GAT AAA GCA ATA TTA CTA AAG TTT GCA AAA CTA    720
226      D   D   S   K   D   K   A   I   L   L   K   F   A   K   L   240

721     GAT TTT TGC ATG CTT CAA GTA ATT CAC CGT AAG GAG TTA AGT ATC    765
241      D   F   C   M   L   Q   V   I   H   R   K   E   L   S   I   255

766     ATC ACA GAG TGG TGG AAA AAT TTA GAT GTT GAA ATA AAT CTC CCA    810
256      I   T   E   W   W   K   N   L   D   V   E   I   N   L   P   270
```

图 8(c)-1

```
811   TAT GCT AGA AAC AGA GTT GTA GAA TGC TAT TTT TGG GCA ATG GGA   855
271    Y   A   R   N   R   V   V   E   C   Y   F   W   A   M   G   285

856   GTG TAT TTT GAG CCT CGA TAC TCC TTT GCA AGA AAG ATA TTG TCC   900
286    V   Y   F   E   P   R   Y   S   F   A   R   K   I   L   S   300

901   AAA GTA ATT GCA ATG GCA TCC ATT TTA GAT GAT ACC TAC GAC GCC   945
301    K   V   I   A   M   A   S   I   L   D   D   T   Y   D   A   315

946   TAT GGC ACA CTT GAA GAA CTT GAG CTC TTT ACA AAT GCT ATC AAA   990
316    Y   G   T   L   E   E   L   E   L   F   T   N   A   I   K   330

991   AGG TGG GAT ATT AGC AAC ATA GAT GTA CTT CCG AAG TAC ATG AAA  1035
331    R   W   D   I   S   N   I   D   V   L   P   K   Y   M   K   345

1036  CTG ATT TAT CAA GGA CTC TTG GAT GTT TTT GGT GAA GCT GAG GAG  1080
346    L   I   Y   Q   G   L   L   D   V   F   G   E   A   E   E   360

1081  GAA ATC TCA AAG GAA GGA CAG ACA TAT TGC ATG TCA TAT GTC ATA  1125
361    E   I   S   K   E   G   Q   T   Y   C   M   S   Y   V   I   375

1126  CAA GCG GTG AAG AAA GTA GTC CAA GCC TAC TTT GAG GAA GCC AAG  1170
376    Q   A   V   K   K   V   V   Q   A   Y   F   E   E   A   K   390

1171  TGG TGC AGT GAA GGT TAT TTT CCA AAA GTG GAG GAG TAT ATG CAA  1215
391    W   C   S   E   G   Y   F   P   K   V   E   E   Y   M   Q   405

1216  GTT TCA CTT GTG ACA ACT TGC TAT CAT ATG CTG GCA ACG GCT TCT  1260
406    V   S   L   V   T   T   C   Y   H   M   L   A   T   A   S   420

1261  TTT CTT GGC ATG GGA AAG ATT GCT GAT AAG CAG GCC TTT GAA TGG  1305
421    F   L   G   M   G   K   I   A   D   K   Q   A   F   E   W   435

1306  ATC TCC AAT TAC CCT AAA ACT GTG AAA GCC TCC CAA GTT ATT TGC  1350
436    I   S   N   Y   P   K   T   V   K   A   S   Q   V   I   C   450

1351  AGA CTT ATG GAT GAT ATA GTG TCT CAC GAG TTT GAA CAA AAA AGA  1395
451    R   L   M   D   D   I   V   S   H   E   F   E   Q   K   R   465

1396  AAG CAT GTT GCC TCG GGT ATT GAA TGT TAC ATG AAG CAG CAT GGC  1440
466    K   H   V   A   S   G   I   E   C   Y   M   K   Q   H   G   480

1441  GTC TCT GAT GAA GAG GTA ATT AAA GTA TTC CGC AAA CAA ATA TCA  1485
481    V   S   D   E   E   V   I   K   V   F   R   K   Q   I   S   495

1486  AAT GGA TGG AAA GAT gTA AAT GAA GGA TTC ATG AAG CCA ACA GAA  1530
496    N   G   W   K   D   V   N   E   G   F   M   K   P   T   E   510

