



(12) 发明专利申请

(10) 申请公布号 CN 103492551 A

(43) 申请公布日 2014. 01. 01

(21) 申请号 201280020706. 3

*C12N 15/09* (2006. 01)

(22) 申请日 2012. 02. 28

(30) 优先权数据

61/447, 286 2011. 02. 28 US

(85) PCT国际申请进入国家阶段日

2013. 10. 28

(86) PCT国际申请的申请数据

PCT/US2012/027003 2012. 02. 28

(87) PCT国际申请的公布数据

W02012/118848 EN 2012. 09. 07

(71) 申请人 诺维信股份有限公司

地址 美国加利福尼亚州

(72) 发明人 A. 费希尔 S. 布朗 S. 鲁特林格

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 封新琴

(51) Int. Cl.

*C12N 1/14* (2006. 01)

*C12P 7/46* (2006. 01)

权利要求书2页 说明书46页

序列表44页 附图14页

(54) 发明名称

用于生产 C4- 二羧酸的微生物

(57) 摘要

本发明涉及具有碳酸氢根转运蛋白活性的分离的多肽, 和编码所述多肽的分离的多核苷酸。本发明亦涉及包含所述多核苷酸的核酸构建体、载体和宿主细胞, 以及产生和使用所述多肽的方法, 和产生 C4- 二羧酸如苹果酸的方法。

1. 包含编码碳酸氢根转运蛋白的异源多核苷酸的重组宿主细胞,其中所述所述异源多核苷酸:

(a) 编码与 SEQ ID NO:2 或 SEQ ID NO:4 具有至少 65%,例如至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性的碳酸氢根转运蛋白;

(b) 在低严格条件,中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件下,与 SEQ ID NO:1, SEQ ID NO:3,或其 cDNA 序列的全长互补链杂交;或

(c) 与 SEQ ID NO:1, SEQ ID NO:3,或其 cDNA 序列具有至少 65%,例如至少 70%,至少 75%,至少 80%,至少 85%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性,

其中所述宿主细胞与没有所述异源多核苷酸的宿主细胞相比,当在相同条件下培养时能够产生更多量的 C4- 二羧酸。

2. 权利要求 1 的重组宿主细胞,其中所述异源多核苷酸编码与 SEQ ID NO:2 或 SEQ ID NO:4 具有至少 90% 序列同一性的碳酸氢根转运蛋白。

3. 权利要求 1 的重组宿主细胞,其中所述异源多核苷酸编码与 SEQ ID NO:2 或 SEQ ID NO:4 具有至少 95% 序列同一性的碳酸氢根转运蛋白。

4. 权利要求 1-3 任一项的重组宿主细胞,其中所述异源多肽在高严格条件下与 SEQ ID NO:1、SEQ ID NO:3、或其 cDNA 序列的全长互补链杂交。

5. 权利要求 1-3 任一项的重组宿主细胞,其中所述异源多肽在非常高严格条件下与 SEQ ID NO:1、SEQ ID NO:3、或其 cDNA 序列的全长互补链杂交。

6. 权利要求 1-5 任一项的重组宿主细胞,其中所述异源多核苷酸与 SEQ ID NO:1、SEQ ID NO:3、或其 cDNA 序列具有至少 90% 序列同一性。

7. 权利要求 1-5 任一项的重组宿主细胞,其中所述异源多核苷酸与 SEQ ID NO:1, SEQ ID NO:3,或其 cDNA 序列具有至少 95% 序列同一性。

8. 权利要求 1-7 任一项的重组宿主细胞,其中所述异源多核苷酸编码包含或组成为 SEQ ID NO:2 或 SEQ ID NO:4 的多肽。

9. 权利要求 1-8 任一项的重组宿主细胞,其中所述编码碳酸氢根转运蛋白的异源多核苷酸可操作地连接于对于该多核苷酸而言是外来的启动子。

10. 权利要求 1-9 任一项的重组宿主细胞,其还包含编码 C4- 二羧酸转运蛋白的异源多核苷酸(例如编码 SEQ ID NO:6, 27, 29, 32, 34, 36, 39, 41, 或 43 的 C4- 二羧酸转运蛋白的异源多核苷酸)。

11. 权利要求 10 的重组宿主细胞,其中所述编码 C4- 二羧酸转运蛋白的异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。

12. 权利要求 1-26 任一项的重组宿主细胞,其还包含编码苹果酸脱氢酶的异源多核苷酸(例如编码 SEQ ID NO:8 或 45 的苹果酸脱氢酶的异源多核苷酸)。

13. 权利要求 12 的重组宿主细胞,其中所述编码苹果酸脱氢酶的异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。

14. 权利要求 1-13 任一项的重组宿主细胞,其还包含编码丙酮酸羧化酶的异源多核苷酸(例如编码 SEQ ID NO:10 的丙酮酸羧化酶的异源多核苷酸)。

15. 权利要求 14 的重组宿主细胞,其中所述编码丙酮酸羧化酶的异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。

16. 权利要求 1-15 任一项的重组宿主细胞,其中所述宿主细胞是真核宿主细胞。

17. 权利要求 16 的重组宿主细胞,其中所述宿主细胞是丝状真菌宿主细胞。

18. 权利要求 17 的重组宿主细胞,其中所述宿主细胞选自下组:枝顶孢霉属 (*Acremonium*)、曲霉属 (*Aspergillus*)、短梗霉属 (*Aureobasidium*)、烟管霉属 (*Bjerkandera*)、拟蜡菌属 (*Ceriporiopsis*)、金孢子菌属 (*Chrysosporium*)、鬼伞属 (*Coprinus*)、革盖菌属 (*Coriolus*)、隐球菌属 (*Cryptococcus*)、*Filibasidium*、镰孢属 (*Fusarium*)、腐质霉属 (*Humicola*)、梨孢菌属 (*Magnaporthe*)、毛霉属 (*Mucor*)、毁丝霉属 (*Myceliophthora*)、新考玛脂霉属 (*Neocallimastix*)、脉孢菌属 (*Neurospora*)、拟青霉属 (*Paecilomyces*)、青霉属 (*Penicillium*)、平革菌属 (*Phanerochaete*)、射脉菌属 (*Phlebia*)、瘤胃壶菌属 (*Piromyces*)、侧耳属 (*Pleurotus*)、根霉属 (*Rhizopus*)、裂褶菌属 (*Schizophyllum*)、踝节菌属 (*Talaromyces*)、嗜热子囊菌属 (*Thermoascus*)、梭孢壳属 (*Thielavia*)、弯颈霉属 (*Tolyposcladium*)、栓菌属 (*Trametes*) 和木霉属 (*Trichoderma*)。

19. 权利要求 17 的重组宿主细胞,其中所述宿主细胞是曲霉属宿主细胞。

20. 权利要求 19 的重组宿主细胞,宿主细胞是米曲霉宿主细胞。

21. 权利要求 19 的重组宿主细胞,宿主细胞是黑曲霉宿主细胞。

22. 权利要求 1-21 任一项的重组宿主细胞,其中所述 C4- 二羧酸是苹果酸。

23. 权利要求 1-22 任一项的重组宿主细胞,其中所述宿主细胞与没有所述编码碳酸氢根转运蛋白的多核苷酸的宿主细胞相比,当在相同条件下培养时,能够产生多至少 25% 的量的 C4- 二羧酸。

24. 产生 C4- 二羧酸的方法,其包括:

(a) 在合适的条件下在培养基中培养权利要求 1-23 任一项的重组宿主细胞以产生所述 C4- 二羧酸;和

(b) 回收所述 C4- 二羧酸。

25. 权利要求 24 的方法,其中所述 C4- 二羧酸是苹果酸。

26. 重组宿主细胞,其包含异源多核苷酸,所述异源多核苷酸编码 SEQ ID NO:2 或 4 的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的碳酸氢根转运蛋白变体;

其中所述宿主细胞与没有所述异源多核苷酸的宿主细胞相比,当在相同条件下培养时,能够产生更多量的 C4- 二羧酸。

## 用于生产 C4- 二羧酸的微生物

[0001] 对相关申请的交叉引用

[0002] 本申请要求 2011 年 2 月 28 日提交的美国临时申请号 61/447, 286 的优先权, 其通过全文提述并入本文。

[0003] 涉及序列列表

[0004] 本申请包含计算机可读形式的序列列表, 其通过提述并入本文。

[0005] 发明背景

### 技术领域

[0006] 本发明涉及用于 C4- 二羧酸 (例如苹果酸) 的重组生产的方法。

### 背景技术

[0007] 有机酸在多种工业中具有悠久历史的商业使用。举例而言, 有机酸用于食品和饲料工业 (柠檬酸、抗坏血酸、乳酸、乙酸和葡糖酸), 作为用于产生多种聚合物的单体 (己二酸、乳酸、丙烯酸和衣康酸), 作为金属螯合剂 (葡糖酸) 和作为“绿色”溶剂 (乙酸) (Sauer 等, 2008, Trends in Biotechnology 26:100-108)。有机酸可以自身是商业产品, 也可以是用于制造其他化学品的化学构件 (building block)。除了专门用途之外, 长久以来公认 C4- 二羧酸亦可充当构件化合物以供产生大量的工业化学品, 如 1, 4- 丁二醇, 四氢呋喃和  $\gamma$ - 丁内酯。由于石油衍生构件的高成本, 通过传统石油化学途径产生这些大体积工业化学品的成本显著增加。

[0008] 有机酸在产业上可通过从石油衍生原料的化学合成 (例如, 延胡索酸、苹果酸、丙烯酸和己二酸) 或通过微生物发酵 (例如柠檬酸、乳酸、葡糖酸和衣康酸) 产生。一些有机酸如延胡索酸和苹果酸亦可通过微生物发酵来产生, 但由于更低的生产成本, 目前在商业上仍通过化学合成从石油化学原料产生。然而, 石油衍生构件化学品的成本的日益升高, 地缘政治的不稳定性对原油价格的影响, 以及对实施利用来源于可再生资源的原料的制造工艺的期望, 再次激发了人们通过微生物发酵产生有机酸和其他化学品的兴趣。

[0009] 虽然现在 C4- 二羧酸如苹果酸通过化学合成从石油化学来源来工业生产, 其亦可通过微生物发酵生产。苹果酸已在基因工程改造的酵母 (酿酒酵母 (*Saccharomyces cerevisiae*)) (Zelle 等, 2008, Appl. Environ. Microbiol. 74:2766-2777) 和天然存在的丝状真菌如曲霉属菌种 (*Aspergillus* spp.) (美国专利号 3, 063, 910; Bercovitz 等, 1990, Appl. Environ. Microbiol. 56:1594-1597) 中获得了高水平生产。Abe 等 (美国专利号 3, 063, 910) 和 Bercovitz 等 (1990, Appl. Environ. Microbiol. 56:1594-1597) 报道了在几种曲霉属菌种中高水平的苹果酸生产。而且, Battat 等 (1991, Biotechnol. Bioengineering, 37:1108-1116) 报道了在优化条件下在搅拌的发酵罐中由黄曲霉 (*Aspergillus flavus*) 产生了高至 113g/L 的苹果酸。在 W02010/003728 中描述了在酵母中通过微生物发酵产生二羧酸。亦在 W02009/011974, W02009/155382 和 W02010/111344 中描述了通过微生物发酵产生苹果酸。通过遗传工程 C4- 二羧酸如苹果酸的产生的改善可使

得能够通过发酵经济地商业性生产苹果酸。

[0010] 在宿主如曲霉属菌种 (*Aspergillus* spp.) 中的苹果酸过量产生在特定的培养条件 (有氧条件和高 C:N 比; 碳酸钙亦可作为中和剂和作为用于苹果酸生物合成的 CO<sub>2</sub> 源添加) 下发生。在这些条件下, 通过胞质溶胶中发生的还原性三羧酸 (TCA) 循环的溢出代谢 (overflow metabolism) 导致的苹果酸生物合成和培养基分泌量增加。已报道了酿酒酵母中通过使用遗传工程增加丙酮酸羧化酶 (Bauer 等, 1999, *FEMS Microbiol Lett.* 179:107-113) 或苹果酸脱氢酶 (Pines 等, 1997, *Appl. Microbiol. Biotechnol.* 48:248-255) 的水平和增加苹果酸转运蛋白的表达 (Zelle 等, 2008, 见上) 可增加苹果酸产生。基于生物化学证据, 提出了在黄曲霉菌株 ATCC13697 中, 苹果酸脱氢酶活性限制苹果酸产生 (Peleg 等, 1988, *Appl. Microbiol. Biotechnol.* 28:69-75)。2010 年 8 月 27 日提交、标题为“Methods for Improving Malic Acid Production in Filamentous Fungi”的美国申请第 12/870, 523 号 (其内容通过提述以其整体并入本文) 和 2010 年 6 月 21 日提交、标题为“Polypeptides Having C4-dicarboxylic acid Transporter Activity and Polynucleotides Encoding Same”的美国临时申请第 61/356, 868 号 (其通过全文提述并入本文) 描述了 C4- 二羧酸生产。

[0011] 在本领域中, 作为使用重组 DNA 技术的遗传工程改造的结果而改善 C4- 二羧酸产生如苹果酸产生会是有益的。本发明提供了用于改善 C4- 二羧酸产生 (例如苹果酸产生) 的方法等。

## 发明内容

[0012] 本发明涉及包含碳酸氢根转运蛋白活性的重组宿主细胞, 其中所述宿主细胞产生 (或能够产生) 增加量的 C4- 二羧酸 (例如苹果酸)。在一个方面, 所述重组宿主细胞包含编码碳酸氢根转运蛋白 (例如硫酸根 - 碳酸氢根转运蛋白) 的异源多核苷酸, 其中所述宿主细胞与不含所述异源核苷酸的宿主细胞相比, 当在相同条件下培养时, 产生 (或能够产生) 和 / 或分泌 (或能够分泌) 更大量的 C4- 二羧酸 (例如苹果酸)。在一些方面, 所述宿主细胞还包含编码 C4- 二羧酸转运蛋白的异源多核苷酸, 编码苹果酸脱氢酶的异源多核苷酸, 和 / 或编码丙酮酸羧化酶的异源多核苷酸。在一些方面, 所述宿主细胞是丝状真菌宿主细胞, 如米曲霉 (*Aspergillus oryzae*) 宿主细胞。

[0013] 本发明亦涉及使用重组宿主细胞用于产生 C4- 二羧酸的方法。在一个方面, 本发明涉及产生 C4- 二羧酸 (例如苹果酸) 的方法, 其包括: (a) 在合适的条件下在培养基中培养具有碳酸氢根转运蛋白活性的重组宿主细胞 (例如丝状真菌宿主细胞) 以产生 C4- 二羧酸; 和 (b) 回收所述 C4- 二羧酸。在一些方面, 所述重组宿主细胞包含编码碳酸氢根转运蛋白 (例如硫酸根 - 碳酸氢根转运蛋白)。在另一个方面, 本发明涉及产生 C4- 二羧酸 (例如苹果酸) 的方法, 其包括: (a) 将本文中所述的编码碳酸氢根转运蛋白的异源多核苷酸转化入宿主细胞 (例如丝状真菌宿主细胞); (b) 在合适的条件下在培养基中培养转化的生物以产生 C4- 二羧酸; 和 (c) 回收所述 C4- 二羧酸。在所述方法的一些方面, 所述重组宿主细胞进一步包含编码 C4- 二羧酸转运蛋白的异源多核苷酸, 编码苹果酸脱氢酶的异源多核苷酸, 和 / 或编码丙酮酸羧化酶的异源多核苷酸。

## 附图说明

[0014] 图 1 显示 pShTh60 的限制图谱。

[0015] 图 2 显示 pAmFs69 的限制图谱。

[0016] 图 3A 和 3B 显示米曲霉 NRRL3488 碳酸氢根转运蛋白基因 (bt1) 的基因组核苷酸构建体序列和推导的氨基酸序列 (分别为 SEQ ID NO:1 和 2)。

[0017] 图 4A 和 4B 显示米曲霉 NRRL3488 碳酸氢根转运蛋白基因 (bt2) 的基因组核苷酸构建体序列和推导的氨基酸序列 (分别为 SEQ ID NO:3 和 4)。

[0018] 图 5 显示 pSaMF36 的限制图谱。

[0019] 图 6 显示棘孢曲霉 (*Aspergillus aculeatus*) C4- 二羧酸转运蛋白基因 (c4t521) 的基因组 DNA 序列和推导的氨基酸序列 (分别为 SEQ ID NO:5 和 6)。

[0020] 图 7 显示米曲霉 NRRL3488 苹果酸脱氢酶基因 (mdh3) 的基因组 DNA 序列和推导的氨基酸序列 (分别为 SEQ ID NO:7 和 8)。

[0021] 图 8 显示 pSaMF21 的限制图谱。

[0022] 图 9A 和 9B 一同显示米曲霉 NRRL3488 丙酮酸羧化酶基因 (pyc) 的基因组 DNA 序列和推导的氨基酸序列 (分别为 SEQ ID NO:9 和 10)。

[0023] 图 10 显示 pRYAN1 的限制图谱。

[0024] 图 11 显示 pShTh77 的限制图谱。

[0025] 图 12 显示 pShTh147 的限制图谱。

[0026] 定义

[0027] 碳酸氢根转运蛋白:术语“碳酸氢根转运蛋白”(bicarbonate transporters) 在本文中定义为能够协助  $\text{HCO}_3^-$  跨生物膜 (如细胞膜和 / 或细胞器的膜) 进行转运的蛋白,如膜整合蛋白(membrane integrated protein)。碳酸氢根转运蛋白的非限定性类型包括阴离子交换蛋白(anion exchanger, AE) 家族的  $\text{Cl}^-/\text{HCO}_3^-$  交换蛋白, NBC 家族的  $\text{Na}^+/\text{HCO}_3^-$  共转运蛋白, 和  $\text{Na}^+$ - 依存性  $\text{Cl}^-/\text{HCO}_3^-$  交换蛋白。在本文中所述的一些方面,所述碳酸氢根转运蛋白是硫酸根-碳酸氢根转运蛋白,其中所述转运蛋白能够协助  $\text{HCO}_3^-$  和  $\text{SO}_4^{2-}$  阴离子二者跨生物膜的转运。碳酸氢根交换活性可如本领域技术中所述来确定,例如如 Sterling 等, 2002, *Am J Physiol Cell Physiol* 283:C1522-1529 中所述确定。

[0028] 所述碳酸氢根转运蛋白具有 SEQ ID NO:2 或 4 的成熟多肽编码序列的碳酸氢根转运活性的至少 20%, 例如至少 40%, 至少 50%, 至少 60%, 至少 70%, 至少 80%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 至少 99%, 或至少 100%。

[0029] C4- 二羧酸转运蛋白:术语“C4- 二羧酸转运蛋白”在本文中定义为可将苹果酸、琥珀酸、草酰乙酸、丙二酸和 / 或延胡索酸转运至细胞外的二羧酸通透酶 (Grobler 等, 1995, *Yeast* 11:1485-1491; Camarasa 等, 2001, *Applied and Environmental Microbiology* 67:4144-4151)。用于预测线粒体输入的蛋白及其靶向序列的计算方法由 Claros 和 Vincens, 1996, *Eur. J. Biochem.* 241:779-786 描述。

[0030] 所述 C4- 二羧酸转运蛋白具有 SEQ ID NO:6 的成熟多肽序列的 C4- 二羧酸转运蛋白活性 (例如苹果酸转运蛋白活性) 的至少 20%, 例如至少 40%, 至少 50%, 至少 60%, 至少 70%, 至少 80%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少

96%, 至少 97%, 至少 98%, 至少 99%, 或至少 100%。

[0031] 苹果酸脱氢酶:术语“苹果酸脱氢酶”在本文中定义为在  $\text{NADH}+\text{H}^+$  的存在下催化将草酰乙酸还原为苹果酸和  $\text{NAD}^+$  的苹果酸: $\text{NAD}^+$  氧还酶 (EC1. 1. 1. 37)。就本发明而言,根据下述步骤确定苹果酸脱氢酶活性。测定溶液由 1mM 草酰乙酸, 100mM Tris pH8.0, 10mM  $\text{NaHCO}_3$ , 5mM  $\text{MgCl}_2$ , 和 0.1mM NADH (Sigma Chemical Co., St. Louis, MO, USA) 组成。将不含草酰乙酸作为底物的测定溶液作为对照运行以测量背景 NADH 降解率。用双蒸水制备每种上清的 1/100, 1/500, 1/2500 和 1/12500 的稀释。将测定溶液的 270  $\mu\text{l}$  的等分试样分配入 96 孔聚苯乙烯平板。添加每种稀释的上清的 30  $\mu\text{l}$  样品以起始测定。使用 SPECTRAMAX® 340PC 读板器 (Molecular Devices, Sunnyvale, CA, USA) 以下述设定监测反应: 340nm, 动力学读取 (kinetic reading)。使用 NADH 的浓度系列以构建标准曲线, 并使用纯化的苹果酸脱氢酶 (Sigma Chemical Co., St. Louis, MO, USA) 的稀释系列作为阳性对照。一个单位的苹果酸脱氢酶活性等于能够在 pH8.0, 25°C 每分钟将 1 微摩尔草酰乙酸和  $\text{NADH}+\text{H}^+$  转化为苹果酸和  $\text{NAD}^+$  的酶量。

[0032] 所述苹果酸脱氢酶具有 SEQ ID NO:8 的成熟多肽序列的苹果酸脱氢酶活性的至少 20%, 例如至少 40%, 至少 50%, 至少 60%, 至少 70%, 至少 80%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 至少 99%, 或至少 100%。

[0033] 丙酮酸羧化酶:术语“丙酮酸羧化酶”在本文中定义为在 ATP 和  $\text{HCO}_3^-$  存在下催化将丙酮酸羧化为草酰乙酸、ADP 和磷酸的丙酮酸:二氧化碳连接酶 (形成 ADP 的) (EC6. 4. 1. 1)。就本发明而言,根据针对丙酮酸羧化酶的 SIGMA® 质量控制测试方法 (Sigma Chemical Co., St. Louis, MO, USA) 的步骤确定丙酮酸羧化酶活性, 只是该测定使用 pH8.0 的 Tris 缓冲液。一个单位的丙酮酸羧化酶活性等于能够在 pH7.8, 30°C 每分钟将 1 微摩尔的丙酮酸和  $\text{CO}_2$  转化为草酰乙酸的酶量。

[0034] 所述丙酮酸羧化酶具有 SEQ ID NO:10 的成熟多肽序列的丙酮酸羧化酶活性的至少 20%, 例如至少 40%, 至少 50%, 至少 60%, 至少 70%, 至少 80%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 至少 99%, 或至少 100%。

[0035] 异源多核苷酸:术语“异源多核苷酸”在本文中定义为对于宿主细胞并非天然的多核苷酸;其中已对编码区进行结构修饰的天然多核苷酸;作为通过重组 DNA 技术 (例如, 不同的 (外源) 启动子) 操纵 DNA 的结果其表达定量改变的天然多核苷酸;或其表达通过将一个或多个 (几个) 额外拷贝的多核苷酸导入宿主细胞而定量改变的天然多核苷酸。

[0036] 分离的 / 纯化的:术语“分离的”和“纯化的”意指从至少一种与其天然结合 (associated) 的成分分离的多肽或多核苷酸。举例而言,如通过 SDS-PAGE 确定的, 该多肽可为至少 1% 纯, 例如至少 5% 纯, 至少 10% 纯, 至少 20% 纯, 至少 40% 纯, 至少 60% 纯, 至少 80% 纯, 至少 90% 纯, 至少 93% 纯, 至少 95% 纯, 至少 97% 纯, 至少 98% 纯, 或至少 99% 纯;且如通过琼脂糖电泳确定的, 该多核苷酸可为至少 1% 纯, 例如至少 5% 纯, 至少 10% 纯, 至少 20% 纯, 至少 40% 纯, 至少 60% 纯, 至少 80% 纯, 至少 90% 纯, 至少 93% 纯, 至少 95% 纯, 至少 97% 纯, 至少 98% 纯, 或至少 99% 纯。

[0037] 编码序列:术语“编码序列”的意思是指定其蛋白产物的氨基酸序列的多核苷酸序列。编码序列的边界通常由开读框确定, 所述开读框通常以 ATG 起始密码子或可供选择的起始密码子如 GTG 和 TTG 开始, 并且以终止密码子如 TAA、TAG 和 TGA 结束。编码序列可以

是基因组 DNA、cDNA、合成多核苷酸和 / 或重组多核苷酸。

[0038] cDNA 序列:术语“cDNA 序列”意指从得自真核细胞的成熟的、已剪接的 mRNA 分子在反转录之后所得的 DNA。来自基因组 DNA 起始的 (initial)、初级的 RNA 转录物是 mRNA 的前体,其通过一系列包括剪接的步骤加工然后作为成熟的已剪接的 mRNA 出现。cDNA 序列缺乏可存在于相应基因组 DNA 序列中的插入的内含子。相应地,短语“SEQ ID NO:X 的 cDNA 序列”意指将 SEQ ID NO:X 中插入的内含子序列 (若存在) 去除之后所得的序列。在一些情况下一当参照的基因组 DNA 序列缺乏插入的内含子序列时—cDNA 序列可与相应的基因组 DNA 序列完全相同。

[0039] 基因组 DNA 序列:术语“基因组 DNA 序列”意指见于来源生物的基因组 (例如真核或原核基因组) 的 DNA 序列。在一些情况下,来自真核基因组的基因组 DNA 序列含有一个或多个插入的内含子序列,其作为 RNA 剪接的结果从初级 RNA 转录物去除。相应地,短语“SEQ ID NO:Y 的基因组 DNA 序列”意指来自来源生物的相应 DNA 序列,其含有在 RNA 剪接之前存在的插入的内含子序列 (若存在)。

[0040] 成熟多肽序列:术语“成熟多肽序列”意指参照的多核苷酸序列在任何翻译后序列修饰 (如 N 端加工和 / 或 C 端截短) 之后的部分。在一些情况下,所述成熟多肽序列可以与整个参照的多肽序列完全相同。在一个方面,基于预测缺乏信号肽的 SignalP 程序 (Nielsen 等,1997,Protein Engineering10:1-6) 和 InterProScan 程序 (The European Bioinformatics Institute),所述成熟多肽编码序列是 SEQ ID NO:2 的核苷酸 1 至 770。在另一个方面,基于预测缺乏信号肽的 SignalP 程序和 InterProScan 程序,所述成熟多肽编码序列是 SEQ ID NO:4 的核苷酸 1 至 843。本领域中已知宿主细胞可产生由相同多核苷酸表达的两种或更多种不同成熟多肽序列 (即具有不同的 C 末端和 / 或 N 末端氨基酸) 的混合物。

[0041] 成熟多肽编码序列:术语“成熟多肽编码序列”意指参照的多核苷酸序列 (例如基因组或 cDNA 序列) 编码成熟多肽序列的部分。在一些情况下,所述成熟多肽编码序列可以与整个参照的多核苷酸序列完全相同。在一个方面,基于预测缺乏信号肽的 SignalP 程序 (Nielsen 等,1997, 见上) 和 InterProScan 程序 (见上),所述成熟多肽编码序列是 SEQ ID NO:1 的核苷酸 1 至 2503。在另一个方面,基于预测缺乏信号肽的 SignalP 程序和 InterProScan 程序,所述成熟多肽编码序列是 SEQ ID NO:3 的核苷酸 1 至 2657。

[0042] 片段:术语“片段”意指从参照的多肽序列的氨基和 / 或羧基末端缺失了一个或多个 (几个) 氨基酸的多肽。在一个方面,所述片段具有碳酸氢根转运蛋白活性。在另一个方面,片段包含 SEQ ID NO:2 的至少 650 个氨基酸残基,例如至少 690 个氨基酸残基或至少 730 个氨基酸残基。在另一个方面,片段包含 SEQ ID NO:4 的至少 720 个氨基酸残基,例如至少 760 个氨基酸残基或至少 800 个氨基酸残基。

[0043] 亚序列:术语“亚序列 (subsequence)”意指从参照的核苷酸序列的 5' 和 / 或 3' 端缺失了一个或多个 (几个) 核苷酸的多核苷酸。在一个方面,所述亚序列编码具有碳酸氢根转运蛋白活性的片段。在另一个方面,亚序列包含 SEQ ID NO:1 的至少 1950 个核苷酸,例如至少 2070 个核苷酸或至少 2190 个核苷酸。在另一个方面,亚序列包含 SEQ ID NO:3 的至少 2160 个核苷酸,例如至少 2280 个核苷酸或至少 2400 个核苷酸。

[0044] 等位变体 (allelic variant):术语“等位变体”意指占据相同染色体基因座的基



因的两种以上可选形式中的任何一种。等位变异通过突变天然地发生,并且可导致种群内的多态性。基因突变可以是沉默的(在编码的多肽中无变化)或可以编码具有改变的氨基酸序列的多肽。多肽的等位变体是由基因的等位变体编码的多肽。

[0045] 序列同一性:两个氨基酸序列间或两个核苷酸序列间的相关性由参数“序列同一性”所描述。

[0046] 就本发明而言,两个氨基酸序列之间的序列同一性程度是使用如EMBOSS 软件包(EMBOSS:欧洲分子生物学开放软件套组(The European Molecular Biology Open Software Suite),Rice等,2000,Trends Genet. 16:276-277)的Needle程序(优选为3.0.0版或之后的版本)中执行的Needleman-Wunsch算法(Needleman和Wunsch,1970,J.Mol. Biol. 48:443-453)确定的。所用的可选参数为缺口开放罚分(gap open penalty)10,缺口延伸罚分(gap extension penalty)0.5,和EBLUSUM62(BLUSUM62的EMBOSS版)取代矩阵。使用标记为“最长同一性”的Needle输出(使用“-nobrief”选项获得)作为百分比同一性并如下计算:

[0047]  $(\text{相同的残基} \times 100) / (\text{比对长度} - \text{比对中缺口总数})$

[0048] 就本发明而言,两个脱氧核糖核苷酸序列之间的序列同一性程度是使用如EMBOSS软件包(EMBOSS:欧洲分子生物学开放软件套组(The European Molecular Biology Open Software Suite),Rice等,2000,见上)的Needle程序(优选为3.0.0版或之后的版本)中执行的Needleman-Wunsch算法(Needleman和Wunsch,1970,见上)确定的。所用的可选参数为缺口开放罚分10,缺口延伸罚分0.5,和EDNAFULL(NCBI NUC4.4的EMBOSS版)取代矩阵。使用标记为“最长同一性”的Needle输出(使用“-nobrief”选项获得)作为百分比同一性并如下计算:

[0049]  $(\text{相同的脱氧核糖核苷酸} \times 100) / (\text{比对长度} - \text{比对中缺口总数})$ 。

[0050] 表达:术语“表达”包括涉及多肽产生的任何步骤,其包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

[0051] 核酸构建体:术语“核酸构建体”意指单链或双链的核酸分子,其分离自天然存在的基因,或经修饰以本来不存在于(not otherwise exist)自然界中的方式含有核酸的区段,或为合成的,其中所述核酸分子包含一个或多个(几个)调控序列。

[0052] 调控序列(control sequence):术语“调控序列”意指对于多肽表达必需的核酸序列。调控序列对于编码所述多肽的多核苷酸可以是天然的或外源的,以及对于彼此可以是天然的或外源的。这些调控序列包括但不限于前导序列、聚腺苷酸化序列、前肽序列、启动子序列、信号肽序列和转录终止子序列。调控序列可以和用于引入特异性限制位点的接头一起提供,所述特异性限制位点促进调控序列与编码多肽的多核苷酸编码区的连接。

[0053] 可操作地连接:术语“可操作地连接”意指这样的构型,其中将调控序列置于相对于多核苷酸的编码序列的适当位置,使得调控序列指导编码序列的表达。

[0054] 表达载体:术语“表达载体”意指线性的或环状的DNA分子,其包含编码多肽的多核苷酸,并与调控序列可操作地连接,其中所述调控序列提供编码所述多肽的多核苷酸的表达。最少的情况,所述表达载体包含启动子序列,和转录和翻译终止信号序列。

[0055] 宿主细胞:术语“宿主细胞”意指任何对于用包含本发明的多核苷酸(例如编码碳酸氢根转运蛋白的多核苷酸)的核酸构建体或表达载体的转化、转染、转导等是易感

(susceptible) 的细胞类型。术语“宿主细胞”涵盖亲本细胞的任何由于在复制过程中发生的突变而不同于亲本细胞的后代。

[0056] 变体：术语“变体”意指在一个或多个（几个）位置包含变化即取代、插入和 / 或缺失一个或多个（几个）氨基酸残基的，具有活性例如碳酸氢根转运蛋白活性的多肽。取代意指将占据某位置的氨基酸用不同氨基酸替换；缺失意指去除占据某位置的氨基酸；而插入意指在占据某位置的氨基酸的相邻处添加一个或多个（几个），例如 1-3 个氨基酸。

[0057] 体积生产力 (volumetric productivity)：术语“体积生产力”指每单位时间每体积（例如，培养基及其中内含物的总体积）使用的系统所产生的参照的产物的量（例如，产生的 C4- 二羧酸的量）。

[0058] 发酵培养基：术语“发酵培养基”指包含一种或多种（几种）糖如葡萄糖、果糖、蔗糖、纤维二糖、木糖、木酮糖、阿拉伯糖、甘露糖、半乳糖和 / 或可溶性寡糖的培养基，其中所述培养基能够部分地通过宿主细胞转化（发酵）为所需产物，如 C4- 二羧酸。在一些情况下，所述发酵培养基源自天然来源，如甘蔗，淀粉，或纤维素，并可为通过酶水解（糖化）而预处理所述来源的结果。

[0059] 在本文中，提及“约”某值或参数包括了涉及该值或参数本身的方面。举例而言，提及“约 X”的描述包括了“X”的方面。

[0060] 如用于本文中和所附权利要求中，除非上下文明确地表示相反，单数形式“一（个 / 种...）(a)”或“所述 / 该 (the)”包括了对复数的提及。应理解本文中所述的发明的各方面包括了“由... 组成 (consisting)”和 / 或“基本上由... 组成 (consisting essentially of)”的方面。

[0061] 除非另行定义或上下文明确指出，本文中使用的所有技术和科学术语具有与本发明所属领域的一般技术人员通常理解的相同的含意。

[0062] 发明详述

[0063] 本发明描述了特定基因在宿主细胞如丝状真菌（例如曲霉属 (*Aspergillus*)）中的过量表达以增强 C4- 二羧酸（例如苹果酸）的产生等。本发明涵盖使用异源基因以供表达碳酸氢根转运蛋白。所述碳酸氢根转运蛋白可为任何适于实践本发明的、描述的碳酸氢根转运蛋白。在一个方面，所述碳酸氢根转运蛋白是在培养条件下过表达，以高效价产生碳酸氢根的转运蛋白。在一个方面，所述碳酸氢根转运蛋白是硫酸根 - 碳酸氢根转运蛋白。所述重组宿主细胞可进一步包含编码 C4- 二羧酸的异源多核苷酸，编码苹果酸脱氢酶的异源多核苷酸和 / 或编码丙酮酸羧化酶的异源多核苷酸。

[0064] 碳酸氢根转运蛋白和编码碳酸氢根转运蛋白的多核苷酸

[0065] 在本文所述的重组宿主细胞和方法的一个方面，所述碳酸氢根转运蛋白选自下组：(a) 碳酸氢根转运蛋白，其与 SEQ ID NO:2 或 SEQ ID NO:4 或其成熟多肽序列具有至少 60% 序列同一性；(b) 碳酸氢根转运蛋白，其由多核苷酸编码，所述多核苷酸在低严格条件下与以下杂交：(i) SEQ ID NO:1 或 3，或其成熟多肽编码序列，(ii) SEQ ID NO:1 或 3 的 cDNA 序列，或其成熟多肽编码序列，(iii) (i) 或 (ii) 的全长互补链；(c) 碳酸氢根转运蛋白，其由多核苷酸编码，所述多核苷酸与 (iv) SEQ ID NO:1 或 3，或其成熟多肽编码序列，(v) SEQ ID NO:1 或 3 的 cDNA 序列，或其成熟多肽编码序列，(vi) (iv) 或 (v) 的全长互补链具有至少 60% 序列同一性；(d) SEQ ID NO:2 或 SEQ ID NO:4 或其成熟多肽序列的包含一个或多个

(几个)氨基酸的取代、缺失和/或插入的碳酸氢根转运蛋白变体;和(e)(a)、(b)、(c)或(d)的多肽具有碳酸氢根转运蛋白活性的片段。

[0066] 在一个方面,所述碳酸氢根转运蛋白包含或组成为氨基酸序列,所述氨基酸序列与SEQ ID NO:2或4,或其成熟多肽序列具有至少60%,例如至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。在一个方面,所述碳酸氢根转运蛋白包含氨基酸序列,所述氨基酸序列与SEQ ID NO:2或4,或其成熟多肽序列相差不多于十个氨基酸,例如不多于五个氨基酸,不多于四个氨基酸,不多于三个氨基酸,不多于两个氨基酸,或一个氨基酸。

[0067] 在一个方面,所述碳酸氢根转运蛋白包含或组成为这样的氨基酸序列,所述氨基酸序列与SEQ ID NO:2或其成熟多肽序列具有至少60%,例如至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。在一个方面,所述碳酸氢根转运蛋白包含这样的氨基酸序列,所述氨基酸序列与SEQ ID NO:2或其成熟多肽序列相差不多于十个氨基酸,例如不多于五个氨基酸,不多于四个氨基酸,不多于三个氨基酸,不多于两个氨基酸,或一个氨基酸。在另一个方面,所述碳酸氢根转运蛋白包含氨基酸序列,所述氨基酸序列与SEQ ID NO:4或其成熟多肽序列具有至少60%,例如至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。在另一个方面,所述碳酸氢根转运蛋白包含这样的氨基酸序列,所述氨基酸序列与SEQ ID NO:4或其成熟多肽序列相差不多于十个氨基酸,例如不多于五个氨基酸,不多于四个氨基酸,不多于三个氨基酸,不多于两个氨基酸,或一个氨基酸。

[0068] 在一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:2的氨基酸序列,SEQ ID NO:2的成熟多肽序列,其等位变体,或前述序列的具有C4-二羧酸转运蛋白活性的序列。在另一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:2的氨基酸序列。在另一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:2的成熟多肽序列。在另一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:2的氨基酸1至770。

[0069] 在一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:4的氨基酸序列,SEQ ID NO:4的成熟多肽序列,其等位变体,或前述序列的具有碳酸氢根转运蛋白活性的序列。在另一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:4的氨基酸序列。在另一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:4的成熟多肽序列。在另一个方面,所述碳酸氢根转运蛋白包含或组成为SEQ ID NO:4的氨基酸1至843。

[0070] 在一个方面,所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸在至少低严格条件,例如中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件与以下杂交:(i)SEQ ID NO:1或3,或其成熟多肽编码序列,(ii)SEQ ID NO:1或3的cDNA序列,或其成熟多肽编码序列,(iii)(i)或(ii)的全长互补链(参见,例如J. Sambrook, E. F. Fritsch和T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 第2版, Cold Spring Harbor, New York)。

[0071] 在一个方面,所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸在至少低严

格条件,例如中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件与以下杂交:(i)SEQ ID NO:1,或其成熟多肽编码序列,(ii)SEQ ID NO:1的cDNA序列,或其成熟多肽编码序列,(iii)(i)或(ii)的全长互补链。在另一个方面,所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸在至少低严格条件,例如中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件与以下杂交:(i)SEQ ID NO:3,或其成熟多肽编码序列,(ii)SEQ ID NO:3的cDNA序列,或其成熟多肽编码序列,(iii)(i)或(ii)的全长互补链。

[0072] 在一个方面,所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸与(iv)SEQ ID NO:1或3,或其成熟多肽编码序列,(v)SEQ ID NO:1或3的cDNA序列,或其成熟多肽编码序列,(vi)(iv)或(v)的全长互补链具有至少65%,例如至少70%,至少75%,至少80%,至少85%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。

[0073] 在一个方面,所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸与(iv)SEQ ID NO:1,或其成熟多肽编码序列,(v)SEQ ID NO:1的cDNA序列,或其成熟多肽编码序列,(vi)(iv)或(v)的全长互补链具有至少65%,例如至少70%,至少75%,至少80%,至少85%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。

[0074] 在一个方面,所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸与(iv)SEQ ID NO:3,或其成熟多肽编码序列,(v)SEQ ID NO:3的cDNA序列,或其成熟多肽编码序列,(vi)(iv)或(v)的全长互补链具有至少65%,例如至少70%,至少75%,至少80%,至少85%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。

[0075] 在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:1或3,或其成熟多肽编码序列编码。在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:1或其成熟多肽编码序列编码。在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:1编码。在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:3或其成熟多肽编码序列编码。在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:3编码。在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:1或3的亚序列编码,其中所述亚序列编码具有碳酸氢根转运蛋白活性的多肽。在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:1的亚序列编码,其中所述亚序列编码具有碳酸氢根转运蛋白活性的多肽。在一个方面,所述碳酸氢根转运蛋白由SEQ ID NO:3的亚序列编码,其中所述亚序列编码具有碳酸氢根转运蛋白活性的多肽。

[0076] 在一个方面,所述碳酸氢根转运蛋白是SEQ ID NO:2或4或其成熟多肽编码序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一个方面,所述碳酸氢根转运蛋白是SEQ ID NO:2的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一个方面,所述碳酸氢根转运蛋白是SEQ ID NO:2的成熟多肽的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一个方面,所述碳酸氢根转运蛋白是SEQ ID NO:4的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一个方面,所述碳酸氢根转运蛋白是SEQ ID NO:4的成熟多肽的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0077] 优选地,氨基酸改变的性质是较不重要的(of a minor nature),即保守的氨基

酸取代或插入,其不显著影响蛋白质的折叠和 / 或活性;通常为 1 至大约 30 个氨基酸的小缺失;小的氨基或羧基末端延伸,例如氨基末端甲硫氨酸残基;多至大约 20-25 个残基的小接头肽;或通过改变净电荷或其它功能来促进纯化的小延伸,如多组氨酸序列 (poly histidine tract)、抗原表位 (antigenic epitope) 或结合域 (binding domain)。

[0078] 保守取代的实例是在以下组之内:碱性氨基酸组(精氨酸、赖氨酸和组氨酸)、酸性氨基酸组(谷氨酸和天冬氨酸)、极性氨基酸组(谷氨酰胺和天冬酰胺)、疏水氨基酸组(亮氨酸、异亮氨酸和缬氨酸)、芳族氨基酸组(苯丙氨酸、色氨酸和酪氨酸)和小氨基酸组(甘氨酸、丙氨酸、丝氨酸、苏氨酸和甲硫氨酸)。通常不改变比活性 (specific activity) 的氨基酸取代是本领域已知的,并且由例如 H. Neurath 和 R. L. Hill, 1979, 于 *The Proteins*, Academic Press, New York 中描述。最普遍发生的交换是 Ala/Ser、Val/Ile、Asp/Glu、Thr/Ser、Ala/Gly、Ala/Thr、Ser/Asn、Ala/Val、Ser/Gly、Tyr/Phe、Ala/Pro、Lys/Arg、Asp/Asn、Leu/Ile、Leu/Val、Ala/Glu 和 Asp/Gly。

[0079] 或者,氨基酸改变具有这样的性质:使多肽的物理化学性质改变。例如,氨基酸改变可改善多肽的热稳定性,改变底物特异性,改变最适 pH 等。

[0080] 能够根据本领域已知的方法,例如定位诱变或丙氨酸分区诱变法 (Cunningham 和 Wells, 1989, *Science* 244:1081-1085) 来鉴定亲本多肽中的必需氨基酸。在后一技术中,将单一丙氨酸突变引入到分子中的每个残基,并且测试所得突变分子的碳酸氢根转运蛋白活性以鉴定对于所述分子的活性关键的氨基酸残基。同样参见 Hilton 等, 1996, *J. Biol. Chem.* 271:4699-4708。酶的活性部位或其它的生物相互作用也能够通过结构的物理分析而测定,如通过以下这些技术:如核磁共振、晶体学、电子衍射或光亲和标记,连同推定的接触位点氨基酸的突变来确定。参见例如 de Vos 等, 1992, *Science* 255:306-312; Smith 等, 1992, *J. Mol. Biol.* 224:899-904; Wlodaver 等, 1992, *FEBS Lett.* 309:59-64。必需氨基酸的身份 (identity) 也能够从与多肽的同一性分析来推断,所述多肽与参照的亲本多肽相关。

[0081] 能够使用已知的诱变、重组和 / 或改组 (shuffling) 方法接着进行有关的筛选方法,例如那些由 Reidhaar-Olson 和 Sauer, 1988, *Science* 241:53-57; Bowie 和 Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86:2152-2156; W095/17413; 或 W095/22625 公开的那些方法来进行并测试单个或多个氨基酸取代、缺失和 / 或插入。能够使用的其它方法包括易错 PCR、噬菌体展示 (例如, Lowman 等, 1991, *Biochemistry* 30:10832-10837; 美国专利 No. 5, 223, 409; W092/06204) 和区域定向的诱变 (Derbyshire 等, 1986, *Gene* 46:145; Ner 等, 1988, *DNA* 7:127)。

[0082] 诱变 / 改组方法能够与高通量、自动化的筛选方法组合以检测由宿主细胞表达的克隆的、诱变的多肽的活性 (Ness 等, 1999, *Nature Biotechnology* 17:893-896)。能够从宿主细胞回收编码活性多肽的诱变的 DNA 分子,并且使用本领域中的标准方法快速测序。这些方法允许快速确定多肽中单个氨基酸残基的重要性。

[0083] 在一些方面,SEQ ID NO:2 或 4,或其成熟多肽序列的氨基酸取代、缺失和 / 或插入的总数不多于 10,例如不多于 1, 2, 3, 4, 5, 6, 7, 8 或 9。

[0084] 在另一个方面,所述碳酸氢根转运蛋白是 SEQ ID NO:2 或 4,或其成熟多肽序列的片段,其中所述片段具有碳酸氢根转运蛋白活性。在一个方面,所述碳酸氢根转运蛋白是