1531  GTG GCA ATG CCT CTC CTT GAG CGC ATT CTC AAT CTT GCA CGA GTG  1575
511    V   A   M   P   L   L   E   R   I   L   N   L   A   R   V   525

1576  ATA GAT GTT ATT TAC AAG GAT GAT GAT GGC TAC ACC AAC TCT TAT  1620
526    I   D   V   I   Y   K   D   D   D   G   Y   T   N   S   Y   540

1621  GTG ATC AAA GAC TAC ATC GCC ACA TTG CTT GAG AAG CCT GTt CCC  1665
541    V   I   K   D   Y   I   A   T   L   L   E   K   P   V   P   555

1666  TTT TGA    1671
556    F   *
```

图 8(c)-2

GFTpsD1

```
1     ATG TCG TCT GGA GAA ACA TTT CGT CGT ACT GCA GAT TTC CAT CCT   45
1      M   S   S   G   E   T   F   R   P   T   A   D   F   H   P   15

46    AGT TTA TGG AGA AAC CAT TTC CTC AAA GGT GCT TCT GAT TTC AAG   90
16     S   L   W   R   N   H   F   L   K   G   A   S   D   F   K   30

91    ACA GTT GAT CAT ACT GCA ACT CAA GAA CGA CAC GAG GCA CTG AAA   135
31     T   V   D   H   T   A   T   Q   E   R   H   E   A   L   K   45

136   GAA GAG GTA AGG AGA ATG ATA ACA GAT GCT GAA GAT AAG CCT GTT   180
46     E   E   V   R   R   M   I   T   D   A   E   D   K   P   V   60

181   CAG AAG TTA CGC TTG ATT GAT GAA GTA CAA CGC CTG GGG GTG GCT   225
61     Q   K   L   R   L   I   D   E   V   Q   R   L   G   V   A   75

226   TAT CAC TTT GAG AAA GAA ATA GAA GAT GCA ATA CAA AAA TTA TGT   270
76     Y   H   F   E   K   E   I   E   D   A   I   Q   K   L   C   90

271   CCA ATC TAT ATT GAC AGT AAT AGA GCT GAT CTC CAC ACC GTT TCC   315
91     P   I   Y   I   D   S   N   R   A   D   L   H   T   V   S   105

316   CTT CAT TTT CGA TTG CTT AGG CAG CAA GGA ATC AAG ATT TCA TGT   360
106    L   H   F   R   L   L   R   Q   Q   G   I   K   I   S   C   120

361   GAT GTG TTT GAG AAG TTC AAA GAT GAT GAG GGT AGA TTC AAG TCA   405
121    D   V   F   E   K   F   K   D   D   E   G   R   F   K   S   135

406   TCG TTG ATA AAC GAT GTT CAA GGG ATG TTA AGT TTG TAC GAG GCA   450
136    S   L   I   N   D   V   Q   G   M   L   S   L   Y   E   A   150

451   GCA TAC ATG GCA GTT CGC GGA GAA CAT ATA TTA GAT GAA GCC ATT   495
151    A   Y   M   A   V   R   G   E   H   I   L   D   E   A   I   165

496   GCT TTC ACT ACC ACT CAC CTG AAG TCA TTG GTA GCT CAG GAT CAT   540
166    A   F   T   T   T   H   L   K   S   L   V   A   Q   D   H   180

541   GTA ACC CCT AAG CTT GCG GAA CAG ATA AAT CAT GCT TTA TAC CGT   585
181    V   T   P   K   L   A   E   Q   I   N   H   A   L   Y   R   195

586   CCT CTT CGT AAA ACC CTA CCA AGA TTA GAG GCG AGG TAT TTT ATG   630
196    P   L   R   K   T   L   P   R   L   E   A   R   Y   F   M   210

631   TCC ATG ATC AAT TCA ACA AGT GAT CAT TTA TAC AAT AAA ACT CTG   675
211    S   M   I   N   S   T   S   D   H   L   Y   N   K   T   L   225

676   CTG AAT TTT GCA AAG TTA GAT TTT AAC ATA TTG CTA GAG CTG CAC   720
226    L   N   F   A   K   L   D   F   N   I   L   L   E   L   H   240

721   AAG GAG GAA CTC AAT GAA TTA ACA AAG TGG TGG AAA GAT TTA GAC   765
241    K   E   E   L   N   E   L   T   K   W   W   K   D   L   D   255

766   TTC ACT ACA AAA CTA CCT TAT GCA AGA GAC AGA TTA GTG GAG TTA   810
256    F   T   T   K   L   P   Y   A   R   D   R   L   V   E   L   270
```

图 8(d)-1

```
811   TAT TTT TGG GAT TTA GGG ACA TAC TTC GAG CCT
271    Y   F   W   D   L   G   T   Y   F   E   P

856   GGG AGA AAG ATA ATG ACC CAA TTA AAT TAC ATA TTA TCC ATC ATA   900
286    G   R   K   I   M   T   Q   L   N   Y   I   L   S   I   I   300

901   GAT GAT ACT TAT GAT GCG TAT GGT ACA CTT GAA GAA CTC AGC CTC   945
301    D   D   T   Y   D   A   Y   G   T   L   E   E   L   S   L   315

946   TTT ACT GAA GCA GTT CAA AGA TGG AAT ATT GAG GCC GTA GAT ATG   990
316    F   T   E   A   V   Q   R   W   N   I   E   A   V   D   M   330

991   CTT CCA GAA TAC ATG AAA TTG ATT TAC AGG ACA CTC TTA GAT GCT  1035
331    L   P   E   Y   M   K   L   I   Y   R   T   L   L   D   A   345

1036  TTT AAT GAA ATT GAG GAA GAT ATG GCC AAG CAA GGA AGA TCA CAC  1080
346    F   N   E   I   E   E   D   M   A   K   Q   G   R   S   H   360

1081  TGC GTA CGT TAT GCA AAA GAG GAG AAT CAA AAA GTA ATT GGA GCA  1125
361    C   V   R   Y   A   K   E   E   N   Q   K   V   I   G   A   375

1126  TAC TCT GTT CAA GCC AAA TGG TTC AGT GAA GGT TAC GTT CCA ACA  1170
376    Y   S   V   Q   A   K   W   F   S   E   G   Y   V   P   T   390

1171  ATT GAG GAG TAT ATG CCT ATT GCA CTA ACA AGT TGT GCT TAC ACA  1215
391    I   E   E   Y   M   P   I   A   L   T   S   C   A   Y   T   405

1216  TTC GTC ATA ACA AAT TCC TTC CTT GGC ATG GGT GAT TTT GCA ACT  1260
406    F   V   I   T   N   S   F   L   G   M   G   D   F   A   T   420

1261  AAA GAG GTT TTT GAA TGG ATC TCC AAT AAC CCT AAG GTT GTA AAA  1305
421    K   E   V   F   E   W   I   S   N   N   P   K   V   V   K   435

1306  GCA GCA TCA GTT ATC TGC AGA CTC ATG GAT GAC ATG CAA GGT CAT  1350
436    A   A   S   V   I   C   R   L   M   D   D   M   Q   G   H   450

1351  GAG TTT GAG CAG AAG AGA GGA CAT GTT GCG TCA GCT ATT GAA TGT  1395
451    E   F   E   Q   K   R   G   H   V   A   S   A   I   E   C   465

1396  TAC ACG AAG CAG CAT GGT GTC TCT AAG GAA GAG GCA ATT AAA ATG  1440
466    Y   T   K   Q   H   G   V   S   K   E   E   A   I   K   M   480

1441  TTT GAA GAA GAA GTT GCA AAT GCA TGG AAA GAT ATT AAC GAG GAG  1485
481    F   E   E   E   V   A   N   A   W   K   D   I   N   E   E   495

1486  TTG ATG ATG AAG CCA ACC GTC GTT GCC CGA CCA CTG CTC GGG ACG  1530
496    L   M   M   K   P   T   V   V   A   R   P   L   L   G   T   510

1531  ATT CTT AAT CTT GCT CGT GCA ATT GAT TTT ATT TAC AAA GAG GAC  1575
511    I   L   N   L   A   R   A   I   D   F   I   Y   K   E   D   525