SEQ ID NO:2 或其成熟多肽序列的片段,其中所述片段具有碳酸氢根转运蛋白活性。在一个方面,所述片段含有 SEQ ID NO:2 的至少 650 个氨基酸残基,例如优选至少 690 个氨基酸残基,或至少 730 个氨基酸残基。在一个方面,所述片段含有碳酸氢根转运蛋白域,例如, SEQ ID NO:2 的氨基酸 280 至 556 的推定的转运蛋白域。在另一个方面,所述碳酸氢根转运蛋白是 SEQ ID NO:4 或其成熟多肽序列的片段,其中所述片段具有碳酸氢根转运蛋白活性。在一个方面,所述片段含有 SEQ ID NO:4 的至少 720 个氨基酸残基,例如优选至少 760 个氨基酸残基,或至少 800 个氨基酸残基。在一个方面,所述片段含有碳酸氢根转运蛋白域,例如, SEQ ID NO:4 的氨基酸 192 至 480 的推定的转运蛋白域。

[0085] 所述碳酸氢根转运蛋白可为融合多肽或可切割的融合多肽,其中另一种多肽融合于本发明多肽的 N 端或 C 端。通过将编码另一个多肽的多核苷酸融合于本发明的多核苷酸来产生融合的多肽。产生融合多肽的技术是本领域已知的,并包括连接编码多肽的编码序列以使它们在阅读框中,并且使融合多肽的表达在相同启动子和终止子的调控下。融合蛋白亦可使用内蛋白 (intein) 技术构建,其中融合物在翻译后产生 (Cooper 等, 1993, EMBO J. 12:2575-2583; Dawson 等, 1994, Science 266:776-779)。

[0086] 融合多肽还可以包含在两个多肽之间的切割位点。一旦分泌了融合多肽,就切割所述位点,释放所述两个多肽。切割位点的实例包括,但不限于,公开于 Martin 等, 2003, J. Ind. Microbiol. Biotechnol. 3:568-76; Svetina 等, 2000, J. Biotechnol. 76:245-251; Rasmussen-Wilson 等, 1997, Appl. Environ. Microbiol. 63:3488-3493; Ward 等, 1995, Biotechnology 13:498-503; 和 Contreras 等, 1991, Biotechnology 9:378-381; Eaton 等, 1986, Biochem. 25:505-512; Collins-Racie 等, 1995, Biotechnology 13:982-987; Carter 等, 1989, Proteins: Structure, Function, and Genetics 6:240-248; 以及 Stevens, 2003, Drug Discovery World 4:35-48 中的位点。

[0087] 用于分离或克隆编码用于本文中提及的任何方面的多核苷酸—例如编码碳酸氢根转运蛋白的多核苷酸—以及其它多核苷酸的技术是本领域内已知的,包括从基因组 DNA 分离,从 cDNA 制备,或其组合。可通过例如使用熟知的聚合酶链式反应 (PCR) 或表达文库的抗体筛选来检测具有共有结构特性的克隆 DNA 片段,从而实现从这种基因组 DNA 克隆多核苷酸。参见,例如, Innis 等, 1990, PCR: A Guide to Methods and Application, Academic Press, New York。可以使用其它核酸扩增方法,如连接酶链式反应 (LCR)、连接活化转录 (ligated activated transcription; LAT) 和基于核苷酸序列的扩增 (NASBA)。可以从曲霉属 (*Aspergillus*) 菌株,或其他或相关生物体克隆所述多核苷酸,并且因此可为例如核苷酸序列的多肽编码区的等位基因变体或种间变体 (species variant)。

[0088] SEQ ID NO:1 或 3 的多核苷酸,或亚序列,以及 SEQ ID NO:2 或 4 的氨基酸序列,或其片段,可用于设计核酸探针,以根据本领域内公知的方法从不同属或种的菌株鉴定和克隆编码碳酸氢根转运蛋白的 DNA。具体而言,根据标准的 Southern 印迹方法,可将这些探针用于与感兴趣的属或种的基因组 DNA 或 cDNA 杂交,以鉴定和分离其中相应的基因。这些探针可明显短于完整序列,例如长度上为至少 14 个核苷酸,至少 25 个核苷酸,至少 35 个核苷酸,或至少 70 个核苷酸。所述核酸探针可以更长,例如至少 100 个核苷酸,至少 200 个核苷酸,至少 300 个核苷酸,至少 400 个核苷酸,或至少 500 个核苷酸的长度,可使用甚至更

长的探针,如至少 600 个核苷酸,例如至少 700 个核苷酸,至少 800 个核苷酸,或至少 900 个核苷酸的长度。DNA 和 RNA 探针二者均可使用。通常将探针标记以供检测相应的基因(例如,用  $^{32}\text{P}$ 、 $^3\text{H}$ 、 $^{35}\text{S}$ 、生物素或抗生物素蛋白(avidin)标记)。这些探针涵盖于本发明中。

[0089] 可从由这些其它菌株制备的基因组 DNA 或 cDNA 文库中筛选与上述探针杂交并且编码具有碳酸氢根转运蛋白活性的多肽的 DNA。可以通过琼脂糖或聚丙烯酰胺凝胶电泳,或通过其它分离技术分离来自这些其它株的基因组或其它 DNA。可以将来自文库的 DNA 或分离的 DNA 转移至硝化纤维素(nitrocellulose)或其它合适的载体材料并且固定于其上。为了鉴定与 SEQ ID NO:1 或 3 的成熟多肽编码序列,或它们的亚序列同源的克隆或 DNA,所述载体材料优选用于 Southern 印迹。

[0090] 就本发明而言,杂交表示多核苷酸在非常低至非常高的严格条件下与标记的核酸探针杂交,所述核酸探针对应于 SEQ ID NO:1 或 3,SEQ ID NO:1 或 3 的成熟多肽编码序列,或其全长互补链;或前述序列的亚序列。可使用例如 X 射线片(X-ray film)检测在这些条件下与核酸探针杂交的分子。

[0091] 在一个方面,核酸探针是 SEQ ID NO:1 或 3。在另一个方面,核酸探针是 SEQ ID NO:1 或 3 的成熟多肽编码序列。在另一个方面,核酸探针是 SEQ ID NO:1 的成熟多肽编码序列。在另一个方面,核酸探针是 SEQ ID NO:1。在另一个方面,核酸探针是 SEQ ID NO:3 的成熟多肽编码序列。在另一个方面,核酸探针是 SEQ ID NO:3。在另一个方面,核酸探针是编码 SEQ ID NO:2 的多肽或其片段的多核苷酸。在另一个方面,核酸探针是编码 SEQ ID NO:4 的多肽或其片段的多核苷酸。

[0092] 对于长度至少 100 个核苷酸的长探针,将非常低至非常高的严格条件定义为在  $42^{\circ}\text{C}$ ,在  $5\times\text{SSPE}$ 、 $0.3\%\text{SDS}$ 、 $200$  微克/ml 已剪切并且变性的鲑精 DNA 中,并且对于非常低和低严格性为 25% 的甲酰胺、对于中和中-高严格性为 35% 的甲酰胺、或对于高和非常高严格性为 50% 的甲酰胺,根据标准的 Southern 印迹法进行预杂交和杂交最佳 12 至 24 小时。使用  $2\times\text{SSC}$ 、 $0.2\%\text{SDS}$  在  $45^{\circ}\text{C}$  (非常低严格性),在  $50^{\circ}\text{C}$  (低严格性),在  $55^{\circ}\text{C}$  (中严格性),在  $60^{\circ}\text{C}$  (中-高严格性),在  $65^{\circ}\text{C}$  (高严格性),和在  $70^{\circ}\text{C}$  (非常高严格性)将载体材料最终洗涤三次,每次 15 分钟。

[0093] 对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针,将严格条件定义为在比使用根据 Bolton 和 McCarthy 计算法(1962, Proc. Natl. Acad. Sci. USA48:1390)计算的  $T_m$  低大约  $5^{\circ}\text{C}$  至大约  $10^{\circ}\text{C}$ ,在  $0.9\text{M NaCl}$ 、 $0.09\text{M Tris-HCl}$  pH7.6、 $6\text{mM EDTA}$ 、 $0.5\%\text{NP-40}$ 、 $1\times\text{Denhardt}$  溶液、 $1\text{mM}$  焦磷酸钠(sodium pyrophosphate)、 $1\text{mM}$  磷酸二氢钠(sodium monobasic phosphate)、 $0.1\text{mM ATP}$  和  $0.2\text{mg}$  每 ml 的酵母 RNA 中,根据标准的 Southern 印迹步骤进行预杂交和杂交最佳 12 至 24 小时。将所述载体材料在  $6\times\text{SSC}$  加  $0.1\%\text{SDS}$  中最终洗涤一次 15 分钟,并用  $6\times\text{SSC}$  在比计算的  $T_m$  低  $5^{\circ}\text{C}$  至  $10^{\circ}\text{C}$  的温度洗涤两次,每次 15 分钟。

[0094] 本发明碳酸氢根转运蛋白可以获得自任何属的微生物。如用于本文,与给定的来源有关的术语“获得自”,意思应为由多核苷酸编码的多肽由所述来源产生,或由其中插入了来自所述来源的多核苷酸的细胞产生。

[0095] 所述碳酸氢根转运蛋白可以是细菌碳酸氢根转运蛋白。例如,所述碳酸氢根转运蛋白可以是革兰氏阳性细菌多肽例如芽孢杆菌属(Bacillus)、链球菌属(Streptococcus)、

链霉菌属 (*Streptomyces*)、葡萄球菌属 (*Staphylococcus*)、肠球菌属 (*Enterococcus*)、乳杆菌属 (*Lactobacillus*)、乳球菌属 (*Lactococcus*)、梭菌属 (*Clostridium*)、地芽孢杆菌属 (*Geobacillus*) 或海洋芽孢杆菌属 (*Oceanobacillus*) 碳酸氢根转运蛋白;或革兰氏阴性细菌多肽,如大肠杆菌 (*E. coli*)、假单胞菌属 (*Pseudomonas*)、沙门氏菌属 (*Salmonella*)、弯曲杆菌属 (*Campylobacter*)、螺杆菌属 (*Helicobacter*)、黄杆菌属 (*Flavobacterium*)、梭杆菌属 (*Fusobacterium*)、泥杆菌属 (*Ilyobacter*)、奈瑟氏菌属 (*Neisseria*) 或脲原体属 (*Ureaplasma*) 碳酸氢根转运蛋白。

[0096] 在一个方面,所述碳酸氢根转运蛋白是嗜碱芽孢杆菌 (*Bacillus alkalophilus*)、解淀粉芽孢杆菌 (*Bacillus amyloliquefaciens*)、短芽孢杆菌 (*Bacillus brevis*)、环状芽孢杆菌 (*Bacillus circulans*)、克劳氏芽孢杆菌 (*Bacillus clausii*)、凝结芽孢杆菌 (*Bacillus coagulans*)、坚强芽孢杆菌 (*Bacillus firmus*)、灿烂芽孢杆菌 (*Bacillus lautus*)、迟缓芽孢杆菌 (*Bacillus lentus*)、地衣芽孢杆菌 (*Bacillus licheniformis*)、巨大芽孢杆菌 (*Bacillus megaterium*)、短小芽孢杆菌 (*Bacillus pumilus*)、嗜热脂肪芽孢杆菌 (*Bacillus stearothermophilus*)、枯草芽孢杆菌 (*Bacillus subtilis*) 或苏云金芽孢杆菌 (*Bacillus thuringiensis*) 碳酸氢根转运蛋白。

[0097] 在另一个方面,所述碳酸氢根转运蛋白是似马链球菌 (*Streptococcus equisimilis*)、酿脓链球菌 (*Streptococcus pyogenes*)、乳房链球菌 (*Streptococcus uberis*) 或马链球菌兽瘟亚种 (*Streptococcus equi* subsp. *Zooepidemicus*) 碳酸氢根转运蛋白。

[0098] 在另一个方面,所述碳酸氢根转运蛋白是不产色链霉菌 (*Streptomyces achromogenes*)、除虫链霉菌 (*Streptomyces avermitilis*)、天蓝链霉菌 (*Streptomyces coelicolor*)、灰色链霉菌 (*Streptomyces griseus*) 或浅青紫链霉菌 (*Streptomyces lividans*) 碳酸氢根转运蛋白。

[0099] 所述碳酸氢根转运蛋白可为真菌碳酸氢根转运蛋白。在一个方面,所述真菌碳酸氢根转运蛋白为酵母碳酸氢根转运蛋白,如假丝酵母属 (*Candida*)、克鲁维酵母属 (*Kluyveromyces*)、毕赤酵母属 (*Pichia*)、酵母属 (*Saccharomyces*)、裂殖酵母属 (*Schizosaccharomyces*) 或西洋著霉属 (*Yarrowia*) 碳酸氢根转运蛋白。

[0100] 在另一个方面,所述碳酸氢根转运蛋白是丝状真菌碳酸氢根转运蛋白,如枝顶孢霉属 (*Acremonium*)、伞菌属 (*Agaricus*)、链格孢属 (*Alternaria*)、曲霉属 (*Aspergillus*)、短梗霉属 (*Aureobasidium*)、*Botryosphaeria*、拟蜡菌属 (*Ceriporiopsis*)、毛喙壳属 (*Chaetomidium*)、金孢子菌属 (*Chrysosporium*)、*Claviceps*、*Cochliobolus*、鬼伞属 (*Coprinopsis*)、*Coptotermes*、棒囊壳属 (*Corynascus*)、隐丛赤壳菌属 (*Cryphonectria*)、隐球菌属 (*Cryptococcus*)、色二孢属 (*Diplodia*)、黑耳属 (*Exidia*)、*Filibasidium*、镰孢属 (*Fusarium*)、赤霉属 (*Gibberella*)、全鞭毛虫属 (*Holomastigotoides*)、腐质霉属 (*Humicola*)、耙齿菌属 (*Irpex*)、蘑菇属 (*Lentinula*)、*Leptosphaeria*、梨孢菌属 (*Magnaporthe*)、*Melanocarpus*、多孔菌属 (*Meripilus*)、毛霉属 (*Mucor*)、毁丝霉属 (*Myceliophthora*)、新考玛脂霉属 (*Neocallimastix*)、脉孢菌属 (*Neurospora*)、拟青霉属 (*Paecilomyces*)、青霉属 (*Penicillium*)、平革菌属 (*Phanerochaete*)、瘤胃壶菌属 (*Piromyces*)、*Poitrasia*、假黑盘菌属 (*Pseudoplectania*)、*Pseudotrichonympha*、



根毛霉属 (*Rhizomucor*)、裂褶菌属 (*Schizophyllum*)、柱顶孢属 (*Scytaalidium*)、踝节菌属 (*Talaromyces*)、嗜热子囊菌属 (*Thermoascus*)、梭孢壳属 (*Thielavia*)、弯颈霉属 (*Tolyposcladium*)、木霉属 (*Trichoderma*)、长毛盘菌属 (*Trichophaea*)、轮枝孢属 (*Verticillium*)、包脚菇属 (*Volvariella*) 或炭角菌属 (*Xylaria*) 碳酸氢根转运蛋白。

[0101] 在另一个方面,所述碳酸氢根转运蛋白是卡尔酵母 (*Saccharomyces carlsbergensis*)、酿酒酵母、糖化酵母 (*Saccharomyces diastaticus*)、道格拉氏酵母 (*Saccharomyces douglasii*)、克鲁弗酵母 (*Saccharomyces kluyveri*)、诺地酵母 (*Saccharomyces norbensis*) 或卵形酵母 (*Saccharomyces oviformis*) 碳酸氢根转运蛋白。

[0102] 在另一个方面,所述碳酸氢根转运蛋白是解纤维枝顶孢霉 (*Acremonium cellulolyticus*)、棘孢曲霉 (*Aspergillus aculeatus*)、泡盛曲霉 (*Aspergillus awamori*)、黄曲霉、烟曲霉 (*Aspergillus fumigatus*)、臭曲霉 (*Aspergillus foetidus*)、日本曲霉 (*Aspergillus japonicus*)、构巢曲霉 (*Aspergillus nidulans*)、黑曲霉 (*Aspergillus niger*)、米曲霉、酱油曲霉 (*Aspergillus sojae*)、嗜角质金孢子菌 (*Chrysosporium keratinophilum*)、*Chrysosporium lucknowense*、热带金孢子菌 (*Chrysosporium tropicum*)、*Chrysosporium merdarium*、*Chrysosporium inops*、毡金孢子菌 (*Chrysosporium pannicola*)、*Chrysosporium queenslandicum*、*Chrysosporium zonatum*、杆孢状镰孢 (*Fusarium bactridioides*)、禾谷镰孢 (*Fusarium cerealis*)、库威镰孢 (*Fusarium crookwellense*)、大刀镰孢 (*Fusarium culmorum*)、禾本科镰孢 (*Fusarium graminearum*)、禾赤镰孢 (*Fusarium graminum*)、异孢镰孢 (*Fusarium heterosporum*)、合欢木镰孢 (*Fusarium negundi*)、尖镰孢 (*Fusarium oxysporum*)、多枝镰孢 (*Fusarium reticulatum*)、粉红镰孢 (*Fusarium roseum*)、接骨木镰孢 (*Fusarium sambucinum*)、肤色镰孢 (*Fusarium sarcochromum*)、拟分枝孢镰孢 (*Fusarium sporotrichioides*)、硫色镰孢 (*Fusarium sulphureum*)、圆镰孢 (*Fusarium torulosum*)、拟丝孢镰孢 (*Fusarium trichothecioides*)、镶片镰孢 (*Fusarium venenatum*)、灰腐质霉 (*Humicola grisea*)、特异腐质霉 (*Humicola insolens*)、疏棉状腐质霉 (*Humicola lanuginosa*)、白耙齿菌 (*Irpex lacteus*)、米黑毛霉 (*Mucor miehei*)、嗜热毁丝霉 (*Myceliophthora thermophila*)、粗糙脉孢菌 (*Neurospora crassa*)、绳状青霉 (*Penicillium funiculosum*)、产紫青霉 (*Penicillium purpurogenum*)、黄孢平革菌 (*Phanerochaete chrysosporium*)、无色梭孢壳 (*Thielavia achromatica*)、*Thielavia albomyces*、*Thielavia albopilosa*、澳洲梭孢壳 (*Thielavia australeinsis*)、*Thielavia fimeti*、小孢梭孢壳 (*Thielavia microspora*)、卵孢梭孢壳 (*Thielavia ovispora*)、*Thielavia peruviana*、瘤孢梭孢壳 (*Thielavia spededonium*)、毛梭孢壳 (*Thielavia setosa*)、*Thielavia subthermophila*、土生梭孢霉 (*Thielavia terrestris*)、哈茨木霉 (*Trichoderma harzianum*)、康宁木霉 (*Trichoderma koningii*)、长枝木霉 (*Trichoderma longibrachiatum*)、里氏木霉 (*Trichoderma reesei*) 或绿色木霉 (*Trichoderma viride*) 碳酸氢根转运蛋白。

[0103] 在一个方面,所述碳酸氢根转运蛋白是曲霉属碳酸氢根转运蛋白,如米曲霉碳酸氢根转运蛋白。在一个方面,所述碳酸氢根转运蛋白是 SEQ ID NO:2 的米曲霉碳酸氢根转运蛋白。在另一个方面,所述碳酸氢根转运蛋白是 SEQ ID NO:4 的米曲霉碳酸氢根转运蛋白。

[0104] 可理解的是对于前述的种,本发明包含完全和不完全阶段(perfect and imperfect states),和其它分类学的等同物(equivalent),例如无性型(anamorph),而无论它们已知的种名。本领域技术人员将容易地识别适合的等同物的身份。

[0105] 这些种的菌株在许多培养物保藏中心对于公众能够容易地取得,所述保藏中心诸如美国典型培养物保藏中心(the American Type Culture Collection)(ATCC)、德意志微生物和细胞培养物保藏中心(Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH)(DSM)、真菌菌种保藏中心(Centraalbureau Voor Schimmelcultures)(CBS)和农业研究机构专利培养物保藏中心北区研究中心(Agricultural Research Service Patent Culture Collection, Northern Regional Research Center)(NRRL)。

[0106] 亦可使用上述的探针从其它来源,包括从自然界(例如,土壤、堆肥、水等)分离的微生物,或直接从自然材料(例如,土壤、堆肥、水等)获得的DNA样品鉴定和获得所述碳酸氢根转运蛋白。用于从天然生境(habitat)分离微生物和DNA的技术是本领域内公知的。然后可通过相似地筛选另一种微生物的基因组DNA或cDNA文库或混合的DNA样品来获得编码碳酸氢根转运蛋白的多核苷酸。一旦用如本文中所述的合适探针检测到编码碳酸氢根转运蛋白的多核苷酸,就能够使用本领域普通技术人员已知的技术将所述序列分离或克隆(参见,例如, J. Sambrook, E. F. Fritsch 和 T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 第2版, Cold Spring Harbor, New York)。

[0107] C4-二羧酸转运蛋白和编码C4-二羧酸转运蛋白的多核苷酸

[0108] 在重组宿主细胞及其使用方法的一些方面,所述宿主细胞具有C4-二羧酸转运蛋白活性。在一些方面,所述宿主细胞包含编码C4-二羧酸转运蛋白的异源多核苷酸。所述C4-二羧酸转运蛋白可为任何适于实践本发明的C4-二羧酸转运蛋白。在一个方面,所述C4-二羧酸转运蛋白存在于宿主细胞胞质溶胶中。

[0109] 在一个方面,所述C4-二羧酸转运蛋白是:(a)C4-二羧酸转运蛋白,其与SEQ ID NO:6或其成熟多肽序列具有至少60%序列同一性;(b)C4-二羧酸转运蛋白,其由多核苷酸编码,所述多核苷酸在低严格条件下与SEQ ID NO:5,其成熟多肽编码序列,或前述的全长互补链杂交;(c)C4-二羧酸转运蛋白,其由多核苷酸编码,所述多核苷酸与SEQ ID NO:5,其成熟多肽编码序列,或前述的全长互补链具有至少60%序列同一性;(d)SEQ ID NO:6或其成熟多肽序列的包含一个或多个氨基酸(例如两个、几个)的取代、缺失和/或插入的C4-二羧酸转运蛋白变体;和(e)(a)、(b)、(c)或(d)的多肽具有C4-二羧酸转运蛋白活性的片段。

[0110] 在一个方面,所述C4-二羧酸转运蛋白包含或组成为氨基酸序列,所述氨基酸序列与SEQ ID NO:6或其成熟多肽序列具有至少60%,例如至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,或至少99%序列同一性。在一个方面,所述C4-二羧酸转运蛋白包含氨基酸序列,所述氨基酸序列与SEQ ID NO:6或其成熟多肽序列相差不多于十个氨基酸,例如不多于五个氨基酸,不多于四个氨基酸,不多于三个氨基酸,不多于两个氨基酸,或一个氨基酸。

[0111] 在一个方面,所述C4-二羧酸转运蛋白包含或组成为SEQ ID NO:6的氨基酸序列,SEQ ID NO:6的成熟多肽序列,其等位变体,或前述的具有C4-二羧酸转运蛋白活性的序列。

在另一个方面,所述 C4- 二羧酸转运蛋白包含或组成为 SEQ ID NO:6 的氨基酸序列。在另一个方面,所述 C4- 二羧酸转运蛋白包含或组成为 SEQ ID NO:6 的成熟多肽序列。在另一个方面,所述 C4- 二羧酸转运蛋白包含或组成为 SEQ ID NO:6 的氨基酸 1 至 418。在另一个方面,所述 C4- 二羧酸转运蛋白包含或组成为 SEQ ID NO:6 的氨基酸 18 至 418。

[0112] 在一个方面,所述 C4- 二羧酸转运蛋白由多核苷酸编码,所述多核苷酸在至少低严格条件,例如中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件与 SEQ ID NO:5,其成熟多肽编码序列,或前述的全长互补链杂交(参见,例如 J. Sambrook, E. F. Fritsch 和 T. Maniatus, 1989, 见上文)。

[0113] 在一个方面,所述 C4- 二羧酸转运蛋白由多核苷酸编码,所述多核苷酸与 SEQ ID NO:5,其成熟多肽编码序列,或前述的全长互补链具有至少 65%,例如至少 70%,至少 75%,至少 80%,至少 85%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性。

[0114] 在一个方面,所述 C4- 二羧酸转运蛋白由 SEQ ID NO:5,或其成熟多肽编码序列编码。在一个方面,所述 C4- 二羧酸转运蛋白由 SEQ ID NO:5 编码。在一个方面,所述 C4- 二羧酸转运蛋白由 SEQ ID NO:5 的成熟多肽编码序列编码。在一个方面,所述成熟多肽编码序列是 SEQ ID NO:5 的核苷酸 1 至 1257。在一个方面,所述成熟多肽编码序列是 SEQ ID NO:5 的核苷酸 52 至 1257。在一个方面,所述 C4- 二羧酸转运蛋白由 SEQ ID NO:5 的亚序列编码,其中所述亚序列编码具有 C4- 二羧酸转运蛋白活性的多肽。在一个方面,所述亚序列含有 SEQ ID NO:7 的至少 885 个核苷酸,例如至少 930 个核苷酸或至少 975 个核苷酸。

[0115] 在一个方面,所述 C4- 二羧酸转运蛋白为 SEQ ID NO:6 或其成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体,如上文所述。在一个方面,所述 C4- 二羧酸转运蛋白为 SEQ ID NO:6 的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一个方面,所述 C4- 二羧酸转运蛋白为 SEQ ID NO:6 的成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一些方面,SEQ ID NO:6 或其成熟多肽序列的氨基酸取代、缺失和/或插入的总数不多于 10,例如不多于 1, 2, 3, 4, 5, 6, 7, 8 或 9。

[0116] 在另一个方面,所述 C4- 二羧酸转运蛋白是 SEQ ID NO:6 或其成熟多肽序列的片段,其中所述片段具有 C4- 二羧酸转运蛋白活性。在一个方面,所述的片段含有 SEQ ID NO:6 的至少 355 个氨基酸残基,例如至少 375 个氨基酸残基,或至少 395 个氨基酸残基。

[0117] 所述 C4- 二羧酸转运蛋白亦可为 C4- 二羧酸转运蛋白的等位变体或人工变体。

[0118] 所述 C4- 二羧酸转运蛋白亦可包括融合多肽或可切割的融合多肽,如上文所述。

[0119] 用于分离或克隆编码 C4- 二羧酸转运蛋白的多核苷酸的技术如上文中所述。

[0120] SEQ ID NO:5 的多核苷酸;或其亚序列;以及 SEQ ID NO:6 的氨基酸序列;或其片段;可用于设计核酸探针以从不同属或种的菌株鉴定和克隆编码 C4- 二羧酸转运蛋白的 DNA,如上文中所述。本发明涵盖此类探针。可对从此类其它生物制备的基因组 DNA 或 cDNA 筛选与如上所述的探针杂交并编码 C4- 二羧酸转运蛋白的 DNA,如上文中所述。

[0121] 在一个方面,所述核酸探针为 SEQ ID NO:5。在另一个方面,所述核酸探针为 SEQ ID NO:5 的成熟多肽编码序列。在另一个方面,所述核酸探针是多核苷酸序列,其编码 SEQ ID NO:6,其成熟多肽序列或前述序列的片段。

[0122] 对于长度至少 100 个核苷酸的长探针,非常低至非常高的严格条件和洗涤条件如上文中所述定义。对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针,严格条件和洗涤条件如上文中所述定义。

[0123] 所述 C4-二羧酸转运蛋白可获得自任何属的微生物。在一个方面,所述 C4-二羧酸转运蛋白可为从本文中所述的微生物获得的细菌、酵母或丝状真菌 C4-二羧酸转运蛋白。在另一个方面,所述 C4-二羧酸转运蛋白是曲霉属 C4-二羧酸转运蛋白,如棘孢曲霉 C4-二羧酸转运蛋白,例如 SEQ ID NO:6 的棘孢曲霉 C4-二羧酸转运蛋白。

[0124] 其它可与本发明中所述的宿主细胞和使用方法使用的 C4-二羧酸转运蛋白包括但不限于,黄曲霉 (*Aspergillus flavus*)C4-二羧酸转运蛋白 (AFLA\_107340), SEQ ID NO:27 的米曲霉 C4-二羧酸转运蛋白 (由 SEQ ID NO:26 的多核苷酸序列编码;参见 US2011/0053233), SEQ ID NO:29 的土曲霉 (*Aspergillus terreus*)C4-二羧酸转运蛋白 (由 SEQ ID NO:28 的多核苷酸序列编码,参见 US2011/0053233), SEQ ID NO:32 的粟酒裂殖酵母 (*Schizosaccharomyces pombe*)C4-二羧酸转运蛋白 (由 SEQ ID NO:30 或 31 的多核苷酸序列编码,参见 US2011/0053233), SEQ ID NO:34 的棘孢曲霉 (*Aspergillus aculeatus*)C4-二羧酸转运蛋白 (由 SEQ ID NO:33 的多核苷酸序列编码,参见 2011 年 6 月 21 日提交的,标题为“Polypeptides Having C4-dicarboxylic acid Transporter Activity and Polynucleotides Encoding Same”的美国申请号 13/165,696), SEQ ID NO:36 的棘孢曲霉 C4-二羧酸转运蛋白 (由 SEQ ID NO:35 的多核苷酸序列编码,参见美国申请号 13/165,696,见上文), SEQ ID NO:39 的日本裂殖酵母 (*Schizosaccharomyces japonicus*)C4-二羧酸转运蛋白 (由 SEQ ID NO:37 或 38 的多核苷酸序列编码,参见 2011 年 6 月 2 日提交的,标题为“C4-dicarboxylic acid Production in Filamentous Fungi”的 PCT/US11/38881), SEQ ID NO:41 的棒曲霉 (*Aspergillus clavatus*)C4-二羧酸转运蛋白 (由 SEQ ID NO:40 的多核苷酸序列编码;参见 2011 年 6 月 21 日提交的,标题为“Methods for Improving C4-dicarboxylic acid Production in Filamentous Fungi”的美国申请号 13/165,719), SEQ ID NO:43 的烟曲霉 (*Aspergillus fumigatus*)C4-二羧酸转运蛋白 (由 SEQ ID NO:42 的多核苷酸编码,参见美国申请号 13/165,719,见上文),或任何相应的参考文献中所述的 C4-二羧酸转运蛋白的任何方面。本文中所述的任何涉及其序列同一性,杂交,氨基酸修饰 (例如取代、缺失和 / 或插入),片段,或亚序列的方面也涵盖上述的 C4-二羧酸转运蛋白。

[0125] 本发明涵盖将本文中所述的序列同一性、杂交、变体和片段的任何方面适用于如上所述的 C4-二羧酸转运蛋白多肽序列和多核苷酸序列。举例而言,在一个方面,所述 C4-二羧酸转运蛋白是 (a)C4-二羧酸转运蛋白,其与 SEQ ID NO:27,29,32,34,36,39,41,或 43 或其成熟多肽序列具有至少 60%,例如至少 65%,至少 70%,至少 75%,至少 80%,至少 85%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性;(b)C4-二羧酸转运蛋白,其由多核苷酸编码,所述多核苷酸在低严格条件,例如中等严格条件,中等-高严格条件,高严格条件或非常高严格条件下与以下杂交:(i)SEQ ID NO:26,28,30,31,33,35,37,38,40,或 42 或其成熟多肽编码序列,(ii)SEQ ID NO:26,28,30,31,33,35,37,38,40,或 42 的 cDNA 序列或其成熟多肽编码序列,或 (iii) (i) 或 (ii) 的全长互补链;(c)C4-二羧酸转运蛋白,其由多核苷酸编码,所述多核苷酸与 (iv)SEQ ID NO:26,28,30,31,33,35,37,38,40,或 42 或其成熟多肽编

码序列, (v) SEQ ID NO:26, 28, 30, 31, 33, 35, 37, 38, 40, 或 42 的 cDNA 序列或其成熟多肽编码序列, 或 (vi) (iv) 或 (v) 的全长互补链具有至少 60%, 例如至少 65%, 至少 70%, 至少 75%, 至少 80%, 至少 85%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 至少 99%, 或 100% 序列同一性; (d) SEQ ID NO:27, 29, 32, 34, 36, 39, 41, 或 43 或其成熟多肽序列的包含一个或多个氨基酸 (几个) 的取代、缺失和 / 或插入的 C4- 二羧酸转运蛋白变体; 和 (e) (a)、(b)、(c) 或 (d) 的多肽具有 C4- 二羧酸转运蛋白活性的片段。

[0126] 所述 C4- 二羧酸转运蛋白亦可从其它来源鉴定和获取, 所述来源包括从自然界 (例如土壤、堆肥、水等) 分离的微生物或直接从自然材料 (例如土壤、堆肥、水等) 获得的 DNA 样品, 如上文中所述。

[0127] 苹果酸脱氢酶和编码苹果酸脱氢酶的多核苷酸

[0128] 在重组宿主细胞及其使用方法的一些方面, 所述宿主细胞具有苹果酸脱氢酶活性。在一些方面, 所述宿主细胞包含编码苹果酸脱氢酶的异源多核苷酸。所述苹果酸脱氢酶可为任何适于实践本发明的苹果酸脱氢酶。在一个方面, 所述苹果酸脱氢酶是存在于宿主细胞胞质溶胶中的酶。

[0129] 在本文所述的重组宿主细胞和方法的一个方面, 所述苹果酸脱氢酶是: (a) 苹果酸脱氢酶, 其与 SEQ ID NO:8 或其成熟多肽序列具有至少 60% 序列同一性; (b) 苹果酸脱氢酶, 其由多核苷酸编码, 所述多核苷酸在低严格条件下与以下杂交: (i) SEQ ID NO:7, 或其成熟多肽编码序列, (ii) SEQ ID NO:7 的 cDNA 序列或其成熟多肽编码序列, 或 (iii) (i) 或 (ii) 的全长互补链; (c) 苹果酸脱氢酶, 其由多核苷酸编码, 所述多核苷酸与以下具有至少 60% 序列同一性: (iv) SEQ ID NO:7, 或其成熟多肽编码序列, (v) SEQ ID NO:7 的 cDNA 序列或其成熟多肽编码序列, 或 (vi) (iv) 或 (v) 的全长互补链; (d) SEQ ID NO:8 或其成熟多肽序列的包含一个或多个 (几个) 氨基酸的取代、缺失和 / 或插入的苹果酸脱氢酶变体; 和 (e) (a)、(b)、(c) 或 (d) 的多肽具有苹果酸脱氢酶活性的片段。

[0130] 在一个方面, 所述苹果酸脱氢酶包含或组成为这样的氨基酸序列, 所述氨基酸序列与 SEQ ID NO:8 或其成熟多肽序列具有至少 60%, 例如至少 65%, 至少 70%, 至少 75%, 至少 80%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 或至少 99% 序列同一性。在一个方面, 所述苹果酸脱氢酶包含氨基酸序列, 所述氨基酸序列与 SEQ ID NO:8 或其成熟多肽序列相差不多于十个氨基酸, 例如不多于五个氨基酸, 不多于四个氨基酸, 不多于三个氨基酸, 不多于两个氨基酸, 或一个氨基酸。

[0131] 在一个方面, 所述苹果酸脱氢酶包含或组成为 SEQ ID NO:8 的氨基酸序列, SEQ ID NO:8 的成熟多肽序列, 其等位变体, 或前述的具有苹果酸脱氢酶活性的序列。在另一个方面, 所述苹果酸脱氢酶包含或组成为 SEQ ID NO:8 的氨基酸序列。在另一个方面, 所述苹果酸脱氢酶包含或组成为 SEQ ID NO:8 的成熟多肽序列。在另一个方面, 所述苹果酸脱氢酶包含或组成为 SEQ ID NO:8 的氨基酸 1 至 330。

[0132] 在一个方面, 所述苹果酸脱氢酶由多核苷酸编码, 所述多核苷酸在至少低严格条件, 例如中等严格条件, 中等 - 高严格条件, 高严格条件, 或非常高严格条件与以下杂交: (i) SEQ ID NO:7 或其成熟多肽编码序列, (ii) SEQ ID NO:7 的 cDNA 序列或其成熟多肽编码序列, 或 (iii) (i) 或 (ii) 的全长互补链 (参见, 例如 J. Sambrook, E. F. Fritsch 和

T. Maniatus, 1989, 见上文)。

[0133] 在一个方面,所述苹果酸脱氢酶由多核苷酸编码,所述多核苷酸与(iv)SEQ ID NO:7或其成熟多肽编码序列,(v)SEQ ID NO:7的cDNA序列或其成熟多肽编码序列,或(vi)(iv)或(v)的全长互补链具有至少65%,例如至少70%,至少75%,至少80%,至少85%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。

[0134] 在一个方面,所述苹果酸脱氢酶由SEQ ID NO:7,或其成熟多肽编码序列编码。在一个方面,所述苹果酸脱氢酶由SEQ ID NO:7编码。在一个方面,所述苹果酸脱氢酶由SEQ ID NO:7的成熟多肽编码序列编码。在一个方面,所述苹果酸脱氢酶由SEQ ID NO:7的亚序列编码,其中所述亚序列编码具有苹果酸脱氢酶活性的多肽。在一个方面,所述亚序列含有SEQ ID NO:7的至少885个核苷酸,例如至少930个核苷酸或至少975个核苷酸。

[0135] 在一个方面,所述苹果酸脱氢酶为SEQ ID NO:8或其成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体,如上文所述。在一个方面,所述苹果酸脱氢酶为SEQ ID NO:8的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一个方面,所述苹果酸脱氢酶为SEQ ID NO:8的成熟多肽序列的包含一个或多个氨基酸的取代、缺失和/或插入的变体。在一些方面,SEQ ID NO:8或其成熟多肽序列的氨基酸取代、缺失和/或插入的总数不多于10,例如不多于1,2,3,4,5,6,7,8或9。

[0136] 在另一个方面,所述苹果酸脱氢酶是SEQ ID NO:8或其成熟多肽序列的片段,其中所述片段具有苹果酸脱氢酶活性。在一个方面,所述的片段含有SEQ ID NO:8的至少295个氨基酸残基,例如至少310个氨基酸残基,或至少325个氨基酸残基。

[0137] 所述苹果酸脱氢酶亦可为苹果酸脱氢酶的等位变体或人工变体。

[0138] 所述苹果酸脱氢酶亦可包括融合多肽或可切割的融合多肽,如上文所述。

[0139] 用于分离或克隆编码苹果酸脱氢酶的多核苷酸的技术如上文中所述。

[0140] SEQ ID NO:7的多核苷酸;或其亚序列;以及SEQ ID NO:8的氨基酸序列;或其片段;可用于设计核酸探针以从不同属或种的菌株鉴定和克隆编码苹果酸脱氢酶的DNA,如上文中所述。本发明涵盖此类探针。可对从此类其它生物制备的基因组DNA或cDNA筛选与如上所述的探针杂交并编码苹果酸脱氢酶的DNA,如上文中所述。

[0141] 在一个方面,所述核酸探针为SEQ ID NO:7。在另一个方面,所述核酸探针为SEQ ID NO:7的成熟多肽编码序列。在另一个方面,所述核酸探针是多核苷酸序列,其编码SEQ ID NO:8,其成熟多肽序列或前述序列的片段。

[0142] 对于长度至少100个核苷酸的长探针,非常低至非常高的严格条件和洗涤条件如上文中所述定义。对于长度大约15个核苷酸至大约70个核苷酸的短探针,严格条件和洗涤条件如上文中所述定义。

[0143] 所述苹果酸脱氢酶可获得自任何属的微生物。在一个方面,所述苹果酸脱氢酶可为从本文中所述的微生物获得的细菌、酵母或丝状真菌苹果酸脱氢酶。在另一个方面,所述苹果酸脱氢酶是米曲霉苹果酸脱氢酶,例如SEQ ID NO:8的米曲霉苹果酸脱氢酶。

[0144] 其它可用于实践本发明的苹果酸脱氢酶包括但不限于:构巢曲霉苹果酸脱氢酶(AN6717.1;SIMS等,2004,Mycol. Res. 108:853-857);黑曲霉苹果酸脱氢酶(An16g00120;Pel等,2007,Nature Biotechnology 25:221-231);Phytophthora infestans 苹果

酸脱氢酶 (PITG13614.1; Calcagno 等, 2009, *Mycological Research* 113:771-781); 酿酒酵母苹果酸脱氢酶 (YKL085W; McAlister-Henn 和 Thompson, 1987, *J Bacteriol.* 169:5157-5166); 埃默森踝节菌苹果酸脱氢酶 (AF439996, AF487682; Maloney 等, 2004, *Eur. J. Biochem.* 271:3115-3126); 以及玉蜀黍黑粉菌 (*Ustilago maydis*) 苹果酸脱氢酶 (um00403, um11161; McCann 和 Snetselaar, 2008, *Fungal Genetics and Biology* 45:S77 - S87), SEQ ID NO:45 的米曲霉苹果酸脱氢酶 (其由 SEQ ID NO:44 的核苷酸序列所编码; 参见美国申请号 12/870, 523, 标题为“Methods for Improving Malic Acid Production in Filamentous Fungi”, 2010 年 8 月 27 日提交), 或任何相应的参考文献中所述的苹果酸脱氢酶的任何方面。本文中所述的任何涉及其序列同一性, 杂交, 氨基酸修饰 (例如取代、缺失和 / 或插入), 片段, 或亚序列的方面也涵盖上述的苹果酸脱氢酶。

[0145] 本发明涵盖将本文中所述的序列同一性、杂交、变体和片段的任何方面适用于如上所述的苹果酸脱氢酶多肽序列和多核苷酸序列。举例而言, 在一个方面, 所述苹果酸脱氢酶是 (a) 苹果酸脱氢酶, 其与 SEQ ID NO:45 或其成熟多肽序列具有至少 60%, 例如至少 65%, 至少 70%, 至少 75%, 至少 80%, 至少 85%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 至少 99%, 或 100% 序列同一性; (b) 苹果酸脱氢酶, 其由多核苷酸编码, 所述多核苷酸在低严格条件, 例如中等严格条件, 中等 - 高严格条件, 高严格条件或非常高严格条件下与以下杂交: (i) SEQ ID NO:44 或其成熟多肽编码序列, (ii) SEQ ID NO:44 的 cDNA 序列或其成熟多肽编码序列, 或 (iii) (i) 或 (ii) 的全长互补链; (c) 苹果酸脱氢酶, 其由多核苷酸编码, 所述多核苷酸与 (iv) SEQ ID NO:44 或其成熟多肽编码序列, (v) SEQ ID NO:44 的 cDNA 序列或其成熟多肽编码序列, 或 (vi) (iv) 或 (v) 的全长互补链具有至少 60%, 例如至少 65%, 至少 70%, 至少 75%, 至少 80%, 至少 85%, 至少 85%, 至少 90%, 至少 91%, 至少 92%, 至少 93%, 至少 94%, 至少 95%, 至少 96%, 至少 97%, 至少 98%, 至少 99%, 或 100% 序列同一性; (d) SEQ ID NO:45 或其成熟多肽序列的包含一个或多个氨基酸 (几个) 的取代、缺失和 / 或插入的苹果酸脱氢酶变体; 和 (e) (a)、(b)、(c) 或 (d) 的多肽具有苹果酸脱氢酶活性的片段。

[0146] 所述苹果酸脱氢酶亦可从其它来源鉴定和获取, 所述来源包括从自然界 (例如土壤、堆肥、水等) 分离的微生物或直接从自然材料 (例如土壤、堆肥、水等) 获得的 DNA 样品, 如上文中所述。

[0147] 丙酮酸羧化酶和编码丙酮酸羧化酶的多核苷酸

[0148] 在重组宿主细胞及其使用方法的一些方面, 所述宿主细胞具有丙酮酸羧化酶活性。在一些方面, 所述宿主细胞包含编码丙酮酸羧化酶的异源多肽。所述丙酮酸羧化酶可为任何适于实践本发明的丙酮酸羧化酶。在一个方面, 所述丙酮酸羧化酶是存在于宿主细胞胞质溶胶中的酶。

[0149] 在本文所述的重组宿主细胞和方法的一个方面, 所述丙酮酸羧化酶是: (a) 丙酮酸羧化酶, 其与 SEQ ID NO:10 或其成熟多肽序列具有至少 60% 序列同一性; (b) 丙酮酸羧化酶, 其由多核苷酸编码, 所述多核苷酸在低严格条件下与以下杂交: (i) SEQ ID NO:9, 或其成熟多肽编码序列, (ii) SEQ ID NO:9 的 cDNA 序列或其成熟多肽编码序列, 或 (iii) (i) 或 (ii) 的全长互补链; (c) 丙酮酸羧化酶, 其由多核苷酸编码, 所述多核苷酸与以下具有至少 60% 序列同一性: (iv) SEQ ID NO:9, 或其成熟多肽编码序列, (v) SEQ ID NO:9 的 cDNA 序

列或其成熟多肽编码序列,或(vi)(iv)或(v)的全长互补链;(d)SEQ ID NO:10或其成熟多肽序列的包含一个或多个氨基酸(几个)的取代、缺失和/或插入的丙酮酸羧化酶变体;和(e)(a)、(b)、(c)或(d)的多肽具有丙酮酸羧化酶活性的片段。

[0150] 在一个方面,所述丙酮酸羧化酶包含或组成为这样的氨基酸序列,所述氨基酸序列与SEQ ID NO:10或其成熟多肽序列具有至少60%,例如至少65%,至少70%,至少75%,至少80%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,或至少99%序列同一性。在一个方面,所述丙酮酸羧化酶包含这样的氨基酸序列,所述氨基酸序列与SEQ ID NO:10或其成熟多肽序列相差不多于十个氨基酸,例如不多于五个氨基酸,不多于四个氨基酸,不多于三个氨基酸,不多于两个氨基酸,或一个氨基酸。

[0151] 在一个方面,所述丙酮酸羧化酶包含或组成为SEQ ID NO:10的氨基酸序列,SEQ ID NO:10的成熟多肽序列,其等位变体,或前述序列的具有丙酮酸羧化酶活性的序列。在另一个方面,所述丙酮酸羧化酶包含或组成为SEQ ID NO:10的氨基酸序列。在另一个方面,所述丙酮酸羧化酶包含或组成为SEQ ID NO:10的成熟多肽序列。在另一个方面,所述丙酮酸羧化酶包含或组成为SEQ ID NO:10的氨基酸1至1193。

[0152] 在一个方面,所述丙酮酸羧化酶由多核苷酸编码,所述多核苷酸在至少低严格条件,例如中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件与以下杂交:(i)SEQ ID NO:9或其成熟多肽编码序列,(ii)SEQ ID NO:9的cDNA序列或其成熟多肽编码序列,或(iii)(i)或(ii)的全长互补链(参见,例如J. Sambrook, E. F. Fritsch和T. Maniatus, 1989, 见上文)。

[0153] 在一个方面,所述丙酮酸羧化酶由多核苷酸编码,所述多核苷酸与(iv)SEQ ID NO:9或其成熟多肽编码序列,(v)SEQ ID NO:9的cDNA序列或其成熟多肽编码序列,或(vi)(iv)或(v)的全长互补链具有至少65%,例如至少70%,至少75%,至少80%,至少85%,至少85%,至少90%,至少91%,至少92%,至少93%,至少94%,至少95%,至少96%,至少97%,至少98%,至少99%,或100%序列同一性。

[0154] 在一个方面,所述丙酮酸羧化酶由SEQ ID NO:9,或其成熟多肽编码序列编码。在一个方面,所述丙酮酸羧化酶由SEQ ID NO:9编码。在一个方面,所述丙酮酸羧化酶由SEQ ID NO:9的成熟多肽编码序列编码。在一个方面,所述丙酮酸羧化酶由SEQ ID NO:9的亚序列编码,其中所述亚序列编码具有丙酮酸羧化酶活性的多肽。在一个方面,所述亚序列含有SEQ ID NO:9的至少3060个核苷酸,例如至少3240个核苷酸或至少3420个核苷酸。

[0155] 在一个方面,所述丙酮酸羧化酶为SEQ ID NO:10或其成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体,如上文所述。在一个方面,所述丙酮酸羧化酶为SEQ ID NO:10的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一个方面,所述丙酮酸羧化酶为SEQ ID NO:10的成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。在一些方面,SEQ ID NO:10或其成熟多肽序列的氨基酸取代、缺失和/或插入的总数不多于10,例如不多于1,2,3,4,5,6,7,8或9。

[0156] 在另一个方面,所述丙酮酸羧化酶是SEQ ID NO:10或其成熟多肽序列的片段,其中所述片段具有丙酮酸羧化酶活性。在一个方面,所述的片段含有SEQ ID NO:10的至少1020个氨基酸残基,例如至少1080个氨基酸残基,或至少1140个氨基酸残基。



[0157] 所述丙酮酸羧化酶亦可为丙酮酸羧化酶的等位变体或人工变体。

[0158] 所述丙酮酸羧化酶亦可包括融合多肽或可切割的融合多肽,如上文所述。

[0159] 所述丙酮酸羧化酶可为线粒体丙酮酸羧化酶的变体,使得体内输入线粒体减少,由此增加所述丙酮酸羧化酶在胞质溶胶中的水平。

[0160] 用于分离或克隆编码丙酮酸羧化酶的多核苷酸的技术如上文所述。

[0161] SEQ ID NO:9的多核苷酸;或其亚序列;以及SEQ ID NO:10的氨基酸序列;或其片段;可用于设计核酸探针以从不同属或种的菌株鉴定和克隆编码丙酮酸羧化酶的DNA,如上文所述。本发明涵盖此类探针。可对从此类其它生物制备的基因组DNA或cDNA筛选与如上所述的探针杂交并编码丙酮酸羧化酶的DNA,如上文所述。

[0162] 在一个方面,所述核酸探针为SEQ ID NO:9。在另一个方面,所述核酸探针为SEQ ID NO:9的成熟多肽编码序列。在另一个方面,所述核酸探针是多核苷酸序列,其编码SEQ ID NO:10,其成熟多肽序列或前述序列的片段。

[0163] 对于长度至少100个核苷酸的长探针,非常低至非常高的严格条件和洗涤条件如上文所述定义。对于长度大约15个核苷酸至大约70个核苷酸的短探针,严格条件和洗涤条件如上文所述定义。

[0164] 所述丙酮酸羧化酶可获得自任何属的微生物。在一个方面,所述丙酮酸羧化酶可为从本文中所述的微生物获得的细菌、酵母或丝状真菌丙酮酸羧化酶。在另一个方面,所述丙酮酸羧化酶是米曲霉丙酮酸羧化酶,例如SEQ IDNO:10的米曲霉丙酮酸羧化酶。

[0165] 其它可用于实施本发明的丙酮酸羧化酶包括但不限于:棒曲霉(*Aspergillus clavatus*)NRRL1丙酮酸羧化酶(XP\_001271664;Direct Submission,Submitted(26-OCT-2006),The Institute for Genomic Research,9712Medical Center Drive,Rockville,MD20850,USA);烟曲霉Af293丙酮酸羧化酶(XP\_752054;Nierman等,2005,Nature438:1151-1156);构巢曲霉FGSC A4丙酮酸羧化酶(XP\_662066;Galagan等,2005,Nature438:1105-1115);黑曲霉丙酮酸羧化酶(An15g02820;Pel等,2007,Nature Biotechnology25:221-231;ASPNG5061;Panneman等,(JUL-1998)提交至EMBL/GenBank/DDBJ数据库);土曲霉丙酮酸羧化酶(093918;Direct Submission,Submitted(OCT-1998)The Institute for Genomic Research,9712Medical Center Drive,Rockville,MD20850,USA);Magnaporthe grisea70-15丙酮酸羧化酶(XP\_367852;Direct Submission,Submitted(26-SEP-2005)Broad Institute of MIT and Harvard,320Charles Street,Cambridge,MA02142,USA);粗糙脉孢菌OR74A丙酮酸羧化酶(XP\_965636;Galagan等,2003,Nature422:859-868);米根霉(*Rhizopus oryzae*)丙酮酸羧化酶(R03G\_06931.1);酿酒酵母丙酮酸羧化酶(NP\_009777;Gaffeau等,1996,Science274:546-547);粟酒裂殖酵母丙酮酸羧化酶(NP\_595900;Direct Submission,Submitted(29-JUN-2007)European Schizosaccharomyces genome sequencing project,Sanger Institute,The Wellcome Trust Genome Campus,Hinxton,Cambridge CB101SA);和玉蜀黍黑粉菌丙酮酸羧化酶(um01054;McCann和Snetselaar,2008,Fungal Genetics and Biology45:S77-S87)。本文中所述的任何涉及其序列同一性,杂交,氨基酸修饰(例如取代、缺失和/或插入),片段,或亚序列的方面也涵盖上述的丙酮酸羧化酶。

[0166] 所述丙酮酸羧化酶亦可从其它来源鉴定和获取,所述来源包括从自然界(例如土

壤、堆肥、水等)分离的微生物或直接从自然材料(例如土壤、堆肥、水等)获得的DNA样品,如上文中所述。

#### [0167] 核酸构建体

[0168] 本发明还涉及使用核酸构建体的重组宿主细胞和方法,所述核酸构建体包含编码碳酸氢根转运蛋白(和/或编码C4-二羧酸转运蛋白,苹果酸脱氢酶和/或丙酮酸羧化酶)的异源多核苷酸与一个或多个(几个)调控序列连接,所述调控序列指导在合适的宿主细胞中在与该调控序列相容的条件下表达。此类核酸构建体可用于本文中所述的任何宿主细胞和方法。本文中所述的多核苷酸可以通过多种方式加以操纵,来为所需多肽的表达提供条件。取决于表达载体,在将多核苷酸插入载体之前对其进行操作可能是理想的或必需的。使用重组DNA方法修饰多核苷酸的技术是本领域熟知的。

[0169] 调控序列可为启动子序列,启动子序列是一种多核苷酸,被用于表达本文中所述的任何多肽(例如碳酸氢根转运蛋白,C4-二羧酸转运蛋白,苹果酸脱氢酶,或丙酮酸羧化酶)的多核苷酸的宿主细胞所识别。启动子序列含有介导多肽的表达的转录调控序列。启动子可以是在所选的宿主细胞中显示转录活性的任何多核苷酸,包括突变的、截短的和杂合的启动子,并且可以从编码与宿主细胞同源或异源的胞外或胞内多肽的基因获得。

[0170] 本文中所述的每个多核苷酸都可以可操作地连接于对于所述多核苷酸而言是外来的启动子。例如,在一个方面,所述编码碳酸氢根转运蛋白的异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。在另一个方面,编码C4-二羧酸的异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。在另一个方面,编码苹果酸脱氢酶的异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。在另一个方面,编码丙酮酸羧化酶的异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。

[0171] 用于指导本发明的核酸构建体在细菌宿主细胞中转录的合适启动子的实例是从下述获得的启动子:解淀粉芽孢杆菌 $\alpha$ -淀粉酶基因(amyQ)、地衣芽孢杆菌 $\alpha$ -淀粉酶基因(amyL)、地衣芽孢杆菌青霉素酶基因(penP)、嗜热脂肪芽孢杆菌产麦芽糖淀粉酶基因(amyM)、枯草芽孢杆菌果聚糖蔗糖酶基因(sacB)、枯草芽孢杆菌xylA和xylB基因、大肠杆菌lac操纵子、大肠杆菌trc启动子(Egon等,1988, Gene69:301-315)、天蓝色链霉菌琼脂糖酶基因(dagA)和原核 $\beta$ -内酰胺酶基因(Villa-Kamaroff等,1978, Proceedings of the National Academy of Sciences USA75:3727-3731),以及tac启动子(DeBoer等,1983, Proc. Natl. Acad. Sci. USA80:21-25)。另外的启动子在“Useful proteins from recombinant bacteria”于Gilbert等,1980, Scientific American, 242:74-94中;和在Sambrook等,1989, 见上文描述。

[0172] 用于指导本发明的核酸构建体在丝状真菌宿主细胞中转录的合适启动子的实例是从下列酶的基因获得的启动子:构巢曲霉乙酰胺酶、黑曲霉中性 $\alpha$ -淀粉酶、黑曲霉酸稳定性 $\alpha$ -淀粉酶、黑曲霉或泡盛曲霉葡糖淀粉酶(glaA)、米曲霉TAKA淀粉酶、米曲霉碱性蛋白酶、米曲霉丙糖磷酸异构酶、尖镰孢胰蛋白酶样蛋白酶(W096/00787)、镶片镰孢淀粉葡糖苷酶(W000/56900)、镶片镰孢Daria(W000/56900)、镶片镰孢Quinn(W000/56900)、曼赫根毛霉(Rhizomucor miehei)脂肪酶、曼赫根毛霉天冬氨酸蛋白酶、里氏木霉 $\beta$ -葡糖苷酶、里氏木霉纤维二糖水解酶I、里氏木霉纤维二糖水解酶II、里氏木霉内切葡聚糖酶I、里

氏木霉内切葡聚糖酶 II、里氏木霉内切葡聚糖酶 III、里氏木霉内切葡聚糖酶 IV、里氏木霉内切葡聚糖酶 V、里氏木霉木聚糖酶 I、里氏木霉木聚糖酶 II、里氏木霉  $\beta$ -木糖苷酶,以及 NA2-tpi 启动子(一种修饰的启动子,其来自曲霉属中编码中性  $\alpha$ -淀粉酶的基因,其中未翻译的前导序列已由曲霉属(Aspergilli)中编码丙糖磷酸异构酶的基因的未翻译的前导序列所替代;非限制性实例包括修饰的启动子,其来自黑曲霉中编码中性  $\alpha$ -淀粉酶的基因,其中未翻译的前导序列已由构巢曲霉或米曲霉中编码丙糖磷酸异构酶的基因的未翻译的前导序列所替代);和它们的突变的、截短的和杂合的启动子。

[0173] 在酵母宿主中,有用的启动子从如下酶的基因获得:酿酒酵母烯醇化酶(ENO-1)、酿酒酵母半乳糖激酶(GAL1)、酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH1, ADH2/GAP)、酿酒酵母丙糖磷酸异构酶(TPI)、酿酒酵母金属硫蛋白(CUP1)和酿酒酵母3-磷酸甘油酸激酶。对于酵母宿主细胞其它有用的启动子由 Romanos 等,1992, *Yeast*8:423-488 描述。

[0174] 调控序列也可以是合适的转录终止子序列,其由宿主细胞识别以终止转录。所述终止子序列与编码所述多肽的多核苷酸的3'末端可操作地连接。可以将所选宿主细胞中有功能的任何终止子用在本发明中。

[0175] 对于丝状真菌宿主细胞优选的终止子从如下酶的基因获得:构巢曲霉邻氨基苯甲酸合酶、黑曲霉葡糖淀粉酶、黑曲霉  $\alpha$ -葡糖苷酶、米曲霉 TAKA 淀粉酶和尖镰孢胰蛋白酶样蛋白酶。

[0176] 对于酵母宿主细胞优选的终止子从如下酶的基因获得:酿酒酵母烯醇化酶、酿酒酵母细胞色素 C(CYC1)和酿酒酵母甘油醛-3-磷酸脱氢酶(gpd)。对于酵母宿主细胞其它有用的终止子由 Romanos 等,1992, 见上文描述。

[0177] 调控序列还可以是合适的前导序列,当被转录时其为对于宿主细胞的翻译重要的 mRNA 非翻译区。前导序列可操作地连接于编码多肽的多核苷酸的5'-末端。可使用在所选宿主细胞中有功能的任何前导序列。

[0178] 对于丝状真菌宿主细胞优选的前导序列从如下酶的基因获得:米曲霉 TAKA 淀粉酶和构巢曲霉丙糖磷酸异构酶。

[0179] 对于酵母宿主细胞合适的前导序列从如下酶的基因获得:酿酒酵母烯醇化酶(ENO-1)、酿酒酵母3-磷酸甘油酸激酶、酿酒酵母  $\alpha$  因子和酿酒酵母醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH2/GAP)。

[0180] 调控序列也可以是聚腺苷酸化序列,其是与多核苷酸的3'末端可操作地连接的序列,并且在转录时,宿主细胞将其识别为将聚腺苷残基添加至转录的 mRNA 的信号。可使用在所选宿主细胞中有功能的任何聚腺苷酸化序列。

[0181] 对于丝状真菌宿主细胞优选的聚腺苷酸化序列从如下酶的基因获得:米曲霉 TAKA 淀粉酶、黑曲霉葡糖淀粉酶、构巢曲霉邻氨基苯甲酸合酶、尖镰孢胰蛋白酶样蛋白酶和黑曲霉  $\alpha$ -葡糖苷酶。

[0182] 对于酵母宿主细胞有用的聚腺苷酸化序列由 Guo 和 Sherman, 1995, *Mol. Cellular Biol.* 15:5983-5990 描述。

[0183] 调控序列还可以是信号肽编码区,其编码与多肽的 N 端相连的信号肽,并且指导多肽进入细胞的分泌途径。多核苷酸的编码序列5'端可固有地包含信号肽编码序列,其与

编码所述多肽的编码序列的区段一起天然地连接在翻译阅读框中。可供选择的是,编码序列 5' 端可含有对于所述编码序列外源的信号肽编码序列。外源信号肽编码序列在编码序列不天然地含有信号肽编码序列时可为必需的。或者,外源信号肽编码序列可以简单地取代天然信号肽编码序列以增强多肽的分泌。然而,可使用指导表达的多肽进入所选宿主细胞的分泌途径的任何信号肽编码序列。

[0184] 对于细菌宿主细胞有效的信号肽编码序列是从如下酶的基因获得的信号肽编码序列:芽孢杆菌属 NCIB11837 产麦芽糖淀粉酶、地衣芽孢杆菌枯草杆菌蛋白酶 (subtilisin)、地衣芽孢杆菌  $\beta$ -内酰胺酶、嗜热脂肪芽孢杆菌  $\alpha$ -淀粉酶、嗜热脂肪芽孢杆菌中性蛋白酶 (nprT, nprS, nprM) 和枯草芽孢杆菌 prsA。另外的信号肽由 Simonen 和 Palva, 1993, *Microbiological Reviews* 57:109-137 描述。

[0185] 对于丝状真菌宿主细胞有效的信号肽编码序列是从如下酶的基因获得的信号肽编码序列:黑曲霉中性淀粉酶、黑曲霉葡糖淀粉酶、米曲霉 TAKA 淀粉酶、特异腐质霉纤维素酶、特异腐质霉内切葡聚糖酶 V、疏棉状腐质霉脂肪酶和曼赫根毛霉天冬氨酸蛋白酶。

[0186] 对于酵母宿主细胞有用的信号肽从酿酒酵母  $\alpha$  因子和酿酒酵母转化酶的基因获得。其它有用的信号肽编码序列由 Romanos 等, 1992, 见上文描述。

[0187] 调控序列还可以是前肽编码序列,其编码位于多肽 N 端的前肽。所得多肽称为酶原 (proenzyme) 或前多肽 (propeptide) (或在某些情况下称为酶原 (zymogen))。前多肽通常是无活性的,并且能够通过前肽的催化或自催化切割从前多肽转化为活性多肽。可以从枯草芽孢杆菌碱性蛋白酶 (aprE), 枯草芽孢杆菌中性蛋白酶 (nprT)、嗜热毁丝霉漆酶 (W095/33836)、曼赫根毛霉天冬氨酸蛋白酶和酿酒酵母  $\alpha$  因子的基因获得前肽编码序列。

[0188] 当信号肽和前肽序列二者均出现在多肽的 N 端时,将前肽序列置于紧接着 (next to) 多肽 N 端,并且将信号肽序列置于紧接着前肽序列的 N 端。

[0189] 同样理想的是添加调节序列,其允许相对于宿主细胞的生长来调节多肽的表达。调节系统的实例是引起基因表达响应化学或物理刺激物,包括调节化合物的存在而开启或关闭的那些系统。原核系统中的调节系统包括 lac、tac 和 trp 操纵基因系统。在酵母中,可使用 ADH2 系统或 GAL1 系统。在丝状真菌中,可以使用黑曲霉葡糖淀粉酶启动子、米曲霉 TAKA  $\alpha$ -淀粉酶启动子和米曲霉葡糖淀粉酶启动子。调节序列的其它实例是那些允许基因扩增的序列。在真核系统中,这些调节序列包括在氨甲蝶呤 (methotrexate) 存在下扩增的二氢叶酸还原酶基因,和以重金属 (with heavy metal) 扩增的金属硫蛋白基因。在这些情况下,编码多肽的多核苷酸将与调节序列可操作地连接。

[0190] 表达载体

[0191] 本发明还涉及利用重组表达载体的重组宿主细胞和方法,所述重组表达载体包含编码碳酸氢根转运蛋白 (和 / 或编码 C4-二羧酸转运蛋白,编码苹果酸脱氢酶或丙酮酸羧化酶的多核苷酸) 的异源多核苷酸,以及启动子;和转录和翻译终止信号。此类重组表达载体可用于任何本文中所述的重组细胞和方法。多种核苷酸和调控序列可以结合在一起以产生重组表达载体,所述表达载体可以包括一个或多个 (几个) 方便的限制位点以允许在这些位点插入或取代编码多肽的多核苷酸。可供选择的是,可以通过在适当的用于表达的载体中插入包含所述序列的多核苷酸或核酸构建体来表达所述多核苷酸。在制备表达载体的过程中,将编码序列置于载体中,从而将该编码序列与适当的表达调控序列可操作地连接。

[0192] 重组表达载体可以是任何载体（例如，质粒或病毒），其能够方便地进行重组 DNA 步骤，并且能够产生多核苷酸的表达。载体的选择将通常依赖于载体与将引入该载体的宿主细胞的相容性。载体可以是线状或闭合环状质粒。

[0193] 在一个方面，每个编码本文中所述的碳酸氢根转运蛋白，苹果酸脱氢酶，和 / 或丙酮酸羧化酶的多核苷酸包含于各自独立的载体上。在一个方面，至少两个所述多核苷酸包含于一个载体上。在一个方面，至少三个所述多核苷酸包含于一个载体上。在一个方面，所有编码碳酸氢根转运蛋白，C4- 二羧酸转运蛋白，苹果酸脱氢酶，和 / 或丙酮酸羧化酶的多核苷酸包含于一个载体上。

[0194] 载体可以是自主复制载体，即，作为染色体外实体 (entity) 存在的载体，其复制独立于染色体复制，例如，质粒、染色体外元件、微型染色体 (minichromosome) 或人工染色体。载体可以含有任何用于确保自复制的手段 (means)。或者，载体可以是一种当被引入宿主细胞中时，整合到基因组中并且与整合了该载体的染色体一起复制的载体。此外，可以使用单独的载体或质粒或两个或更多个载体或质粒，其共同含有待引入宿主细胞基因组的完整 DNA (total DNA)，或可以使用转座子 (transposon)。

[0195] 载体优选含有一个或多个（几个）选择性标记，其允许简单选择经转化、转染、转导等的细胞。选择性标记是基因，其产物提供杀生物剂或病毒抗性、对重金属的抗性、对营养缺陷型的原养性 (prototrophy to auxotrophs) 等。

[0196] 细菌选择性标记的实例是来自枯草芽孢杆菌或地衣芽孢杆菌的 *dal* 基因，或赋予抗生素抗性的标记，所述抗生素抗性如氨苄青霉素、氯霉素、卡那霉素或四环素抗性。对于酵母宿主细胞合适的标记是 *ADE2*、*HIS3*、*LEU2*、*LYS2*、*MET3*、*TRP1* 和 *URA3*。用于丝状真菌宿主细胞的选择性标记包括但不限于 *amdS* (乙酰胺酶)、*argB* (鸟氨酸氨甲酰基转移酶)、*bar* (草铵膦 (phosphinothricin) 乙酰转移酶)、*hph* (潮霉素磷酸转移酶)、*niaD* (硝酸还原酶) (nitrate reductase)、*pyrG* (乳清酸核苷 -5'-磷酸脱羧酶) (orotidine-5'-phosphate decarboxylase)、*sC* (硫酸腺苷酰转移酶) 和 *trpC* (邻氨基苯甲酸合酶 (anthranilate synthase)) 以及它们的等同物。优选用在曲霉属细胞中的是构巢曲霉或米曲霉的 *amdS* 和 *pyrG* 基因和吸水链霉菌 (*Streptomyces hygroscopicus*) 的 *bar* 基因。

[0197] 载体优选含有允许载体整合入宿主细胞基因组或载体在细胞中独立于基因组的自主复制的元件。

[0198] 为了整合入宿主细胞基因组，载体可依赖编码多肽的多核苷酸的序列、或任何其他载体元件，来通过同源或非同源重组整合入基因组。或者，载体可以含有额外的多核苷酸，用于指导通过同源重组整合入宿主细胞基因组染色体中的精确位置。为了增加在精确位置整合的可能性，整合元件应含有足够数量的与相应的目标序列具有高度序列同一性的核酸，如 100 至 10,000 个碱基对，400 至 10,000 个碱基对，800 至 10,000 个碱基对，以增强同源重组的概率。整合元件可以是任何，其宿主细胞基因组中的目标序列同源的序列。此外，整合元件可以是非编码或编码的多核苷酸。另一方面，可以将载体通过非同源重组整合到宿主细胞的基因组中。

[0199] 为了自主复制，载体可以进一步包含复制起点。复制起点使载体能够在所述的宿主细胞中自主地复制。复制起点可以是在细胞中发挥作用，介导自主复制的任何质粒复制子 (replicator)。术语“复制起点”或“质粒复制子”意指能够使质粒或载体发生体内复制

的多核苷酸。

[0200] 细菌复制起点的实例有：允许在大肠杆菌中复制的质粒 pBR322、pUC19、pACYC177 和 pACYC184 的复制起点，和允许在芽孢杆菌属中复制的质粒 pUB110、pE194、pTA1060 和 pAM $\beta$ 1 的复制起点。

[0201] 用于酵母宿主细胞中的复制起点的实例有 2 微米复制起点，ARS1、ARS4、ARS1 和 CEN3 的组合，和 ARS4 和 CEN6 的组合。

[0202] 在丝状真菌细胞中有用的复制起点的实例有 AMA1 和 ANS1 (Gems 等, 1991, Gene 98:61-67; Cullen 等, 1987, Nucleic Acids Res. 15:9163-9175; W000/24883)。AMA1 基因的分离和包含该基因的质粒或载体的构建可以根据公开于 W000/24883 中的方法完成。

[0203] 可以将多于一个拷贝的本文中所述的多核苷酸插入宿主细胞以增加多肽的产生。多核苷酸拷贝数的增加可通过如下方法实现：使至少一个额外拷贝的序列整合入宿主细胞基因组，或使多核苷酸包含可扩增的选择性标记基因，其中可通过在合适的选择剂 (selectable agent) 存在下培养细胞来选出含有选择性标记基因的扩增拷贝，且由此选出含有多核苷酸的额外拷贝的细胞。

[0204] 用于连接上述元件以构建本发明的重组表达载体的方法是本领域技术人员熟知的 (参见, 例如, Sambrook 等, 1989, 见上文)。

[0205] 宿主细胞

[0206] 如本文中所述, 本发明还涉及重组宿主细胞, 其包含本文中所述的一种或几种多核苷酸, 所述多核苷酸可以可操作地连接于一个或多个 (几个) 调控序列, 所述调控序列指导一种或多种所述的多肽的表达, 用于重组产生 C4- 二羧酸转运蛋白。本发明还涵盖使用此类宿主细胞产生 C4- 二羧酸的方法。所述宿主细胞可包含任何一种或多种本文中所述的多核苷酸的组合。举例而言, 在一个方面, 所述重组宿主细胞包含编码碳酸氢根转运蛋白的异源多肽, 并任选地包含编码 C4- 二羧酸转运蛋白的异源多核苷酸, 苹果酸脱氢酶的异源多核苷酸, 和 / 或编码丙酮酸脱羧酶的异源多核苷酸; 其中所述宿主细胞与不含有所述编码碳酸氢根转运蛋白的异源多核苷酸的宿主细胞相比, 在相同条件下培养时, 产生 (或能够产生) 更多量的 C4- 二羧酸。

[0207] 在一个方面, 所述重组宿主细胞包含:

[0208] (1) 异源多核苷酸, 其编码碳酸氢根转运蛋白, 如选自下组的 C4 碳酸氢根转运蛋白: (a) 碳酸氢根转运蛋白, 其与 SEQ ID NO:2 或 SEQ ID NO:4 或其成熟多肽序列具有至少 60% 序列同一性; (b) 碳酸氢根转运蛋白, 其由多核苷酸编码, 所述多核苷酸在低严格条件下与以下杂交: (i) SEQ ID NO:1 或 3, 或其成熟多肽编码序列, (ii) SEQ ID NO:1 或 3 的 cDNA 序列, 或其成熟多肽编码序列, (iii) (i) 或 (ii) 的全长互补链; (c) 碳酸氢根转运蛋白, 其由多核苷酸编码, 所述多核苷酸与 (iv) SEQ ID NO:1 或 3, 或其成熟多肽编码序列, (v) SEQ ID NO:1 或 3 的 cDNA 序列, 或其成熟多肽编码序列, (vi) (iv) 或 (v) 的全长互补链具有至少 60% 序列同一性; (d) SEQ ID NO:2 或 SEQ ID NO:4 或其成熟多肽序列的包含一个或多个 (几个) 氨基酸的取代、缺失和 / 或插入的碳酸氢根转运蛋白变体; 和 (e) (a)、(b)、(c) 或 (d) 的多肽具有碳酸氢根转运蛋白活性的片段;

[0209] (2) 任选的异源第二多核苷酸, 其编码 C4- 二羧酸转运蛋白, 如选自下组的 C4- 二

羧酸转运蛋白：(a) C4- 二羧酸转运蛋白，其与 SEQ ID NO:6 或其成熟多肽序列具有至少 60% 序列同一性；(b) C4- 二羧酸转运蛋白，其由多核苷酸编码，所述多核苷酸在低严格条件下与 SEQ ID NO:5，或其成熟多肽编码序列，或前述序列的全长互补链杂交；(c) C4- 二羧酸转运蛋白，其由多核苷酸编码，所述多核苷酸与 SEQ ID NO:5，其成熟多肽编码序列，或前述序列的全长互补链具有至少 60% 序列同一性；(d) SEQ ID NO:6 或其成熟多肽序列的包含一个或多个氨基酸（例如两个、几个）的取代、缺失和 / 或插入的 C4- 二羧酸转运蛋白变体；和 (e) (a)、(b)、(c) 或 (d) 的多肽具有 C4- 二羧酸转运蛋白活性的片段；

[0210] (3) 任选的异源第三多核苷酸，其编码苹果酸脱氢酶，如选自下组的苹果酸脱氢酶：(a) 苹果酸脱氢酶，其与 SEQ ID NO:8 或其成熟多肽序列具有至少 60% 序列同一性；(b) 苹果酸脱氢酶，其由多核苷酸编码，所述多核苷酸在低严格条件下与以下杂交：(i) SEQ ID NO:7 或其成熟多肽编码序列，(ii) SEQ ID NO:7 的 cDNA 序列或其成熟多肽编码序列，或 (iii) (i) 或 (ii) 的全长互补链；(c) 苹果酸脱氢酶，其由多核苷酸编码，所述多核苷酸与以下具有至少 60% 序列同一性：(iv) SEQ ID NO:7 或其成熟多肽编码序列，(v) SEQ ID NO:7 的 cDNA 序列或其成熟多肽编码序列，或 (vi) (iv) 或 (v) 的全长互补链；(d) SEQ ID NO:8 或其成熟多肽序列的包含一个或多个氨基酸（几个）的取代、缺失和 / 或插入的苹果酸脱氢酶变体；和 (e) (a)、(b)、(c) 或 (d) 的多肽具有苹果酸脱氢酶活性的片段；和

[0211] (4) 任选的异源第四多核苷酸，其编码丙酮酸羧化酶，如选自下组的丙酮酸羧化酶：(a) 丙酮酸羧化酶，其与 SEQ ID NO:10 或其成熟多肽序列具有至少 60% 序列同一性；(b) 丙酮酸羧化酶，其由多核苷酸编码，所述多核苷酸在低严格条件下与以下杂交：(i) SEQ ID NO:9 或其成熟多肽编码序列，(ii) SEQ ID NO:9 的 cDNA 序列或其成熟多肽编码序列，或 (iii) (i) 或 (ii) 的全长互补链；(c) 丙酮酸羧化酶，其由多核苷酸编码，所述多核苷酸与以下具有至少 60% 序列同一性：(iv) SEQ ID NO:9 或其成熟多肽编码序列，(v) SEQ ID NO:9 的 cDNA 序列或其成熟多肽编码序列，或 (vi) (iv) 或 (v) 的全长互补链；(d) SEQ ID NO:10 或其成熟多肽序列的包含一个或多个氨基酸（几个）的取代、缺失和 / 或插入的丙酮酸羧化酶变体；和 (e) (a)、(b)、(c) 或 (d) 的多肽具有丙酮酸羧化酶活性的片段；

[0212] 其中所述宿主细胞与不含有所述一种或多种（几种）多核苷酸（例如不含有编码碳酸氢根转运蛋白的异源多核苷酸）的宿主细胞相比在相同条件下培养时产生（或能够产生）更多量的 C4- 二羧酸（例如苹果酸）。

[0213] 将包含一种或多种（几种）多核苷酸的构建体或载体（或多个构建体或载体）导入宿主细胞，使所述构建体或载体如前所述作为染色体整合体或者作为自复制的染色体外载体维持。术语“宿主细胞”涵盖任何亲本细胞的后代，其由于在复制中发生的突变而与亲本细胞不同。在一些情况下，宿主细胞的选择会很大程度上依赖于编码多肽的基因及其来源。下文中所述的方面适用于宿主细胞本身，以及使用宿主细胞的方法。

[0214] 所述宿主细胞可为任何能够重组产生本发明的多肽的细胞，例如原核细胞或真核细胞，和 / 或任何能够重组产生 C4- 二羧酸（例如苹果酸）的细胞。

[0215] 原核宿主细胞可以是任何革兰氏阳性或革兰氏阴性细菌。革兰氏阳性细菌包括但不限于，芽孢杆菌属、梭菌属、肠球菌属、地芽孢杆菌属、乳杆菌属、乳球菌属、海洋芽孢杆菌属、葡萄球菌属、链球菌属和链霉菌属。革兰氏阴性细菌包括但不限于，弯曲杆菌属、大肠杆菌、黄杆菌属、梭杆菌属、螺杆菌属、泥杆菌属、奈瑟氏菌属、假单胞菌属、沙门氏菌属和脲原

体属。

[0216] 细菌宿主细胞可以是任何芽孢杆菌属细胞,包括但不限于嗜碱芽孢杆菌、解淀粉芽孢杆菌、短芽孢杆菌、环状芽孢杆菌、克劳氏芽孢杆菌、凝结芽孢杆菌、坚强芽孢杆菌、灿烂芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、巨大芽孢杆菌、短小芽孢杆菌、嗜热脂肪芽孢杆菌、枯草芽孢杆菌和苏云金芽孢杆菌细胞。

[0217] 细菌宿主细胞还可以是任何链球菌属细胞,包括但不限于,似马链球菌、酿脓链球菌、乳房链球菌和马链球菌兽瘟亚种细胞。

[0218] 细菌宿主细胞还可以是任何链霉菌属细胞,包括但不限于,不产色链霉菌、除虫链霉菌、天蓝链霉菌、灰色链霉菌和浅青紫链霉菌细胞。

[0219] 可通过如下方法实现将 DNA 引入到芽孢杆菌属细胞:例如原生质体转化(参见,例如,Chang 和 Cohen, 1979, Mol. Gen. Genet. 168:111-115),使用感受态细胞(参见,例如,Young 和 Spizizen, 1961, J. Bacteriol. 81:823-829 或 Dubnau 和 Davidoff-Abelson, 1971, J. Mol. Biol. 56:209-221),电穿孔(参见,例如,Shigekawa 和 Dower, 1988, Biotechniques 6:742-751)或接合(参见,例如,Koehler 和 Thorne, 1987, J. Bacteriol. 169:5771-5278)。可通过如下方法实现将 DNA 引入到大肠杆菌细胞:例如原生质体转化(参见,例如,Hanahan, 1983, J. Mol. Biol. 166:557-580)或电穿孔(参见,例如,Dower 等, 1988, Nucleic Acids Res. 16:6127-6145)。可通过如下方法实现将 DNA 引入到链霉菌属细胞:例如原生质体转化和电穿孔(参见,例如,Gong 等, 2004, Folia Microbiol. (Praha) 49:399-405),接合(参见,例如,Mazodier 等, 1989, J. Bacteriol. 171:3583-3585),或转导(参见,例如,Burke 等, 2001, Proc. Natl. Acad. Sci. USA 98:6289-6294)。可通过如下方法实现将 DNA 引入到假单胞菌属细胞:例如电穿孔(参见,例如,Choi 等, 2006, J. Microbiol. Methods 64:391-397)或接合(参见,例如,Pinedo 和 Smets, 2005, Appl. Environ. Microbiol. 71:51-57)。可通过如下方法实现将 DNA 引入到链球菌属细胞:例如天然感受态(natural competence)(参见,例如,Perry 和 Kuramitsu, 1981, Infect. Immun. 32:1295-1297),原生质体转化(参见,例如,Catt 和 Jollick, 1991, Microbios. 68:189-207),电穿孔(参见,例如,Buckley 等, 1999, Appl. Environ. Microbiol. 65:3800-3804)或接合(参见,例如,Clewell, 1981, Microbiol. Rev. 45:409-436)。然而,可以使用本领域已知的将 DNA 引入宿主细胞的任何方法。

[0220] 宿主细胞还可以是真核生物,如哺乳动物、昆虫、植物或真菌细胞。

[0221] 所述宿主细胞可为真菌细胞。“真菌”用在本文包括以下门:子囊菌门(Ascomycota)、担子菌门(Basidiomycota)、壶菌门(Chytridiomycota)和接合菌门(Zygomycota)(如由 Hawksworth 等,于 Ainsworth and Bisby's Dictionary of The Fungi, 第 8 版, 1995, CAB International, University Press, Cambridge, UK 中所定义)以及卵菌门(Oomycota)(如 Hawksworth 等, 1995, 见上, 171 页中所引用)和所有有丝分裂孢子真菌(mitosporic fungi)(Hawksworth 等, 1995, 见上)。

[0222] 所述真菌宿主细胞可为酵母细胞。“酵母”用在本文包括产子囊酵母(ascosporogenous yeast)(内孢霉目(Endomycetales))、产担子酵母(basidiosporogenous yeast)和属于半知菌类(Fungi Imperfecti)(芽孢纲(Blastomycetes))的酵母。由于酵母的分类在未来可能改变,就本发明而言,将酵母定义为



如 *Biology and Activities of Yeast* (Skinner, F. A., Passmore, S. M., 和 Davenport, R. R. 编, Soc. App. Bacteriol. Symposium Series No. 9, 1980) 中所述。

[0223] 所述酵母宿主细胞可为假丝酵母属、汉逊酵母属 (*Hansenula*)、克鲁维酵母属、毕赤酵母属、酵母属、裂殖酵母属或西洋蓍霉属细胞, 如乳酸克鲁维酵母 (*Kluyveromyces lactis*)、卡尔酵母、酿酒酵母、糖化酵母、道格拉氏酵母、克鲁弗酵母、诺地酵母、卵形酵母、或解脂西洋蓍霉 (*Yarrowia lipolytica*) 细胞。

[0224] 所述真菌宿主细胞可为丝状真菌细胞。“丝状真菌”包括真菌门 (*Eumycota*) 和卵菌门的亚门 (如由 Hawksworth 等, 1995, 见上文, 所定义) 的所有丝状形式。丝状真菌通常的特征在于由壳多糖 (chitin)、纤维素、葡聚糖、壳聚糖 (chitosan)、甘露聚糖和其它复杂多糖组成的菌丝体壁。通过菌丝延伸进行营养生长, 而碳分解代谢是专性需氧的。相反, 酵母例如酿酒酵母的营养生长通过单细胞菌体的出芽生殖 (budding) 进行, 而碳分解代谢可以是发酵性的。

[0225] 所述丝状真菌宿主细胞可为枝顶孢霉属、曲霉属、短梗霉属、烟管霉属 (*Bjerkandera*)、拟蜡菌属、金孢子菌属、鬼伞属 (*Coprinus*)、革盖菌属 (*Coriolus*)、隐球菌属、*Filibasidium*、镰孢属、腐质霉属、梨孢菌属、毛霉属、毁丝霉属、新考玛脂霉属、脉孢菌属、拟青霉属、青霉属、平革菌属 (*Phanerochaete*)、射脉菌属 (*Phlebia*)、瘤胃壶菌属、侧耳属 (*Pleurotus*)、裂褶菌属、踝节菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属 (*Trametes*) 或木霉属细胞。

[0226] 所述丝状真菌宿主细胞可为棘孢曲霉、泡盛曲霉、烟曲霉、臭曲霉、日本曲霉、构巢曲霉、黑曲霉、米曲霉、黑刺烟管菌 (*Bjerkandera adusta*)、干拟蜡菌 (*Ceriporiopsis aneirina*)、*Ceriporiopsis caregiea*、*Ceriporiopsis gilvescens*、*Ceriporiopsis pannocinta*、*Ceriporiopsis rivulosa*、*Ceriporiopsis subrufa*、虫拟蜡菌 (*Ceriporiopsis subvermispora*)、*Chrysosporium inops*、嗜角质金孢子菌、*Chrysosporium lucknowense*、*Chrysosporium merdarium*、毡金孢子菌、*Chrysosporium queenslandicum*、热带金孢子菌、*Chrysosporium zonatum*、灰盖鬼伞 (*Coprinus cinereus*)、毛革盖菌 (*Coriolus hirsutus*)、杆孢状镰孢、禾谷镰孢、库威镰孢、大刀镰孢、禾本科镰孢、禾赤镰孢、异孢镰孢、合欢木镰孢、尖镰孢、多枝镰孢、粉红镰孢、接骨木镰孢、肤色镰孢、拟分枝孢镰孢、硫色镰孢、圆镰孢、拟丝孢镰孢、镶片镰孢、特异腐质霉、疏棉状腐质霉、米黑毛霉、嗜热毁丝霉、粗糙脉孢菌、产紫青霉、黄孢平革菌、辐射射脉菌 (*Phlebia radiata*)、刺芹侧耳 (*Pleurotus eryngii*)、土生梭孢霉、长绒毛栓菌 (*Trametes villosa*)、变色栓菌 (*Trametes versicolor*)、哈茨木霉、康宁木霉、长枝木霉、里氏木霉或绿色木霉细胞。

[0227] 在一个方面, 所述宿主细胞是曲霉属宿主细胞。在另一个方面, 所述宿主细胞是米曲霉。

[0228] 可以将真菌细胞通过涉及原生质体形成、原生质体转化和细胞壁再生的方法以本身公知的方式转化。用于转化曲霉属和木霉属宿主细胞的合适方法在 EP238023 和 Yelton 等, 1984, Proc. Natl. Acad. Sci. USA81:1470-1474 中描述。用于转化镰孢属菌种的合适方法由 Malardier 等, 1989, Gene78:147-156 和 W096/00787 描述。可以使用由如下文献描述的方法转化酵母: Becker 和 Guarente, 于 Abelson, J. N. 和 Simon, M. I. 编, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume194, pp182-187, Acade

mic Press, Inc., New York ;Ito等, 1983, J. Bacteriol. 153:163 ;和Hinnen等, 1978, Proc. Natl. Acad. Sci. USA75:1920。

[0229] 在一些方面,所述宿主细胞包含一种或多种(几种)本文中所述的多核苷酸,其中所述宿主细胞与不含有所述一种或多种(几种)多核苷酸的宿主细胞相比在相同条件下培养时分泌(或能够分泌)增加水平的C4-二羧酸。在一些方面,所述宿主细胞与不含有所述一种或多种(几种)多核苷酸(例如不含有编码碳酸氢根转运蛋白的异源多核苷酸)的宿主细胞相比,在相同条件下培养时,分泌(或能够分泌)的C4-二羧酸(例如苹果酸)增加至少5%,例如至少10%,至少15%,至少20%,至少25%,至少50%,至少100%,至少150%,至少200%,至少300%,或500%水平。

[0230] 在本文中所述的重组宿主细胞和方法的任何方面中,所述C4-二羧酸可为苹果酸、琥珀酸、草酰乙酸、丙二酸或延胡索酸,或其组合。在一些方面中,所述C4-二羧酸是苹果酸、琥珀酸或延胡索酸,或其组合。在一些方面中,所述C4-二羧酸是苹果酸或延胡索酸,或苹果酸和延胡索酸的组合。在一些方面中,所述C4-二羧酸是苹果酸。

[0231] 在任何这些方面,所述宿主细胞产生(和/或能够产生)高于理论值至少10%,例如至少20%,至少30%,至少40%,至少50%,至少60%,至少70%,至少80%或至少90%的产率。

[0232] 在任何这些方面,所述重组宿主具有大于约0.1g/L每小时,例如,大于约0.2g/L每小时,0.5g/L每小时,0.6g/L每小时,0.7g/L每小时,0.8g/L每小时,0.9g/L每小时,1.0g/L每小时,1.1g/L每小时,1.2g/L每小时,1.3g/L每小时,1.5g/L每小时,1.75g/L每小时,2.0g/L每小时,2.25g/L每小时,2.5g/L每小时,或3.0g/L每小时;或约0.1g/L每小时至约2.0g/L每小时,例如,约0.3g/L每小时至约1.7g/L每小时,约0.5g/L每小时至约1.5g/L每小时,约0.7g/L每小时至约1.3g/L每小时,约0.8g/L每小时至约1.2g/L每小时,或约0.9g/L每小时至约1.1g/L每小时的C4-二羧酸体积生产力(volumetric productivity)(例如苹果酸体积生产力)。

[0233] 可将所述重组宿主细胞在适于产生碳酸氢根转运蛋白、C4-二羧酸转运蛋白、苹果酸脱氢酶或丙酮酸羧化酶的营养培养基中使用本领域中公知的方法进行培养。例如,可以通过在合适培养基中在允许表达和/或分离合意的多肽的条件下进行的摇瓶培养或实验室或工业发酵罐中的小规模或大规模发酵(包括连续、分批、补料分批或固态发酵)来培养细胞。使用本领域已知的方法在合适的营养培养基中进行培养,所述营养培养基包含碳源和氮源和无机盐。合适的培养基能够从商业供应商获得或可以根据公开的组成制备(例如,在美国典型培养物保藏中心的目录中),或可从商业上可获得的成分制备。

[0234] 所述碳酸氢根转运蛋白、C4-二羧酸转运蛋白、苹果酸脱氢酶和丙酮酸羧化酶,及其活性,可使用本领域中公知的方法检测。这些检测方法可包括使用特异性抗体,酶产物的形成,或酶底物的消失。参见,例如 Sambrook等, *Molecular Cloning: A Laboratory Manual*, 第3版, Cold Spring Harbor Laboratory, New York (2001); Ausubel等, *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, MD (1999); 及 Hanai等, *Appl. Environ. Microbiol.* 73:7814-7818 (2007)。

[0235] 方法

[0236] 本发明亦涉及使用本文中所述的重组宿主细胞产生C4-二羧酸的方法。在一个方面,本发明涵盖产生C4-二羧酸(例如苹果酸)的方法,其包括:(a)在合适的条件下在

培养基中培养任一种本文中所述的重组宿主细胞（例如任何具有碳酸氢根转运蛋白活性，C4- 二羧酸转运蛋白活性，苹果酸脱氢酶活性，和 / 或丙酮酸羧化酶活性的宿主细胞）以产生 C4- 二羧酸；和 (b) 回收所述 C4- 二羧酸。在一个方面，本发明涵盖产生 C4- 二羧酸（例如苹果酸）的方法，其包括：(a) 在合适的条件下在培养基中培养任一种本文中所述的重组宿主细胞，其中所述宿主细胞包含编码碳酸氢根转运蛋白的异源多核苷酸，以及任选地，编码 C4- 二羧酸转运蛋白的异源多核苷酸，编码苹果酸脱氢酶的异源多核苷酸，和 / 或编码丙酮酸羧化酶的异源多核苷酸，以产生 C4- 二羧酸；和 (b) 回收所述 C4- 二羧酸。在一个方面，所述培养基是发酵培养基。

[0237] 在所述方法的一个方面，以大于约 10g/L，例如，大于约 25g/L, 50g/L, 75g/L, 100g/L, 125g/L, 150g/L, 160g/L, 170g/L, 180g/L, 190g/L, 200g/L, 210g/L, 225g/L, 250g/L, 275g/L, 300g/L, 325g/L, 350g/L, 400g/L, 或 500g/L；或约 10g/L 至约 500g/L，例如，约 50g/L 至约 350g/L, 约 100g/L 至约 300g/L, 约 150g/L 至约 250g/L, 约 175g/L 至约 225g/L, 或约 190g/L 至约 210g/L 的效价产生所述 C4- 二羧酸（例如苹果酸）。