1576  GAC GGC TAT ACG CAT TCT TAC CTA ATT AAA GAT CAA ATT GCT TCT  1620
526    D   G   Y   T   H   S   Y   L   I   K   D   Q   I   A   S   540

1621  GTG CTA GGA GAC CAC GTT CCA TTT TGA    1647
541    V   L   G   D   H   V   P   F   *
```

图 8(d)-2

GFTpsD2

```
1     ATG TCG TCT GGA GAA ACA TTT CGT CCT ACT GCA GAT TTC CAT CCT   45
1      M   S   S   G   E   T   F   R   P   T   A   D   F   H   P   15

46    AGT TTA TGG AGA AAC CAT TTC CTC AAA GGT GCT TCT GAT TTC AAG   90
16     S   L   W   R   N   H   F   L   K   G   A   S   D   F   K   30

91    ACA GTT GAT CAT ACT GCA ACT CAA GAA CGA CAC GAG GCA CTG AAA   135
31     T   V   D   H   T   A   T   Q   E   R   H   E   A   L   K   45

136   GAA GAG GTA AGG AGA ATG ATA ACA GAT GCT GAA GAT AAG CCT GTT   180
46     E   E   V   R   R   M   I   T   D   A   E   D   K   P   V   60

181   CAG AAG TTA CGC TTG ATT GAT GAA GTA CAA CGC CTG GGG GTG GCT   225
61     Q   K   L   R   L   I   D   E   V   Q   R   L   G   V   A   75

226   TAT CAC TTT GAG AAA GAA ATA GAA GAT GCA ATA CAA AAA TTA TGT   270
76     Y   H   F   E   K   E   I   E   D   A   I   Q   K   L   C   90

271   CCA AAC TAT ATT CAC AGT AAT AGC CCT GAT CTT CAC ACC GTT TCT   315
91     P   N   Y   I   H   S   N   S   P   D   L   H   T   V   S   105

316   CTT CAT TTT CGA TTG CTT AGG CAG CAA GGA ATC AAG ATT TCA TGT   360
106    L   H   F   R   L   L   R   Q   Q   G   I   K   I   S   C   120

361   GAT GTG TTT GAG AAG TTC AAA GAT GAT GAG GGT AGA TTC AAG TCA   405
121    D   V   F   E   K   F   K   D   D   E   G   R   F   K   S   135

406   TCG TTG ATA AAC GAT GTT CAA GGG ATG TTA AGT TTG TAC GAG GCA   450
136    S   L   I   N   D   V   Q   G   M   L   S   L   Y   E   A   150

451   GCA TAC ATG GCA GTT CGC GGA GAA CAT ATA TTA GAT GAA GCC ATT   495
151    A   Y   M   A   V   R   G   E   H   I   L   D   E   A   I   165

496   GCT TTC ACT ACC ACT CAC CTG AAG TCA TTG GTA GCT CAG GAT CAT   540
166    A   F   T   T   T   H   L   K   S   L   V   A   Q   D   H   180

541   GTA ACC CCT AAG CTT GCG GAA CAG ATA AAT CAT GCT TTA TAC CGT   585
181    V   T   P   K   L   A   E   Q   I   N   H   A   L   Y   R   195

586   CCT CTT CGT AAA ACC CTA CCA AGA TTA GAG GCG AGG TAT TTT ATG   630
196    P   L   R   K   T   L   P   R   L   E   A   R   Y   F   M   210

631   TCC ATG ATC AAT TCA ACA AGT GAT CAT TTA TAC AAT AAA ACT CTG   675
211    S   M   I   N   S   T   S   D   H   L   Y   N   K   T   L   225

676   CTG AAT TTT GCA AAG TTA GAT TTT AAC ATA TTG CTA GAG CTG CAC   720
226    L   N   F   A   K   L   D   F   N   I   L   L   E   L   H   240

721   AAG GAG GAA CTC AAT GAA TTA ACA AAG TGG TGG AAA GAT TTA GAC   765
241    K   E   E   L   N   E   L   T   K   W   W   K   D   L   D   255

766   TTC ACT ACA AAA CTA CCT TAT GCA AGA GAC AGA TTA GTG GAG TTA   810
256    F   T   T   K   L   P   Y   A   R   D   R   L   V   E   L   270
```

图 8(e)-1

```
811   TAT TTT TGG GAT TTA GGG ACA TAC TTC GAG CCT CAA TAT GCA TTT   855
271    Y   F   W   D   L   G   T   Y   F   E   P   Q   Y   A   F   285

856   GGG AGA AAG ATA ATG ACC CAA TTA AAT TAC ATA TTA TCC ATC ATA   900
286    G   R   K   I   M   T   Q   L   N   Y   I   L   S   I   I   300

901   GAT GAT ACT TAT GAT GCG TAT GGT ACA CTT GAA GAA CTC AGC CTC   945
301    D   D   T   Y   D   A   Y   G   T   L   E   E   L   S   L   315

946   TTT ACT GAA GCA GTT CAA AGA TGG AAT ATT GAG GCC GTA GAT ATG   990
316    F   T   E   A   V   Q   R   W   N   I   E   A   V   D   M   330

991   CTT CCA GAA TAC ATG AAA TTG ATT TAC AGG ACA CTC TTA GAT GCT  1035
331    L   P   E   Y   M   K   L   I   Y   R   T   L   L   D   A   345

1036  TTT AAT GAA ATT GAG GAA GAT ATG GCC AAG CAA GGA AGA TCA CAC  1080
346    F   N   E   I   E   E   D   M   A   K   Q   G   R   S   H   360

1081  TGC GTA CGT TAT GCA AAA GAG GAG AAT CAA AAA GTA ATT GGA GCA  1125
361    C   V   R   Y   A   K   E   E   N   Q   K   V   I   G   A   375

1126  TAC TCT GTT CAA GCC AAA TGG TTC AGT GAA GGT TAC GTT CCA ACA  1170
376    Y   S   V   Q   A   K   W   F   S   E   G   Y   V   P   T   390

1171  ATT GAG GAG TAT ATG CCT ATT GCA CTA ACA AGT TGT GCT TAC ACA  1215
391    I   E   E   Y   M   P   I   A   L   T   S   C   A   Y   T   405

1216  TTC GTC ATA ACA AAT TCC TTC CTT GGC ATG GGT GAT TTT GCA ACT  1260
406    F   V   I   T   N   S   F   L   G   M   G   D   F   A   T   420

1261  AAA GAG GTT TTT GAA TGG ATC TCC AAT AAC CCT AAG GTT GTA AAA  1305
421    K   E   V   F   E   W   I   S   N   N   P   K   V   V   K   435

1306  GCA GCA TCA GTT ATC TGC AGA CTC ATG GAT GAC ATG CAA GGT CAT  1350
436    A   A   S   V   I   C   R   L   M   D   D   M   Q   G   H   450

1351  GAG TTT GAG CAG AAG AGA GGA CAT GTT GCG TCA GCT ATT GAA TGT  1395
451    E   F   E   Q   K   R   G   H   V   A   S   A   I   E   C   465

1396  TAC ACG AAG CAG CAT GGT GTC TCT AAG GAA GAG GCA ATT AAA ATG  1440
466    Y   T   K   Q   H   G   V   S   K   E   E   A   I   K   M   480

1441  TTT GAA GAA GAA GTT GCA AAT GCA TGG AAA GAT ATT AAC GAG GAG  1485
481    F   E   E   E   V   A   N   A   W   K   D   I   N   E   E   495

1486  TTG ATG ATG AAG CCA ACC GTC GTT GCC CGA CCA CTG CTC GGG ACG  1530
496    L   M   M   K   P   T   V   V   A   R   P   L   L   G   T   510

1531  ATT CTT AAT CTT GCT CGT GCA ATT GAT TTT ATT TAC AAA GAG GAC  1575
511    I   L   N   L   A   R   A   I   D   F   I   Y   K   E   D   525

1576  GAC GGC TAT ACG CAT TCT TAC CTA ATT AAA GAT CAA ATT GCT TCT  1620
526    D   G   Y   T   H   S   Y   L   I   K   D   Q   I   A   S   540