[0238] 在所述方法的一个方面，产生的 C4- 二羧酸（例如苹果酸）的量在相同条件下培养时与不含有编码所述碳酸氢根转运蛋白的多核苷酸的宿主细胞相比，高至少 5%，例如至少 10%，至少 15%，至少 20%，至少 25%，至少 30%，至少 50%，或至少 100%。

[0239] 在所述方法的一个方面，所述 C4- 二羧酸选自下组：苹果酸、琥珀酸、草酰乙酸、丙二酸和延胡索酸。在一个方面，所述 C4- 二羧酸是苹果酸。

[0240] 所述重组的 C4- 二羧酸可任选地从发酵培养基使用任何本领域中已知的方法回收（参见例如，W01998/022611 和 U. S. 7, 601, 865），所述方法包括但不限于层析（例如大小排阻层析，吸附层析，离子交换层析），电泳方法，差示溶解度，渗透，蒸馏，提取（例如液液萃取），渗透蒸发，萃取性过滤（extractive filtration），膜过滤，膜分离，反相（reverse）或超滤。在一个实例中，所述 C4- 二羧酸从发酵培养基中的其它材料通过过滤来回收。

[0241] 在所述方法的一些方面，在任选地纯化之前和 / 或之后的重组 C4- 二羧酸是基本上纯的。对于产生 C4- 二羧酸（或其具体 C4- 二羧酸，如苹果酸）的方法，“基本上纯的”意指含有不超过 15% 杂质的回收的 C4- 二羧酸制备物，其中杂质已知除了 C4- 二羧酸以外的化合物。在一种变化中，提供基本上纯的 C4- 二羧酸制备物，其中所述制备物含有不超过 25% 杂质，或不超过 20% 杂质，或不超过 10% 杂质，或不超过 5% 杂质，或不超过 3% 杂质，或不超过 1% 杂质，或不超过 0.5% 杂质。

[0242] 可使用本领域中已知的方法进行对本文中所述的产生方法和宿主细胞合适的测试 C4- 二羧酸的产生的测定法。举例而言，最终 C4- 二羧酸产物（例如苹果酸）和其它有机化合物，可通过如 HPLC（高效液相色谱），GC-MS（气相色谱 - 质谱），和 LC-MS（液相色谱 - 质谱）的方法或其它合适的分析方法，用本领域中公知的常规步骤进行分析。发酵液中 C4- 二羧酸的释放亦可用培养上清进行测试。发酵培养基中的副产物和残余糖（例如葡萄糖）可通过 HPLC，例如使用供葡萄糖和醇类用的折射率检测器，和供有机酸用的 UV 检测器（Lin 等，Biotechnol. Bioeng. 90:775-779(2005)）的 HPLC，或其它本领域中公知的合适的测定和检测方法来定量。

[0243] 本发明进一步通过下述实施例描述，其不应视为对本发明范围的限制。

## 实施例

[0244] 用作缓冲液和底物是至少试剂级的商品。

[0245] 真菌菌株

[0246] 使用米曲霉 NRRL3488 (或 ATCC56747) 作为碳酸氢根转运蛋白基因 (bt1) 丙酮酸羧化酶基因 (pyc)、苹果酸脱氢酶基因 (mdh3) 的来源,并用于产生 C4- 二羧酸。棘孢曲霉用作 C4- 二羧酸转运蛋白基因 (c4t521) 的来源。

[0247] 培养基

[0248] YEG 培养基组成为:20g 葡萄糖,5g 酵母提取物,和去离子水加至 1 升。

[0249] COVE 平板组成为:1M 蔗糖,2%COVE 盐溶液,10mM 乙酰胺,15mM CsCl,和 25g/1 高贵级琼脂 (Agar Noble)。

[0250] COVE 盐溶液组成为:26g KCl,26g MgSO<sub>4</sub> · 7H<sub>2</sub>O,76g KH<sub>2</sub>PO<sub>4</sub>,50ml COVE 痕量元素溶液,和去离子水加至 1 升。

[0251] COVE 痕量元素溶液组成为:0.04g Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub> · 10H<sub>2</sub>O,0.04g CuSO<sub>4</sub> · 5H<sub>2</sub>O,1.2g FeSO<sub>4</sub> · 7H<sub>2</sub>O,0.7g MnSO<sub>4</sub> · H<sub>2</sub>O,0.8g Na<sub>2</sub>MoO<sub>2</sub> · 2H<sub>2</sub>O,10g ZnSO<sub>4</sub> · 7H<sub>2</sub>O 和去离子水加至 1 升。

[0252] 种子培养基组成为:40g 葡萄糖,6g 细菌蛋白胨 (Bacto peptone),750mg KH<sub>2</sub>PO<sub>4</sub>,750mg K<sub>2</sub>HPO<sub>4</sub>,100mg MgSO<sub>4</sub> · 7H<sub>2</sub>O,100mg CaCl<sub>2</sub> · H<sub>2</sub>O,5mg FeSO<sub>4</sub> · 7H<sub>2</sub>O,5mg NaCl,和去离子水加至 1 升。

[0253] 种子培养基 B 组成为:30g 葡萄糖,3g 细菌蛋白胨,560mg KH<sub>2</sub>PO<sub>4</sub>,560mg K<sub>2</sub>HPO<sub>4</sub>,925mg NaH<sub>2</sub>PO<sub>4</sub> · H<sub>2</sub>O,820mg Na<sub>2</sub>HPO<sub>4</sub>,75mg MgSO<sub>4</sub> · 7H<sub>2</sub>O,75mg CaCl<sub>2</sub> · H<sub>2</sub>O,0.75ml 的 1000X 微量营养物溶液 (Micronutrient Solution),和去离子水加至 1 升。

[0254] 产酸培养基 C 组成为:100g 葡萄糖,80g CaCO<sub>3</sub>,6g 细菌蛋白胨,150mg KH<sub>2</sub>PO<sub>4</sub>,150mg K<sub>2</sub>HPO<sub>4</sub>,100mg MgSO<sub>4</sub> · 7H<sub>2</sub>O,100mg CaCl<sub>2</sub> · H<sub>2</sub>O,1ml1000X 微量营养物溶液,和去离子水加至 1 升。

[0255] 发酵罐分批培养基组成为:60g 葡萄糖,120g CaCO<sub>3</sub>,9g Bacto-peptone,150mg KH<sub>2</sub>PO<sub>4</sub>,150mg K<sub>2</sub>HPO<sub>4</sub>,100mg MgSO<sub>4</sub> · 7H<sub>2</sub>O,100mg CaCl<sub>2</sub> · 2H<sub>2</sub>O,5mg FeSO<sub>4</sub> · 7H<sub>2</sub>O,5mg NaCl,5ml Pluronic L61,和去离子水加至 1 升。

[0256] 1000X 微量营养物溶液组成为:5g NaCl,5g FeSO<sub>4</sub> · 7H<sub>2</sub>O,1g 柠檬酸,和去离子水加至 1 升。

[0257] PDA 平板组成为:39g/1 马铃薯右旋糖琼脂。

[0258] 2XYT+amp 平板组成为 16g 胰蛋白胨,10g 酵母提取物,5g NaCl,100mg 氨苄青霉素,15g 细菌琼脂 (Bacto agar),和去离子水加至 1 升。

[0259] 实施例 1:米曲霉碳酸氢根转运蛋白基因 (bt1) 的克隆和表达载体 pAmFs69 的构建

[0260] 通过 PCR 扩增从米曲霉 NRRL3488 基因组 DNA 克隆碳酸氢根转运蛋白基因 bt1 (A0090012000782),PCR 扩增使用与在已公开的米曲霉 ATCC42149 基因组序列 (Galagan 等,2005,Nature438:1105-1115) 中发现的预测的米曲霉碳酸氢根转运蛋白基因模型号 A0090012000782 同源的引物。

[0261] 来自米曲霉 NRRL3488 的基因组 DNA 的分离通过下述进行:用 2X10<sup>6</sup> 个孢子接种摇瓶中的 100ml YEG 培养基,并在 37°C 200rpm 振荡下将烧瓶温育过夜。将菌丝体收集在衬

有MIRACLOTH® (Calbiochem, San Diego, CA, USA) 的漏斗中,并将大约 2 克的组织冻结于液氮。在冷的杵和研钵中研磨来破坏菌丝体。使用DNeasy®植物大提试剂盒 (Plant Maxi Kit) (QIAGEN Inc., Valencia, CA, USA), 根据生产商的指示从粉末化的菌丝体分离基因组 DNA。使用下示的正向引物 069824 和反向引物 069825 扩增米曲霉 bt1 基因:

[0262] 引物 069824:

[0263] 5' -GTGATAGAACATCGTCCATAATGGAATCCAGCGCTGTACA-3' (SEQ ID NO:11)

[0264] 引物 069825:

[0265] 5' -GTGTCAGTCACCTCTAGTTAAtcagatttcaatctcgtctt-3' (SEQ ID NO:12)

[0266] 扩增反应使用 Phusion® Hot Start 高保真 DNA 聚合酶 (Hot Start High-Fidelity DNA Polymerase) (Finnzymes OY, Finland) 根据生产商的指示进行。每个 PCR 反应体系包含 47ng 米曲霉 NRRL3488 基因组 DNA, dNTPs 各 200 μM, 50pM 正向引物, 50pM 反向引物, 1X Phusion® GC Buffer 反应缓冲液 (Finnzymes OY, Finland), 和 50 个单位的 Phusion® 热启动高保真 DNA 聚合酶, 终体积为 50 μl。将扩增反应体系在 EPPENDORF® MASTERCYCLER® (Eppendorf Scientific Inc., Westbury, NY, USA) 中温育, 其编程如下: 1 个循环, 在 98°C 进行 30 秒; 35 个循环, 每个在 98°C 进行 10 秒, 66°C 进行 30 秒, 和 72°C 进行 2.5 分钟; 和 1 个循环, 在 72°C 进行 10 分钟。PCR 产物通过 50mM Tris 碱 -50mM 乙酸 -0.1mM EDTA 二钠盐 (TAE) 中的 1% 琼脂糖凝胶电泳进行纯化。从凝胶切出大约 2.5kb 的片段, 并使用 QIAQUICK® 凝胶提取试剂盒 (QIAGEN Inc., Valencia, CA, USA) 从琼脂糖提取出来。

[0267] 将质粒 pShTh60 (图 1, 亦参见 2010 年 8 月 27 日提交的 PCR 申请号 PCT/US10/47002) 用 Sex AI 和 Pac I 消化, 通过在 TBE 缓冲液 (10.8g/L Tris 碱, 5.5g/L 硼酸, 2mM EDTA, pH8.0) 中的 0.8% 琼脂糖凝胶电泳分离, 并使用 QIAQUICK® 凝胶提取试剂盒 (QIAGEN Inc., Valencia, CA, USA) 纯化。然后使用 In-Fusion™ Advantage 反应试剂盒 (Clontech, Mountain View, CA, USA) 根据生产商的指示将上述纯化的 PCR 产物插入经消化的 pShTh60 片段, 得到质粒 pAmFs69 (图 2)。

[0268] 将 2.5 μl 等份的连接反应体系根据生产商的指示转化至 ONE SHOT® TOP10 化学感受态大肠杆菌细胞。将转化体铺板于 2XYT+amp 平板上并在 37°C 温育过夜。

[0269] 对所得的转化体使用 DNA 序列分析以确认 bt1 编码序列的完整性。使用下示的引物 610849, 610851, 610853, 610855, 610857, 610859, 和 610861, 以及 ABI3130XL DNA 分析仪 (Applied Biosystems, Inc., Foster City, CA, USA) 和带染料终止物化学的引物步移技术 (Giesecke 等, 1992, J. Virol. Methods 38:47-60)。

[0270] 引物 610849 :5' -GAACAGGAAGAAATCCAAAA-3' (SEQ ID NO:13)

[0271] 引物 610851 :5' -GTCGGCATAGCCACTGCAAT-3' (SEQ ID NO:14)

[0272] 引物 610853 :5' -TGTTGCCGCCAAGGGACTTA-3' (SEQ ID NO:15)

[0273] 引物 610855 :5' -CCGAGAGCGTTGAGTTAATC-3' (SEQ ID NO:16)

[0274] 引物 610857 :5' -AGCATTAGGGCTAGCTCCGT-3' (SEQ ID NO:17)

[0275] 引物 610859 :5' -CCAAGATGCCATGTCAGGAC-3' (SEQ ID NO:18)

[0276] 引物 610861 :5' -TCACAAAAGAGTAGAGGCCA-3' (SEQ ID NO:19)

[0277] 米曲霉 bt1 基因的基因组 DNA 序列的核苷酸构建体 (SEQ ID NO:1) 和推导的氨基酸序列 (SEQ ID NO:2) 示于图 3A 和 3B。2503bp 的基因组编码序列 (包含一个终止密码子) 由三个 78bp (465-542), 51bp (1173-1223), 和 61bp (1747-1807) 的内含子分隔。相应的 cDNA 序列 (图 3A 和 3B 中所示的粗体核苷酸序列) 为 2313bp, 包含一个终止密码子。预测的编码蛋白为 770 个氨基酸, 具有 83.9kDa 的预测的分子量和 6.9 的等电点 pH。

[0278] 实施例 2 :米曲霉碳酸氢根转运蛋白基因 A0090003000798 的克隆和相应的表达载体的构建

[0279] 通过 PCR 扩增从米曲霉 NRRL3488 基因组 DNA 克隆碳酸氢根转运蛋白基因 bt2 (A0090003000798), PCR 扩增使用与在已公布的米曲霉 ATCC42149 基因组序列 (Galagan 等, 2005, 见上) 中发现的米曲霉预测的曲霉碳酸氢根转运蛋白基因模型号 A0090003000798 同源的引物。

[0280] 来自米曲霉 NRRL3488 的基因组 DNA 的分离和菌丝体的收获和处理如实施例 1 中所述。使用下示的正向引物 0614058 和反向引物 0614057 扩增米曲霉 bt2 基因:

[0281] 引物 0614058:

[0282] 5' -gtgatagaacatcgctccataatgccggggcgatctcaaaacc-3' (SEQ ID NO:52)

[0283] 引物 0614057:

[0284] 5' -gtgtcagtcacctctagttactatgcatcaaggacattc-3' (SEQ ID NO:53)

[0285] 扩增反应使用 **Phusion®** Hot Start 高保真 DNA 聚合酶 (Finnzymes) 根据生产商的指示进行。每个 PCR 反应体系包含 47ng 米曲霉 NRRL3488 基因组 DNA, dNTPs 各 200  $\mu$  M, 50pM 正向引物, 50pM 反向引物, 1X **Phusion®** GC Buffer 反应缓冲液 (Finnzymes), 和 50 个单位的 **Phusion®** Hot Start 高保真 DNA 聚合酶, 终体积为 50  $\mu$  l。将扩增反应体系在 **EPPENDORF® MASTERCYCLER®** (Eppendorf Scientific Inc.) 中温育, 其编程如下: 1 个循环, 在 98°C 进行 2 分钟; 35 个循环, 每个在 98°C 进行 15 秒, 65°C 进行 15 秒, 和 74°C 进行 1 分钟; 和 1 个循环, 在 74°C 进行 1 分钟。将 PCR 产物通过 50mM Tris 碱 -50mM 乙酸 -0.1mM EDTA 二钠盐 (TAE) 中的 1% 琼脂糖凝胶电泳进行纯化。将大约 2.7kb 的片段从凝胶切出, 并使用 **QIAQUICK®** 凝胶提取试剂盒 (QIAGEN Inc.) 从琼脂糖提取出来。

[0286] 将质粒 pShTh77 (图 11) 用 Sex AI 和 Pac I 消化, 通过在 TBE 缓冲液 (10.8g/L Tris 碱, 5.5g/L 硼酸, 2mM EDTA, pH8.0) 中的 0.8% 琼脂糖凝胶电泳分离, 并使用 **QIAQUICK®** 凝胶提取试剂盒 (QIAGEN Inc., Valencia, CA, USA) 纯化。然后将上述纯化的 PCR 产物使用 In-Fusion™ Advantage 反应试剂盒 (Clontech) 根据生产商的指示插入消化的 pShTh77 片段, 得到质粒 pShTh147 (图 12)。

[0287] 将 2.5  $\mu$  l 等份的连接反应体系根据生产商的指示转化至 ONE **SHOT®** TOP10 化学感受态大肠杆菌细胞。将转化体铺板于 2XYT+amp 平板上并在 37°C 温育过夜。

[0288] 对所得的转化体使用 DNA 序列分析以确认 bt2 编码序列的完整性。使用下示的引物 0614313, 0614314, 996270, 和 0611428, 以及 ABI3130XL DNA 分析仪 (Applied Biosystems, Inc., Foster City, CA, USA) 和带染料终止子化学的引物步移技术 (Giesecke

等, 1992, J. Virol. Methods 38:47-60)。

[0289] 引物 0614313 :5' -gattgagatcggcatttact-3' (SEQ ID NO:54)

[0290] 引物 0614314 :5' -acgcggaacagcagaatggc-3' (SEQ ID NO:55)

[0291] 引物 996270 :5' -CTATAGCGAAATGGATTGATTGTCT-3' (SEQ ID NO:56)

[0292] 引物 0611428 :5' -TTCACCGTGAAACGTATTGA-3' (SEQ ID NO:57)

[0293] 米曲霉 bt1 基因的基因组 DNA 序列的核苷酸构建体 (SEQ ID NO:3) 和推导的氨基酸序列 (SEQ ID NO:4) 示于图 4A 和 4B。2657bp 的基因组编码序列 (包含终止密码子) 由两个 64bp (302-365) 和 61bp (512-572) 的内含子分隔。相应的 cDNA 序列 (图 4A 和 4B 中所示的粗体核苷酸序列) 为 2532bp, 包含一个终止密码子。预测的编码蛋白为 843 个氨基酸, 具有 92.5kDa 的预测的分子量和 8.4 的等电点 pH。

[0294] 实施例 3 :棘孢曲霉 C4- 二羧酸转运蛋白基因的克隆和载体 pSaMF36 的构建

[0295] 通过用  $2 \times 10^6$  个孢子接种摇瓶中的 100ml YEG 培养基, 并在 34°C 在 160rpm 振荡下将烧瓶温育过夜来分离来自棘孢曲霉的基因组 DNA。通过使用衬有 **MIRACLOTH®** (Calbiochem, San Diego, CA, USA) 的漏斗过滤来收获菌丝体, 并回收大约 2 克的菌丝体并冻结于液氮。在冻结的菌丝体包裹在 **MIRACLOTH®** 内时, 迅速用锤子敲碎来将其打断。然后将经打断的菌丝体转移至 50ml 聚丙烯锥形离心管中, 离心管装有 10ml 的 1X 裂解缓冲液 (100mM EDTA, 10mM Tris pH8.0, 1% Triton® X-100, 0.5M 盐酸胍, 200mM NaCl) 和 3  $\mu$ l 的 RNase A (QIAGEN Inc., Valencia, CA, USA, 100mg/ml)。通过轻柔地涡旋试管来混合, 然后在室温温育 5 分钟, 接着添加 150  $\mu$ l 蛋白酶 K (QIAGEN Inc., Valencia, CA, USA; 20mg/ml)。通过倒置试管来混合, 并在 50°C 温育试管 1 小时。然后将试管在 7240x g 离心 20 分钟。然后将上清加到经预平衡的 QIAGEN-tip100 (QIAGEN Inc., Valencia, CA, USA) 上, 根据生产商的指示来进行其余的 DNA 提取步骤。将 DNA 重悬于 100  $\mu$ l TE 缓冲液 (10mM Tris 碱, 1mM EDTA, pH8.0)。

[0296] 使用下示的引物 069700 和 069701 从分离的棘孢曲霉基因组 DNA 扩增了 1257bp 的 C4- 二羧酸转运蛋白基因 c4t521。

[0297] 引物 069700 :

[0298] 5' -TGTGATAGAACATCGTCCATAATGCACGACCACAGC-3' (SEQ ID NO:20)

[0299] 引物 069701 :

[0300] 5' -GTGTCAGTCACCTCTAGTTATCATTCGAACAACCTCGGACA-3' (SEQ ID NO:21)

[0301] PCR 反应体系由 10  $\mu$ l 15X 反应缓冲液, 1  $\mu$ l 棘孢曲霉基因组 DNA 模板 (105ng/ $\mu$ l), 1  $\mu$ l 引物 069700 (100ng/ $\mu$ l), 1  $\mu$ l 引物 069701 (100ng/ $\mu$ l), 1  $\mu$ l dNTP 混合物 (10mM), 35.5  $\mu$ l 去离子水, 和 0.5  $\mu$ l Phusion™ Hot Start 高保真聚合酶 (Finnzymes, Inc, Massachusetts, USA) 构成。将扩增反应体系在 **EPPENDORF® MASTERCYCLER®** 中温育, 其编程如下 :1 个循环, 在 98°C 进行 30 秒 ;30 个循环, 每个在 98°C 进行 10 秒, 60°C 进行 30 秒, 72°C 进行 1 分钟, 和 1 个循环, 在 72°C 进行 10 分钟。将 PCR 产物用 Dpn I 消化 1 小时以降解任何质粒 DNA 模板。

[0302] 将质粒 pShTh60 (图 1) 用 Sex AI 和 Pac I 消化, 通过在 TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳分离, 并使用 **QIAQUICK®** 凝胶提取试剂盒纯化。然后使用 In-Fusion™

Advantage 反应试剂盒将上述纯化的 PCR 产物插入该经消化的 pShTh60 片段, 组成为: 2  $\mu$ l 15X 缓冲液、3  $\mu$ l 纯化的 PCR 产物 (26ng/ $\mu$ l)、1.5  $\mu$ l 凝胶纯化的 Sex AI 和 Pac I 消化的和凝胶纯化的 pShTh60 (132ng/ $\mu$ l)、1  $\mu$ l In-Fusion<sup>TM</sup> 酶和 2.5  $\mu$ l 去离子水。将反应体系在 37°C 温育 15 分钟, 50°C 温育 15 分钟, 置于冰上 5 分钟, 并用 40  $\mu$ l TE 缓冲液稀释, 得到 pSaMF36 (图 5)。

[0303] 将 2.5  $\mu$ l 等份的连接反应体系根据生产商的指示转化至 ONE SHOT<sup>®</sup> TOP10 化学感受态大肠杆菌细胞。将转化体铺板于 2XYT+amp 平板上并在 37°C 温育过夜。挑取所得的转化体对其进行 DNA 测序以确认 mat521 基因成功地整合入载体。

[0304] 棘孢曲霉 c4t521 基因的基因组 DNA 序列的核苷酸构建体 (SEQ ID NO:5) 和推导的氨基酸序列 (SEQ ID NO:6) 示于图 6。1257bp 的基因组编码序列 (包括终止密码子) 不含内含子。预测的编码蛋白为 418 个氨基酸, 具有 46.8kDa 的预测的分子量和 6.36 的等电点 pH。使用 SignalP 程序 (Nielsen 等, 1997, Protein Engineering 10:1-6), 预测了一条 17 个残基的信号肽。基于该程序, 预测的成熟蛋白含有 401 个氨基酸, 具有 44.9kDa 的预测的分子量和 6.89 的等电点 pH。

[0305] 实施例 4: 米曲霉苹果酸脱氢酶基因的克隆和表达载体 pSaMF21 的构建

[0306] 构建了质粒 pSaMF21, 该质粒含有 NAD 依存性苹果酸脱氢酶 (mdh3) 基因序列 (DOGAN:A0090701000013), 其是来自米曲霉的一个 1430bp 片段, 如 2010 年 8 月 27 日提交的 PCT 申请号 PCT/US10/47002 中所述。米曲霉 NRRL3488 苹果酸脱氢酶 mdh3 基因的基因组 DNA 序列的核苷酸构建体 (SEQ ID NO:7) 和推导的氨基酸序列 (SEQ ID NO:8) 示于图 7。1430bp 的基因组编码序列 (包括终止密码子) 被 57bp (14-70bp), 70bp (103-172bp), 74bp (284-357bp), 68bp (446-513bp), 58bp (892-949bp), 48bp (1035-1082bp), 和 62bp (1228-1289bp) 的 7 个内含子分隔开来。mdh3 基因的编码区的 G+C 含量为 50.3%。对应的 cDNA 序列 (图 7 中所示的粗体核苷酸序列) 为 993 个 bp, 包含一个终止密码子。预测的编码蛋白为 330 个氨基酸, 具有 34.5kDa 的预测的分子量, 和 6.79 的等电点 pH。

[0307] 简言之, 该质粒是通过用 Sex AI 和 Pac I 进行的限制性消化使 pShTh60 直链化 (图 1) 而构建的。将经消化的载体通过 TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳分离, 并使用 QIAQUICK<sup>®</sup> 凝胶提取试剂盒纯化。使用引物 067522 和 067525 从 pShTh71 (2010 年 8 月 27 日提交的 PCT 申请号 PCT/US10/47002) 扩增 mdh3 基因。

[0308] 引物 067522:

[0309] 5'-AGAACATCGTCCATAATGGTCAAAGCTGGTGAGTTA-3' (SEQ ID NO:22)

[0310] 引物 067525:

[0311] 5'-GTGTCAGTCACCTCTAGTTATTACTTTGGTGGTGGTTCT-3' (SEQ ID NO:23)

[0312] PCR 反应体系由 5  $\mu$ l 10X 反应缓冲液, 1  $\mu$ l pShTh71 模板 (87ng/ $\mu$ l), 1  $\mu$ l 引物 067522 (100ng/ $\mu$ l), 1  $\mu$ l 引物 067525 (100ng/ $\mu$ l), 1  $\mu$ l dNTP 混合物 (10mM), 45.5  $\mu$ l 去离子水, 和 0.5  $\mu$ l **Herculase<sup>®</sup>** HotStart DNA 聚合酶 (Stratagene, La Jolla, CA, USA) 构成。将扩增反应体系在 **EPPENDORF<sup>®</sup> MASTERCYCLER<sup>®</sup>** 中温育, 其编程为: 1 个循环, 在 95°C 进行 2 分钟; 10 个循环, 每循环在 95°C 进行 10 秒, 58°C 进行 30 秒, 和 72°C 进行 1.5 分钟; 20 个循环, 每循环在 95°C 进行 10 秒, 50°C 进行 30 秒, 和 72°C 进行 1.5 分钟且每



个循环增加 10 秒。对 PCR 反应体系用 Dpn I 进行限制性消化 1 小时,以降解任何质粒 DNA 模板。然后使用 MinElute® PCR 纯化试剂盒 (QIAGEN Inc., Valencia, CA, USA) 纯化 PCR 产物。将纯化的 PCR 产物使用 In-Fusion™ Advantage 反应插入载体,其中所述反应体系由 2  $\mu$ l 15X 缓冲液,0.5  $\mu$ l 纯化的 PCR 产物 (110ng/ $\mu$ l),1.7  $\mu$ l 凝胶纯化的经 Sex AI 和 Pac I 限制性消化的 pShTh60 (图 1;78ng/ $\mu$ l),1  $\mu$ l In-Fusion™ 酶和 4.8  $\mu$ l 去离子水构成。将反应体系在 37°C 温育 15 分钟,接着在 50°C 温育 15 分钟,之后将其置于冰上 5 分钟,并用 40  $\mu$ l TE 缓冲液稀释,得到 pSaMF21 (图 8)。将 2  $\mu$ l 等份的连接反应体系依照生产商的指示转化入 ONE SHOT® TOP10 化学感受态大肠杆菌细胞 (Invitrogen, San Diego, CA, USA)。将转化体铺板于 2XYT+amp 平板上,并在 37°C 温育过夜。挑取所得的转化体,并对其进行 DNA 测序以确认 mdh3 基因成功地整合入载体。

[0313] 实施例 5:米曲霉丙酮酸羧化酶基因的克隆和表达载体 pRyan1 的构建

[0314] 构建了质粒 pRyan1,该质粒含有丙酮酸羧化酶 (pyc) 基因序列 (DOGAN:A0090023000801),其是来自米曲霉的 3646bp 片段 (包含两个终止密码子),如 2010 年 8 月 27 日提交的 PCT 申请号 PCT/US10/47002 中所述。米曲霉丙酮酸羧化酶基因的基因组 DNA 序列的核苷酸构建体 (SEQ ID NO:9) 和推导的氨基酸序列 (SEQ ID NO:10) 示于图 9A 和 9B。米曲霉 NRRL3488 和 ATCC56747 丙酮酸羧化酶基因具有相同的核苷酸序列。基因编码区的 G+C 含量是 57.1%。3643bp 的基因组编码序列 (包含一个终止密码子) 被 1 个 61bp (3475-3535bp) 的内含子分隔。基因编码区的 G+C 含量是 57.1%。相应的 cDNA 序列 (图 9A 和 9B 中所示的粗体核苷酸序列) 为 3582bp,包含一个终止密码子。预测的编码蛋白为 1193 个氨基酸,具有 131kDa 的预测分子量。

[0315] 简言之,该质粒是通过用 Sex AI 和 Pac I 进行的限制性消化使 pShTh60 直链化 (图 1) 而构建的。消化的载体通过 TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳分离,并使用 QIAQUICK® 凝胶提取试剂盒纯化。使用如下所示的引物 066549 和 067388 从米曲霉 NRRL3488 基因组 DNA 扩增 pyc 基因。

[0316] 引物 066549:

[0317] 5' -TAGAACATCGTCCATAATGGCGGCTCCGTTTCGTCA-3' (SEQ ID NO:24)

[0318] 引物 067388:

[0319] 5' -GTGTCAGTCACCTCTAGTTATTATTACGCTTTGACGATCT-3' (SEQ ID NO:25)

[0320] PCR 反应体系由 5  $\mu$ l 110X 反应缓冲液,1  $\mu$ l 米曲霉 NRRL3488 基因组 DNA (110ng/ $\mu$ l),1  $\mu$ l 引物 066549 (100ng/ $\mu$ l),1  $\mu$ l 引物 067388 (100ng/ $\mu$ l),1  $\mu$ l dNTP 混合物 (10mM),45.5  $\mu$ l 去离子水,和 0.5  $\mu$ l Herculase® HotStart DNA 聚合酶构成。将扩增反应体系在 EPPENDORF® MASTERCYCLER® 中温育,其编程为:1 个循环,在 95°C 进行 2 分钟;10 个循环,每循环在 95°C 进行 10 秒,58°C 进行 30 秒,和 72°C 进行 3.5 分钟;20 个循环,每循环在 95°C 进行 10 秒,58°C 进行 30 秒,和 72°C 进行 3.5 分钟且每个循环增加 10 秒。然后使用 MinElute® PCR 纯化试剂盒纯化 PCR 产物。

[0321] 使用 In-Fusion™ Advantage 反应物将纯化的 PCR 产物插入载体,其中反应体系由 2  $\mu$ l 15X 缓冲液,1  $\mu$ l 纯化的 PCR 产物 (144ng/ $\mu$ l),2  $\mu$ l 凝胶纯化的、Sex AI 和 Pac I 限制性消化的 pShTh60 (图 1;78ng/ $\mu$ l),1  $\mu$ l In-Fusion™ 酶和 4  $\mu$ l 去离子水构成。将反应

体系在 37°C 温育 15 分钟,接着在 50°C 温育 15 分钟,之后将其置于冰上 5 分钟,并用 40  $\mu$  l TE 缓冲液稀释,得到 pRYAN1(图 10)。将 2  $\mu$  l 等份的连接反应体系依照生产商的指示转化入 ONE SHOT® TOP10 化学感受态大肠杆菌细胞。将转化体铺板于 2XYT+amp 平板上,并在 37°C 温育过夜。挑取所得的转化体,并对其进行 DNA 测序以确认 pyc 基因成功地整合入载体。核苷酸 1308 从 C 变为 T,但并不影响蛋白序列。

[0322] 实施例 6:将 pAmFs69, pRyan1, pSaMf21, pSaMf36 表达载体片段转化入米曲霉 NRRL3488 (ShTh6900)

[0323] 米曲霉 NRRL3488 的原生质体制备和转化如下进行:将大约  $2 \times 10^7$  个孢子接种入 100ml 的 YEG 培养基,并将烧瓶在 27°C 以 140rpm 温育 16-18 小时。将培养物倾倒通过衬有 MIRACLOTH® 的经无菌漏斗,并用 50ml 的 0.7M KCl 漂洗,来收集菌丝体。将经洗涤的菌丝体重悬于含有 20ml 原生质体化溶液 (protoplasting solution) 的 125ml 烧瓶中,其中所述原生质体化溶液的组成为:5mg 的 GLUCANEX™ (Novozymes A/S, Bagsværd, Denmark) 和 0.5mg 的壳多糖酶 (chitinase) (Sigma Chemical Co., St. Louis, MO, USA) 每毫升 0.7M KCl (经过滤灭菌),然后在 34°C 以 80rpm 的混合进行温育 30 分钟。将所述原生质体化溶液倾倒经过衬有 MIRACLOTH® 的无菌漏斗,并用 50ml 的 STC 缓冲液 (由 1M 山梨醇 -10mM Tris-HCl pH6.5-10mM CaCl<sub>2</sub> 构成) 漂洗。流过物收集在两个 50ml 聚丙烯管中。将管以 1300x g 在室温离心 10 分钟。弃去上清,并将原生质体离心沉淀重悬于 20ml 的 STC 缓冲液。通过将离心沉淀在 20ml 的 STC 缓冲液中重悬后以 1300x g 在室温离心 10 分钟两轮来洗涤原生质体。将最终的离心沉淀重悬于 2ml 的 STC 缓冲液。移出 10  $\mu$  l 样品并将其在血细胞计数器 (VWR, West Chester, PA, USA) 中计数来对原生质体进行计数。用 STC 缓冲液调整体积以获得  $2 \times 10^7$  每 ml 的原生质体浓度。

[0324] 通过用 Pme I 在 37°C 限制性消化 4 小时分别制备了质粒表达载体 pAmFs69 (实施例 1), pSaMF36 (实施例 3), pSaMF21 (实施例 4) 和 pRyan1 (实施例 5) 以供转化。通过 TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳将来自每个转化体的大约 5-6kb 的表达盒与载体序列分离,并使用 QIAQUICK® 凝胶提取试剂盒根据生产商的指示纯化出来。

[0325] 通过将上述 100  $\mu$  l 原生质体制备物添加至四个 12ml 聚丙烯管来制备四个转化反应体系。对于每个试管,将 2 微克的经消化的 pRyan1pyc 片段,和各 1 微克的经消化的 pAmFs69bt1 片段,经消化的 pSaMF36C4T521 片段,和经消化的 pSaMF21mdh 片段添加至 250  $\mu$  l 的聚乙二醇 (PEG) 溶液 (60%w/v 聚乙二醇 (PEG), 10mM Tris6.5, 10mM CaCl), 接着轻柔地混合并在 37°C 温育 30 分钟。用 6ml 的 STC 缓冲液稀释每个转化体系,接着将三个单独的等分试样铺板于 COVE 平板。然后将每个平板在 34°C 温育 7-10 日。将所得的转化体中的六十个转移至单独的 COVE 平板,并在 34°C 温育 5 日。通过制备每个培养物的甘油储备物 (800  $\mu$  l 孢子储液, 200  $\mu$  l 10.1% TWEEN® 80) 并冷冻于 -80°C 来储藏培养物。

[0326] 将转化体在摇瓶中培育,并根据上述描述分离基因组 DNA。用于测试四种表达载体片段中的每一个的存在的单独 PCR 反应由 5  $\mu$  l 110X 反应缓冲液, 0.5  $\mu$  l 模板 (80-300ng/ $\mu$  l); 1.0  $\mu$  l 正向引物 (50pM, 见下); 1.0  $\mu$  l 反向引物 (50pM; 见下); 0.5  $\mu$  l dNTP 混合物 (10mM), 16.75  $\mu$  l 去离子水, 和 0.25  $\mu$  l Phusion® DNA 聚合酶。

[0327] 正向引物 065067 (用于 pRyan1pyc, pSaMf21mdh, 和 pSaMf36C4T521 片段):

[0328] 5'-tgaccttccacgctgaccac-3' (SEQ ID NO:46)

[0329] 正向引物 0610854 (用于 pAmFs69bt1 片段):

[0330] 5'-GGCTGAGAAAATATGTTGCA-3' (SEQ ID NO:47)

[0331] 反向引物 0611365 (用于 pSaMF36C4T521 片段):

[0332] 5'-gatagaccactaatcatggtggcgatggag-3' (SEQ ID NO:48)

[0333] 反向引物 061752 (用于 pRyan1pyc 片段)

[0334] 5'-TGCGGTCCTGAGTCAGGCCAGTTGCTCGA-3' (SEQ ID NO:49)

[0335] 反向引物 062400 (用于 pSaMF21mdh 片段)

[0336] 5'-GGGATTTGAACAGCAGAAGG-3' (SEQ ID NO:50)

[0337] 反向引物 996270 (用于 pAmFs69bt1 片段)

[0338] 5'-tcacaaaagagtagaggcca-3' (SEQ ID NO:51)

[0339] 扩增反应体系在 **EPENDORF® MASTERCYCLER®** 中温育,其编程为:1 个循环,在 98°C 进行 30 秒;35 个循环,每循环在 98°C 进行 10 秒,66°C (对于 pRyan1pyc 片段) 或 58°C (对于 pAmFs69bt1, pSaMF21mdh, 和 pSaMF36C4T521 片段) 进行 10 秒;72°C 进行 15 秒;和 1 个循环,在 72°C 进行 10 分钟。使用米曲霉 NRRL3488 基因组 DNA (110ng/ $\mu$ l) 作为阴性对照模板,使用每种质粒 (pRyan1, pAmFs69, pSaMF21, 或 pSaMF36, 稀释至 20ng/ $\mu$ l) 作为阳性对照模板。扩增反应混合物通过凝胶电泳,使用 2  $\mu$ l 的每种反应混合物在 0.8% 琼脂糖凝胶上进行分析。然后对得到预期 PCR 片段大小,从而确认了整合的转化体,测试苹果酸的产生。

[0340] 制备了含有 pSaMF36, pSaMF21, 和 pRyan1 但缺乏 pAmFs69 的表达载体片段的对照转化体 (命名为 SaMF3603 转化体), 并如上所述在类似的步骤中进行验证。

[0341] 实施例 7: 在含有 pAmFs69, pRyan1, pSaMF21, 和 pSaMF36 的表达载体片段的米曲霉转化体 (ShTh6900) 的摇瓶培养中产生苹果酸

[0342] 将来自实施例 6 中所述的 ShTh6900 转化体和作为对照的米曲霉 NRRL3488 的孢子铺板于单独的 PDA 平板上, 并允许在 34°C 孢子形成 5 至 7 日。将孢子收集在 0.1% **TWEEN®** 80 中, 并使用血细胞计数器计数。在含有 100ml 种子培养基 B 的 250ml 烧瓶中制备种子培养物, 并接种 300  $\mu$ l 孢子悬液。将种子培养物在 30°C 在 200rpm 振荡下培育大约 17 个小时。在含有 50ml 产酸培养基 C 和 3ml 的 17 小时种子培养物的 250ml 不带挡板的烧瓶中制备产酸培养物。将培养物在 30°C 在 200rpm 振荡下温育 2 至 10 日。

[0343] 对于摇瓶培养物的苹果酸定量通过使用 1200 系列二元 LC 系统 (Binary LC System) 和 1200 系列二极管阵列检测器 (Diode Array Detector) (DAD) (Agilent Technologies, Santa Clara, CA USA) 的反相高效液相色谱进行。使用 Aqua5  $\mu$  C18 **125Å** 205x4.6mm ID 柱和 AQ C184x3.0mm 预柱 (Security Guard Cartridge) (Phenomenex, Inc., Torrance, CA, USA) 进行反相分离。流动相由 10% 甲醇 (HPLC 级) 和 90%145mM 磷酸盐 pH1.5 缓冲液组成。

[0344] 移出全培养物样品, 并将其在 HPLC 运行缓冲液中以 1:10 稀释, 所述 HPLC 运行缓冲液组成为 850ml 的 64mM 磷酸盐缓冲液和 150ml 的甲醇 pH1.65。然后将样品通过 25mm 的 0.45 微米聚醚砜膜 (Whatman, Florham Park, NJ, USA) 过滤, 并将 1.5ml 的滤过物置于 HPLC

小瓶中以供酸分析。将剩余量的摇瓶培养物通过三层粗滤布 (cheese cloth) 过滤,并用 10 体积的双蒸灭菌水漂洗三次以去除不溶性  $\text{CaCO}_3$ 。从粗滤布收获细胞沉淀,置于 15ml 培养管中,并储藏于  $-20^\circ\text{C}$ 。

[0345] RP-HPLC 运行的使用的条件为:注入体积  $10\ \mu\text{l}$ ,流速  $0.7\text{ml}/\text{分钟}$ (等度洗脱),柱温  $25^\circ\text{C}$ ,运行时间 11 分钟。检测设定为  $210\text{nm}$ ,  $8\text{nm}$  带宽,参照为  $360\text{nm}$ ,  $40\text{nm}$  带宽。空白时间 (void time) 确定为 3.8 分钟。通过进行浓度范围为  $49.2$  至  $3.93\text{mM}$  的系列稀释的苹果酸标样的重复注入,确定了该反相法对于苹果酸的定量能力。对于重复注入的相对标准偏差 (RSD) 为  $\leq 5\%$ 。苹果酸显示  $R^2 \geq 0.9999$ 。

[0346] 含有 pAmFs69, pRyan1, pSaMf21, 和 pSaMf36 的表达载体片段的米曲霉 ShTh6900 转化体与米曲霉 NRRL3488 对照株相比显示高两倍以上苹果酸效价,且高于在单独实验中用各 SaMf3603 转化体(含有 pSaMF36, pSaMF21, 和 pRyan1 的表达载体片段,但缺乏 pAmFs69 的表达载体片段)中观察到的效价。

[0347] 实施例 8:含有 pAmFs69, pRyan1, pSaMf21, 和 pSaMf36 的表达载体片段的米曲霉转化体 (ShTh6900) 的发酵

[0348] 将在实施例 7 中所述的米曲霉 ShTh6900 转化体和对照转化体 SaMf3603(含有 pSaMF36, pSaMF21, 和 pRyan1 的表达载体片段,但缺乏 pAmFs69 的表达载体片段)在  $34^\circ\text{C}$  在 PDA 平板上培育大约 7 日。将  $5\text{--}6\text{ml}$  体积的含有  $0.2\%$  TWEEN® 80 的无菌的  $50\text{mM}$  磷酸钠缓冲液 ( $50\text{mM}$ ,  $\text{pH}6.8$ ) 添加至每个平板,并通过用接种环刮擦使孢子悬浮。通过移液将每个悬液转移至一个  $50\text{ml}$  锥形管中。对于每个管,将  $25\text{ml}$  的无菌的含  $0.2\%$  TWEEN® 80 的磷酸钠缓冲液 ( $50\text{mM}$ ,  $\text{pH}6.8$ ) 添加至含有  $75\text{ml}$  的种子培养基的  $500\text{ml}$  不带挡板的烧瓶,然后将其用  $2\text{ml}$  的孢子悬液接种。然后将烧瓶在  $32^\circ\text{C}$  和  $180\text{rpm}$  温育约 24 小时。合并各种种子烧瓶以供应每罐所需的  $144\text{ml}$  接种物。

[0349] 通过导入  $144\text{ml}$  ( $8\%$ ) 的来自 ShTh6900 转化体或米曲霉 SaMf3603 转化体的三个合并的种子烧瓶的种子培养液来分别接种含有  $1.8$  升的发酵罐分批培养基的三升发酵罐。将发酵罐平衡于  $32 \pm 0.1^\circ\text{C}$  并以  $500\text{rpm}$  搅拌。输入空气流维持在  $1\text{v}/\text{v}/\text{m}$ 。从发酵约 20 小时开始,以大约  $7.3\text{g}/\text{hr}$  的速率施加  $25\%$  葡萄糖流。在大约第 5 日时添加无菌的  $\text{CaCO}_3$  (约  $100\text{g}$ ) 以保持发酵 pH 在 6 至 7 的范围。每日移出样品,并如实施例 6 中所述分析苹果酸产生。发酵在 7 或 8 日之后完成。

[0350] ShTh6900 转化体与 SaMF3603 对照转化体相比显示更高的苹果酸效价,具有更快的产生速率(特别是在最初的 72 小时内),和更快的葡萄糖消耗。

[0351] 实施例 9:将 pShTh147 的表达载体片段转化入米曲霉 M727 (ShTh147)

[0352] 通过将大约  $2 \times 10^7$  个孢子接种入  $100\text{ml}$  的 YEG 培养基,并将烧瓶在  $27^\circ\text{C}$  以  $140\text{rpm}$  温育  $16\text{--}18$  小时来制备和转化米曲霉 M727 (ShTh6900 的突变株,其通过用 NTG 标准诱变并选择增加的 C4 酸产生而制成)的原生质体。通过将培养物倾倒通过衬有 MIRACLOTH® 的无菌漏斗,并用  $50\text{ml}$  的  $0.7\text{M}$  KCl 漂洗来收集菌丝体。将经洗涤的菌丝体重悬于含有  $20\text{ml}$  原生质体化溶液 (protoplasting solution) 的  $125\text{ml}$  烧瓶中,其中所述溶液由  $5\text{mg}$  的 GLUCANEX™ (Novozymes A/S, Bagsvaerd, Denmark) 和  $0.5\text{mg}$  的壳多糖酶 (chitinase) (Sigma Chemical Co., St. Louis, MO, USA) 每毫升  $0.7\text{M}$  KCl (经过滤灭菌) 构成,然后在  $34^\circ\text{C}$  以  $80\text{rpm}$  的混合进行温育 30 分钟。将所述原生质体化溶液倾倒通过衬有 MIRACLOTH® 的

无菌漏斗,并用 50ml 的 STC 缓冲液(由 1M 山梨醇-10mM Tris-HCl pH6.5-10mM CaCl<sub>2</sub> 构成)漂洗。将流过物收集在两个 50ml 聚丙烯管中。将管以 1300x g 在室温离心 10 分钟。弃去上清,并将原生质体离心沉淀重悬于 20ml 的 STC 缓冲液。通过将沉淀在 20ml 的 STC 缓冲液中重悬再以 1300x g 在室温离心 10 分钟两轮来洗涤原生质体。将最终的离心沉淀重悬于 2ml 的 STC 缓冲液中。通过移出 10 μl 样品并将其在血细胞计数器中计数(VWR)来对原生质体进行计数。用 STC 缓冲液调整体积以获得 2x10<sup>7</sup> 每 ml 的原生质体浓度。

[0353] 通过用 PmeI 在 37°C 限制性消化 4 小时制备质粒表达载体 pShTh147(实施例 2)以供转化。通过在 TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳将大约 5.2kb 表达盒与载体序列分离,并使用 QIAQUICK® 凝胶提取试剂盒根据生产商的指示纯化出来。

[0354] 通过将上述 100 μl 原生质体制备物添加至四个 12ml 聚丙烯管来制备四个转化反应体系。对于每个试管,将 2 微克消化的 pShTh147bt2 片段添加至 250 μl 的聚乙二醇(PEG)溶液(60%w/v 聚乙二醇(PEG), 10mM Tris6.5, 10mM CaCl),接着轻柔地混合并在 37°C 温育 30 分钟。用 6ml 的 STC 缓冲液稀释每个转化反应体系,接着将三个单独的等分试样铺板于 COVE 平板上。然后将每个平板在 34°C 温育 7-10 日。将所得的转化体中的四十个转移至单独的 COVE 平板,并在 34°C 温育 5 日。通过制备每个培养物的甘油储液(800 μl 孢子储液, 200 μl 0.1% TWEEN® 80)并冷冻于 -80°C 来储藏培养物。

[0355] 实施例 10:在含有 pAmFs69, pRyan1, pSaMf21, pSaMf36, 和 pShTh147 的表达载体片段的米曲霉转化体(ShTh147)的摇瓶培养中产生苹果酸

[0356] 将来自实施例 9 中所述的 ShTh147 转化体和作为对照的米曲霉 NRRL3488 的孢子铺板于单个 PDA 平板上,并允许在 34°C 孢子形成 5 至 7 日。将孢子在 0.1% TWEEN® 80 中收集,并使用血细胞计数器计数。在含有 100ml 种子培养物 B 的 250ml 烧瓶中制备种子培养物,并接种 1ml 收获的孢子。将种子培养物在 30°C 在 200rpm 振荡下培育大约 22 个小时。在含有 50ml 产酸培养基 C 和 3ml 的 22 小时种子培养物的 250ml 不带挡板的烧瓶中制备产酸培养物。将培养物在 30°C 在 200rpm 振荡下温育 3 日。

[0357] 对于摇瓶培养物的苹果酸定量通过使用 1200 系列二元 LC 系统和 1200 系列二极管阵列检测器(DAD)(Agilent Technologies, Santa Clara, CA USA)的反相高效液相色谱(RP-HPLC)进行。使用 Aqua5 μ C18 125Å 205x4.6mm ID 柱和 AQ C184x3.0mm 预柱(Security Guard Cartridge)(Phenomenex, Inc., Torrance, CA, USA)进行反向分离。流动相由 10% 甲醇(HPLC 级)和 90%145mM 磷酸盐 pH1.5 缓冲液组成。

[0358] 移出全培养物样品,并将其在 HPLC 运行缓冲液中以 1:20 稀释,所述 HPLC 缓冲液由为 900ml 的 145mM 磷酸盐缓冲液和 100ml 的甲醇 pH1.50 构成。然后将样品通过 96 孔 0.45 微米 Durapore PVDF 膜过滤到 96 孔板中以供酸分析。

[0359] 使用下述条件进行 RP-HPLC:注入体积 10 μl,流速 0.7ml/分钟(等度洗脱),柱温 20°C。检测为 210nm,8nm 带宽,参照为 360nm,40nm 带宽。运行时间为 13 分钟。空白时间(void time)确定为 3.8 分钟。通过进行浓度范围为 49.2 至 3.93mM 的系列稀释的苹果酸标样的重复注入,对于苹果酸确定了反相方法的定量能力。对于重复注入的相对标准偏差(RSD)为 ≤ 5%。苹果酸显示 R<sup>2</sup> ≥ 0.9999。

[0360] 在摇瓶测试之后,鉴定出了六个米曲霉 ShTh147 转化体,其以高于 M727 对照的水

平产生苹果酸,包括两个改善了 1.15x 和 1.14x 的。

[0361] 本发明可通过下述编号段落进一步描述:

[0362] [1] 包含编码碳酸氢根转运蛋白的异源多核苷酸的重组宿主细胞,其中所述宿主细胞与不含所述异源核苷酸的宿主细胞相比,当在相同条件下培养时,能够产生更大量的 C4-二羧酸。

[0363] [2] 段 [1] 的重组宿主细胞,其中所述碳酸氢根转运蛋白是硫酸根-碳酸氢根转运蛋白。

[0364] [3] 段 [1] 或 [2] 的重组宿主细胞,其中所述碳酸氢根转运蛋白选自下组:

[0365] (a) 多肽,其与 SEQ ID NO:2 或 SEQ ID NO:4 或其成熟多肽序列具有至少 65%,例如至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性;

[0366] (b) 多肽,其由多核苷酸编码,所述多核苷酸在低严格条件,中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件下,与以下杂交:(i) SEQ ID NO:1 或 3,或其成熟多肽编码序列,(ii) SEQ ID NO:1 或 3 的 cDNA 序列,或其成熟多肽编码序列,(iii) (i) 或 (ii) 的全长互补链;

[0367] (c) 多肽,其由多核苷酸编码,所述多核苷酸与 (iv) SEQ ID NO:1 或 3,或其成熟多肽编码序列,(v) SEQ ID NO:1 或 3 的 cDNA 序列,或其成熟多肽编码序列,(vi) (iv) 或 (v) 的全长互补链具有至少 65%,例如至少 70%,至少 75%,至少 80%,至少 85%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性;

[0368] (d) SEQ ID NO:2 或 4 或其成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体;和

[0369] (e) (a)、(b)、(c) 或 (d) 的多肽具有碳酸氢根转运蛋白活性的片段。

[0370] [4] 段 [1]-[3] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白与 SEQ ID NO:2 或 4 或其成熟多肽序列具有至少 65%,例如至少 70%,至少 75%,至少 80%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性。

[0371] [5] 段 [1]-[4] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸在低严格条件,中等严格条件,中等-高严格条件,高严格条件,或非常高严格条件下,与以下杂交:(i) SEQ ID NO:1 或 3,或其成熟多肽编码序列,(ii) SEQ ID NO:1 或 3 的 cDNA 序列,或其成熟多肽编码序列,(iii) (i) 或 (ii) 的全长互补链;

[0372] [6] 段 [1]-[5] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白由多核苷酸编码,所述多核苷酸与 (iv) SEQ ID NO:1 或 3,或其成熟多肽编码序列,(v) SEQ ID NO:1 或 3 的 cDNA 序列,或其成熟多肽编码序列,或 (vi) (iv) 或 (v) 的全长互补链具有至少 65%,例如至少 70%,至少 75%,至少 80%,至少 85%,至少 85%,至少 90%,至少 91%,至少 92%,至少 93%,至少 94%,至少 95%,至少 96%,至少 97%,至少 98%,至少 99%,或 100% 序列同一性。

[0373] [7] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白包含或组成为 SEQ ID NO:2 或 4,或其成熟多肽序列。

[0374] [8] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白包含或组成

为 SEQ ID NO:2 或 4。

[0375] [9] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白包含或组成为 SEQ ID NO:2。

[0376] [10] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白包含或组成为 SEQ ID NO:4。

[0377] [11] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白包含或组成为 SEQ ID NO:2 或 4 的成熟多肽序列。

[0378] [12] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白包含或组成为 SEQ ID NO:2 的成熟多肽序列。

[0379] [13] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白包含或组成为 SEQ ID NO:4 的成熟多肽序列。

[0380] [14] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:2 或 4 或其成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0381] [15] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:2 或 4 的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0382] [16] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:2 的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0383] [17] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:4 的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0384] [18] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:2 或 4 的成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0385] [19] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:2 的成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0386] [20] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:4 的成熟多肽序列的包含一个或多个(几个)氨基酸的取代、缺失和/或插入的变体。

[0387] [21] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:2 或 4 的片段,其中所述片段具有碳酸氢根转运蛋白活性。

[0388] [22] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:2 的片段,其中所述片段具有碳酸氢根转运蛋白活性。

[0389] [23] 段 [1]-[6] 任一项的重组宿主细胞,其中所述碳酸氢根转运蛋白为 SEQ ID NO:4 的片段,其中所述片段具有碳酸氢根转运蛋白活性。

[0390] [24] 段 [1]-[23] 任一项的重组宿主细胞,其中所述异源多核苷酸可操作地连接于对于所述多核苷酸而言是外来的启动子。

[0391] [25] 段 [1]-[24] 任一项的重组宿主细胞,其还包含编码 C4-二羧酸转运蛋白的异源第二多核苷酸(例如编码 SEQ ID NO:6, 27, 29, 32, 34, 36, 39, 41, 或 43 或其任何相关方面的异源多核苷酸)。

[0392] [26] 段 [25] 的重组宿主细胞,其中所述异源第二多核苷酸可操作地连接于对于

所述多核苷酸而言是外来的启动子。

[0393] [27] 段 [1]-[26] 任一项的重组宿主细胞,其还包含编码苹果酸脱氢酶的异源第三多核苷酸(例如编码 SEQ ID NO:8 或 45,或其任何相关方面的异源多核苷酸)。

[0394] [28] 段 [27] 的重组宿主细胞,其中所述异源第三多核苷酸可操作地连接于对于所述多核苷酸是外来的启动子。

[0395] [29] 段 [1]-[28] 任一项的重组宿主细胞,其还包含编码丙酮酸羧化酶的异源第四多核苷酸(例如编码 SEQ ID NO:10,或其任何相关方面的异源多核苷酸)。

[0396] [30] 段 [29] 的重组宿主细胞,其中所述异源第四多核苷酸可操作地连接于对于所述多核苷酸是外来的启动子。

[0397] [31] 段 [1]-[24] 任一项的重组宿主细胞,其还包含编码 C4-二羧酸转运蛋白的异源第二多核苷酸,编码苹果酸脱氢酶的异源第三多核苷酸,和编码丙酮酸羧化酶的异源第四多核苷酸。

[0398] [32] 段 [1]-[31] 任一项的重组宿主细胞,其中所述宿主细胞是真核宿主细胞。

[0399] [33] 段 [32] 的重组宿主细胞,其中所述宿主细胞是丝状真菌宿主细胞。

[0400] [34] 段 [33] 的重组宿主细胞,其中所述宿主细胞选自下组:枝顶孢霉属(Acremonium)、曲霉属(Aspergillus)、短梗霉属(Aureobasidium)、烟管霉属(Bjerkandera)、拟蜡菌属(Ceriporiopsis)、金孢子菌属(Chrysosporium)、鬼伞属(Coprinus)、革盖菌属(Coriolus)、隐球菌属(Cryptococcus)、Filibasidium、镰孢属(Fusarium)、腐质霉属(Humicola)、梨孢菌属(Magnaporthe)、毛霉属(Mucor)、毁丝霉属(Myceliophthora)、新考玛脂霉属(Neocallimastix)、脉孢菌属(Neurospora)、拟青霉属(Paecilomyces)、青霉属(Penicillium)、平革菌属(Phanerochaete)、射脉菌属(Phlebia)、瘤胃壶菌属(Piromyces)、侧耳属(Pleurotus)、根霉属(Rhizopus)、裂褶菌属(Schizophyllum)、踝节菌属(Talaromyces)、嗜热子囊菌属(Thermoascus)、梭孢壳属(Thielavia)、弯颈霉属(Tolypocladium)、栓菌属(Trametes) 和木霉属(Trichoderma) 细胞。

[0401] [35] 段 [34] 的重组宿主细胞,其中所述宿主细胞是曲霉属宿主细胞。

[0402] [36] 段 [35] 的重组宿主细胞,其中所述宿主细胞是米曲霉宿主细胞。

[0403] [37] 段 [35] 的重组宿主细胞,其中所述宿主细胞是黑曲霉宿主细胞。

[0404] [38] 段 [1]-[37] 任一项的重组宿主细胞,其中所述 C4-二羧酸选自苹果酸、琥珀酸、草酰乙酸、丙二酸和延胡索酸。

[0405] [39] 段 [38] 的重组宿主细胞,其中所述 C4-二羧酸是苹果酸。

[0406] [40] 段 [1]-[39] 任一项的重组宿主细胞,其中所述细胞能够具有大于约 0.1g/L 每小时,例如,大于约 0.2g/L 每小时,0.5g/L 每小时,0.6g/L 每小时,0.7g/L 每小时,0.8g/L 每小时,0.9g/L 每小时,1.0g/L 每小时,1.1g/L 每小时,1.2g/L 每小时,1.3g/L 每小时,1.5g/L 每小时,1.75g/L 每小时,2.0g/L 每小时,2.25g/L 每小时,2.5g/L 每小时,或 3.0g/L 每小时;或约 0.1g/L 每小时至约 2.0g/L 每小时,例如,约 0.3g/L 每小时至约 1.7g/L 每小时,约 0.5g/L 每小时至约 1.5g/L 每小时,约 0.7g/L 每小时至约 1.3g/L 每小时,约 0.8g/L 每小时至约 1.2g/L 每小时,或约 0.9g/L 每小时至约 1.1g/L 每小时的 C4-二羧酸体积生产力。



[0407] [41] 段 [1]-[40] 任一项的重组宿主细胞,其中所述宿主细胞与不含所述异源多核苷酸的宿主细胞相比,当在相同条件下培养时,能够产生多至少 5%,例如至少 10%,至少 15%,至少 20%,至少 25%,至少 30%,至少 50%,或至少 100% 的量的 C4- 二羧酸。

[0408] [42] 组合物,其包含段 [1]-[41] 任一项的重组宿主细胞。

[0409] [43] 段 [42] 的组合物,其中所述培养基是可发酵培养基。

[0410] [44] 段 [42] 或 [43] 的组合物,其还包含 C4- 二羧酸。

[0411] [45] 段 [44] 的组合物,其中所述 C4- 二羧酸选自苹果酸、琥珀酸、草酰乙酸、丙二酸和延胡索酸。

[0412] [46] 段 [45] 的组合物,其中所述 C4- 二羧酸是苹果酸。

[0413] [47] 段 [42]-[46] 任一项的组合物,其中所述 C4- 二羧酸的效价大于约 10g/L,例如,大于约 25g/L,50g/L,75g/L,100g/L,125g/L,150g/L,160g/L,170g/L,180g/L,190g/L,200g/L,210g/L,225g/L,250g/L,275g/L,300g/L,325g/L,350g/L,400g/L,或 500g/L;或约 10g/L 至约 500g/L,例如,约 50g/L 至约 350g/L,约 100g/L 至约 300g/L,约 150g/L 至约 250g/L,约 175g/L 至约 225g/L,或约 190g/L 至约 210g/L。

[0414] [48] 产生 C4- 二羧酸的方法,其包括:

[0415] (a) 在合适的条件下在培养基中培养段 [1]-[41] 任一项的重组宿主细胞以产生所述 C4- 二羧酸;和

[0416] (b) 回收所述 C4- 二羧酸。

[0417] [49] 段 [48] 的方法,其中所述培养基是可发酵的培养基。

[0418] [50] 段 [48] 或 [49] 的方法,其中所述 C4- 二羧酸的效价大于约 10g/L,例如,大于约 25g/L,50g/L,75g/L,100g/L,125g/L,150g/L,160g/L,170g/L,180g/L,190g/L,200g/L,210g/L,225g/L,250g/L,275g/L,300g/L,325g/L,350g/L,400g/L,或 500g/L;或约 10g/L 至约 500g/L,例如,约 50g/L 至约 350g/L,约 100g/L 至约 300g/L,约 150g/L 至约 250g/L,约 175g/L 至约 225g/L,或约 190g/L 至约 210g/L。

[0419] [51] 段 [48]-[50] 任一项的方法,其中产生的 C4- 二羧酸的量与相同条件下培养没有编码碳酸氢根转运蛋白的多核苷酸的宿主细胞相比高至少 5%,例如至少 10%,至少 15%,至少 20%,至少 25%,至少 30%,至少 50%,或至少 100%。

[0420] [52] 段 [48]-[51] 任一项的方法,其中所述 C4- 二羧酸选自苹果酸、琥珀酸、草酰乙酸、丙二酸和延胡索酸。

[0421] [53] 段 [52] 的方法,其中所述 C4- 二羧酸是苹果酸。

[0422] 本文描述和要求保护的本发明并不局限于本文公开的具体方面的范围内,因为这些方面旨在作为本发明几个方面的说明。旨在将任何等同的方面包含于本发明的范围内。实际上,从前面的说明中,除本文所显示和描述的之外,本发明的多种修改对于本领域的技术人员来说是显而易见的。这些修改也旨在落入所附的权利要求的范围内。在冲突的情况下,将以包括定义部分的本公开为准。

[0001]

序列表

- <110> 诺维信公司 (Novozymes, Inc.)  
Fischer, Amanda  
Brown, Stephen  
Luttringer, Sheryl
- <120> 用于生产 C4-二羧酸的微生物
- <130> 12128-WO-PCT
- <150> US 61/447,286
- <151> 2011-02-28
- <160> 57
- <170> PatentIn version 3.5
- <210> 1
- <211> 2503
- <212> DNA
- <213> 米曲霉 (Aspergillus oryzae)

```

<400> 1
atggaatcca gcgctgtaca ggagccgact caacagcgcct ctttgcggga tgcattttt      60
aacctctttc gtacctcttc ctcaaatgat gccccgggic tccggcgaag actcgtaac      120
gctgagagcg cagcgcaaaa cgaagggicg gcgttaactct atccgccacg ggagcctgat      180
gcaaggactc gctctctoga atcgtacgat cgcggggaac ggggtctgag gaactccggc      240
gttcattggga cttttcttc acgacctgaa caggaagaaa tccaaaaatg ggaatcgaagc      300
tctttgcaga atgctggtaa cgaagaanga tctcagtcctc caggaggagc agaaggccat      360
atgggtcttc ccggcgacct ctccagatac ccacagggac cagagaatat accatcgcfa      420
gacitctctt tcacagcalt gcacatgaag aatcaataat ctctgtaggc ttataatcac      480
gttcgcccctg cttttcaaca caattgttat ctccatcgtg gagacaacta acgttcatca      540
aggtataatc ctactacat eccatttttc aattggatia ctcaataccg gtggctgtac      600
attcgaggtg aattggttgc tgcgacaacc attgcgtcca tctatactcc tatggctttg      660
tccitafctt caaatctctc ccacgcacct cctateaaig gcctctactc ttttgtgac      720
aacccittea tctatgcgat ctccgggagc agcccctgtt taatagtggg eccagaagca      780
gcaggctctc tgcitactgg cagattgtc aaaaactagt tcagaccagc cccaictggt      840
gaggacgacg aagiagcgaa tgccatcgtg gtccgcaatg ccaatgcaat ggcggcgccc      900
atgatactga tcgctgggct tacacggctg ggatttctgg acaatgtctt gagccggccc      960
tttcttaggg gtttreattnc agcgatcggf ttgtgafit ttgttgatca actcatcccc      1020
gaagtgggat tgaccgact agcaaaggaa gctgggtgta cccatgggac tacagttgac      1080
aagctcatgt tccitataag aaacatagga ggttgccaig cgtttacaac cgcggtggct      1140
tttgggagct tigtattat aaigtattt cgttagtgti ggtgactcg gaagccigt      1200
gcttagactg attaccatta caggactctc aagaaaatgc tccagccgag gtatcctcag      1260
gtgatttatc ttccggaccg aattctcgtg gttattcttt cagccgicct gacatggcat      1320
cttggttggg atgacaanng gttggagatt cttgggcccc tgaacaaaaa tgecaatggc      1380
ctttttgcgt tcaaatggcc tttccagitt agccagatga agcatgtacg cgttgcfaat      1440

```

[0002]

```

agtacttctt tcgcatcgc gttacttggc tttttcgagt cttctgttgc cgccaagga 1500
ctfagtggcg aggccagaca agaaggigic cagggaaatgc ctgfcagtgc taacagagag 1560
atggtggcgc tgggtcttgc taatactgtg gggggctgtt tcaigccgtc tectgcgitt 1620
ggfggctatg caagaagcaa agtcaacgct tcaactggag ctccggtctc gaigagcagc 1680
alllccctga gcallattac ctllgtttgl atcatggtgc llllgccgia ellatactaf 1740
ctfccggiga gictcgacc caaatacttc cgagcgaagg ctgagaaaat atgttgcaat 1800
aatlcagaaa gccgttcttt ctftctatgat atctgtctgc gcattcagtc tcattgaaga 1860
atglectcac gacgigcctt tctttatccg actgcgcgga tggacggagc tagccctaat 1920
gccttcctac ttgtctcga ctatlttcta ttctctagag ctgggaattg cccctgggat 1980
tggccttctc alettgatcc llailegcea ttclacgcag cctcggatcc aaatlcggg 2040
laagatagca ggcactaccg accgtttcga taacgcigaa ctccaccocg agagcgttga 2100
gttaatcgaa ggcgcgctta ttgtaagat cccggaaccg ctccaccttg ccaatactgg 2160
tgagctcaag aatgccttc ggcggttga attataiggc agtagccgag cgcacccttc 2220
tcctccccc accgcaccc ccgaacataa caagaafatt atattgatg ttcatggtgt 2280
tactagcatic galggtccg glacgcaagi ctlatagag allgtggacg gatatgcaga 2340
ccagggggtc agcgtctct tctgccgctt ccgaactcgc aatgtttcc gcaigtitga 2400
acgaagtgga atgttgaac gatgcggtgg gataacgcac ttcttctatg gtgtcgaoga 2460
agcccctcgc ctggccgaat cggaagacga gatgaaatc tga 2503

```

```

<210> 2
<211> 770
<212> PRT
<213> 米曲霉(Aspergillus oryzae)
<400> 2

```

```

Met Glu Ser Ser Ala Val Gln Glu Pro Thr Gln Gln Arg Ser Leu Arg
1          5          10          15

Asp Arg Ile Phe Asn Leu Phe Arg Thr Ser Ser Ser Asn Asp Ala Pro
20          25          30

Gly Leu Pro Ala Arg Leu Val Thr Ala Glu Ser Ala Ala Gln Asn Glu
35          40          45

Gly Ser Ala Leu Ile Tyr Pro Pro Arg Glu Pro Asp Ala Arg Thr Arg
50          55          60

Leu Leu Glu Ser Tyr Asp Arg Gly Glu Arg Gly Leu Arg Asn Ser Gly
65          70          75          80

Val His Gly Thr Phe Ser Ser Arg Pro Glu Gln Glu Glu Ile Gln Lys
85          90          95

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

```

[0003]

Ser Pro Gly Gly Ala Asp Gly His Ile Gly Ser Pro Gly Asp Val Ser  
 115 120 125  
 Gly Tyr Pro Gln Gly Pro Glu Asn Ile Pro Ser Leu Asp Ser Ser Phe  
 130 135 140  
 Thr Ala Leu His Met Lys Asn His Lys Ser Leu Tyr Ile Ser Tyr Tyr  
 145 150 155 160  
 Ile Pro Phe Phe Asn Trp Ile Thr Gln Tyr Arg Trp Ser Tyr Ile Arg  
 165 170 175  
 Gly Asp Leu Val Ala Ala Thr Thr Ile Ala Ser Ile Tyr Ile Pro Met  
 180 185 190  
 Ala Leu Ser Leu Ser Ser Asn Leu Ala His Ala Pro Pro Ile Asn Gly  
 195 200 205  
 Leu Tyr Ser Phe Val Ile Asn Pro Phe Ile Tyr Ala Ile Phe Gly Ser  
 210 215 220  
 Ser Pro Leu Leu Ile Val Gly Pro Glu Ala Ala Gly Ser Leu Leu Thr  
 225 230 235 240  
 Gly Thr Ile Val Lys Thr Ser Val Arg Pro Gly Pro Ser Gly Glu Asp  
 245 250 255  
 Asp Glu Val Ala Asn Ala Ile Val Val Gly Ile Ala Thr Ala Met Ala  
 260 265 270  
 Gly Ala Met Ile Leu Ile Ala Gly Leu Thr Arg Leu Gly Phe Leu Asp  
 275 280 285  
 Asn Val Leu Ser Arg Pro Phe Leu Arg Gly Phe Ile Thr Ala Ile Gly  
 290 295 300  
 Phe Val Ile Phe Val Asp Gln Leu Ile Pro Glu Val Gly Leu Thr Glu  
 305 310 315 320  
 Leu Ala Lys Glu Ala Gly Val Thr His Gly Thr Thr Val Asp Lys Leu  
 325 330 335  
 Met Phe Leu Ile Arg Asn Ile Gly Gly Cys His Ala Leu Thr Thr Ala  
 340 345 350  
 Val Ala Phe Gly Ser Phe Ala Ile Ile Met Val Phe Arg Thr Leu Lys  
 355 360 365  
 Lys Met Leu Gln Pro Arg Tyr Pro Gln Val Ile Tyr Leu Pro Asp Arg  
 370 375 380  
 Ile Leu Val Val Ile Leu Ser Ala Val Leu Thr Trp His Leu Gly Trp  
 385 390 395 400

[0004]

Asp Asp Lys Gly Leu Glu Ile Leu Gly Pro Leu Lys Gln Asn Ala Asn  
 405 410 415

Gly Leu Phe Ala Phe Lys Trp Pro Phe Gln Phe Ser Gln Met Lys His  
 420 425 430

Val Arg Ala Ala Met Ser Thr Ser Phe Val Ile Ala Leu Leu Gly Phe  
 435 440 445

Phe Glu Ser Ser Val Ala Ala Lys Gly Leu Ser Gly Glu Ala Arg Gln  
 450 455 460

Glu Gly Val Gln Gly Met Pro Val Ser Ala Asn Arg Glu Met Val Ala  
 465 470 475 480

Leu Gly Leu Ala Asn Thr Val Gly Gly Cys Phe Met Ala Leu Pro Ala  
 485 490 495

Phe Gly Gly Tyr Ala Arg Ser Lys Val Asn Ala Ser Thr Gly Ala Arg  
 500 505 510

Ser Pro Met Ser Ser Ile Phe Leu Ser Ile Ile Thr Phe Val Cys Ile  
 515 520 525

Met Val Leu Leu Pro Tyr Leu Tyr Tyr Leu Pro Lys Ala Val Leu Ser  
 530 535 540

Ser Met Ile Ser Val Val Ala Phe Ser Leu Ile Glu Glu Cys Pro His  
 545 550 555 560

Asp Val Ala Phe Phe Ile Arg Leu Arg Gly Trp Thr Glu Leu Ala Leu  
 565 570 575

Met Leu Leu Ile Phe Val Ser Thr Ile Phe Tyr Ser Leu Glu Leu Gly  
 580 585 590

Ile Ala Leu Gly Ile Gly Leu Ser Ile Leu Ile Leu Ile Arg His Ser  
 595 600 605

Thr Gln Pro Arg Ile Gln Ile Leu Gly Lys Ile Ala Gly Thr Thr Asp  
 610 615 620

Arg Phe Asp Asn Ala Glu Leu His Pro Glu Ser Val Glu Leu Ile Glu  
 625 630 635 640

Gly Ala Leu Ile Val Lys Ile Pro Glu Pro Leu Thr Phe Ala Asn Thr  
 645 650 655

Gly Glu Leu Lys Asn Arg Leu Arg Arg Leu Glu Leu Tyr Gly Ser Ser  
 660 665 670

Arg Ala His Pro Ser Leu Pro Pro Thr Arg Thr Pro Glu His Asn Lys  
 675 680 685

[0005]

Asn Ile Ile Phe Asp Val His Gly Val Thr Ser Ile Asp Gly Ser Gly  
690 695 700

Thr Gln Val Leu Tyr Glu Ile Val Asp Gly Tyr Ala Asp Gln Gly Val  
705 710 715 720

Ser Val Phe Phe Cys Arg Val Ala Thr Arg Asn Val Phe Arg Met Phe  
725 730 735

Glu Arg Ser Gly Ile Val Glu Arg Cys Gly Gly Ile Thr His Phe Val  
740 745 750

His Gly Val Asp Glu Ala Leu Arg Leu Ala Glu Ser Glu Asp Glu Ile  
755 760 765

Glu Ile  
770

- <210> 3
- <211> 2657
- <212> DNA
- <213> 米曲霉(Aspergillus oryzae)

```

<400> 3
atgccggcgc atctcaaac caaaattggt cacggcgcgg ccaaggcctt ggggatcaag      60
atcccctacc gtagctctct cggagttcat gctgaccag tcacacgagg cgagtcgatg      120
ttctccgtcg gaacgatcga cacatactcc tatctcgagc ccgaaccac tcccgctgaa      180
tggctgaagg aagctgccc tagctggcat caggtgggce gttatttita caaccttttc      240
ccttctctct cgtggattac gaggiacaac ttgcaatggt tgcctggaga tatgatggcc      300
gglaagagcc tttccatgt gtttgatttg atcgacaagt agacaacata ctccatggaa      360
tgcaggcgtc acggtcggtg ctgtggtcgt tccgcaggga atggcctacg ctaaaatggc      420
aaacctacci glagagtag gictctatc ctctctcatg ggtgttctca ttatitggtt      480
ttttgccacc tcaaggata tcaccattgg tgaagtcct tctgaccca tgtcagcatg      540
taicttgeta atatagtatc ttcccgttc agccgggggc tgtcatgtct acctttacag      600
glaagatagi tccgaggcgc caaacgaagc tcccagatgt cgaagggcct glaalcgcc      660
ccigtittgce tatcatttgt ggagccgttg ttgcgctat gggcctgctt cggctgggat      720
ttatcgttga ttctattct ctgccggcaa ttctagcttt catgacgggt tccgccatca      780
atatcigtct cggacaggtc aaagacatgc tgggagagac ggccgacttc tgcacgaaag      840
attctaceta tctggttatc atcaacacce tcaagcactc tcccctcgca aaaatcgatg      900
ccgccatggg tctcagtgtc ttagctatgc tctacattat ccgttcgggt tgcaattatg      960
ggcgaagaa gttccccctg catgccaagg ttgggtcttt cgtttcgact ttgcgcacag     1020
tgttctgtai ctgtttctat aegatgatca gtgccctgt gaacttgcac cggcggtcta     1080
accgcgggtt caagctcctg ggtaaagtcc ctctgtgttt ccaacaigcg gctgtccctc     1140
aggtaaatte gaggatcctc agcgcatttg ctagcgaact tctgtcttcg attatigtcc     1200
tgcctatcga acacatcctc atctcgaatc cctttggccg tgtcaacaac tacacaattg     1260
    
```

[0006]

```

atccctctca ggagctgggt gotattggig tgcgaactt gcttggaccg itccttggig 1320
gttaccaccg gacitggatcg tctcccga aatcgaatcaa atcgaagcgg ggtgtccgca 1380
ccccacttgc egggttatt acitcggttg ttgtcctcct cgcattttac gctctgcccc 1440
ctgtctctct ttacatcccg aaagcttccc ttgctgggtg catcattcat gcagtcggig 1500
acctcattac cccaccaaac accgittacc agttctggcg cgtgtcccct ctggatgcca 1560
tcattttctt tatcgggtgt atcgtgactg tcttcaccac gattgagatc ggcatttact 1620
gtaccgtttg tgtgtctgtt gccattctgc tgttcccgct cgtccaaggcc cgcggtcaat 1680
tcttaggaag agtcaclatc caclcggiga tgggigacca tctggtagac galgatggga 1740
aatatgggtc tgcgaactcc cctaatgctg ccagcgaatg caaagatgaa ttgagccggt 1800
ctatcttctt gctatcaac cacacggacg gatcgaatcc ccatgtcagc gtcagcaac 1860
cttactctgg tatcttcatc taccgattct cgtgaaggatt caactacccc aatgccaatc 1920
actacaccga ttatttggtc cagactatct tcaagcatac acgtcgcaca aatccgttct 1980
cctacggtaa accgggtgat cggccatgga ataatccctg cctctgcagg ggcaagtctg 2040
aagatgacga gtgcatttg ccttacttgc aggcctgcat tcttgacitc tcatccgca 2100
acaatgttga tctgacctcg giccagaacc tcatcgaigt ccgcaatcaa ctgacctct 2160
acgtctgcc taagaactg cagttggact ttgtctatat taacaaccgc tggacgaac 2220
gagcccttgc agcagcaggt ttccgcttcc catctccgga ctccgatgaa ggattccaga 2280
galggaagcc aalltctgc gttgctgaga tcaaggcag tgcctctgcc gcagctcag 2340
cagagatggt gaacaacaga cacaccaccg ataacatcaa gagcgaagac ctgagcatg 2400
gctcaagca cgttcagag accaccgagc gtgagacaca cggcctcga gaatcctccg 2460
aigccagcag caccgggag gacaagtgc aacgggacct gaaggatagc aagcttacc 2520
gcagtcgcc aagggtcct atgggtcagg gctcaaccg gccattcttc cacatgacc 2580
tgactagtgc actgcagagt gcttggcca acgctggcga gcagccggac cctaaaatga 2640
atgtcttga tgcctag 2657

```

```

<210> 4
<211> 843
<212> PRT
<213> 米曲霉(Aspergillus oryzae)

```

```

<400> 4
Met Pro Gly Asp Leu Lys Thr Lys Ile Gly His Gly Ala Ala Lys Ala
1           5           10           15

Leu Gly Ile Lys Ile Pro Tyr Arg Asp Pro Leu Gly Val His Ala Asp
                20           25           30

Pro Val Thr Arg Gly Glu Ser Met Phe Ser Val Gly Thr Ile Asp Thr
35           40           45

Tyr Ser Tyr Leu Glu Pro Glu Pro Thr Pro Ala Glu Trp Leu Lys Glu
50           55           60

```

[0007]

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

[0008]



Pro Ala Ser Ile Ile Val Leu Leu Ile Glu His Ile Ala Ile Ser Lys  
 355 360 365  
 Ser Phe Gly Arg Val Asn Asn Tyr Thr Ile Asp Pro Ser Gln Glu Leu  
 370 375 380  
 Val Ala Ile Gly Val Ser Asn Leu Leu Gly Pro Phe Leu Gly Gly Tyr  
 385 390 395 400  
 Pro Ala Thr Gly Ser Phe Ser Arg Thr Ala Ile Lys Ser Lys Ala Gly  
 405 410 415  
 Val Arg Thr Pro Leu Ala Gly Val Ile Thr Ala Val Val Val Leu Leu  
 420 425 430  
 Ala Ile Tyr Ala Leu Pro Ala Val Phe Phe Tyr Ile Pro Lys Ala Ser  
 435 440 445  
 Leu Ala Gly Val Ile Ile His Ala Val Gly Asp Leu Ile Thr Pro Pro  
 450 455 460  
 Asn Thr Val Tyr Gln Phe Trp Arg Val Ser Pro Leu Asp Ala Ile Ile  
 465 470 475 480  
 Phe Phe Ile Gly Val Ile Val Thr Val Phe Thr Thr Ile Glu Ile Gly  
 485 490 495  
 Ile Tyr Cys Thr Val Cys Val Ser Val Ala Ile Leu Leu Phe Arg Val  
 500 505 510  
 Ala Lys Ala Arg Gly Gln Phe Leu Gly Arg Val Thr Ile His Ser Val  
 515 520 525  
 Ile Gly Asp His Leu Val Gln Asp Asp Gly Lys Tyr Gly Ser Ala Asn  
 530 535 540  
 Ser Pro Asn Ala Ala Ser Asp Asp Lys Asp Glu Leu Ser Arg Ser Ile  
 545 550 555 560  
 Phe Leu Pro Ile Asn His Thr Asp Gly Ser Asn Pro Asp Val Glu Val  
 565 570 575  
 Gln Gln Pro Tyr Pro Gly Ile Phe Ile Tyr Arg Phe Ser Glu Gly Phe  
 580 585 590  
 Asn Tyr Pro Asn Ala Asn His Tyr Thr Asp Tyr Leu Val Gln Thr Ile  
 595 600 605  
 Phe Lys His Thr Arg Arg Thr Asn Pro Phe Ser Tyr Gly Lys Pro Gly  
 610 615 620  
 Asp Arg Pro Trp Asn Asn Pro Gly Pro Arg Arg Gly Lys Ser Glu Asp  
 625 630 635 640

[0009]

Asp Glu Ser His Leu Pro Leu Leu Gln Ala Val Ile Leu Asp Phe Ser  
645 650 655

Ser Val Asn Asn Val Asp Val Thr Ser Val Gln Asn Leu Ile Asp Val  
660 665 670

Arg Asn Gln Leu Asp Leu Tyr Ala Ser Pro Lys Thr Val Gln Trp His  
675 680 685

Phe Ala His Ile Asn Asn Arg Trp Thr Lys Arg Ala Leu Ala Ala Ala  
690 695 700

Gly Phe Gly Phe Pro Ser Pro Asp Ser Asp Glu Gly Phe Gln Arg Trp  
705 710 715 720

Lys Pro Ile Phe Ser Val Ala Glu Ile Glu Gly Ser Ala Ser Ala Ala  
725 730 735

Ala His Ala Glu Met Val Asn Asn Arg His Thr Gln His Asn Ile Lys  
740 745 750

Ser Glu Asp Leu Glu His Gly Leu Lys His Asp Ser Glu Thr Thr Glu  
755 760 765

Arg Glu Thr His Gly Ile Glu Glu Ser Ser Asp Ala Ser Ser Thr Arg  
770 775 780

Glu Asp Lys Leu Gln Arg Asp Leu Lys Asp Ser Lys Ala Tyr Arg Ser  
785 790 795 800

Arg Arg Arg Val Ala Met Val Gln Gly Leu Asn Arg Pro Phe Phe His  
805 810 815

Ile Asp Leu Thr Ser Ala Leu Gln Ser Ala Leu Ala Asn Ala Gly Glu  
820 825 830

Gln Pro Asp Pro Lys Met Asn Val Leu Asp Ala  
835 840

<210> 5  
 <211> 1257  
 <212> DNA  
 <213> 核抱曲霉 (Aspergillus aculeatus)

<400> 5  
 atgcacgacc acagcactgg aictagtcca tacatctcgg acgtggaaac citgaaccac 60  
 gctgcgaga agtcctgcaa ccccgagacc aaagtctccc agcctcagga atctcccaat 120  
 atcagcaata atgaacatca ggagtgtggt aagctgggca tccgccaacg gctgcgcat 180  
 ttacacctggg cctgggtata cctaaaccatg agcgcaggig gaetggccct tctctctcgc 240  
 aaccagccgt atcaattcaa ggggttgaag gagataggcc iggigtgata catagccaat 300  
 ctctctctct ttactatcat oggetctct atgatacaca ggtttgtict tiacaacaac 360  
 ctaiggaact ctctccgaca cgaccgagaa ggtttctict tccaaccit ctgctctcc 420

[0010]

atcgcaacca tgattagtagg tctatctgce tacttctcta ctgaagacac gcaccgcctc 480  
aattaigctc tggagggtct ctctggggcg tactgtatct tcacgtttgc ctgagcagtg 540  
atccagtact cctttgtctt ctctatcac acgttccctc tgcaactat gatgcatca 600  
tggatcttac cggcattccc tacaigtg agcggaacca ttgcctctgc cgcctccagc 660  
taaccgctg cgggtctgca cagcctatg atgttgcg gcatacagtt ccagggactc 720  
ggattctgca tcagcttcat gatgtacgc cactacatcg ggctctgat ggagacgggc 780  
atccccttga cgcagcaccg tcttggtag ttcctctgig tgggcccccc tgccttcacg 840  
ctgctggcta tcatggcat ggccaacgc ctcccgagg gcttcagtat cctggggcat 900  
gggtgcatgg acgaccgca cactatgga gtaactggcg tetgctgggg catgttctc 960  
tggctctga gcatttgggt cttctgtgic gctctgggct cagttgtgag ggcgcctccc 1020  
catgatttcc acctcaactg gttggctatg gttctcccta acaccggaet cactctcgcc 1080  
accatcacc cggccaagtc actggacagt gccgcgttga aatgggtggg cgtgggcatg 1140  
tcccctgagc tgatctgcat gttcatttc gttctctgta gcaccattag ggtgttctc 1200  
ttgaagagga tcaigtggcc aggtcgggat gaggatgigt ccgagttgtt cgaatga 1257

<210> 6  
<211> 418  
<212> PRT  
<213> 棘孢曲霉

<400> 6

Met His Asp His Ser Thr Gly Ser Ser Pro Tyr Ile Ser Asp Val Glu  
1 5 10 15  
Thr Leu Asn His Ala Cys Glu Lys Ser Val Asn Pro Glu Ala Lys Val  
20 25 30  
Ser Gln Pro Gln Glu Ser Pro Ile Ile Ser Asn Asn Glu His Gln Glu  
35 40 45  
Phe Val Lys Leu Gly Ile Arg Gln Arg Leu Arg His Phe Thr Trp Ala  
50 55 60  
Trp Tyr Thr Leu Thr Met Ser Ala Gly Gly Leu Ala Leu Leu Leu Arg  
65 70 75 80  
Asn Gln Pro Tyr Gln Phe Lys Gly Leu Lys Glu Ile Gly Leu Val Val  
85 90 95  
Tyr Ile Ala Asn Leu Val Phe Phe Thr Ile Ile Gly Ser Leu Met Ile  
100 105 110  
Thr Arg Phe Val Leu Tyr Asn Asn Leu Met Asp Ser Leu Arg His Asp  
115 120 125  
Arg Glu Gly Phe Phe Phe Pro Thr Phe Trp Leu Ser Ile Ala Thr Met  
130 135 140

[0011]

Ile Ser Gly Leu Ser Ala Tyr Phe Ser Thr Glu Asp Thr His Arg Leu  
 145 150 155 160

Asn Tyr Ala Leu Glu Gly Leu Phe Trp Ala Tyr Cys Ile Phe Thr Phe  
 165 170 175

Ala Ser Ala Val Ile Gln Tyr Ser Phe Val Phe Ser Tyr His Thr Phe  
 180 185 190

Pro Leu Gln Thr Met Met Pro Ser Trp Ile Leu Pro Ala Phe Pro Ile  
 195 200 205

Met Leu Ser Gly Thr Ile Ala Ser Ala Ala Ser Ser Tyr Gln Pro Ala  
 210 215 220

Val Ser Ala Thr Pro Met Ile Val Ala Gly Ile Thr Phe Gln Gly Leu  
 225 230 235 240

Gly Phe Cys Ile Ser Phe Met Met Tyr Ala His Tyr Ile Gly Arg Leu  
 245 250 255

Met Glu Thr Gly Ile Pro Ser Ser Glu His Arg Pro Gly Met Phe Ile  
 260 265 270

Cys Val Gly Pro Pro Ala Phe Thr Leu Leu Ala Ile Ile Gly Met Ala  
 275 280 285

Asn Gly Leu Pro Glu Gly Phe Ser Ile Leu Gly Asp Gly Gly Met Asp  
 290 295 300

Asp Arg His Ile Met Arg Val Leu Ala Val Cys Ala Gly Met Phe Leu  
 305 310 315 320

Trp Ala Leu Ser Ile Trp Phe Phe Cys Val Ala Leu Gly Ser Val Val  
 325 330 335

Arg Ala Pro Pro His Asp Phe His Leu Asn Trp Trp Ala Met Val Phe  
 340 345 350

Pro Asn Thr Gly Leu Thr Leu Ala Thr Ile Thr Leu Ala Lys Ser Leu  
 355 360 365

Asp Ser Ala Ala Leu Lys Trp Val Gly Val Gly Met Ser Leu Cys Val  
 370 375 380

Ile Cys Met Phe Ile Phe Val Phe Val Ser Thr Val Arg Ala Val Leu  
 385 390 395 400

Leu Lys Arg Ile Met Trp Pro Gly Arg Asp Glu Asp Val Ser Glu Leu  
 405 410 415

Phe Glu

[0012]

<210> 7  
 <211> 1430  
 <212> DNA  
 <213> 米曲霉 (*Aspergillus oryzae*)