1621  GTG CTA GGA GAC CAC GTT CCA TTT TGA    1647
541    V   L   G   D   H   V   P   F   *
```

图 8(e)-2

GFTpsE

```
1     ATG TCT TTG GAA GTT TCA GCC TCT CCT GCT AAA GTT ATC CAA AAT   45
1      M   S   L   E   V   S   A   S   P   A   K   V   I   Q   N   15

46    GCT GGG AAA GAT TCT ACT CGT CGC TCT GCA AAT TAT CAT CCA AGC   90
16     A   G   K   D   S   T   R   R   S   A   N   Y   H   P   S   30

91    ATC TGG GGG GAT CAT TTC CTT CAA TAT ACT TGT GAC ACC CAG GAA   135
31     I   W   G   D   H   F   L   Q   Y   T   C   D   T   Q   E   45

136   ACT GAT GAT GGC AGC AAT GTA AAG CAT CTA GAG CTG AAG AAA GAA   180
46     T   D   D   G   S   N   V   K   H   L   E   L   K   K   E   60

181   ATT AGA AGA ATG CTA AAA GCT GAT AAC AAG CCT TCA CGT ACA CTT   225
61     I   R   R   M   L   K   A   D   N   K   P   S   R   T   L   75

226   CAA TTG ATT GAT GCA ATT CAG CGT TTA GGA GTG TCT TAC CAT TTT   270
76     Q   L   I   D   A   I   Q   R   L   G   V   S   Y   H   F   90

271   GAA AGT GAG ATT GAT GAA ATA TTG GGA AAG ATG CAT AAG GCT TCC   315
91     E   S   E   I   D   E   I   L   G   K   M   H   K   A   S   105

316   CAA GAC TCT GAT CTT TGT GAT AAT GAA AAT GAT GAG CTC TAT TAT   360
106    Q   D   S   D   L   C   D   N   E   N   D   E   L   Y   Y   120

361   ATC TCT CTT CAT TTT CGA TTA CTT AGA CAA AAT GGC TAT AAA ATT   405
121    I   S   L   H   F   R   L   L   R   Q   N   G   Y   K   I   135

406   TCC GCT GAT GTG TTC AAA AAG TTC AAA GAC ACG GAT GGG AAC TTT   450
136    S   A   D   V   F   K   K   F   K   D   T   D   G   N   F   150

451   AAA ACA TCT CTT GCG AAA GAT GTT CGA GGA ATG TTA AGC TTG TAT   495
151    K   T   S   L   A   K   D   V   R   G   M   L   S   L   Y   165

496   GAA GCT ACG CAT CTC GGG GTA CAT GAA GAA GAT ATA CTA GAT GAA   540
166    E   A   T   H   L   G   V   H   E   E   D   I   L   D   E   180

541   GCG CTT GCT TTC ACC ACT AGT CAC CTA GAG TCA ATA GCG ACT CAT   585
181    A   L   A   F   T   T   S   H   L   E   S   I   A   T   H   195

586   CAA ATC AGG TCT CCA CTT GTT GAA CAA GTC AAA CAT GCC TTA GTT   630
196    Q   I   R   S   P   L   V   E   Q   V   K   H   A   L   V   210

631   CAG CCT ATC CAC AGG GGC TTC CAA AGG CTT GAG GCA AGA CAG TAC   675
211    Q   P   I   H   R   G   F   Q   R   L   E   A   R   Q   Y   225

676   ATT CCT ATC TAT CAA GAA GAA TCT CCC CAC AAT GAA GCT CTG TTA   720
226    I   P   I   Y   Q   E   E   S   P   H   N   E   A   L   L   240

721   ACT TTT GCA AAG TTA GAT TTT AAC AAA TTG CAA AAG CCT CAC CAG   765
241    T   F   A   K   L   D   F   N   K   L   Q   K   P   H   Q   255

766   AAG GAA CTC GGT GAT ATT TCA AGG TGG TGG AAA GAA TTA GAC TTT   810
256    K   E   L   G   D   I   S   R   W   W   K   E   L   D   F   270
```

图 8(f)-1

```
811   GCA CAT AAG CTA CCT TTC ATA AGA GAT AGA GTT GCG GAG TGC TAC   855
271    A   H   K   L   P   F   I   R   D   R   V   A   E   C   Y   285