<400> 7  
 atggtc<sup>1</sup>aaag ctgggtg<sup>2</sup>agtt agcaatc<sup>3</sup>ctt aacagatgac actctca<sup>4</sup>atag gta<sup>5</sup>ctaacte 60  
 gaaacgt<sup>6</sup>tiag cgg<sup>7</sup>tiacttgg agcttct<sup>8</sup>gggt ggcatt<sup>9</sup>ggcc aggtat<sup>10</sup>ggat atccc<sup>11</sup>caegc 120  
 ctta<sup>12</sup>caaecc tgg<sup>13</sup>icacaat atgac<sup>14</sup>cttgt t<sup>15</sup>cgatactga ctatct<sup>16</sup>ccca agcc<sup>17</sup>acigt<sup>18</sup>c 180  
 tctcct<sup>19</sup>gttg aagac<sup>20</sup>ctgtc ccttag<sup>21</sup>tiga agag<sup>22</sup>cttgc<sup>23</sup>t ctctac<sup>24</sup>gatg ttgt<sup>25</sup>gaacac 240  
 cctt<sup>26</sup>gtgtgt gct<sup>27</sup>gtgatc tate<sup>28</sup>ccacat ct<sup>29</sup>gtctate gct<sup>30</sup>gtactgt act<sup>31</sup>gccacaa 300  
 tgcga<sup>32</sup>attgc ccgat<sup>33</sup>gaag aggc<sup>34</sup>gaaaa iggtat<sup>35</sup>cttg ct<sup>36</sup>tacctgg<sup>37</sup>g cgatt<sup>38</sup>agaaa 360  
 atct<sup>39</sup>ctgggt tct<sup>40</sup>gtccca agat<sup>41</sup>gatgg<sup>42</sup>g ctga<sup>43</sup>agcag<sup>44</sup>g cct<sup>45</sup>tiactg<sup>46</sup>g t<sup>47</sup>gotaa<sup>48</sup>tatt 420  
 gtt<sup>49</sup>gtcatec cg<sup>50</sup>gtt<sup>51</sup>gat tccc<sup>52</sup>cgtaag tcc<sup>53</sup>ctacc<sup>54</sup>ct ttc<sup>55</sup>gcatt<sup>56</sup>gc tct<sup>57</sup>ctgtatg 480  
 tt<sup>58</sup>ctgtggg gcc<sup>59</sup>agt<sup>60</sup>ttc t<sup>61</sup>gatagtga tag<sup>62</sup>gcaag<sup>63</sup>cc t<sup>64</sup>ggtatg<sup>65</sup>acc cgt<sup>66</sup>ga<sup>67</sup>gacc 540  
 tct<sup>68</sup>caagat ca<sup>69</sup>agcc<sup>70</sup>ggc atag<sup>71</sup>tgcgag act<sup>72</sup>tgg<sup>73</sup>tcaa g<sup>74</sup>ggtat<sup>75</sup>cgcc gag<sup>76</sup>tct<sup>77</sup>gcc 600  
 cca<sup>78</sup>agccct t<sup>79</sup>gttct<sup>80</sup>ggt atct<sup>81</sup>caaac cc<sup>82</sup>gita<sup>83</sup>ttc tact<sup>84</sup>gtt<sup>85</sup>ctt att<sup>86</sup>gtt<sup>87</sup>gcag 660  
 ag<sup>88</sup>gtctcaa ag<sup>89</sup>ccgt<sup>90</sup>ggc g<sup>91</sup>cttt<sup>92</sup>gacc cga<sup>93</sup>agc<sup>94</sup>ctt ct<sup>95</sup>ttgg<sup>96</sup>tgc acc<sup>97</sup>acact<sup>98</sup>gg 720  
 ac<sup>99</sup>gtcgt<sup>100</sup>gc tgc<sup>101</sup>agag<sup>102</sup>act t<sup>103</sup>tcacca<sup>104</sup>ag ag<sup>105</sup>ttct<sup>106</sup>ggg cc<sup>107</sup>aga<sup>108</sup>ggat cct<sup>109</sup>ttc<sup>110</sup>gtc 780  
 t<sup>111</sup>caaat<sup>112</sup>ccc ag<sup>113</sup>tgt<sup>114</sup>ttgt g<sup>115</sup>gccact<sup>116</sup>ct g<sup>117</sup>agag<sup>118</sup>ccat t<sup>119</sup>gtccc<sup>120</sup>ctc t<sup>121</sup>tcag<sup>122</sup>caaga 840  
 ct<sup>123</sup>acc<sup>124</sup>cccgc aat<sup>125</sup>tcag<sup>126</sup>ata ccc<sup>127</sup>gagg<sup>128</sup>aga ag<sup>129</sup>tatg<sup>130</sup>acc act<sup>131</sup>gatcc<sup>132</sup>ac cgt<sup>133</sup>agg<sup>134</sup>ttgt 900  
 ccc<sup>135</sup>aaaga<sup>136</sup>at ct<sup>137</sup>catg<sup>138</sup>aata tct<sup>139</sup>tgt<sup>140</sup>gta ag<sup>141</sup>cact<sup>142</sup>aa<sup>143</sup>ct at<sup>144</sup>gtct<sup>145</sup>cagg cgt<sup>146</sup>cca<sup>147</sup>att 960  
 gg<sup>148</sup>tggag<sup>149</sup>atg ag<sup>150</sup>gtgg<sup>151</sup>lcca ag<sup>152</sup>ct<sup>153</sup>aagg<sup>154</sup>ac gg<sup>155</sup>gtct<sup>156</sup>ggt ccc<sup>157</sup>ccact<sup>158</sup>tt g<sup>159</sup>ctat<sup>160</sup>ggcc 1020  
 tat<sup>161</sup>gcc<sup>162</sup>gggt ac<sup>163</sup>aggt<sup>164</sup>agg<sup>165</sup>g at<sup>166</sup>gtc<sup>167</sup>gta cc<sup>168</sup>gtg<sup>169</sup>agag<sup>170</sup>c act<sup>171</sup>ctc<sup>172</sup>ggct aac<sup>173</sup>atg<sup>174</sup>ccat 1080  
 ag<sup>175</sup>gttct<sup>176</sup>ctg ag<sup>177</sup>agtg<sup>178</sup>aat caa<sup>179</sup>agct<sup>180</sup>ca aag<sup>181</sup>gt<sup>182</sup>caaa cgg<sup>183</sup>gtatt<sup>184</sup>gt cga<sup>185</sup>gcc<sup>186</sup>fac 1140  
 tt<sup>187</sup>ctct<sup>188</sup>acc tgc<sup>189</sup>ctg<sup>190</sup>aat tccc<sup>191</sup>ggc<sup>192</sup>gt gat<sup>193</sup>gag<sup>194</sup>atc t<sup>195</sup>taagg<sup>196</sup>caac tgg<sup>197</sup>ctg<sup>198</sup>gaa 1200  
 tt<sup>199</sup>ctct<sup>200</sup>ctc ct<sup>201</sup>ctg<sup>202</sup>aac ct<sup>203</sup>tagg<sup>204</sup>gta ag<sup>205</sup>att<sup>206</sup>cat<sup>207</sup>ct cct<sup>208</sup>cac<sup>209</sup>gaa tct<sup>210</sup>ctg<sup>211</sup>lca 1260  
 tat<sup>212</sup>ca<sup>213</sup>gccca gg<sup>214</sup>ctaac<sup>215</sup>ct att<sup>216</sup>aac<sup>217</sup>aga ct<sup>218</sup>aat<sup>219</sup>ggc<sup>220</sup>gc ag<sup>221</sup>aga<sup>222</sup>agg<sup>223</sup>ct ag<sup>224</sup>ca<sup>225</sup>ag<sup>226</sup>t<sup>227</sup>c 1320  
 tt<sup>228</sup>aggg<sup>229</sup>ctg gacc<sup>230</sup>gaga<sup>231</sup>ag gaaa<sup>232</sup>aga<sup>233</sup>gc tt<sup>234</sup>ctc<sup>235</sup>gag<sup>236</sup>gc tt<sup>237</sup>gcac<sup>238</sup>gaaa ggc<sup>239</sup>ctt<sup>240</sup>aagg 1380  
 g<sup>241</sup>taalat<sup>242</sup>cca gaa<sup>243</sup>agc<sup>244</sup>atc gac<sup>245</sup>ttc<sup>246</sup>gta aga<sup>247</sup>acc<sup>248</sup>acc acc<sup>249</sup>aa<sup>250</sup>gtaa 1430

<210> 8  
 <211> 330  
 <212> PRT  
 <213> 米曲霉 (*Aspergillus oryzae*)

<400> 8  
 Met Val Lys Ala Ala Val Leu Gly Ala Ser Gly Gly Ile Gly Gln Pro  
 1 5 10 15  
 Leu Ser Leu Leu Leu Lys Thr Cys Pro Leu Val Glu Glu Leu Ala Leu  
 20 25 30  
 Tyr Asp Val Val Asn Thr Pro Gly Val Ala Ala Asp Leu Ser His Ile  
 35 40 45

[0013]

Ser Ser Ile Ala Lys Ile Ser Gly Phe Leu Pro Lys Asp Asp Gly Leu  
 50 55 60

Lys Gln Ala Leu Thr Gly Ala Asn Ile Val Val Ile Pro Ala Gly Ile  
 65 70 75 80

Pro Arg Lys Pro Gly Met Thr Arg Asp Asp Leu Phe Lys Ile Asn Ala  
 85 90 95

Gly Ile Val Arg Asp Leu Val Lys Gly Ile Ala Glu Phe Cys Pro Lys  
 100 105 110

Ala Phe Val Leu Val Ile Ser Asn Pro Val Asn Ser Thr Val Pro Ile  
 115 120 125

Ala Ala Glu Val Leu Lys Ala Ala Gly Val Phe Asp Pro Lys Arg Leu  
 130 135 140

Phe Gly Val Thr Thr Leu Asp Val Val Arg Ala Glu Thr Phe Thr Gln  
 145 150 155 160

Glu Phe Ser Gly Gln Lys Asp Pro Ser Ala Val Gln Ile Pro Val Val  
 165 170 175

Gly Gly His Ser Gly Glu Thr Ile Val Pro Leu Phe Ser Lys Thr Thr  
 180 185 190

Pro Ala Ile Gln Ile Pro Glu Glu Lys Tyr Asp Ala Leu Ile His Arg  
 195 200 205

Val Gln Phe Gly Gly Asp Glu Val Val Gln Ala Lys Asp Gly Ala Gly  
 210 215 220

Ser Ala Thr Leu Ser Met Ala Tyr Ala Gly Tyr Arg Phe Ala Glu Ser  
 225 230 235 240

Val Ile Lys Ala Ser Lys Gly Gln Thr Gly Ile Val Glu Pro Thr Phe  
 245 250 255

Val Tyr Leu Pro Gly Ile Pro Gly Gly Asp Glu Ile Val Lys Ala Thr  
 260 265 270

Gly Val Glu Phe Phe Ser Thr Leu Val Thr Leu Gly Thr Asn Gly Ala  
 275 280 285

Glu Lys Ala Ser Asn Val Leu Glu Gly Val Thr Glu Lys Glu Lys Lys  
 290 295 300

Leu Leu Glu Ala Cys Thr Lys Gly Leu Lys Gly Asn Ile Glu Lys Gly  
 305 310 315 320

Ile Asp Phe Val Lys Asn Pro Pro Pro Lys  
 325 330

[0014]

<210> 9  
 <211> 3643  
 <212> DNA  
 <213> 米曲霉 (*Aspergillus oryzae*)

<400> 9  
 atggcgctc cgttcgtca gccigaggag ggggtcgaig acaccgagti catcgaigac 60  
 caccatgaac acclecgiga faccgigcaac catcgggtgc ggcceaatic caccatlatg 120  
 cacttcacaga agatcctcgt cgccaacogt ggigagatcc ccattcgtat ctfcagaacg 180  
 gccccacgagc tgcatttga gacggttgct atctactctc atgaggatcg actgtcaatg 240  
 caccgtcaaa agggcgtatg ggctacatg atiggccaac ggggtcagta caccctgtc 300  
 ggigcgtacc tggcgggga tgagatcacc aagatcgccc tggagcacgg tgtccagctg 360  
 atccaccggg gctacggitt ctgtccgag aacggcgaat tcgcccga ggttggagaac 420  
 gccggcattg tctttgtggg acccactccc gataccattg acagcttggg tgacaaggtg 480  
 tcggcccgtc gactggccat taagtgcgag gtcctgtcg ttccgggtac ggaggcccc 540  
 gtcgagcgct atgaggaggi caaggcgttc acagacacct atggcttccc cactatcatic 600  
 aaggctgcct ttggcgggtg tggcgttgg atcgtgttg tccgtgacca ggccgagctg 660  
 cgtgactcgt tcgagcgagc cactctgag gcccgctccg ccttcggcaa tggtagcttc 720  
 ttctcgcagc gcttctctga caaaccaag cacattgaag tccagcttct gggtagacagc 780  
 caccgcaacg ttgtccatct gtttagcgt gactgtccc tgcagcgtcg tcaccagaag 840  
 gtcgttgagg ttgtccggc taaggacctg ccagccgatg tccgggaccg cactctggcc 900  
 gatgcttga agctggcaa gtcctcaac taccgtaac ccggtacagc tgagtctctg 960  
 gtggaccagc agaacggcca ctacttcatt gaaatcaatc ctctatacca agtcgagcac 1020  
 accatcacgg aagagattac tggatcgat atcgtggctg cacagatcca gatigtctgt 1080  
 ggtgcaagcc tcgagcaact gggcctgact caggaccgca tctccgccg cggatttgcc 1140  
 attcaatgtc gtaicaccac ggaagatccc gccaaagggf tctctccgga tactggtaa 1200  
 attgaggttt atcgttccgc tggigglaac ggtagccgic tggatggtag taacggttcc 1260  
 gctggtgcta tcatcaccoc tcactacgac tccatgctgg tcaaggttac ctgccgtggt 1320  
 tcgacctatg aaatcgtctg tcgcaagggt gtgcgtgctt tggtaggtt ccgtattctg 1380  
 ggtgtgaaga ccaacattcc ctctctgact tcgcttctga gccaccgac ctctcgtgat 1440  
 ggaaactgct ggaccacttt catcgacgac aacctgaat tgttctctct tgtcggcagt 1500  
 cagaaccgtg ccagaagct gctcgcaaac ctccgggatg tagctgtcaa cggtagtagc 1560  
 atcaagggcc aatitggcga gcccaagctc aagggtgag tcatcaagcc gaagcttttc 1620  
 gatgccgagg gcaagccctc tgacgttccc gcccccgtca ccaagggtg gaagcagatt 1680  
 ctggaccggg aaggccccgc tgcctttgcg aaggccgtgc gtagcaaca gggitgcttg 1740  
 atcaiggata ctacctggcg tgacgccca cagcttttgc tggccaccgc tgtgcgtacc 1800  
 atcgaactgt tgaacatcgc ccattgagacc agctacgctt actccaatgc gtacagittg 1860  
 gaaigtgagg gttgtgttac ctctgatgtg gccatgctt tctctatga ggaccccgtg 1920

[0015]

gaccgccicg gcaagaicgc taaggcigtg cctaacaicc cattccagat gfigciccgt 1980  
 ggfgccaacg gigtgccta etctccctc ccagacaacg ccactacca effctgtaag 2040  
 caggctaaga agtgcggigt cgacatttc cgtgttttcg acgcccicaa cgatgctgat 2100  
 cagctcgagg tggtafca ggcigtatc gtgcccagg gigtgtoga ggcaccaig 2160  
 tgcacacgc gtgacatgct gaacccccac aagaagtaca acctggagta ctacatggcc 2220  
 ttggtggata agattgtagc catgaagcct cacatccttg gtafcaagga taaggccgtt 2280  
 gtgctgaagc eccaggccgc tgcctgttg gtggctcca tccgtcagcg ctaccctgac 2340  
 ctcccaicc acgtccacac ccacgactcc gctggtactg gigtacttc catgattgcc 2400  
 tggcccagg cgggtgccga cgcctggac gcccgacog acagcaigtc cggatgacc 2460  
 tcccagcta gcattggcgc cttctggcc tctcttgagg gcactgagca agaccocgtt 2520  
 ctcaacctcg cccacgtcgc cgtatfcat agctactggg cacagctcgc ctigtctac 2580  
 tctctcttcg aggcgggtct cactggccc gaccctgagg tctacgagca cgagatccct 2640  
 ggtggicagt tgaccaacct tatctccag gccagtcagc tggcctggg ccagcagtg 2700  
 gccgaaacca agaagccca tgaggcggct aatgatfca tggcgacat tgaaggctc 2760  
 actcccacct coaggtgtg cggtagctg gctcagfca tggctcga caaactgact 2820  
 ccgaggatg ttgtgagcg tgcctggag ctgactcc ctggtctgt gctcgaatc 2880  
 ctcaaggctc tcatgggaca gccctcgtt ggattcccc agccattcgc ctcccgcgc 2940  
 ctgcgcgac gccgcaagct cgagaagcgt ccagctctc acctcgacc ttggatttg 3000  
 gctaagatca agagccagat ccgtgagaag ttccgtgctg ctactgagta tgactggcc 3060  
 agctatgcca tgaafccca ggtcttcgag gactacaaga agtctctca gaagttcgt 3120  
 gatctctccg tctgcccac accgacttc ttggccaagc ctgagattgg cgaggagtc 3180  
 caegttagc tggagaagg taaggtctc atctgaagt tgttggcat cggccctct 3240  
 tcagagcaga ctggtcagcg tgggtcttc tacgaagca accgtgaggt ggcaccgct 3300  
 gctgttagt acaacaagg tccgtggac aacacttcc gccctaaggc cgaigtgggt 3360  
 gacagcacc aggtcggfct cctatgagc ggtgtggtt tgaatccg tctccagat 3420  
 ggtctggagg ttaagaagg taccacctt gccgtctga gtccatgaa gatgtaagt 3480  
 tcaifccga tcatfctct cactggcaa ctacagatc taacagctta tccaggaaat 3540  
 ggttatctct gctctcaca gtgaaaagt ctccagctt cgttcaagg agggcattc 3600  
 tgtggaigcc caggatctc tctcaagat cgtcaagcg taa 3643

<210> 10  
 <211> 1193  
 <212> PRT  
 <213> 米曲霉 (*Aspergillus oryzae*)

<400> 10  
 Met Ala Ala Pro Phe Arg Gln Pro Glu Glu Ala Val Asp Asp Thr Glu  
 1 5 10 15  
 Phe Ile Asp Asp His His Glu His Leu Arg Asp Thr Val His His Arg  
 20 25 30

[0016]



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

[0017]

Val Asp Gln Gln Asn Arg His Tyr Phe Ile Gln Ile Asn Pro Arg Ile  
 325 330 335

Gln Val Glu His Thr Ile Thr Glu Glu Ile Thr Gly Ile Asp Ile Val  
 340 345 350

Ala Ala Gln Ile Gln Ile Ala Ala Gly Ala Ser Leu Glu Gln Leu Gly  
 355 360 365

Leu Thr Gln Asp Arg Ile Ser Ala Arg Gly Phe Ala Ile Gln Cys Arg  
 370 375 380

Ile Thr Thr Glu Asp Pro Ala Lys Gly Phe Ser Pro Asp Thr Gly Lys  
 385 390 395 400

Ile Glu Val Tyr Arg Ser Ala Gly Gly Asn Gly Val Arg Leu Asp Gly  
 405 410 415

Gly Asn Gly Phe Ala Gly Ala Ile Ile Thr Pro His Tyr Asp Ser Met  
 420 425 430

Leu Val Lys Cys Thr Cys Arg Gly Ser Thr Tyr Glu Ile Ala Arg Arg  
 435 440 445

Lys Val Val Arg Ala Leu Val Glu Phe Arg Ile Arg Gly Val Lys Thr  
 450 455 460

Asn Ile Pro Phe Leu Thr Ser Leu Leu Ser His Pro Thr Phe Val Asp  
 465 470 475 480

Gly Asn Cys Trp Thr Thr Phe Ile Asp Asp Thr Pro Glu Leu Phe Ser  
 485 490 495

Leu Val Gly Ser Gln Asn Arg Ala Gln Lys Leu Leu Ala Tyr Leu Gly  
 500 505 510

Asp Val Ala Val Asn Gly Ser Ser Ile Lys Gly Gln Ile Gly Glu Pro  
 515 520 525

Lys Leu Lys Gly Asp Val Ile Lys Pro Lys Leu Phe Asp Ala Glu Gly  
 530 535 540

Lys Pro Leu Asp Val Ser Ala Pro Cys Thr Lys Gly Trp Lys Gln Ile  
 545 550 555 560

Leu Asp Arg Glu Gly Pro Ala Ala Phe Ala Lys Ala Val Arg Ala Asn  
 565 570 575

Lys Gly Cys Leu Ile Met Asp Thr Thr Trp Arg Asp Ala His Gln Ser  
 580 585 590

Leu Leu Ala Thr Arg Val Arg Thr Ile Asp Leu Leu Asn Ile Ala His  
 595 600 605

[0018]

Glu Thr Ser Tyr Ala Tyr Ser Asn Ala Tyr Ser Leu Glu Cys Trp Gly  
 610 615 620

Gly Ala Thr Phe Asp Val Ala Met Arg Phe Leu Tyr Glu Asp Pro Trp  
 625 630 635 640

Asp Arg Leu Arg Lys Met Arg Lys Ala Val Pro Asn Ile Pro Phe Gln  
 645 650 655

Met Leu Leu Arg Gly Ala Asn Gly Val Ala Tyr Ser Ser Leu Pro Asp  
 660 665 670

Asn Ala Ile Tyr His Phe Cys Lys Gln Ala Lys Lys Cys Gly Val Asp  
 675 680 685

Ile Phe Arg Val Phe Asp Ala Leu Asn Asp Val Asp Gln Leu Glu Val  
 690 695 700

Gly Ile Lys Ala Val His Ala Ala Glu Gly Val Val Glu Ala Thr Met  
 705 710 715 720

Cys Tyr Ser Gly Asp Met Leu Asn Pro His Lys Lys Tyr Asn Leu Glu  
 725 730 735

Tyr Tyr Met Ala Leu Val Asp Lys Ile Val Ala Met Lys Pro His Ile  
 740 745 750

Leu Gly Ile Lys Asp Met Ala Gly Val Leu Lys Pro Gln Ala Ala Arg  
 755 760 765

Leu Leu Val Gly Ser Ile Arg Gln Arg Tyr Pro Asp Leu Pro Ile His  
 770 775 780

Val His Thr His Asp Ser Ala Gly Thr Gly Val Ala Ser Met Ile Ala  
 785 790 795 800

Cys Ala Gln Ala Gly Ala Asp Ala Val Asp Ala Ala Thr Asp Ser Met  
 805 810 815

Ser Gly Met Thr Ser Gln Pro Ser Ile Gly Ala Ile Leu Ala Ser Leu  
 820 825 830

Glu Gly Thr Glu Gln Asp Pro Gly Leu Asn Leu Ala His Val Arg Ala  
 835 840 845

Ile Asp Ser Tyr Trp Ala Gln Leu Arg Leu Leu Tyr Ser Pro Phe Glu  
 850 855 860

Ala Gly Leu Thr Gly Pro Asp Pro Glu Val Tyr Glu His Glu Ile Pro  
 865 870 875 880

Gly Gly Gln Leu Thr Asn Leu Ile Phe Gln Ala Ser Gln Leu Gly Leu  
 885 890 895

[0019]

Gly Gln Gln Trp Ala Glu Thr Lys Lys Ala Tyr Glu Ala Ala Asn Asp  
 900 905 910

Leu Leu Gly Asp Ile Val Lys Val Thr Pro Thr Ser Lys Val Val Gly  
 915 920 925

Asp Leu Ala Gln Phe Met Val Ser Asn Lys Leu Thr Pro Glu Asp Val  
 930 935 940

Val Glu Arg Ala Gly Glu Leu Asp Phe Pro Gly Ser Val Leu Glu Phe  
 945 950 955 960

Leu Glu Gly Leu Met Gly Gln Pro Phe Gly Gly Phe Pro Glu Pro Leu  
 965 970 975

Arg Ser Arg Ala Leu Arg Asp Arg Arg Lys Leu Glu Lys Arg Pro Gly  
 980 985 990

Leu Tyr Leu Glu Pro Leu Asp Leu Ala Lys Ile Lys Ser Gln Ile Arg  
 995 1000 1005

Glu Lys Phe Gly Ala Ala Thr Glu Tyr Asp Val Ala Ser Tyr Ala  
 1010 1015 1020

Met Tyr Pro Lys Val Phe Glu Asp Tyr Lys Lys Phe Val Gln Lys  
 1025 1030 1035

Phe Gly Asp Leu Ser Val Leu Pro Thr Arg Tyr Phe Leu Ala Lys  
 1040 1045 1050

Pro Glu Ile Gly Glu Glu Phe His Val Glu Leu Glu Lys Gly Lys  
 1055 1060 1065

Val Leu Ile Leu Lys Leu Leu Ala Ile Gly Pro Leu Ser Glu Gln  
 1070 1075 1080

Thr Gly Gln Arg Glu Val Phe Tyr Glu Val Asn Gly Glu Val Arg  
 1085 1090 1095

Gln Val Ala Val Asp Asp Asn Lys Ala Ser Val Asp Asn Thr Ser  
 1100 1105 1110

Arg Pro Lys Ala Asp Val Gly Asp Ser Ser Gln Val Gly Ala Pro  
 1115 1120 1125

Met Ser Gly Val Val Val Glu Ile Arg Val His Asp Gly Leu Glu  
 1130 1135 1140

Val Lys Lys Gly Asp Pro Leu Ala Val Leu Ser Ala Met Lys Met  
 1145 1150 1155

Glu Met Val Ile Ser Ala Pro His Ser Gly Lys Val Ser Ser Leu  
 1160 1165 1170

[0020]

Leu Val Lys Glu Gly Asp Ser Val Asp Gly Gln Asp Leu Val Cys 1175 1180 1185	
Lys Ile Val Lys Ala 1190	
<210> 11 <211> 40 <212> DNA <213> 米曲霉( <i>Aspergillus oryzae</i> )	
<400> 11 gtgatagaac atcgccata atggaatcca gcgctgtaca	40
<210> 12 <211> 40 <212> DNA <213> 米曲霉( <i>Aspergillus oryzae</i> )	
<400> 12 gtgicagtea cctctagtta tcagatttca atctcgtcct	40
<210> 13 <211> 20 <212> DNA <213> 米曲霉( <i>Aspergillus oryzae</i> )	
<400> 13 gaacaggaag aaatcaaaa	20
<210> 14 <211> 20 <212> DNA <213> 米曲霉( <i>Aspergillus oryzae</i> )	
<400> 14 gtcggcatag ccactgcaat	20
<210> 15 <211> 20 <212> DNA <213> 米曲霉( <i>Aspergillus oryzae</i> )	
<400> 15 tgttgccgcc aaggactta	20
<210> 16 <211> 20 <212> DNA <213> 米曲霉( <i>Aspergillus oryzae</i> )	
<400> 16 ccgagagcgt tgagttaate	20
<210> 17 <211> 20 <212> DNA <213> 米曲霉( <i>Aspergillus oryzae</i> )	
<400> 17 agcattaggg ctagctccgt	20

[0021]

<210> 18		
<211> 20		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 18	ccaagatgcc atgtcaggac	20
<210> 19		
<211> 20		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 19	tcacaaaaga gtagaggcca	20
<210> 20		
<211> 36		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 20	tgtgataaaa catcgtccat aatgcacgac cacagc	36
<210> 21		
<211> 40		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 21	gtgtcagtc a cctctagtta tcattcgaac aactcggaca	40
<210> 22		
<211> 36		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 22	agaacatcgt ccataatggt caaagciggt gaggta	36
<210> 23		
<211> 40		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 23	gtgtcagtc a cctctagtta ttactttggt ggiggattct	40
<210> 24		
<211> 36		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 24	tagaacatcg tccataaalgg cggctccggt tcgtca	36
<210> 25		
<211> 40		
<212> DNA		
<213> 米曲霉( <i>Aspergillus oryzae</i> )		
<400> 25	gtgtcagtc a cctctagtta ttattacgct ttgacgatct	40

[0022]

<210> 26  
 <211> 1143  
 <212> DNA  
 <213> 米曲霉(*Aspergillus oryzae*)

<400> 26  
 atgctgacac ctcecaagtt tgaggatgag aagcagctgg gccccgtggg tatccgggag 60  
 aggcctcgcc atttcaacttg ggcctggfac acattaacga tgagtgagg agggctggcc 120  
 gtcctcatca tcagccagcc ctctgggttc cggcgattga gagagatcgg catcctgttc 180  
 tatacctca acctgatcct ctctgccctt gctgctcta ccattgctat aaggttcaic 240  
 ctgcacggca acctcttggg gtccctcgt catgaccgag agggctctct ctcccgacc 300  
 tictggctct ccgtcgcaac catcctctgc gcttctctc gctacttcgg tgaagaatcg 360  
 aatgagtcct tccaactagc cctcgaagcc ctctcttggg tctactcgt ctgcacctta 420  
 ctgctcgcaa tcaaccaala ctctgtctc tctcctccc acaagtcagg ccttcaaac 480  
 atgatgcctt catggtcct tccagcttc cccatcctg tcagcggcac catcctctc 540  
 gtcctgggtg aacaacaacc cgtctcgca gccctccca tcatggggc cggcgtcacc 600  
 ttcaggggcc tggctcttc catcagcttc atgatgtac cccactacat cggccgacig 660  
 atggagtcgc gccctccca cagcagacc agaccagca tgttcctct cgtcggacc 720  
 ccgccttca cagcctctgc cctctctggc atgagcaaag gccctccca agacttcaag 780  
 ctgtccacg acgcccagc ctctgaagat ggcccatca tcagctctg gccatctcc 840  
 gccggctct tctctgggc ctctgctct tggctctct gctcggcat tctctcctc 900  
 atcctctgc ccccagagg cttcaccct aactgggtgg ccattggtct ccccaacacc 960  
 ggcttacc cggccaccat caccctgggc aagctctca acagtacgg cgtgaaggcc 1020  
 gtcctctcg ccattctat ctctctctg tcatgtaca tctctctct tctcaacaat 1080  
 gtcctcggc ttatccggaa ggatataig taccctggta aagatgagga tctatctgat 1140  
 tag 1143

<210> 27  
 <211> 380  
 <212> PRT  
 <213> 米曲霉(*Aspergillus oryzae*)

<400> 27

Met Leu Thr Pro Pro Lys Phe Glu Asp Glu Lys Gln Leu Gly Pro Val  
 1 5 10 15

Gly Ile Arg Glu Arg Leu Arg His Phe Thr Trp Ala Trp Tyr Thr Leu  
 20 25 30

Thr Met Ser Gly Gly Gly Leu Ala Val Leu Ile Ile Ser Gln Pro Phe  
 35 40 45

Gly Phe Arg Gly Leu Arg Glu Ile Gly Ile Ala Val Tyr Ile Leu Asn  
 50 55 60

Leu Ile Leu Phe Ala Leu Val Cys Ser Thr Met Ala Ile Arg Phe Ile

[0023]

65		70		75		80
Leu His Gly Asn	Leu Leu Glu Ser	Leu Arg His Asp Arg Glu Gly Leu				
	85	90				95
Phe Phe Pro Thr	Phe Trp Leu Ser	Val Ala Thr Ile Ile Cys Gly Leu				
	100	105				110
Ser Arg Tyr Phe	Gly Glu Glu Ser	Asn Glu Ser Phe Gln Leu Ala Leu				
	115	120				125
Glu Ala Leu Phe	Trp Ile Tyr Cys Val Cys Thr	Leu Leu Val Ala Ile				
	130	135				140
Ile Gln Tyr Ser	Phe Val Phe Ser Ser His Lys Tyr	Gly Leu Gln Thr				
	145	150				160
Met Met Pro Ser	Trp Ile Leu Pro Ala Phe Pro	Ile Met Leu Ser Gly				
	165	170				175
Thr Ile Ala Ser	Val Ile Gly Glu Gln Gln Pro	Ala Arg Ala Ala Leu				
	180	185				190
Pro Ile Ile Gly	Ala Gly Val Thr Phe Gln Gly	Leu Gly Phe Ser Ile				
	195	200				205
Ser Phe Met Met	Tyr Ala His Tyr Ile Gly Arg	Leu Met Glu Ser Gly				
	210	215				220
Leu Pro His Ser	Asp His Arg Pro Gly Met Phe	Ile Cys Val Gly Pro				
	225	230				240
Pro Ala Phe Thr	Ala Leu Ala Leu Val Gly Met	Ser Lys Gly Leu Pro				
	245	250				255
Glu Asp Phe Lys	Leu Leu His Asp Ala His Ala	Leu Glu Asp Gly Arg				
	260	265				270
Ile Ile Glu Leu	Leu Ala Ile Ser Ala Gly Val	Phe Leu Trp Ala Leu				
	275	280				285
Ser Leu Trp Phe	Phe Cys Ile Ala Ile Val Ala	Val Ile Arg Ser Pro				
	290	295				300
Pro Glu Ala Phe	His Leu Asn Trp Trp Ala Met	Val Phe Pro Asn Thr				
	305	310				320
Gly Phe Thr Leu	Ala Thr Ile Thr Leu Gly Lys	Ala Leu Asn Ser Asn				
	325	330				335
Gly Val Lys Gly	Val Gly Ser Ala Met Ser	Ile Cys Ile Val Cys Met				
	340	345				350
Tyr Ile Phe Val	Phe Val Asn Asn Val Arg Ala	Val Ile Arg Lys Asp				

[0024]



	355	360	365	
Ile Met Tyr Pro Gly Lys Asp Glu Asp Val Ser Asp				
	370	375	380	
<210>	28			
<211>	1182			
<212>	DNA			
<213>	土曲霉			
<400>	28			
atgitttgaga acactgcccc tccaggagc tcccgcctccg actctggcab cctggaccaa			60	
gaaticgaga agcagccggg ttcggigggc atgctgtaac gcaiccacca tiitaccctgg			120	
gccigtgata ctctcacaat gagtgcctgt ggcttggccc tctctcttgg gagccagcca			180	
aacaccttca ccggcctgag ggagattgga ctgcctgtgt acctgctcaa cctgctcttc			240	
tiigccctgg tctgtctgac caiggccggc eggttcafcc tgcacggagg gcigtctgac			300	
tctctccggc acgaacgaga gggaatttc tcccacact tctggctctc gatgccacc			360	
atcaatcacag gctgtaccg ctacttggc gaagaagccg gaagccctt cgtctctgcc			420	
ctcgaagccc tcttctggat ctactggct tgcaacctcc tcttgcctgt cctccaatac			480	
tcctgctctc tctccggccc caaatacgc ctccaaaccg ccatgcccg ctggatctc			540	
cccgccttcc ctgtcatgct ctctggcacc atcgcctccg tcatgccga gcagagccg			600	
gcccgcgccc ccattcccat catctctcc gccaccact tccagggcct gggcttctcc			660	
atcagcatga tcatgtaccg ccaactagtc ggcgcctca tggagtccg cctgcccgtc			720	
cgcgagcacc gcccgggcat gttcatgcc gtccggccc cggctttcac ggccttgcc			780	
ctctctggca tgaccaaggg gctccgcac gacttccagc tcatggcga tgaattgcc			840	
tctgagatg cccgcactct gcagctctg gcgatcggc tccgctgtt tctctggcg			900	
ctgagctgt ggttcttctg caitgcggcc atgcccctg tgcctcccc gccaacggc			960	
tccacctga gcigtgggc caigtcttc cccaacacgg gcttaccct cgcacagtc			1020	
aacctggga ccgcccctaa gacgagggt atccagggtg tggggacggc caigtctgt			1080	
ggaattgtgt ctatttctt gtttctgtt atcagccatg tccggctgt cctcaggaaa			1140	
gacatttgt atctgggaa agacaggat gtggtggagt aa			1182	
<210>	29			
<211>	393			
<212>	PRT			
<213>	土曲霉			
<400>	29			
Met Phe Glu Asn Thr Ala Pro Pro Gly Ser Ser Arg Ser Asp Ser Gly				
1	5	10	15	
Ile Leu Asp His Glu Phe Glu Lys Gln Pro Gly Ser Val Gly Met Arg				
20	25	30		
Glu Arg Ile Arg His Phe Thr Trp Ala Trp Tyr Thr Leu Thr Met Ser				
35	40	45		

[0025]

Ala Gly Gly Leu Ala Leu Leu Leu Gly Ser Gln Pro Asn Thr Phe Thr  
 50 55 60

Gly Leu Arg Glu Ile Gly Leu Ala Val Tyr Leu Leu Asn Leu Leu Phe  
 65 70 75 80

Phe Ala Leu Val Cys Ser Thr Met Ala Gly Arg Phe Ile Leu His Gly  
 85 90 95

Gly Leu Val Asp Ser Leu Arg His Glu Arg Glu Gly Ile Phe Phe Pro  
 100 105 110

Thr Phe Trp Leu Ser Ile Ala Thr Ile Ile Thr Gly Leu Tyr Arg Tyr  
 115 120 125

Phe Gly Glu Asp Ala Gly Arg Pro Phe Val Leu Ala Leu Glu Ala Leu  
 130 135 140

Phe Trp Ile Tyr Cys Ala Cys Thr Leu Leu Val Ala Val Ile Gln Tyr  
 145 150 155 160

Ser Trp Leu Phe Ser Gly Pro Lys Tyr Arg Leu Gln Thr Ala Met Pro  
 165 170 175

Gly Trp Ile Leu Pro Ala Phe Pro Val Met Leu Ser Gly Thr Ile Ala  
 180 185 190

Ser Val Ile Ala Glu Gln Gln Pro Ala Arg Ala Ala Ile Pro Ile Ile  
 195 200 205

Val Ala Gly Thr Thr Phe Gln Gly Leu Gly Phe Ser Ile Ser Met Ile  
 210 215 220

Met Tyr Ala His Tyr Val Gly Arg Leu Met Glu Ser Gly Leu Pro Cys  
 225 230 235 240

Arg Glu His Arg Pro Gly Met Phe Ile Ala Val Gly Pro Pro Ala Phe  
 245 250 255

Thr Ala Leu Ala Leu Val Gly Met Thr Lys Gly Leu Pro His Asp Phe  
 260 265 270

Gln Leu Ile Gly Asp Asp Phe Ala Phe Glu Asp Ala Arg Ile Leu Gln  
 275 280 285

Leu Leu Ala Ile Ala Val Gly Val Phe Leu Trp Ala Leu Ser Leu Trp  
 290 295 300

Phe Phe Cys Ile Ala Ala Ile Ala Val Val Arg Ser Pro Pro Thr Ala  
 305 310 315 320

Phe His Leu Ser Trp Trp Ala Met Val Phe Pro Asn Thr Gly Phe Thr  
 325 330 335

[0026]

Leu Ala Thr Ile Asn Leu Gly Thr Ala Leu Lys Ser Glu Gly Ile Gln  
340 345 350

Gly Val Gly Thr Ala Met Ser Ile Gly Ile Val Ser Ile Phe Leu Phe  
355 360 365

Val Phe Ile Ser His Val Arg Ala Val Ile Arg Lys Asp Ile Met Tyr  
370 375 380

Pro Gly Lys Asp Glu Asp Val Val Glu  
385 390

<210> 30  
<211> 1317  
<212> DNA  
<213> 粟酒裂殖酵母

<400> 30  
atgggagaat tgaaggaat tctcaagcag cctaccatg aattgctega ctggaacgtc 60  
aaagcacccc acgtccctct ctccagagg ttgaagcatt tcacatggtc gtggctcgg 120  
tgtacgatgg caaccggagg cgtcggactc atcatcggat cctcccttt cagattctac 180  
ggactcaaca cgtcggcaa gattgtgtac atctccaga tttcctctt ctcttgttc 240  
ggctcgtgta tctcttcag gttcatcaag tctcgtcca caatcaagga ctctggaac 300  
catcatctcg agaaactct cttcggact tctctctct cagtttcgac atcatcgat 360  
atgtggcga tctacgccta ccccgacaca ggcgagtgga tgggtgggt catccgaatc 420  
ctctactaca tctacgtgc ggtctcttc atttacigtg tgatggcgtt ctccacgatc 480  
ttcaacaacc acgtctaac catgaaacc gctcgcctg catggatctt cctatcttc 540  
cctccgatga tctgggtgt cttgcccgt gcggtgaact ccaaccagcc tgcgcaccag 600  
ctcaaaaaca tgggtatctt cggatctct tccagggat tgggtttctg ggtctactg 660  
ctctgtctg cagtcaact gctccggtc tccacggtc gcttggcaaa gccccaggac 720  
cgacctggca tgtcatgtt cgtgggacct cctgcgitt cggcttggc actcaacaac 780  
atcgcgaggg gtgccatggg ctccaggccg tacatcttcg tgggagcaaa ctctcggaa 840  
tacttgggtt tctgtctgac gttcatggc atttcatct gggccttggc agcatgggtt 900  
tattgtctcg ccatgggtg ctctctcga ggtctctca cagcgcacc ttggaagttc 960  
ggctgtggtt ggttcgcat catcttccc aacgtgggt tctgtaactg tacgatlgag 1020  
atcggcaaga tgatcgaac caaaccttc cagatgttc gccacalia cgggtgcatc 1080  
ctctglatcc agtggatct gctcatgial ttgatggtc gtcgtctct ggtcaacgac 1140  
ttgtgtatc ccgtaaaga cagggagccc cctcgcctc ccaaaccaa cacaggctc 1200  
ctcaacecca ccttccctc cgaaaaaga cctgcctcc tcaaaaagt cgtatcacat 1260  
gtcacltcca ctggcggaga gtcggalcc cgtctctcg aacacgagtc ggctca 1317

<210> 31  
<211> 1317  
<212> DNA  
<213> 粟酒裂殖酵母

[0027]

```

<400> 31
atgggtgaac tcaaggaaat ctgaaacag aggtatcatg agttgcttga ctggaatgtc 60
aaagcccctc atgtccctct cagtcancga ctgaagcatt ttacatggtc ttggtttgca 120
igtactatgg caactggtag tgttggfifg attattgggt ctttcccttt tctattttat
ggcttfaata caattggcaa aattgtttat attcttcaaa tctttttgtt ttctctcttt 240
ggafcaatga tgcctttttc cttttaitaa tatcttcaa ctatcaagga ttcttggaac
catcatttgg aaaagctttt cattgctact tgcctcttt caatatecac gttcatcgac 360
atgcttgcca tafacgcta tctgatacc ggcgagtgga tgggtgggt cattcgaatc
ctttattaca ttacgttgc agtatcttt atatactgg taatggcttt ttttacaatt 480
ttcaacaacc atgtatatac cattgaaacc gcactctctg ctggattct tctattttc
ctctcatga ttgtgtgtgt cattgtctgc gccgtcaat ctacacaacc cgcctatcaa 600
ttaaaaaata tggttatctt tggtaacttc ttcaaggac ttggttttgg gggtttatct
ttactgtttg ccgtcaatgt cttacggttt ttactgtag gccctggcaa accccaagat 720
cgacctggta tgtttatgtt tctcgggtca ccagcttctt caggttttgg cttaattaat
attgcgcgtg gtgctatggg cagtcgctt tatattttg ttggcgcaa ctcatcgag 840
tacttgggtt ttgtttctac cttatggtt atttttattt ggggtcttgc tgccttgggt 900
facgtctctg ccattggttag ctttttagcg ggtttttca ctcgagcccc tctcaagttt
gcttctggat ggtttgcatl caatttcccc aacgtgggtt ttgitaattg taccattgag 1020
alaggiaaaa tcatagatlc caaagcttc caaigtitg gacatacat tgggtcatl 1080
ctttgtatct agtggatctt cctaatgtat ttaatggctc gtgcatttct cgtcaatgat 1140
ctttgctatc ctggcaaaga cgaagatgcc catctccac caaaacaaa tacagggtgc 1200
cttaacceta ccttccacc tgaaaaagca cctgcatttt tggaaaaagt cgatacacat 1260
gtcacatcta ctgggttga atcggatctt cctagtatgt aacatgaaag cgttttaa 1317
    
```

```

<210> 32
<211> 438
<212> PRT
<213> 粟酒裂殖酵母
    
```

```

<400> 32
Met Gly Glu Leu Lys Glu Ile Leu Lys Gln Arg Tyr His Glu Leu Leu
1          5          10          15

Asp Trp Asn Val Lys Ala Pro His Val Pro Leu Ser Gln Arg Leu Lys
20          25          30

His Phe Thr Trp Ser Trp Phe Ala Cys Thr Met Ala Thr Gly Gly Val
35          40          45

Gly Leu Ile Ile Gly Ser Phe Pro Phe Arg Phe Tyr Gly Leu Asn Thr
50          55          60

Ile Gly Lys Ile Val Tyr Ile Leu Gln Ile Phe Leu Phe Ser Leu Phe
65          70          75          80
    
```

[0028]

Gly Ser Cys Met Leu Phe Arg Phe Ile Lys Tyr Pro Ser Thr Ile Lys  
 85 90 95

Asp Ser Trp Asn His His Leu Glu Lys Leu Phe Ile Ala Thr Cys Leu  
 100 105 110

Leu Ser Ile Ser Thr Phe Ile Asp Met Leu Ala Ile Tyr Ala Tyr Pro  
 115 120 125

Asp Thr Gly Glu Trp Met Val Trp Val Ile Arg Ile Leu Tyr Tyr Ile  
 130 135 140

Tyr Val Ala Val Ser Phe Ile Tyr Cys Val Met Ala Phe Phe Thr Ile  
 145 150 155 160

Phe Asn Asn His Val Tyr Thr Ile Glu Thr Ala Ser Pro Ala Trp Ile  
 165 170 175

Leu Pro Ile Phe Pro Pro Met Ile Cys Gly Val Ile Ala Gly Ala Val  
 180 185 190

Asn Ser Thr Gln Pro Ala His Gln Leu Lys Asn Met Val Ile Phe Gly  
 195 200 205

Ile Leu Phe Gln Gly Leu Gly Phe Trp Val Tyr Leu Leu Leu Phe Ala  
 210 215 220

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

Arg Pro Gly Met Phe Met Phe Val Gly Pro Pro Ala Phe Ser Gly Leu  
 245 250 255

Ala Leu Ile Asn Ile Ala Arg Gly Ala Met Gly Ser Arg Pro Tyr Ile  
 260 265 270

Phe Val Gly Ala Asn Ser Ser Glu Tyr Leu Gly Phe Val Ser Thr Phe  
 275 280 285

Met Ala Ile Phe Ile Trp Gly Leu Ala Ala Trp Cys Tyr Cys Leu Ala  
 290 295 300

Met Val Ser Phe Leu Ala Gly Phe Phe Thr Arg Ala Pro Leu Lys Phe  
 305 310 315 320

Ala Cys Gly Trp Phe Ala Phe Ile Phe Pro Asn Val Gly Phe Val Asn  
 325 330 335

Cys Thr Ile Glu Ile Gly Lys Met Ile Asp Ser Lys Ala Phe Gln Met  
 340 345 350

Phe Gly His Ile Ile Gly Val Ile Leu Cys Ile Gln Trp Ile Leu Leu  
 355 360 365

[0029]

Met Tyr Leu Met Val Arg Ala Phe Leu Val Asn Asp Leu Cys Tyr Pro  
370 375 380

Gly Lys Asp Glu Asp Ala His Pro Pro Pro Lys Pro Asn Thr Gly Val  
385 390 395 400

Leu Asn Pro Thr Phe Pro Pro Glu Lys Ala Pro Ala Ser Leu Glu Lys  
405 410 415

Val Asp Thr His Val Thr Ser Thr Gly Gly Glu Ser Asp Pro Pro Ser  
420 425 430

Scr Glu His Glu Scr Val  
435

<210> 33

<211> 1194

<212> DNA

<213> 株孢曲霉

<400> 33

```

atgctcgggc aacatccgcc tcccgacacc tctgctcgg accitacaac ataccagcat. 60
gagctcaaag cctccaaata cctctagtcc accaafgtgt cctctacggga cegtctcgt. 120
caftttaccf gggcgiggta lactctgacl atgagcaccg gcggtctagc cctctctctg. 180
gccagccagc cctactcctt cctccgactg caacagatcg ggcttgcagt ctacatcatc 240
aacctggcct tctttggtt gctgtgtagc ctcatggccg caccgttcat tctccacggc 300
aacftctctg actcctctcg acacgaccgc gagggtcttt tctttcctac tttctgctt 360
tctattgcaa ctatcatcac eggcctgtac cgtactctcg gcgacaccac acagcctgca 420
ttcatttacg ctcttgaggi gctctctctg cctactctg ccttcaact ct gatgaccgct. 480
attatccaat actcctttgt ctttaccgcc caccactacc ctctacaaac gatgatgcc 540
tcaiggatcc tcccgcart cctatcatg ctctagcgca cgategcctc cgtcattgcc 600
gaacagcagc cccgctctc tcttattecc atgatctctg cccgcaccac ctccaagge 660
cttggcttct ccatcagttt cctcaigtac gcgcactata tccggcgctf catggagacg 720
ggccttccgt cccgggaaca ccgaccggg atgttcatct gcgttggccc cccggcttcc 780
accgcccctg ccttaatcgg catgaccaac ggcttctctg aggatattca agtcccttca 840
gaccgcacc cctttcaaga ccgcacatc ctccgactcc ttgccatcgc cacgggcgcc 900
ttctctctgg cctcagctct ctgttctctt agcattgcca tcatcgcac catccgcctc 960
ccacctacag ccttccact caactggigg gccatggttt ttccaaacac gggttttact 1020
ctcgcgacca tcacgctggg caagccctc gatagccctg gactcaaggc cgtcggatct 1080
gccatgtcca ttgcatctg ggggagtgg ctgttctgtt ttccgagcaa tctccgtgcc 1140
gttgtcaaac gggatattgt tttccctggg aaggacgagg atgtatcgga gtaa 1194

```

<210> 34

<211> 397

<212> PRT

[0030]

<213> 棘孢曲霉

<400> 34

Met Leu Gly Gln His Pro Pro Pro Asp Thr Ser Cys Ser Asp Leu Thr  
1 5 10 15

Thr Tyr Gln His Glu Leu Lys Ala Ser Lys Tyr Ser Ser Ser Thr Asn  
20 25 30

Val Ser Leu Arg Asp Arg Leu Arg His Phe Thr Trp Ala Trp Tyr Thr  
35 40 45

Leu Thr Met Ser Thr Gly Gly Leu Ala Leu Leu Leu Ala Ser Gln Pro  
50 55 60

Tyr Ser Phe Ser Gly Leu Gln Gln Ile Gly Leu Ala Val Tyr Ile Ile  
65 70 75 80

Asn Leu Ala Phe Phe Ala Leu Leu Cys Ser Leu Met Ala Ala Arg Phe  
85 90 95

Ile Leu His Gly Asn Phe Leu Asp Ser Leu Arg His Asp Arg Glu Gly  
100 105 110

Leu Phe Phe Pro Thr Phe Trp Leu Ser Ile Ala Thr Ile Ile Thr Gly  
115 120 125

Leu Tyr Arg Tyr Phe Gly Asp Thr Thr Gln Pro Ala Phe Ile Tyr Ala  
130 135 140

Leu Glu Val Leu Phe Trp Leu Tyr Cys Ala Phe Thr Leu Met Thr Ala  
145 150 155 160

Ile Ile Gln Tyr Ser Phe Val Phe Thr Ala His His Tyr Pro Leu Gln  
165 170 175

Thr Met Met Pro Ser Trp Ile Leu Pro Ala Phe Pro Ile Met Leu Ser  
180 185 190

Gly Thr Ile Ala Ser Val Ile Ala Glu Gln Gln Pro Ala Arg Ser Ala  
195 200 205

Ile Pro Met Ile Val Ala Gly Thr Thr Phe Gln Gly Leu Gly Phe Ser  
210 215 220

Ile Ser Phe Leu Met Tyr Ala His Tyr Ile Gly Arg Leu Met Glu Thr  
225 230 235 240

Gly Leu Pro Ser Arg Glu His Arg Pro Gly Met Phe Ile Cys Val Gly  
245 250 255

Pro Pro Ala Phe Thr Ala Leu Ala Leu Ile Gly Met Thr Asn Gly Leu  
260 265 270

[0031]

Pro Glu Asp Phe Gln Val Leu Gln Asp Pro His Pro Phe Gln Asp Ala  
 275 280 285

His Ile Leu Arg Leu Leu Ala Ile Ala Thr Gly Ala Phe Leu Trp Ala  
 290 295 300

Leu Ser Leu Trp Phe Phe Ser Ile Ala Ile Ile Ala Thr Ile Arg Leu  
 305 310 315 320

Pro Pro Thr Ala Phe His Leu Asn Trp Trp Ala Met Val Phe Pro Asn  
 325 330 335

Thr Gly Phe Thr Leu Ala Thr Ile Thr Leu Gly Lys Ala Phe Asp Ser  
 340 345 350

Pro Gly Val Lys Gly Val Gly Ser Ala Met Ser Ile Cys Ile Val Gly  
 355 360 365

Met Trp Leu Phe Val Phe Ala Ser Asn Ile Arg Ala Val Val Lys Arg  
 370 375 380

Asp Ile Val Phe Pro Gly Lys Asp Glu Asp Val Ser Glu  
 385 390 395

<210> 35  
 <211> 1194  
 <212> DNA  
 <213> 棘孢曲霉

<400> 35  
 atgctcgggc aacactcgcc tcccgccacc tccgctcgg accitacaac ataccaacat 60  
 gagetiaaag cctccaaaata ctctagticc accaatgigt ctctacggga cegictcgei 120  
 cattttacct gggectggta tactcigact atgagcaecg gggccctagc gcttctgetg 180  
 gccagccage cctacacell ctccggacig caacagatcg ggttgcagt ctatcatcctc 240  
 aacctggctc tctttgcttt gctgtgcage ctcatggcca cgcgcitcat tctccaagcc 300  
 aaellccctc actccctccg acacgaccgc gagggtcttt tcttcccac ttcttgctt 360  
 tccattgcaa ctatcafcac cggactctac cgtactctcg ggcacaccac acagccigca 420  
 ttcalttacg cctttaggtt gcttctcigg ctctacigig ccttcacact gatgaccgt 480  
 atcaiceaat actctttigt ctttactgce caccactacc ctctacaaac gatgatgccc 540  
 tcttggatcc tcccgcatt ccccatcatg ctaagcggca cgategecte tgcctatgcc 600  
 gaacagcago cccgcccctc tgcctatccc atgatcgtcg ccggcaccac ctccaagge 660  
 cttggctctc ccatcagttt cctcatgtac gcgcactata tgggacgct catggagacg 720  
 ggcttccgt cccgggaaca ccgaccggg atgttcatct gcttggccc ccttgccttc 780  
 accgcccctg cctaatcgg catgaccac gcccttccctg aggatitca agtctctcaa 840  
 gaccgcacc ccttcaaga ccgcataic ctccgactcc ttgcatcgc caccggcgc 900  
 ttctctggg cctcagctc ctggtcttc agcatigcca ttatcgcac caccgcctc 960  
 ccacctacg ccttcacct caactgggtg gccatggtt ttccaacac gggtttact 1020

[0032]



```

ctcgcgacca tcacgcctggg caaagcccttc gaaagcccttg gagtcaaggg cgtcggatct 1080
gccatgiccaa ttgcaicgt ggggatgtgg ctgttcgtgt ttgcgagcaa taicccgccc 1140
gttgcacaaac gggataatgt gtttcctggc aaggacgagg atgatacggg gtaa 1194

<210> 36
<211> 397
<212> PRT
<213> 棘孢曲霉

<400> 36
Met Leu Gly Gln His Ser Pro Pro Gly Thr Ser Cys Ser Asp Leu Thr
1 5 10 15

Thr Tyr Gln His Glu Leu Lys Ala Ser Lys Tyr Ser Ser Ser Thr Asn
20 25 30

Val Ser Leu Arg Asp Arg Leu Arg His Phe Thr Trp Ala Trp Tyr Thr
35 40 45

Leu Thr Met Ser Thr Gly Gly Leu Ala Leu Leu Leu Ala Ser Gln Pro
50 55 60

Tyr Thr Phe Ser Gly Leu Gln Gln Ile Gly Leu Ala Val Tyr Ile Ile
65 70 75 80

Asn Leu Val Phe Phe Ala Leu Leu Cys Ser Leu Met Ala Thr Arg Phe
85 90 95

Ile Leu His Gly Asn Phe Leu Asp Ser Leu Arg His Asp Arg Glu Gly
100 105 110

Leu Phe Phe Pro Thr Phe Trp Leu Ser Ile Ala Thr Ile Ile Thr Gly
115 120 125

Leu Tyr Arg Tyr Phe Gly Asp Thr Thr Gln Pro Ala Phe Ile Tyr Ala
130 135 140

Leu Glu Val Leu Phe Trp Leu Tyr Cys Ala Phe Thr Leu Met Thr Ala
145 150 155 160

Ile Ile Gln Tyr Ser Phe Val Phe Thr Ala His His Tyr Pro Leu Gln
165 170 175

Thr Met Met Pro Ser Trp Ile Leu Pro Ala Phe Pro Ile Met Leu Ser
180 185 190

Gly Thr Ile Ala Ser Val Ile Ala Glu Gln Gln Pro Ala Arg Ser Ala
195 200 205

Ile Pro Met Ile Val Ala Gly Thr Thr Phe Gln Gly Leu Gly Phe Ser
210 215 220

Ile Ser Phe Leu Met Tyr Ala His Tyr Ile Gly Arg Leu Met Glu Thr
225 230 235 240

```

[0033]

Gly Leu Pro Ser Arg Glu His Arg Pro Gly Met Phe Ile Cys Val Gly  
 245 250 255

Pro Pro Ala Phe Thr Ala Leu Ala Leu Ile Gly Met Thr Asn Gly Leu  
 260 265 270

Pro Glu Asp Phe Gln Val Leu Gln Asp Pro His Pro Phe Gln Asp Ala  
 275 280 285

His Ile Leu Arg Leu Leu Ala Ile Ala Thr Gly Ala Phe Leu Trp Ala  
 290 295 300

Leu Ser Leu Trp Phe Phe Ser Ile Ala Ile Ile Ala Thr Ile Arg Leu  
 305 310 315 320

Pro Pro Thr Ala Phe His Leu Asn Trp Trp Ala Met Val Phe Pro Asn  
 325 330 335

Thr Gly Phe Thr Leu Ala Thr Ile Thr Leu Gly Lys Ala Phe Asp Ser  
 340 345 350

Pro Gly Val Lys Gly Val Gly Ser Ala Met Ser Ile Cys Ile Val Gly  
 355 360 365

Met Trp Leu Phe Val Phe Ala Ser Asn Ile Arg Ala Val Val Lys Arg  
 370 375 380

Asp Ile Val Phe Pro Gly Lys Asp Glu Asp Val Ser Glu  
 385 390 395

<210> 37  
 <211> 1353  
 <212> DNA  
 <213> 日本裂殖酵母

<400> 37  
 atgtctcgg agacaccgac atactcctcc tggcatccc agaggtacaa cgaatgatt 60  
 gcaiggaacg tcaagggtcc gaggtgcccc atcgcacaga ggctcaagca cttcacgtgg 120  
 tctgtgttca cgtgtaefat ggcaacagcc ggtgtcggaa tgatcctcgc gtcctcctcc 180  
 tafaggttca cgggcttgaa caccatcgga aaggctgtct tcatittcca ggigtctctg 240  
 ttggccatct tctgttcgce catggccttc aggttcattc gctacccgga gactttcaag 300  
 aagtegatct atcaccactt ggagaaaitg ttcateggta caitcttgot ctcgatgtcg 360  
 accttcacgc atatgctcgc agcctacggc tatccttcca cgggtgaatg gatggigtac 420  
 ttgatcegaa tcttctactg gatgtacttc gccgtctcct tctctctacc gatctctgca 480  
 ttcgcaacta ctttccatct gcatccttat acctctgaaa cggcatcgcc tgcctggatc 540  
 ctcccgattt tccctcgcat gatctccgga gcagtggcag gaacctgtgc attcactcag 600  
 cctccccatc agctcaaaaa cctcgtggig tgtggcattt tgttccaggg ttgggcttc 660  
 tgggtctaca tcatgttgtt cgcggteaac atgctcaaat tgttcacaaa gggcatgatg 720

[0034]

ggagcctcgg aacgaccggg ttgttcatg ttctctggac ctccggcata cacaggcctc 780  
 gccctcatcg gfatgggcaa gaccgccatg gattccaaaa tctccatgtt ctccgccact 840  
 cccgtctcct ccgaacaccl ccgattcatg tgtaccttca tggcactcct catgtggggt 900  
 ctccgagcgt gggttattg tglggcgaig gtdgttttcg cagcagggtt catgltcagg 960  
 gcacctatcc agttcaagtt gggatggttc gggttcatct tccctgtctg ggctttctgt 1020  
 aacgtcacca tgaagatcgg cgagatgatt gactcgcag ccttcaaaa ctctggccac 1080  
 gtcctcggag ccatgttggc catccagtgg atgtctgta tgttcttcat ggtcgcagcg 1140  
 gtcctgttcg aggaatcat gtaacctgga cgggacgagg acgtcaaac accgccitga 1200  
 gccacaectc ctccgacctc cgtgacctcc cctctctctt tccatccct ccaggatgic 1260  
 aaggatggac accccatcca ggtgaaggic tcccgacctc gggatcggtc gaaacagcac 1320  
 atgtcccagg gctcggacga gaaaagatt taa 1353

<210> 38  
 <211> 1353  
 <212> DNA  
 <213> 日本裂殖酵母

<400> 38  
 atgtcttcgg aaacgccgac ttacagctct tggcgcagcc aacggtacaa cgaattgate 60  
 gcttggaaac tcaagggtcc tgcctatccc attgccccagc gctcaaaaa ctctacttgg 120  
 tctctgttta ccgttacct gccaccgggt ggtgtcggta tgaattctgg atctctgccc 180  
 taccgatcca caggcttaaa cacgatgggt aaagtcgtgt tcaatttcca ggtctgtttg 240  
 ctggcgattt tctgttcgac gatggccttt cggttcattc gttaccacca aaccttcaaa 300  
 aagctatatt accatcattt ggagaagctc ttcatggta ccttctctgt tccaatgctc 360  
 acgttcatcg atatgctcgc cgcctacgga taaccagca ctggcgagtg gatgctgtac 420  
 ctaatttcca tttttactg gatgtacttt gcgcctctgt tctataccg catcttccga 480  
 ttgtctacca cctttacat gcatccctac accctggaga cggcttcccc agcatggatt 540  
 ctgctatatt tcccagctat gattagcggc gctgtcggcg gtaactgtgc cttcacacaa 600  
 ccgcccacc aattgaagan ttggctcgtg tgggtatca tgttccaggg ctgtggtttc 660  
 tgggtgtaca tcaigtctgt cccgtgaac atgtctcagg tgtttacgaa gggtaigtat 720  
 ggigcctctg aacgccctgg tctttttatg ttctgtggc ctccggccta taccggcttg 780  
 gctttaatcg gtatgggtaa aactgctatg gactccaaga tctccatgtt ttctgcaacc 840  
 cccgtttctt ctgaacaccl tgcctttatg tgtaccttca tggccttgit fatgtggggt 900  
 ctgtctctgt ggigtctatt tctggccatg gctctgtttg ctgtgggtt catgtctctg 960  
 gctcctatcc aaitcaaacl cggctggttc gcatttatit tcccagtcgt tggttttgtc 1020  
 aacgllacta tgaagattgg tgagalgat gattcggccc cgttcaagal ctltggicat 1080  
 gtcattgttg caatgcttgc caticagtgg atgtttgtga tgttcttcat ggtccgcgcc 1140  
 gcttctatcg aagagatcat gtaccgggac cgcgacgaag atgtcaagac acctcccgtt 1200  
 gccactctc ctcccactit ggtgacgagt cctttgtcct ttgttctct gcaagacgta 1260  
 aaagatggcc atcccattca ggtaccggig tcccgacctc gagacagaag caaacagcac 1320

[0035]

```

atgtcgcagg gctctgatga agaaaaaatc tag                                1353

<210> 39
<211> 450
<212> PRT
<213> 日本裂殖酵母

<400> 39

Met Ser Ser Glu Thr Pro Thr Tyr Ser Ser Trp Ala Ser Gln Arg Tyr
1          5          10          15

Asn Glu Leu Ile Ala Trp Asn Val Lys Gly Pro Arg Leu Pro Ile Ala
20          25          30

Gln Arg Leu Lys His Phe Thr Trp Ser Trp Phe Thr Cys Thr Met Ala
35          40          45

Thr Gly Gly Val Gly Met Ile Leu Ala Ser Leu Pro Tyr Arg Phe Thr
50          55          60

Gly Leu Asn Thr Ile Gly Lys Val Val Phe Ile Phe Gln Val Val Leu
65          70          75          80

Leu Ala Ile Phe Cys Ser Ala Met Ala Phe Arg Phe Ile Arg Tyr Pro
85          90          95

Glu Thr Phe Lys Lys Ser Ile Tyr His His Leu Glu Lys Leu Phe Ile
100         105         110

Gly Thr Phe Leu Leu Ser Met Ser Thr Phe Ile Asp Met Leu Ala Ala
115         120         125

Tyr Gly Tyr Pro Ser Thr Gly Glu Trp Met Val Tyr Leu Ile Arg Ile
130         135         140

Phe Tyr Trp Met Tyr Phe Ala Val Ser Phe Val Tyr Ala Ile Phe Ala
145         150         155         160

Phe Ala Thr Thr Phe His Met His Pro Tyr Thr Leu Glu Thr Ala Ser
165         170         175

Pro Ala Trp Ile Leu Pro Ile Phe Pro Ala Met Ile Ser Gly Ala Val
180         185         190

Ala Gly Thr Val Ala Phe Thr Gln Pro Pro His Gln Leu Lys Asn Leu
195         200         205

Val Val Cys Gly Ile Met Phe Gln Gly Leu Gly Phe Trp Val Tyr Ile
210         215         220

Met Leu Phe Ala Val Asn Met Leu Lys Leu Phe Thr Lys Gly Met Met
225         230         235         240

Gly Ala Ser Glu Arg Pro Gly Leu Phe Met Phe Val Gly Pro Pro Ala

```

[0036]

	245		250		255
Tyr Thr Gly Leu Ala Leu Ile Gly Met Gly Lys Thr Ala Met Asp Ser					
	260		265		270
Lys Ile Ser Met Phe Ser Ala Thr Pro Val Ser Ser Glu His Leu Ala					
	275		280		285
Phe Met Cys Thr Phe Met Ala Leu Phe Met Trp Gly Leu Ala Ala Trp					
	290		295		300
Cys Tyr Cys Val Ala Met Val Cys Phe Ala Ala Gly Phe Met Ser Arg					
	305		310		315
Ala Pro Ile Gln Phe Lys Leu Gly Trp Phe Ala Phe Ile Phe Pro Val					
	325		330		335
Val Gly Phe Val Asn Val Thr Met Lys Ile Gly Glu Met Ile Asp Ser					
	340		345		350
Ala Ala Phe Lys Ile Phe Gly His Val Ile Gly Ala Met Leu Ala Ile					
	355		360		365
Gln Trp Met Phe Val Met Phe Phe Met Val Arg Ala Val Leu Leu Gln					
	370		375		380
Glu Ile Met Tyr Pro Gly Arg Asp Glu Asp Val Lys Thr Pro Pro Gly					
	385		390		395
Ala Thr Pro Pro Pro Thr Leu Val Thr Ser Pro Leu Ser Phe Ala Ser					
	405		410		415
Leu Gln Asp Val Lys Asp Gly His Pro Ile Gln Val Thr Val Ser Arg					
	420		425		430
Thr Arg Asp Arg Ser Lys Gln His Met Ser Gln Gly Ser Asp Glu Glu					
	435		440		445
Lys Ile					
	450				

<210> 40  
 <211> 1179  
 <212> DNA  
 <213> 棒曲霉 (Aspergillus clavatus)

<400> 40  
 atgttcgaaa atcgtataacc gccgacctcg tctcagtcag actcttgctt cctcgagaac 60  
 cagctggaaa aacaacatcg actcagccctc cgtgagaggt taaggcactt tacctgggcc 120  
 tggtaacat tgaccatgag cacaggtggg ttgctctcc tgatagcgag ccagccatac 180  
 accttcaagg ggttgaagac cattggactg gttgcttaca tctgaactt gatctgttt 240  
 ggcttgtct gttcccttat gccactagg tcatccctcc acggtgctt cctcgactcc 300  
 ctctgccaig agcgcgaggg tctttcttt cctacctctt ggetatccgt agcaaccatc 360

[0037]

atcacccggct tgcacgccta ctccggctcc gatgctcgag aatcgtaect gatfgcactc 420  
 gaagtaectct tctgggctca ctgtgctgt acactggcca cagcagtgat ccagfactcc 480  
 ttcatctctc ctggcacaag atacggcctc cagaccatga tgcctctctg gattctccca 540  
 gcttcccca tcatgctcag tggcagcatt gcctccgtca tggcgaage tcaaccgca 600  
 cggicatcga tcccgcctat catggccgga gicacctcc agggcctggg gtctctgatc 660  
 agcttcata tgtacgccca ctatatcggc cgctgatgg aatcagggt cccctgccc 720  
 gacacagac cggcaatgtt cafcctggct ggccccccgg ctctcacage cctcgcctca 780  
 gtccggatgg ccaagggcct gcccgccgag ttcaagctca tcaacgacgc acacgccctc 840  
 gaagacgcgc ggatcctcga gctgctcgea atcacccggg gcattctctt ctgggccctg 900  
 agtctgtggt tctctctcat cggcctcctc gccgtctctc ggccccccgc tactctctc 960  
 cactcctaact ggtgggctct ggtcttcccg aacacgggct tcactttggc caccatcag 1020  
 ctggaaagg catgggcag tcccggaac ttggcgcttg gttctgcat gcccctggc 1080  
 atcgttggca tgggctggt tgttttctg agccatactc gtcctatcat caaccaggat 1140  
 atcatgtatc cgggcaaaga tgaggatgct gcagactag 1179

<210> 41  
 <211> 392  
 <212> PRT  
 <213> 棒曲霉(Aspergillus clavatus)

<400> 41

Met Phe Glu Asn Arg Ile Pro Pro Thr Ser Ser Gln Ser Asp Ser Gly  
 1 5 10 15

Phe Leu Glu Asn Gln Leu Glu Lys Gln His Arg Leu Ser Leu Arg Glu  
 20 25 30

Arg Leu Arg His Phe Thr Trp Ala Trp Tyr Thr Leu Thr Met Ser Thr  
 35 40 45

Gly Gly Leu Ala Leu Leu Ile Ala Ser Gln Pro Tyr Thr Phe Lys Gly  
 50 55 60

Leu Lys Thr Ile Gly Leu Val Val Tyr Ile Val Asn Leu Ile Leu Phe  
 65 70 75 80

Gly Leu Val Cys Ser Leu Met Ala Thr Arg Phe Ile Leu His Gly Gly  
 85 90 95

Phe Leu Asp Ser Leu Arg His Glu Arg Glu Gly Leu Phe Phe Pro Thr  
 100 105 110

Phe Trp Leu Ser Val Ala Thr Ile Ile Thr Gly Leu His Arg Tyr Phe  
 115 120 125

Gly Ser Asp Ala Arg Glu Ser Tyr Leu Ile Ala Leu Glu Val Leu Phe  
 130 135 140

[0038]

Trp Val Tyr Cys Ala Cys Thr Leu Ala Thr Ala Val Ile Gln Tyr Ser  
145 150 155 160

Phe Ile Phe Ser Ala His Arg Tyr Gly Leu Gln Thr Met Met Pro Ser  
165 170 175

Trp Ile Leu Pro Ala Phe Pro Ile Met Leu Ser Gly Thr Ile Ala Ser  
180 185 190

Val Ile Gly Glu Ala Gln Pro Ala Arg Ser Ser Ile Pro Val Ile Met  
195 200 205

Ala Gly Val Thr Phe Gln Gly Leu Gly Phe Ser Ile Ser Phe Met Met  
210 215 220

Tyr Ala His Tyr Ile Gly Arg Leu Met Glu Ser Gly Leu Pro Cys Arg  
225 230 235 240

Glu His Arg Pro Gly Met Phe Ile Cys Val Gly Pro Pro Ala Phe Thr  
245 250 255

Ala Leu Ala Leu Val Gly Met Ala Lys Gly Leu Pro Ala Glu Phe Lys  
260 265 270

Leu Ile Asn Asp Ala His Ala Leu Glu Asp Ala Arg Ile Leu Glu Leu  
275 280 285

Leu Ala Ile Thr Ala Gly Ile Phe Leu Trp Ala Leu Ser Leu Trp Phe  
290 295 300

Phe Phe Ile Ala Val Ile Ala Val Leu Arg Ser Pro Pro Thr Ser Phe  
305 310 315 320

His Leu Asn Trp Trp Ala Leu Val Phe Pro Asn Thr Gly Phe Thr Leu  
325 330 335

Ala Thr Ile Thr Leu Gly Lys Ala Leu Gly Ser Pro Gly Ile Leu Gly  
340 345 350

Val Gly Ser Ala Met Ser Leu Gly Ile Val Gly Met Trp Leu Phe Val  
355 360 365

Phe Val Ser His Ile Arg Ala Ile Ile Asn Gln Asp Ile Met Tyr Pro  
370 375 380

Gly Lys Asp Glu Asp Ala Ala Asp  
385 390

- <210> 42
- <211> 1182
- <212> DNA
- <213> 烟曲霉 (Aspergillus fumigatus)

<400> 42  
atgttcaacg atcatgatca tgttcacca acatcatcac agteggattc tggctttttt 60

[0039]

gaacaagaaa tgaagaaatc tccctcgacta agccttcgtg agcgcctacg gcacttcacc 120  
 tggcgctggg ataccttgac gatgagfacg ggtggactgg ctctctgat igctagtcag 180  
 ccglatacct tcaatggcat gaaggcacc gggatggctg tttatatcct caatctctcg 240  
 ttatctctc tttctcttctc tttgatggg ctgagattcg ttttgcattg cggtttctt 300  
 gagacttgc gccaccctcg cgaaggctc tcttcccta ccttctggt aiccattgca 360  
 acgatcatca ctgcttga tcttacttc ggcctccgac acctagagtc gtaacctc 420  
 gcactcgaag tctctctctg ggtctactgt agttgcacc tcgccacagc tctgattcag 480  
 tactcattcc tcttctgccc ccaactctac ggctgcaga caatgatcc atcattgac 540  
 ctaccagcct tcccattcal gctcagcggg accatccct cggctcagc cgaatccag 600  
 cccgcggat ccgcgatccc cactcactt gccggcgta ccttccaggg cctcggctc 660  
 tcaatcagct tcataatga cgcaccctac atcgccgac tcatcagtc agggcttccc 720  
 tgcgcgaac acagaccagc catgttcait tgcgtgggc ctccgtctt caccgctt 780  
 gcgctagtac ggtggcca gggctgccc gacgaattca agataatcaa agaagcacac 840  
 gctgaggacg cccgatcct cgaactgat gctattatcg tcggctgtt cctgtggcc 900  
 ctgactctct ggttctctt cattgcttct gttctctctg tccggtgccc gccactg 960  
 ttcacacta gctgggggc catggtctt cccaacactg ggttccgct ggcactatt 1020  
 accctgggga gggcattgg gagccctggc gcttgggagc tggcctggc catgtcggc 1080  
 ggtgtgtct gcaitgggt cttcttctt gctaccaca ttcgtctgt cactaggca 1140  
 gacatcatgt acccgggcaa agacaggat gctctagatt aa 1182

<210> 43  
 <211> 393  
 <212> PRT  
 <213> 烟曲霉 (Aspergillus fumigatus)  
 <400> 43

Met Phe Asn Asp His Asp His Val Pro Pro Thr Ser Ser Gln Ser Asp  
 1 5 10 15  
 Ser Gly Phe Phe Gln Gln Glu Met Lys Lys Ser Pro Arg Leu Ser Leu  
 20 25 30  
 Arg Gln Arg Leu Arg His Phe Thr Trp Ala Trp Tyr Thr Leu Thr Met  
 35 40 45  
 Ser Thr Gly Gly Leu Ala Leu Leu Ile Ala Ser Gln Pro Tyr Thr Phe  
 50 55 60  
 Asn Gly Met Lys Gly Ile Gly Met Val Val Tyr Ile Leu Asn Leu Leu  
 65 70 75 80  
 Leu Phe Ala Leu Val Cys Ser Leu Met Val Leu Arg Phe Val Leu His  
 85 90 95  
 Gly Gly Phe Leu Asp Ser Leu Arg His Pro Arg Gln Gly Leu Phe Phe

[0040]





385 390

<210> 44  
 <211> 1294  
 <212> DNA  
 <213> 米曲霉(*Aspergillus oryzae*)

<400> 44  
 atgttcgcig ctccagcagc ttcaaacctc cccagaagc ggccttctc cgcctctgcc 60  
 agccagggtg gtgattgaat ggatccattg gacctcggag ctagnetctg aacatcaaca 120  
 aaaactaaca acfaacttat cttcttcata ggttccaag gttgccttic ttggtgccgc 180  
 tgggtggcatt ggcagcctc tctccctct cctcaagctc aacccccgig ttctgagct 240  
 tgcctctctc galatccgag gtggcccagg tatgttttg cacagcttgc aacatctccg 300  
 acctcggiga ttcaagacag ggctaacata aggaatacaat agggilgcc gcigacciga 360  
 gccacataca caccacacag accgtctctg gctacgagge taccctctct ggctcccgig 420  
 atgcctctca ggctccgag atcgtctctc tccctgcagg tgttctctgc aagcccggca 480  
 tgaccctgga cggtaigaac cgttaacttg tcaatggcac tgggaatiga atactaaita 540  
 taatatgcc agaccigtg aacaccaacg cctccatgtt ccgagacctt gctaaagccg 600  
 ccgcccagge ttcccccgag gccaacatcc tctctctctc caacctctga tgaagcttcc 660  
 caccactctc taccagtta cctcgcctaa ttgcaatcag gtaacttcca ccgtccctat 720  
 cgtctctgag gctctcaagt ccaagggigt ctacaacccc aagcgtctct tgggtgtcac 780  
 tacccttgac gtgtccgtg cctctctctt cctctctcag gtcagaaga ccgacctctc 840  
 caacagagcc gtcactctg tgggtgtgca ctcgggtgtg accatgtctc cctctctctc 900  
 ccagctccag caccaccaga ttgagggtaa gaccctcctg gactctctca accctctca 960  
 gttcgggtgt gatgaggtg tcaaggccaa ggatgggtct ggtctctcca cctctctcat 1020  
 ggccatggct gggtctctga tggctgagtc cctctctgaag gccgcccagg gtgagaaggg 1080  
 tgtctgttag cccacttctg tctctctcag gaccaggtgt ttgacttctt 1140  
 cgcctccaag gctgagctg gccccaacgg lgttgagaag atcctccccg ttggccaggt 1200  
 caacgctctc gaggagaagc tctctcaggg ctccttgggt gacctcaaga agaacaacca 1260  
 gaaggttatt gactctctca agccaaccc ttaa 1294

<210> 45  
 <211> 340  
 <212> PRT  
 <213> 米曲霉(*Aspergillus oryzae*)

<400> 45

Met Phe Ala Ala Arg Gln Ser Phe Asn Leu Leu Gln Lys Arg Ala Phe  
 1 5 10 15

Ser Ala Ser Ala Ser Gln Ala Ser Lys Val Ala Val Leu Gly Ala Ala  
 20 25 30

Gly Gly Ile Gly Gln Pro Leu Ser Leu Leu Leu Lys Leu Asn Pro Arg  
 35 40 45

[0042]

Val Ser Glu Leu Ala Leu Tyr Asp Ile Arg Gly Gly Pro Gly Val Ala  
 50 55 60

Ala Asp Leu Ser His Ile Asn Thr Asn Ser Thr Val Ser Gly Tyr Glu  
 65 70 75 80

Ala Thr Pro Ser Gly Leu Arg Asp Ala Leu Lys Gly Ser Glu Ile Val  
 85 90 95

Leu Ile Pro Ala Gly Val Pro Arg Lys Pro Gly Met Thr Arg Asp Asp  
 100 105 110

Leu Phe Asn Thr Asn Ala Ser Ile Val Arg Asp Leu Ala Lys Ala Ala  
 115 120 125

Ala Glu Ala Ser Pro Glu Ala Asn Ile Leu Val Ile Ser Asn Pro Val  
 130 135 140

Asn Ser Thr Val Pro Ile Val Ser Glu Val Phe Lys Ser Lys Gly Val  
 145 150 155 160

Tyr Asn Pro Lys Arg Leu Phe Gly Val Thr Thr Leu Asp Val Val Arg  
 165 170 175

Ala Ser Arg Phe Ile Ser Gln Val Gln Lys Thr Asp Pro Ser Asn Glu  
 180 185 190

Ala Val Thr Val Val Gly Gly His Ser Gly Val Thr Ile Val Pro Leu  
 195 200 205

Leu Ser Gln Ser Ser His Pro Ser Ile Glu Gly Lys Thr Arg Asp Glu  
 210 215 220

Leu Val Asn Arg Ile Gln Phe Gly Gly Asp Glu Val Val Lys Ala Lys  
 225 230 235 240

Asp Gly Ala Gly Ser Ala Thr Leu Ser Met Ala Met Ala Gly Ala Arg  
 245 250 255

Met Ala Glu Ser Leu Leu Lys Ala Ala Gln Gly Glu Lys Gly Val Val  
 260 265 270

Glu Pro Thr Phe Val Asp Ser Pro Leu Tyr Lys Asp Gln Gly Val Asp  
 275 280 285

Phe Phe Ala Ser Lys Val Glu Leu Gly Pro Asn Gly Val Glu Lys Ile  
 290 295 300

Leu Pro Val Gly Gln Val Asn Ala Tyr Glu Glu Lys Leu Leu Gln Ala  
 305 310 315 320

Cys Leu Gly Asp Leu Lys Lys Asn Ile Gln Lys Gly Ile Asp Phe Val  
 325 330 335

[0043]

Lys Ala Asn Pro  
340

<210> 46  
<211> 20  
<212> DNA  
<213> 米曲霉(*Aspergillus oryzae*)

<400> 46  
tgaccttcca cgcctgaccac 20

<210> 47  
<211> 20  
<212> DNA  
<213> 米曲霉(*Aspergillus oryzae*)

<400> 47  
ggctgagaaa aiatgttgca 20

<210> 48  
<211> 30  
<212> DNA  
<213> 米曲霉(*Aspergillus oryzae*)

<400> 48  
gatagaccac taatcatggt ggcgatggag 30

<210> 49  
<211> 30  
<212> DNA  
<213> 米曲霉(*Aspergillus oryzae*)

<400> 49  
tgcggctctg agtcaggccc agttgctcga 30

<210> 50  
<211> 20  
<212> DNA  
<213> 米曲霉(*Aspergillus oryzae*)

<400> 50  
gggatttgaa cagcagaagg 20

<210> 51  
<211> 20  
<212> DNA  
<213> 米曲霉(*Aspergillus oryzae*)

<400> 51  
tcacaaaaga gtagaggcca 20

<210> 52  
<211> 41  
<212> DNA  
<213> 人工序列

<220>  
<223> 人工 DNA 引物

<400> 52  
gtgatagaac atcgtccata atgccgggeg atcicaaaac c 41

<210> 53

[0044]

<211> 39	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工 DNA 引物	
<400> 53	
gtgtcagfca cctctagtia ctatgcatca aggacattc	39
<210> 54	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工 DNA 引物	
<400> 54	
gattgagaic ggcatttact	20
<210> 55	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工 DNA 引物	
<400> 55	
acgcggaaca gcagaatggc	20
<210> 56	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工 DNA 引物	
<400> 56	
ctatagcga aatggattgat tgiat	25
<210> 57	
<211> 20	
<212> DNA	
<213> 人工序列	
<220>	
<223> 人工 DNA 引物	
<400> 57	
ttcacctga aacgtattga	20

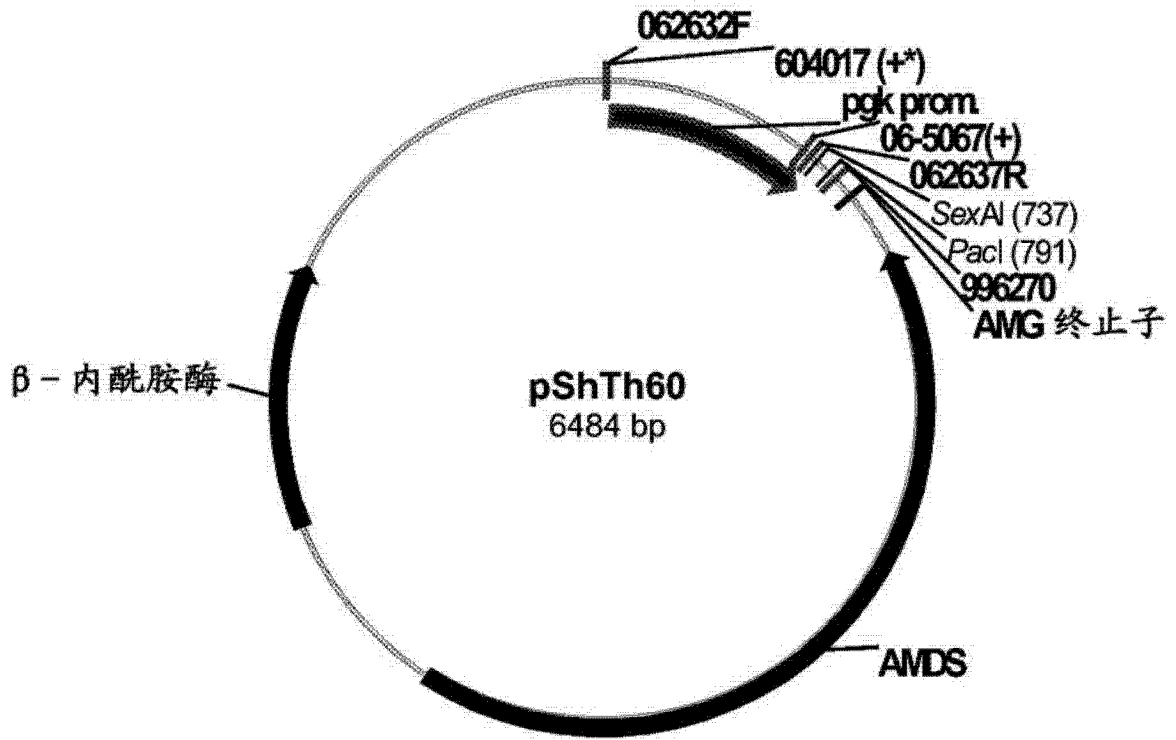


图 1

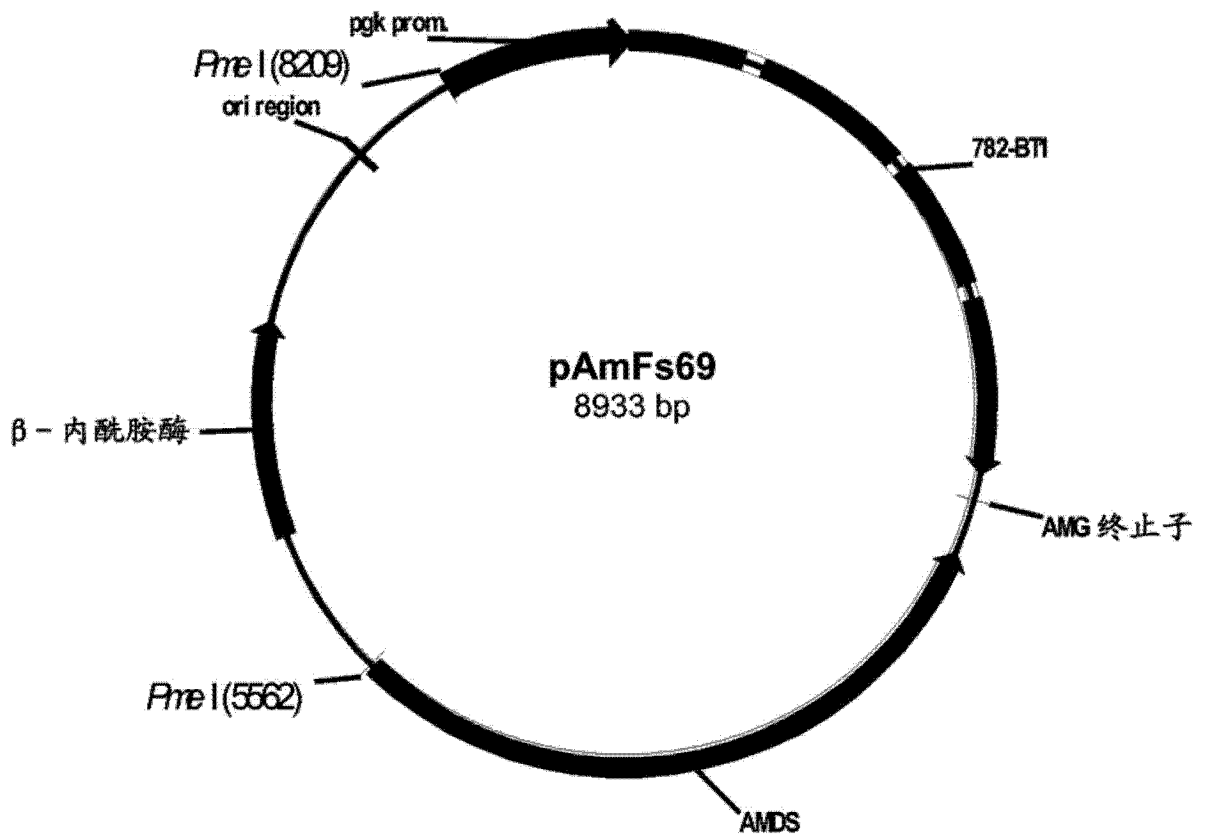


图 2

```

M E S S A V Q E P T Q Q R S L R D R I F
1 ATGGAATCCA GCGCTGTACA GGAGCCGACT CAACAGCGCT CTTTGGGGGA TCGCATTTTT
N L F R T S S S N D A P G L P A R L V T
61 AACCTCTTTC GTACCTCTTC CTCAAATGAT GCCCGGGTC TTCCGGCAAG ACTCGTAACC
A E S A A Q N E G S A L I Y P P R E P D
121 GCTGAGAGCG CAGCGCAAAA CGAAGGGTCG GCGTTAATCT ATCCGCCACG GGAGCCTGAT
A R T R L L E S Y D R G E R G L R N S G
181 GCAAGGACTC GTCTTCTCGA ATCGTACGAT CGCGGGGAAC GTGGTCTGAG GAACTCCGGC
V H G T F S S R P E Q E E I Q K W D A S
241 GTTCATGGGA CTTTTCTTTC ACGACCTGAA CAGGAAGAAA TCCAAAAATG GGATGCAAGC
S L Q N A G N E E R S Q S P G G A D G H
301 TCCTTGCAGA ATGCTGGTAA CGAAGAAAGA TCTCAGTCCC CAGGAGGAGC AGACGGCCAT
I G S P G D V S G Y P Q G P E N I P S L
361 ATTGGGTCTC CCGGCGACGT CTCAGGATAC CCACAGGGAC CAGAGAATAT ACCATCGCTA
D S S F T A L H M K N H K S L
421 GACTCCTCTT TCACAGCATT GCACATGAAG AATCATAAAT CTCTGTAGGT TTATAATCAC

481 GTTCGCCCTG CTTTCTAACA CAATTGTTAT CTCCATCGTG GAGACAATA ACCTTCATCA
Y I S Y Y I P F F N W I T Q Y R W S Y
541 AGGTATATAT CTTACTACAT CCCATTTTTT AATTGGATTA CTCAATACCG GTGGTCTGAT
I R G D L V A A T T I A S I Y I P M A L
601 ATTCGAGGTG ATTTGGTTGC TCGCAACACC ATTGCGTCCA TCTATATCCC TATGGCTTTG
S L S S N L A H A P P I N G L Y S F V I
661 TCCTTATCCT CAAATCTCGC CCACGCACCT CCTATCAATG GCCTCTACTC TTTTGTGATC
N P F I Y A I F G S S P L L I V G P E A
721 AACCCTTTCA TCTATGCGAT CTTCCGGGAGC AGCCCGCTGT TAATAGTGGG CCCAGAAGCA
A G S L L T G T I V K T S V R P G P S G
781 GCAGGCTCCT TGCTTACTGG CACGATTGTC AAAACTAGTG TCAGACCAGC CCCATCTGGT
E D D E V A N A I V V G I A T A M A G A
841 GAGGACGACG AAGTAGCGAA TGCCATCGTG GTCGGCATAG CCACTGCAAT GCGGGGCGCC
M I L I A G L T R L G F L D N V L S R P
901 ATGATACTGA TCGCTGGGCT TACACGGCTG GGATTTCTGG ACAATGTGCT GAGCCGGCCC
F L R G F I T A I G F V I F V D Q L I P
961 TTTCTTAGGG GTTTCATTAC AGCGATCGGT TTTGTGATTT TTGTGGATCA ACTCATCCCC
E V G L T E L A K E A G V T H G T V D
1021 GAAGTCGGAT TGACCGAGCT AGCAAAGGAA GCTGGTGTTA CCCATGGGAC TACAGTTGAC
K L M F L I R N I G G C H A L T T A V A
1081 AAGCTCATGT TCCTTATAAG AAACATAGGA GGTGCCATG CGCTTACAAC CGCGGTGGCT
F G S F A I I M V F R
1141 TTTGGGAGCT TTGCTATTAT AATGGTATTT CGGTTAGTGT TGGTGACTCG GAAGCCTGGT
T L K K M L Q P R Y P Q
1201 GCTTAGACTG ATTACCATTA CAGGACTCTC AAGAAAATGC TCCAGCCGCG GTATCCTCAG
V I Y L P D R I L V V I L S A V L T W H
1261 GTGATTTATC TTCCGGACCG AATTCTCGTA GTTATTCTTT CAGCCGTCCT GACATGGCAT
L G W D D K G L E I L G P L K Q N A N G
1321 CTTGGTTGGG ATGACAAAGG GTTGGAGATT CTTGGGCCCT TGAACAATA TGCCAATGGC
L F A F K W P F Q F S Q M K H V R A A M
1381 CTTTTGCGT TCAAATGGCC TTTCCAGTTT AGCCAGATGA AGCATGTACG CGCTGCAATG
S T S F V I A L L G F F E S S V A A K G
1441 AGTACTTCTT TCGTCATCGC GTTACTTGGC TTTTTCGAGT CTTCTGTTGC CGCCAAGGGA
L S G E A R Q E G V Q G M P V S A N R E
1501 CTTAGTGGCG AGGCCAGACA AGAAGGTGTC CAGGGAATGC CTGTCAAGTC TAACAGAGAG
M V A L G L A N T V G G C F M A L P A F
1561 ATGGTGGCGC TGGTCTTGC TAATACTGTG GGGGCTGTT TCATGGCGCT TCCTGCGTTT