856   TTT TGG ATA TTA GGA GTG TAT TTC GAG CCC CAA TAT TCA TTT GCA   900
286    F   W   I   L   G   V   Y   F   E   P   Q   Y   S   F   A   300

901   AGA AGA ATA TTG ACG AAA GTG ATC TCC ATG ACT TCT GTT ATT GAT   945
301    R   R   I   L   T   K   V   I   S   M   T   S   V   I   D   315

946   GAT ATC TAT GAT GTG TAT GGC AAA ATT GAA GAA CTT GAG CTT TTT   990
316    D   I   Y   D   V   Y   G   K   I   E   E   L   E   L   F   330

991   ACT TCA GCT ATT GAG AGG TGG GAT ATC AGT GCC ATA GAT CAA CTT  1035
331    T   S   A   I   E   R   W   D   I   S   A   I   D   Q   L   345

1036  CCT GAG TAT ATG AAA TTG TGT TAT AGG GCC CTT CTT GAT GTT TTT  1080
346    P   E   Y   M   K   L   C   Y   R   A   L   L   D   V   F   360

1081  AGT GAA GCA GAG AAG GAT TTG GCC CCC CAA GGA AAA TCA TAC CGC  1125
361    S   E   A   E   K   D   L   A   P   Q   G   K   S   Y   R   375

1126  CTC TAT TAT GCA AAA GAA GCG ATG AAG AAT ATG GTT AAG AAT TAC  1170
376    L   Y   Y   A   K   E   A   M   K   N   M   V   K   N   Y   390

1171  TTC TAC GAA GCT AAA TGG TGT CTT CAG AAT TAT GTA CCT ACA GTG  1215
391    F   Y   E   A   K   W   C   L   Q   N   Y   V   P   T   V   405

1216  GAT GAG TAC ATG ACG GTT GCA TTA GTT ACA TCT GGC TCC CCA ATG  1260
406    D   E   Y   M   T   V   A   L   V   T   S   G   S   P   M   420

1261  TTG TCA ACC ACA TCC TTT GTT GGC ATG GGA GAC ATT GTA ACT AAA  1305
421    L   S   T   T   S   F   V   G   M   G   D   I   V   T   K   435

1306  GAA TCT TTT GAG TGG TTA TTC AGC AAT CCT AGA TTT ATT AGG GCT  1350
436    E   S   F   E   W   L   F   S   N   P   R   F   I   R   A   450

1351  TCT TCT ATA GTT TGC CGA CTC ATG GAT GAC ATA GTG TCA CAC AAG  1395
451    S   S   I   V   C   R   L   M   D   D   I   V   S   H   K   465

1396  TTT GAA CAA AGC AGA GGG CAC GTT GCC TCA AGC GTT GAG TGT TAC  1440
466    F   E   Q   S   R   G   H   V   A   S   S   V   E   C   Y   480

1441  ATG AAA CAA CAT GGA GCA ACA GAA GAG GAA GCA TGC AAT GAG TTT  1485
481    M   K   Q   H   G   A   T   E   E   E   A   C   N   E   F   495

1486  CGG AAA CAA GTT TCA AAT GCC TGG AAG GAT ATA AAT GAG GAC TGC  1530
496    R   K   Q   V   S   N   A   W   K   D   I   N   E   D   C   510

1531  CTA CGC CCA ACG GTT GTG CCA ATG CCA CTT CTG ATG CGA ATT CTC  1575
511    L   R   P   T   V   V   P   M   P   L   L   M   R   I   L   525

1576  AAT CTT ACA CGC GTT ATA GAT GTC ATT TAC AAG TAT GAA GAT GGC  1620
526    N   L   T   R   V   I   D   V   I   Y   K   Y   E   D   G   540

1621  TAC ACT CAT TCC GCA GTT GTG CTG AAA GAT TTT GTT GCT TCT TTG  1665
541    Y   T   H   S   A   V   V   L   K   D   F   V   A   S   L   555

1666  TTT ATT AAT CCT GTG CCG ATA TGA    1689
556    F   I   N   P   V   P   I   *
```

图 8(f)-2