```

图 3A

```

      G G Y A R S K V N A S T G A R S P M S S
1621 GGTGGCTATG CAAGAAGCAA AGTCAACGCT TCAACTGGAG CTCGGTCTCC GATGAGCAGC
      I F L S I I T F V C I M V L L P Y L Y Y
1681 ATTTTCCTGA GCATTATTAC CTTTGTGTTGT ATCATGGTGC TTTTGCCGTA CTTATACTAT
      L P
1741 CTTCGGTGA GTCTCGACCC CAAATACTTC CGAGCGAAGG CTGAGAAAAT ATGTTGCAAT
      K A V L S S M I S V V A F S L I E E
1801 AATTCAGAAA GCCGTTCTTT CTTCTATGAT ATCTGTCGTC GCATTCAGTC TCATTGAAGA
      C P H D V A F F I R L R G W T E L A L M
1861 ATGTCCTCAC GACGTGGCTT TCTTTATCCG ACTGCGCGGA TGGACGGAGC TAGCCCTAAT
      L L I F V S T I F Y S L E L G I A L G I
1921 GCTTCTCATC TTTGTCTCGA CTATTTTCTA TTCTCTAGAG CTGGGAATTG CCCTTGGTAT
      G L S I L I L I R H S T Q P R I Q I L G
1981 TGGCCTTTCT ATCTTGATCC TTATTCGCCA TTCTACGCAG CCTCGGATCC AAATTCTGGG
      K I A G T T D R F D N A E L H P E S V E
2041 TAAGATAGCA GGCCTACCG ACCGTTTTCGA TAACGCTGAA CTCCACCCCG AGAGCGTTGA
      L I E G A L I V K I P E P L T F A N T G
2101 GTTAATCGAA GCGCGCTTA TTGTTAAGAT CCGGAACCG CTCACCTTTG CCAATACTGG
      E L K N R L R R L E L Y G S S R A H P S
2161 TGAGCTCAAG AATCGTCTTC GCGGTTTGA ATTATATGGC AGTAGCCGAG CGCACCTTTC
      L P P T R T P E H N K N I I F D V H G V
2221 TCTTCCCCC ACGCGCACCC CCGAACATAA CAAGAATATT ATATTTGATG TTCATGGTGT
      T S I D G S G T Q V L Y E I V D G Y A D
2281 TACTAGCATC GATGGTCCG GTACGCAAGT CTTATATGAG ATTGTGGACG GATATGCAGA
      Q G V S V F F C R V A T R N V F R M F E
2341 CCAGGGGGTC AGCGTCTTCT TCTGCCCGT CGCAACTCGC AATGTTTTCC GCATGTTTGA
      R S G I V E R C G G I T H F V H G V D E
2401 ACGAAGTGGA ATTGTGGAAC GATGCGGTGG GATAACGCAC TTCGTTTCATG GTGTCGACGA
      A L R L A E S E D E I E I *
2461 AGCCCTCCGC CTTGCCGAAT CGGAAGACGA GATTGAAATC TGA

```

图 3B



```

M P G D L K T K I G H G A A K A L G I K
1 ATGCCGGGCG ATCTCAAAC CAAATTGGT CACGGCGCGG CCAAGGCCCT GGGGATCAAG
I P Y R D P L G V H A D P V T R G E S M
61 ATCCCCTACC GTGATCCTCT CGGAGTTCAT GCTGACCCAG TCACACGAGG CGAGTCGATG
F S V G T I D T Y S Y L E P E P T P A E
121 TTCTCCGTCG GAACGATCGA CACATACTCC TATCTCGAGC CCGAACCCAC TCCCGCTGAA
W L K E V C P S W H Q V G R Y F Y N L F
181 TGGCTGAAGG AAGTCTGCCC TAGCTGGCAT CAGGTGGGCC GTTATTTTTTA CAACCTTTTC
P F L S W I T R Y N L Q W L L G D M I A
241 CCTTTCCTCT CGTGGATTAC GAGGTACAAC TTGCAATGGT TGCTGGGAGA TATGATTGCC

301 GGTAAGAGCC TTTCCACTGT GTTTGATTG ATCGACAAGT AGACAACATA CTCATTGGAA
G V T V G A V V V P Q G M A Y A K L A
361 TGCAAGCGTC ACGGTCCGGT CTGTGGTCGT TCCGAGGGA ATGGCCTACG CTAAACTGGC
N L P V E Y G L Y S S F M G V L I Y W F
421 AAACCTACCT GTAGAGTATG GTCTCTATTC CTCGTTTCATG GGTGTTCTCA TTTATTGGTT
F A T S K D I T I G
481 TTTTGCCACC TCAAAGGATA TCACCATTGG TGTAAGTCAT TCTGCACCCA TGTCAGCATG
P V A V M S T L T

541 TATCTTGCTA ATATAGTATC TTCCCTGTTC AGCCGGTGGC TGTCATGTCT ACCCTTACAG
G K I V A E A Q T K L P D V E G H V I A
601 GTAAGATAGT TGCCGAGGCG CAAACGAAGC TCCCAGATGT CGAAGGGCAT GTAATCGCCT
S C L A I I C G A V V C A M G L L R L G
661 CCTGTTTGGC TATCATTGT GGAGCCGTGG TTTGCGCTAT GGGCCTGCTT CCGGCTGGAT
F I V D F I P L P A I S A F M T G S A I
721 FTATCGTGA TTTCATTCCCT CTGCCGGCAA TTTCAGCTTT CATGACGGGT TCCGCCATCA
N I C S G Q V K D M L G E T A D F S T K
781 ATATCTGCTC CGGACAGGTC AAAGACATGC TGGGAGAGAC GGCCGACTTC TCGACGAAAG
D S T Y L V I I N T L K H L P S A K I D
841 ATTCTACCTA TCTGGTTATC ATCAACACCC TCAAGCATCT TCCCTCCGCA AAAATCGATG
A A M G V S A L A M L Y I I R S G C N Y
901 CGCCATGGG TGTCAGTGCT TTAGCTATGC TGTACATTAT CCGTTCGGGT TGCAATTATG
G A K K F P R H A K V W F F V S T L R T
961 GCGGAAGAA GTTCCCCCGT CATGCCAAGG TTTGGTTCTT CGTTTCGACT TTGCGCACAG
V F V I L F Y T M I S A A V N L H R R S
1021 TGTTCTGAT CTTGTTCTAT ACGATGATCA GTGCCGCTGT GAACTTGCAC CGGCGGTCTA
N P R F K L L G K V P R G F Q H A A V P
1081 ACCCGCGGTT CAAGCTCCTG GGTAAGTTC CTCGTGGTTT CCAACATGCG GCTGTCCCTC
Q V N S R I I S A F A S E L P A S I I V
1141 AGGTAAATTC GAGGATCATC AGCGCATTGG CTAGCGAACT TCCTGCTTCG ATTATTGTCC
L L I E H I A I S K S F G R V N N Y T I
1201 TGCTTATCGA ACACATCGCT ATCTCGAAAT CCTTTGGCCG TGTCACAAC TACACAATTG
D P S Q E L V A I G V S N L L G P F L G
1261 ATCCCTCTCA GGAGCTGGTT GCTATTGGTG TGTCGAACTT GCTTGGACCG TTCCTTGGTG
G Y P A T G S F S R T A I K S K A G V R
1321 GTTACCCAGC GACTGGATCG TTCTCCCGAA CTGCAATCAA ATCGAAAGCG GGTGTCCGCA
T P L A G V I T A V V V L L A I Y A L P
1381 CCCCACTTGC CGGTGTATT ACTGCGGTTG TTGTCCTCCT CGCCATTTAC GCTCTGCCCG
A V F F Y I P K A S L A G V I I H A V G
1441 CTGTCTTCTT TTACATCCCG AAAGCTTCCC TTGCTGGTGT CATCATTCAT GCAGTCGGTG
D L I T P P N T V Y Q F W R V S P L D A
1501 ACCTCATTAC CCCACCAAAC ACCGTTTACC AGTTCGGCG CGTGTCCCCT CTGGATGCGA
I I F F I G V I V T V F T T I E I G I Y
1561 TCATTTTCTT TATCGGTGTT ATCGTGACTG TCTCACCAC GATTGAGATC GGCATTTACT
C T V C V S V A I L L F R V A K A R G Q
1621 GTACCGTTTG TGTGTCTGTT GCCATTCTGC TGTTCGCGT CGCCAAGGCC CGCGGTCAAT

```

图 4A

```

      F L G R V T I H S V I G D H L V Q D D G
1741 TCTTAGGAAG AGTCACTATC CACTCGGTGA TCGGTGACCA TCTGGTACAG GATGATGGGA
      K Y G S A N S P N A A S D D K D E L S R
1801 AATATGGGTC TGCCAACTCC CCTAATGCTG CCAGCGATGA CAAAGATGAA TTGAGCCGGT
      S I F L P I N H T D G S N P D V E V Q Q
1861 CTATCTTCTT GCCTATCAAC CACACGGACG GATCGAATCC CGATGTCGAG GTGCAGCAAC
      P Y P G I F I Y R F S E G F N Y P N A N
1921 CTTATCCTGG TATCTTCATC TACCGATTCT CGGAAGGATT CAACTACCCC AATGCCAATC
      H Y T D Y L V Q T I F K H T R R T N P F
1981 ACTACACCGA TTATTTGGTC CAGACTATCT TCAAGCATA CCGTCGCACA AATCCGTTCT
      S Y G K P G D R P W N N P G P R R G K S
2041 CCTACGGTAA ACCGGGTGAT CGGCCATGGA ATAATCCTGG CCCTCGCAGG GGCAAGTCTG
      E D D E S H L P L L Q A V I L D F S S V
2101 AAGATGACGA GTCGCATTG CCCTTACTGC AGGCTGTCAT TCTTGACTTC TCATCCGTCA
      N N V D V T S V Q N L I D V R N Q L D L
2161 ACAATGTTGA TGTGACCTG GTCCAGAACC TCATCGATGT CCGCAATCAA CTCGACCTCT
      Y A S P K T V Q W H F A H I N N R W T K
2221 ACGCTTCGCC TAAGACTGTG CAGTGGCACT TTGCTCATAT TAACAACCGC TGGACGAAAC
      R A L A A A G F G F P S P D S D E G F Q
2281 GAGCCCTTGC AGCAGCAGGT TTCGGCTTCC CATCTCCGGA CTCGGATGAA GGATTCCAGA
      R W K P I F S V A E I E G S A S A A A H
2341 GATGGAAGCC AATTTTCAGC GTGGCTGAGA TCGAAGGCAG TGCCTCTGCC GCAGCTCATG
      A E M V N N R H T Q H N I K S E D L E H
2401 CAGAGATGGT GAACAACAGA CACACCCAGC ATAACATCAA GAGCGAAGAC CTCGAGCATG
      G L K H D S E T T E R E T H G I E E S S
2461 GCCTCAAGCA CGATTCAGAG ACCACCGAGC GTGAGACACA CGGCATCGAA GAATCCTCCG
      D A S S T R E D K L Q R D L K D S K A Y
2521 ATGCCAGCAG CACCCGGGAG GACAAGTTGC AACCGGACCT GAAGGATAGC AAGGCTTACC
      R S R R R V A M V Q G L N R P F F H I D
2581 GCAGTCGCCG AAGGGTCGCT ATGGTGCAGG GCCTCAACCG GCCATTCTTC CACATCGACC
      L T S A L Q S A L A N A G E Q P D P K M
2641 TGACTAGTGC ACTGCAGAGT GCCTTGGCCA ACGCGGGCGA GCAGCCGGAC CCTAAAATGA
      N V L D A *
2701 ATGTCCTTGA TGCATAG

```

图 4B

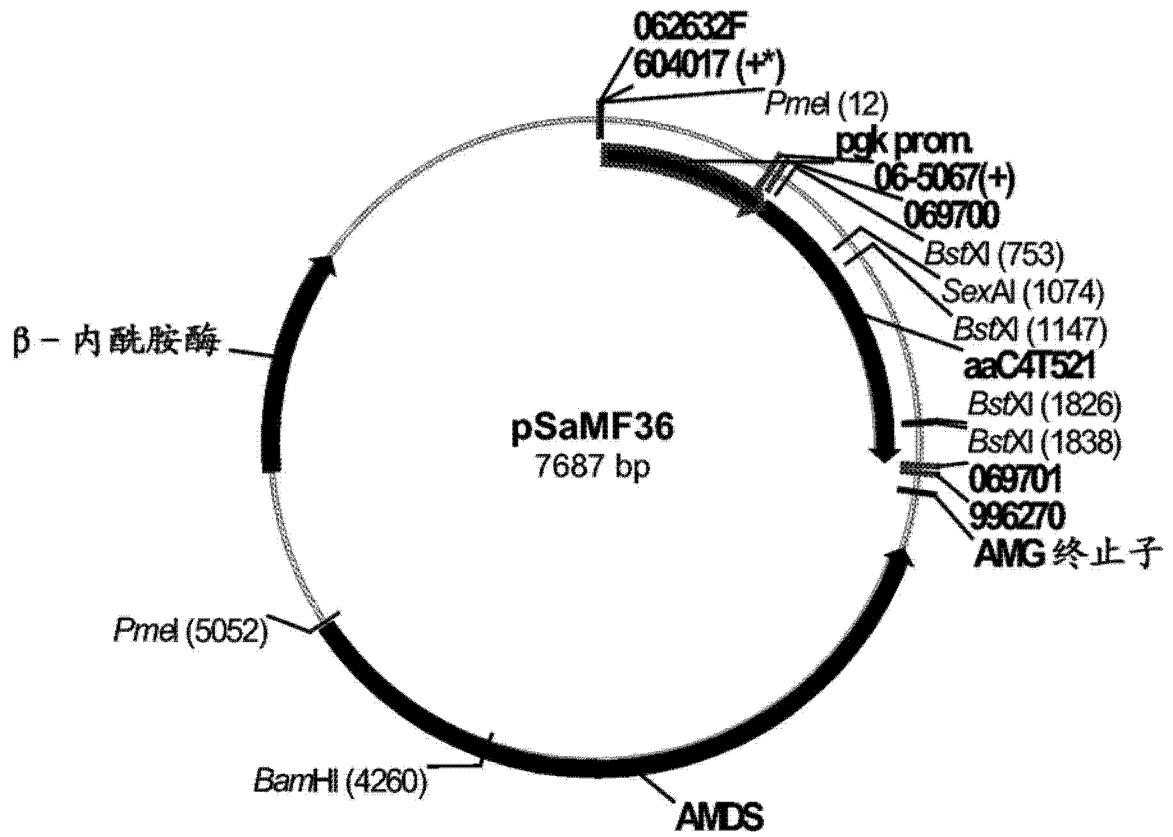


图 5

```

      M H D H S T G S S P Y I S D V E T L N H
1  ATGCACGACC ACAGCACTGG ATCTAGTCCA TACATCTCGG ACGTGGAAC CTTGAACCAC
   A C E K S V N P E T K V S Q P Q E S P I
61 GCCTGCGAGA AGTCCGTCAA CCCCAGAGACC AAAGTCTCCC AGCCTCAGGA ATCTCCATT
   I S N N E H Q E F V K L G I R Q R L R H
121 ATCAGCAATA ATGAACATCA GGAGTTTGTT AAGCTGGGCA TCCGCCAACG GCTGCGTCAT
   F T W A W Y T L T M S A G G L A L L L R
181 TTCACCTGGG CCFGGTATAC CCTAACCATG AGCGCAGGTG GACTGGCCCT TCTTCTCCGC
   N Q P Y Q F K G L K E I G L V V Y I A N
241 AACCAGCCGT ATCAATTCAA GGGGTTGAAG GAGATAGGCC TGGTGGTATA CATAGCCAAT
   L V F F T I I G S L M I T R F V L Y N N
301 CTGCTCTTCT TTACTATCAT CGGCTCTCTT ATGATCACCA GGTGTTGTTCT TTACAACAAC
   L M D S L R H D R E G F F F P T F W L S
361 CTGATGGACT CTCTCCGCCA CGACCGAGAA GGTTCCTTCT TTCCAACCTT CTGGCTCTCC
   I A T M I S G L S A Y F S T E D T H R L
421 ATCGCCACCA TGATTAGTGG TCTATCTGCC TACTTCTCTA CTGAAGACAC GCACCGCCTC
   N Y A L E G L F W A Y C I F T F A S A V
481 AATTATGCTC TCGAGGGTCT CTTCTGGGCG TACTGTATCT TCACGTTTGC CTCAGCAGTG
   I Q Y S F V F S Y H T F P L Q T M M P S
541 ATCCAGTACT CCTTTGCTTT CTCCTATCAC ACGTTCCTC TCAAACATAT GATGCCATCA
   W I L P A F P I M L S G T I A S A A S S
601 TGGATCTTAC CGCATTCCC TATCATGCTG AGCGGAACCA TTGCCTCTGC CGCTTCCAGC
   Y Q P A V S A T P M I V A G I T F Q G L
661 TACCAGCCTG CGGTGTCTGC CACGCCTATG ATTGTTGCCG GCATCACGTT CCAGGGACTC
   G F C I S F M M Y A H Y I G R L M E T G
721 GGATTCGCA TCAGCTTCAT GATGTACGCC CACTACATCG GCGTCTGAT GGAGACGGGC
   I P S S E H R P G M F I C V G P P A F T
781 ATCCCTTCGA GCGAGCACCG TCCPGGTATG TTCATCTGTG TCGGCCCCCC TGCCTTCACG
   L L A I I G M A N G L P E G F S I L G D
841 CTGCTGGCTA TCATCGGCAT GGCCAACGGC CTTCCCAGAG GCTTCAGTAT CCTGGGCGAT
   G G M D D R H I M R V L A V C A G M F L
901 GGTGGCATGG ACGACCGTCA CATCATGCGA GFACTGGCCG TCTGCGCGGG CATGTTCCTC
   W A L S I W F F C V A L G S V V R A P P
961 TGGGCTCTGA GCATTTGGTT CTTCTGTGTC GCTCTGGGCT CAGTTGTGCG GCGCCCTCCC
   H D F H L N W W A M V F P N T G L T L A
1021 CATGATTTCC ACCTCAACTG GTGGGCTATG GTCCTCCCTA ACACCGGACT CACTCTCGCC
   T I T L A K S L D S A A L K W V G V G M
1081 ACCATCACCC TGGCCAAGTC ACTGGACAGT GCCCGTTGA AATGGGTGGG CGTGGGCATG
   S L C V I C M F I F V F V S T I R A V L
1141 TCCCTCTGCG TGATCTGCAT GTTCATCTTC GTCCTCGTGA GCACCATTAG GGCTGTCTCT
   L K R I M W P G R D E D V S E L F E *
1201 TTGAAGAGGA TCATGTGGCC AGGTCGGGAT GAGGATGTGT CCGAGTTGTT CGAATGA

```

图6

```

      M V K A
1  ATGGTCAAAG CTGGTGAGTT AGCAATCCTT AACAGATGAC ACTCTCATAG GTACTAACTC
      A V L G A S G G I G Q
61  GAAACGTTAG CGGTACTTGG AGCTTCTGGT GGCATTGGCC AGGTATGGAT ATCCCCACGC
      P L S
121 CTTACAACCC TGGTCACAAT ATGACCTTGT TCGATACTGA CTATCTCCCA AGCCACTGTC
      L L L K T C P L V E E L A L Y D V V N T
181 TCTCCTGTTG AAGACCTGTC CCTTAGTTGA AGAGCTTGCT CTCTACGATG TTGTGAACAC
      P G V A A D L S H I S S I A
241 CCCTGGTGTG GCTGCTGATC TATCCCACAT CTCGTCTATC GCTGTACGTT ACTGCCACAA
      K
301 TCGGAATTGC CCGATGGAAG AGGCGAAAAA TGGTATCTTG CTTACCTGGG CGATTAGAAA
      I S G F L P K D D G L K Q A L T G A N I
361 ATCTCTGGTT TTCTGCCCAA AGATGATGGG CTGAAGCAGG CCCTTACTGG TGCTAATATT
      V V I P A G I P
421 GTFGTATCC CGGCTGGTAT TCCCCGTAAG TCCCTACCCT TCGCATTGC TCCCTGGTATG
      R K P G M T R D D
481 TTCGCTGGTG GCCAGTTTTT TGATAGTTGA TAGGCAAGCC TGGTATGACC CGTGACGACC
      L F K I N A G I V R D L V K G I A E F C
541 TCTTCAAGAT CAAAGCCGGC ATAGTGCAGG ACTTGGTCAA GGGTATCGCC GAGTTCTGCC
      P K A F V L V I S N P V N S T V P I A A
601 CCAAGGCCTT TGTCTGGTT ATCTCAAACC CCGTTAATC TACTGTTCTT ATTGACTCAG
      E V L K A A G V F D P K R L F G V T T L
661 AGGTGCTCAA AGCCGCTGGC GTCTTTGACC CGAAGCGCCT CTTTGGTGTG ACCACACTGG
      D V V R A E T F T Q E F S G Q K D P S A
721 ACGTCGTTTC TGCAGAGACT TTCACCCAAG AGTTCTCGGG CCAGAAGGAT CCTTCTGCTG
      V Q I P V V G G H S G E T I V P L F S K
781 TTCAAATCCC AGTTGTTGGT GGCCACTCTG GAGAGCCAT TGTCCCCCTC TTCAGCAAGA
      T T P A I Q I P E E K Y D A L I H
841 CTACCCCGC AATTCAGATA CCCGAGGAGA AGTATGACGC ACTGATCCAC CGTAGGTTGT
      R V Q F
901 CCCAAAGAAT CTCATGAATA TCTTGCTGTA AGCACTAACT ATGCTTCAGG CGTCCAATTT
      G G D E V V Q A K D G A G S A T L S M A
961 GGTGGAGATG AGGTGGTCCA AGCTAAGGAC GGTGCTGGTT CCGCCACCTT GTCTATGGCC
      Y A G Y R
1021 TATGCCGGTT ACAGGTAGGG ATGCTGCGTA CCGTGAGAGC ACTCGCGGGT AACATGCCAT
      F A E S V I K A S K G Q T G I V E P T
1081 AGGTTCCGCTG AGAGTGTAAT CAAAGCTTCA AAGGGTCAAA CGGGTATTGT CGAGCCTACC
      F V Y L P G I P G G D E I V K A T G V E
1141 TTCGTCTACC TGCTTGAAT TCCCGCGGGT GATGAGATCG TTAAGGCAAC TGGCGTGGAA
      F F S T L V T L G
1201 TTCTTCTCTA CTCTTGTAAC CTTAGGAGTA AGATTCATCT CCTCACAGAA TCTTCGTTCA
      T N G A E K A S N V
1261 TATCACGCCA GGCTAACGCT ATTAAACAGA CTAATGGCGC AGAGAAGGCT AGCAACGTTT
      L E G V T E K E K K L L E A C T K G L K
1321 TTGAGGGCGT GACCGAGAAG GAAAAGAAGC TTCTCGAGGC TTGCACGAAA GGCCTTAAGG
      G N I E K G I D F V K N P P P K *
1381 GTAATATCGA GAAAGGCATC GACTTCGTTA AGAACCACC ACCAAAGTAA

```

图7

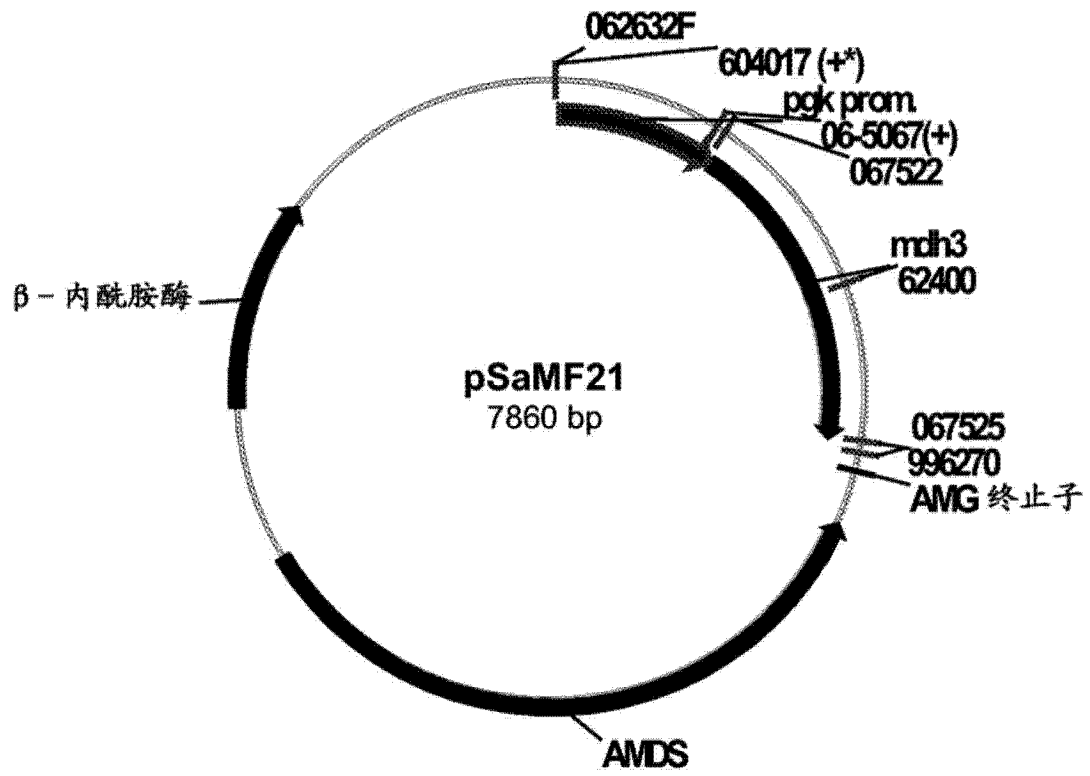


图 8

```

M A A P F R Q P E E A V D D T E F I D D
1 ATGGCGGCTC CGTTTCGTCA GCCTGAGGAG GCGGTGATG ACACCGAGTT CATCGATGAC
H H E H L R D T V H H R L R A N S S I M
61 CACCATGAAC ACCTCCGFGA TACCGTGCAC CATCGGTTGC GCGCCAATTC CTCATTATG
H F Q K I L V A N R G E I P I R I F R T
121 CACTTCCAGA AGATCCTCGT CGCCAACCGT GGTGAGATCC CCATTGCTAT CTTCAGAACG
A H E L S L Q T V A I Y S H E D R L S M
181 GCCCAGGAGC TGTCCCTTGA GACGGTTGCT ATCTACTCTC ATGAGGATCG ACTGTCAATG
H R Q K A D E A Y M I G H R G Q Y T P V
241 CACCGTCAAA AGGCCGATGA GGCCTACATG ATTGGCCACC GCGGTGAGTA CACCCCTGTC
G A Y L A G D E I I K I A L E H G V Q L
301 GTGCGTACC TGGCGGGCGA TGAGATCATC AAGATCGCCC TGGAGCACGG TGTCCAGCTG
I H P G Y G F L S E N A D F A R K V E N
361 ATCCACCCGG GCTACGGTTT CTTGTCCGAG AACGCCGACT TCGCCCGCAA GGTGAGAAC
A G I V F V G P T P D T I D S L G D K V
421 GCCCGCATTG TCTTTGTGG ACCCCATCCC GATACCATTG ACAGCTTGGG TGCCAGCTG
S A R R L A I K C E V P V V P G T E G P
481 TCGGCCCGTC GGCTGGCCAT TAAGTGGGAG GTCCCTGTCT TCCGGGTAC GGAGGGCCCC
V E R Y E E V K A F T D T Y G F P I I I
541 GTCGAGCGCT ATGAGGAGT CAAGCGTTC ACAGACACTC ATGGCTTCCC CATCATCATC
K A A F G G G G R G M R V V R D Q A E L
601 AAGGCTGCCT TTGGCGGTGG TGGCGGTGG ATGCGTGTGG TCCGTGACCA GGCCGAGCTG
R D S F E R A T S E A R S A F G N G T V
661 CGTACTCGT TCGAGCGAG CACCTCTGAG GCCCGCTCCG CCTTCGGCAA TGTCCAGCTC
F V E R F L D K P K H I E V Q L L G D S
721 TTCGTCGAGC GCTTCCTCGA CAAACCCAAG CACATTGAAG TCCAGTTCTT GGSTGACAGC
H G N V V H L F E R D C S V Q R R H Q K
781 CACGGCAACG TTGTCCATCT GTTTGAGCGT GACTGCTCCG TGCAGCGTGC TCACAGAGAG
V V E V A P A K D L P A D V R D R I L A
841 GTCGTTGAGG TTGCTCCGGC TAAGGACCTG CCAGCCGATG TCCGGGACCG CATCCTGGCC
D A V K L A K S V N Y R N A G T A E F L
901 GATGCTGTGA AGCTGGCCAA GTCCGTCAAC TACCSTAACG CCGGTACAGC TGAGTTCCTG
V D Q Q N R H Y F I E I N P R I Q V E H
961 GTGGACCAGC AGAACCGCCA CTACTTCATT GAAATCAATC CTCGTATCCA AGTCGAGCAC
T I T E E I T G I D I V A A Q I Q I A A
1021 ACCATCACCG AAGAGATTAC TGTTTCGAT ATCGTGGCTG CACAGATCCA GATTGCTGCT
G A S L E Q L G L T Q D R I S A R G F A
1081 GGTCAAGCC TCGAGCAACT GGGCCTGACT CAGGACCGCA TCTCCGCCCC CGGATTTGCC
I Q C R I T T E D P A K G F S P D T G K
1141 ATTCAATGTC GTATCACCAC GGAAGATCCC GCCAAGGGGT TCTCTCCGGA TACTGGTAAG
I E V Y R S A G G N G V R L D G G N G F
1201 ATTGAGGTTT ATCGTCCGCG TGTTGGTAAC GGTGTCCGTC TGGATGGTGG TAACGGTTTC
A G A I I T P H Y D S M L V K C T C R G
1261 GCTGGTGCTA TCATCACCCC TCACTACGAC TCCATGCTGG TCAAGTGCAC CTGCCGTGGT
S T Y E I A R R K V V R A L V E F R I R
1321 TCGACCTATG AAATCGCTCG TCGCAAGGTT GTGCGTGCCCT TGGTGGAGTT CCGTATTGCT
G V K T N I P F L T S L L S H P T F V D
1381 GGTGTGAAGA CCAACATTCC CTTCTGACT TCGCTTCTGA GCCACCCGAC CTTCTGCTGAT
G N C W T T F I D D T P E L F S L V G S
1441 GGAABACTGCT GGACCACITTT CATCGACGAC ACCCCTGAAT TGTTCTCTCT TGTCGGCAGT
Q N R A Q K L L A Y L G D V A V N G S S
1501 CAGAACCCTG CCCAGAAGCT GCTCGCATAC CTCGGCGATG TAGCTGTCAA CCGTAGTAGC
I K G Q I G E P K L K G D V I K P K L F
1561 ATCAAGGGCC AAATTGGCGA GCCCAAGCTC AAGGGTGATG TCATCAAGCC GAAGCTTTTC
D A E G K P L D V S A P C T K G W K Q I
1621 GATGCCGAGG GCAAGCCGCT TGACGTTTCC GCCCCCTGCA CCAAGGGTTG GAAGCAGATT
L D R E G P A A F A K A V R A N K G C L
1681 CTGACCCGGG AGGGTCCGGC TGCCTTTGCG AAGGCCGTGC GTGCCAACAA GGGTTGCTTG
I M D T T W R D A H Q S L L A T R V R T
1741 ATCATGGATA CTACCTGGCG TGACGCCAC CAGTCTTTGC TGGCCACCCG TGTGCGTACC
I D L L N I A H E T S Y A Y S N A Y S L
1801 ATCGACTTGT TGAACATCGC CCATGAGACC AGCTACGCCCT ACTCCAATGC GTACAGTTTG

```

图 9A

```

E C W G G A T F D V A M R F L Y E D P W
1861 GAATGCTGGG GTGGTGCTAC CTTCGATGTG GCCATGCGGT TCCTCTATGA GGACCCTGG
D R L R K M R K A V P N I P F Q M L L R
1921 GACCGCCTGC GCAAGATCG TAAGGCTGTT CCTAACATCC CATTCCAGAT GTTGCTCCGT
G A N G V A Y S S L P D N A I Y H F C K
1981 GGTGCCAAGG GTGTCGCCA CTCTCCCTC CCAGACAAGC CCATCTACCA CTTCGTAAAG
Q A K K C G V D I F R V F D A L N D V D
2041 CAGGCTAAGA AGTGC GG TGT CGACATTTTC CGTGT TTTTCG ACGCCCTCAA CGATGTCGAT
Q L E V G I K A V H A A E G V V E A T M
2101 CAGCTCGAGG TCGGTATCAA GGCTGTTCAT GCTGCCGAGG GTGTTGTCGA GGCCACCATG
C Y S G D M L N P H K K Y N L E Y Y M A
2161 TGCTACAGCG GTGACATGCT GAACCCCCAC AAGAAGTACA ACCTGGAGTA CTACATGGCC
L V D K I V A M K P H I L G I K D M A G
2221 TTGGTGGATA AGATTGTAGC CATGAAGCCT CACATCCTTG GTATCAAGGA TATGGCCGGT
V L K P Q A A R L L V G S I R Q R Y P D
2281 GTGCTGAAGC CCCAGGCCGC TCGCCTGTTG GTGGGCTCCA TCCGTCAGCG CTACCTGAC
L P I H V H T H D S A G T G V A S M I A
2341 CTCCCATCC ACGTCCACAC CCACGACTCC GCTGGTACTG GTGTGCTTC CATGATTCC
C A Q A G A D A V D A A T D S M S G M T
2401 TGTGCCCAGG CGGGTGCCGA CGCCGTGGAC GCCGCGACCG ACAGCATGTC CGGTATGACC
S Q P S I G A I L A S L E G T E Q D P G
2461 TCCGAGCCTA GCATTGGTGC CATTCTGGCC TCTCTTGAGG GCACTGAGCA AGACCCCGGT
L N L A H V R A I D S Y W A Q L R L L Y
2521 CTCACCTCG CCCACGTGCG CGCTATTGAT AGCTACTGGG CACAGCTGCG CTTGTCTAC
S P F E A G L T G P D P E V Y E H E I P
2581 TCTCCTTTTC AGCGGGTCT CACTGGCCCC GACCCTGAGG TCTACGAGCA CGAGATCCCT
G G Q L T N L I F Q A S Q L G L G Q Q W
2641 GGTGGTCAGT TGACCAACCT TATCTTCCAG GCCAGTCAGC TCGGCTGGG CCAGCAGTGG
A E T K K A Y E A A N D L L G D I V K V
2701 GCCGAAACCA AGAAGGCCTA TGAGCGGGCT AATGATTAC TCGGCGACAT TGTAAAGTTC
T P T S K V V G D L A Q F M V S N K L T
2761 ACTCCACCT CCAAGGTGGT CGGTGACTTG GCTCAGTCA TGGTCTCGAA CAACTGACT
P E D V V E R A G E L D F P G S V L E F
2821 CCAGAGGATG TTGTTGAGCG TGCTGGTGAG CTGGACTTCC CTGGTCTGT GCTCGAATTC
L E G L M G Q P F G G F P E P L R S R A
2881 CTCGAAGGTC TCATGGGACA GGCCTTCGGT GGATTCCCCG AGCCATTGCG CTCGCCGCC
L R D R R K L E K R P G L Y L E P L D L
2941 CTGCGCGATC GCCGCAAGCT CGAGAAGCGT CCAGGTCTCT ACCTCGAGCC TTTGGATTG
A K I K S Q I R E K F G A A T E Y D V A
3001 GCTAAGATCA AGAGCCAGAT CCGTGAGAAG TTCGGTGCTG CTA CTGAGTA TGACGTGGCC
S Y A M Y P K V F E D Y K K F V Q K F G
3061 AGCTATGCCA TGTATCCCAA GGTCTTCGAG GACTACAAGA AGTTCGTCCA GAAGTCCGTT
D L S V L P T R Y F L A K P E I G E E F
3121 GATCTCTCCG TCTTGCCAC ACGGTACTTC TTGGCCAAGC CTGAGATTGG CGAGGAGTTC
H V E L E K G K V L I L K L L A I G P L
3181 CACGTTGAGC TGGAGAAGG TAAGGTGCTC ATCCTGAAGT TGTGGCCAT CGCCCTCTT
S E Q T G Q R E V F Y E V N G E V R Q V
3241 TCAGAGCAGA CTGGTCAGCG TGAGGTCTTC TACGAAGTCA ACGGTGAGGT GCGCCAGGTC
A V D D N K A S V D N T S R P K A D V G
3301 GCTGTGATG ACAACAAGG TTCCGTGGAC AACACTTCAC GCCCTAAGGC CGATGTGGGT
D S S Q V G A P M S G V V V E I R V H D
3361 GACAGCAGCC AGGTCGGTGC TCCTATGAGC GGTGTGGTTG TTGAAATCCG TGTCCACGAT
G L E V K K G D P L A V L S A M K M
3421 GGTCTGGAGG TTAAGAAGG TGACCCACTT GCCGTCTCGA GTGCCATGAA GATGTAAGT
E M
3481 TCATTCCGAA TCATTTTCT CACTGGTCAA CTACAGATGC TAACAGCTTA TCCAGGAAT
V I S A P H S G K V S S L L V K E G D S
3541 GGTATCTCT GCTCCTCACA GTGGAAGGT CTCCAGCTTG CTGGTCAAGG AGGGCGATT
V D G Q D L V C K I V K A *
3601 TGTGGATGCC CAGGATCTCG TCTGCAAGAT CGTCAAAGCG TAA

```

图 9B



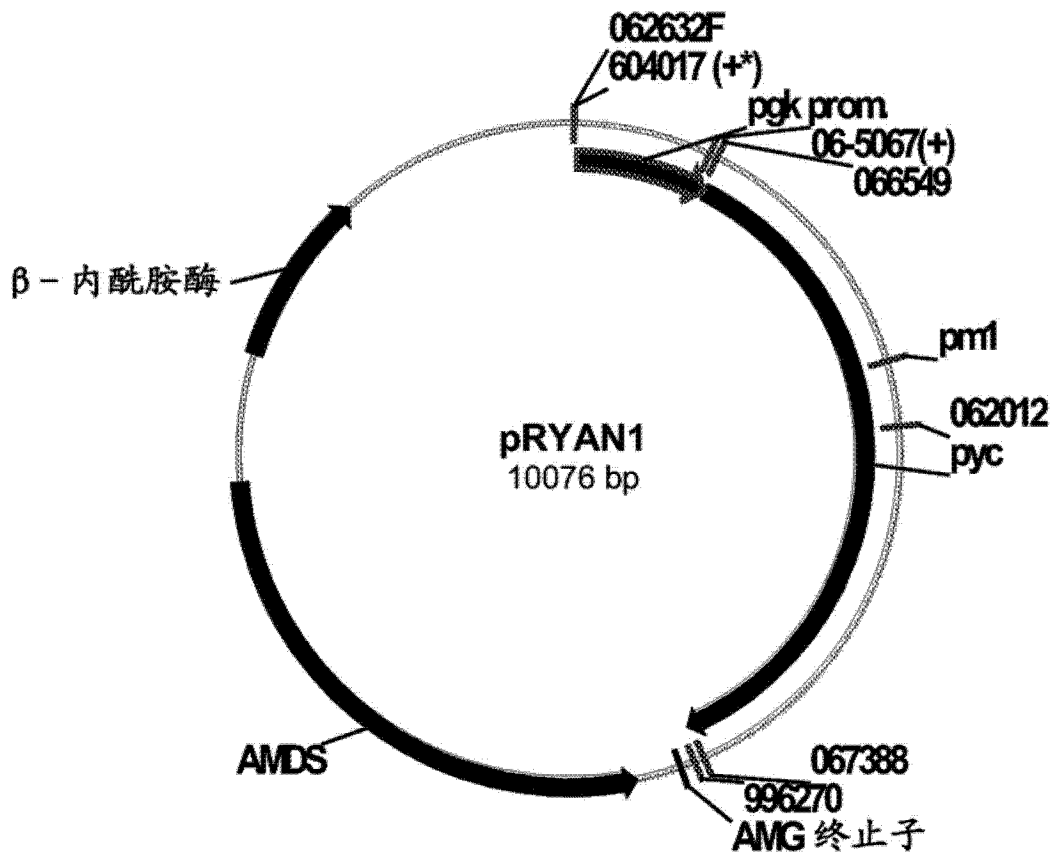


图 10

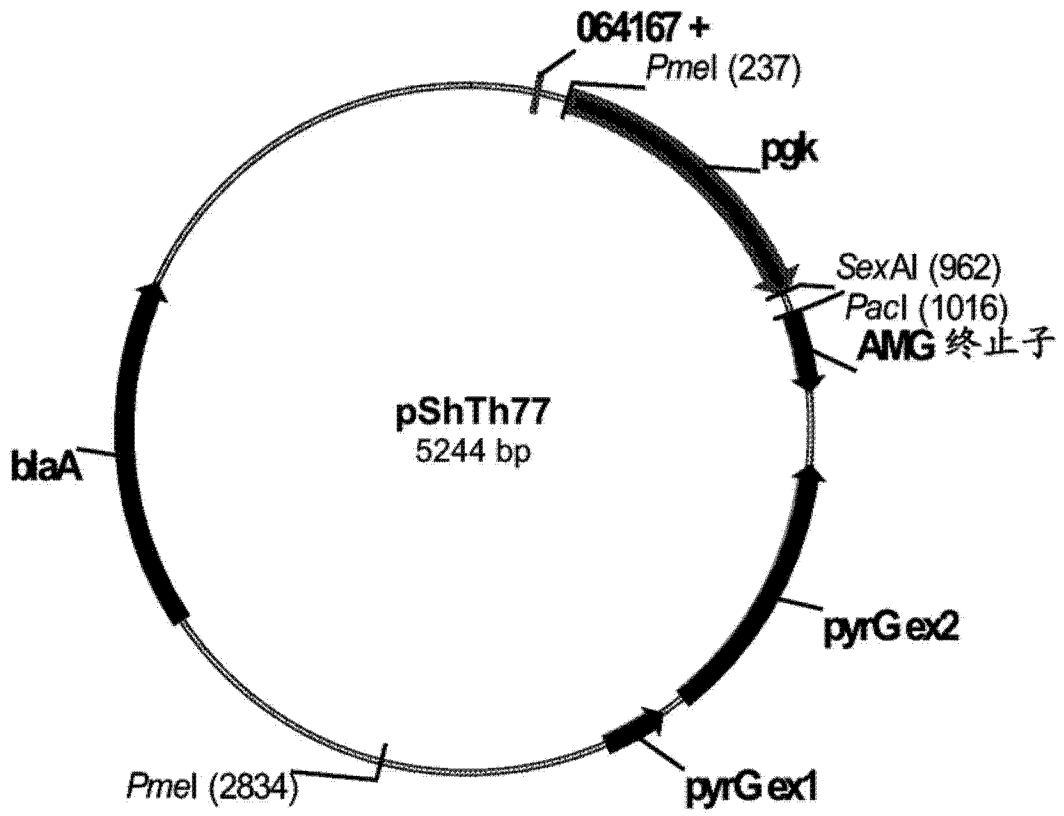


图 11

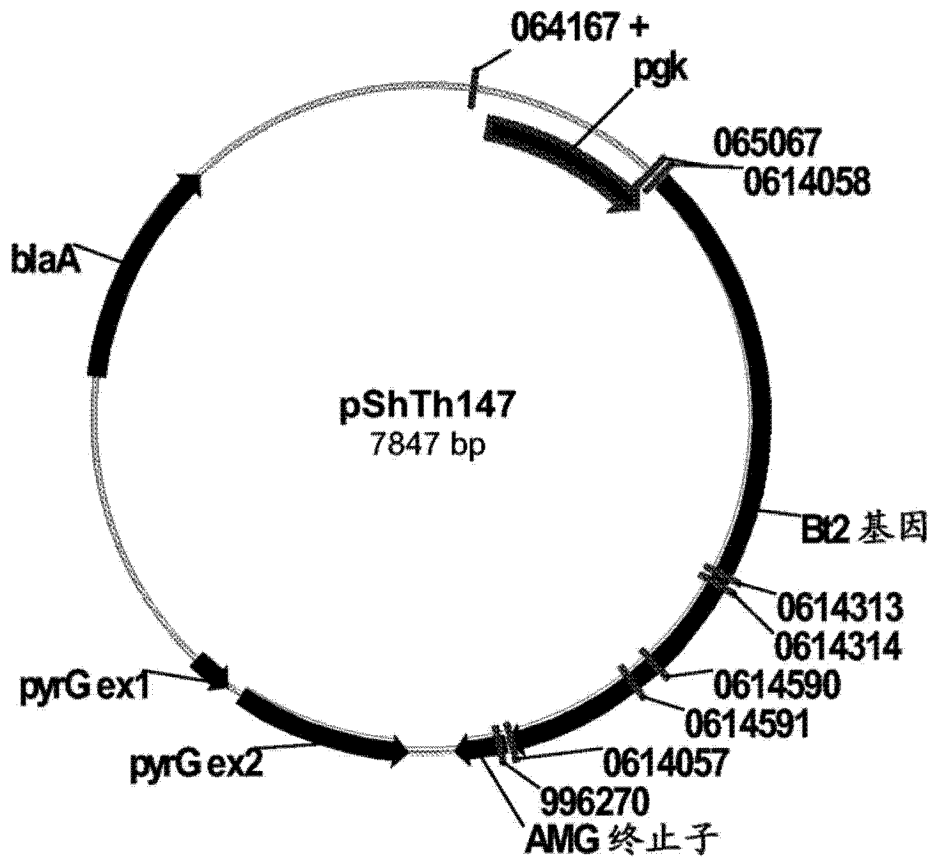


图 12