

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12N 15/75 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200780051674.2

[43] 公开日 2010年2月24日

[11] 公开号 CN 101657538A

[22] 申请日 2007.12.19

[21] 申请号 200780051674.2

[30] 优先权

[32] 2006.12.21 [33] US [31] 60/877,053

[86] 国际申请 PCT/US2007/088186 2007.12.19

[87] 国际公布 WO2008/079895 英 2008.7.3

[85] 进入国家阶段日期 2009.8.21

[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 兰迪·伯卡 米歇尔·马兰塔

玛丽亚·唐 巴巴拉·彻里

[74] 专利代理机构 北京市柳沈律师事务所

代理人 史悦

权利要求书 11 页 说明书 58 页 序列表 27 页
附图 24 页

[54] 发明名称

在芽孢杆菌属细胞中获得遗传感受态的方法

[57] 摘要

本发明涉及为用外源 DNA 转化而在非感受态芽孢杆菌属细胞中获得遗传感受态的方法。

1. 获得芽孢杆菌属转化体的方法，包括：

(a) 将外源 DNA 转化进通过至少一个拷贝导入的第一核酸构建体而成为感受态的芽孢杆菌属宿主细胞，所述第一核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区，其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的，所述细胞在导入第一核酸构建体之前为非感受态的；和

(b) 分离包含所述 DNA 的芽孢杆菌属宿主细胞的转化体。

2. 权利要求 1 的方法，其中编码 ComS 多肽的多核苷酸选自下组：(i) 多核苷酸，其编码的 ComS 多肽包含与 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 优选具有至少 60% 同一性的氨基酸序列，更优选至少 65% 同一性，甚至更优选至少 70% 同一性，甚至更优选至少 75% 同一性，甚至更优选至少 80% 同一性，甚至更优选至少 85% 同一性，最优选至少 90% 同一性，和甚至最优选至少 95% 同一性；(ii) 多核苷酸，其包含与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9 优选具有至少 60% 同一性的核苷酸序列，更优选至少 65% 同一性，甚至更优选至少 70% 同一性，甚至更优选至少 75% 同一性，甚至更优选至少 80% 同一性，甚至更优选至少 85% 同一性，最优选至少 90% 同一性，和甚至最优选至少 95% 同一性；(iii) 多核苷酸，其在优选至少中严紧性条件下，更优选至少中-高严紧性条件下，和最优选至少高严紧性条件下，与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9, 或其全长互补链杂交；和(iv)多核苷酸，其编码 ComS 变体，该 ComS 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10。

3. 权利要求 1 的方法，其中 ComS 多肽包含或其组成为 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10; 或其保持 ComS 多肽活性的片段。

4. 权利要求 1-3 中任一项的方法，其中感受态芽孢杆菌属宿主细胞进一步包含至少一个拷贝的导入的第二核酸构建体，所述第二核酸构建体包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区，赋予芽孢杆菌属宿主细

胞更进一步的感受态。

5. 权利要求 4 的方法, 其中编码 ComK 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComK 多肽包含与 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 优选具有至少 60%同一性的氨基酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 优选具有至少 60%同一性的核苷酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (iii) 多核苷酸, 其在优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 和最优选至少高严紧性条件下, 与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49, 或其全长互补链杂交; 和(iv)多核苷酸, 其编码 ComK 变体, 该 ComK 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50。

6. 权利要求 4 的方法, 其中 ComK 多肽包含或其组成为 SEQ ID NO: 12,

SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50; 或其保持 ComK 多肽活性的片段。

7. 获得感受态芽孢杆菌属宿主细胞的方法, 包括:

(a) 向非感受态芽孢杆菌属宿主细胞导入至少一个拷贝的第一核酸构建体, 其包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的; 和

(b) 分离包含编码 ComS 多肽的多核苷酸的感受态芽孢杆菌属宿主细胞。

8. 权利要求 7 的方法, 其中编码 ComS 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComS 多肽包含与 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 优选具有至少 60%同一性的氨基酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9 优选具有至少 60%同一性的核苷酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (iii) 多核苷酸, 其在优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 和最优选至少高严紧性条件下, 与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9, 或其全长互补链杂交; 和(iv)多核苷酸, 其编码 ComS 变体, 该 ComS 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10。

9. 权利要求 7 的方法, 其中 ComS 多肽包含或其组成为 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10; 或其保持 ComS 多肽活性的片段。

10. 权利要求 7-9 中任一项的方法, 其中感受态芽孢杆菌属宿主细胞进一步包含至少一个拷贝的导入的第二核酸构建体, 所述核酸构建体包含与编码

ComK 多肽的多核苷酸可操作连接的启动子区，赋予芽孢杆菌属宿主细胞更进一步的感受态。

11. 权利要求 10 的方法，其中编码 ComK 多肽的多核苷酸选自下组：(i) 多核苷酸，其编码的 ComK 多肽包含与 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 优选具有至少 60%同一性的氨基酸序列，更优选至少 65%同一性，甚至更优选至少 70%同一性，甚至更优选至少 75%同一性，甚至更优选至少 80%同一性，甚至更优选至少 85%同一性，最优选至少 90%同一性，和甚至最优选至少 95%同一性；(ii) 多核苷酸，其包含与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 优选具有至少 60%同一性的核苷酸序列，更优选至少 65%同一性，甚至更优选至少 70%同一性，甚至更优选至少 75%同一性，甚至更优选至少 80%同一性，甚至更优选至少 85%同一性，最优选至少 90%同一性，和甚至最优选至少 95%同一性；(iii) 多核苷酸，其在优选至少中严紧性条件下，更优选至少中-高严紧性条件下，和最优选至少高严紧性条件下，与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 或其全长互补链杂交；和(iv)多核苷酸，其编码 ComK 变体，该 ComK 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50。

12. 权利要求 10 的方法, 其中 ComK 多肽包含或其组成为 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50; 或其保持 ComK 多肽活性的片段。

13. 产生生物物质的方法, 包含:

(a) 在有利于产生所述物质的条件下培养用外源 DNA 转化的芽孢杆菌属宿主细胞, 所述 DNA 编码或参与具有生物活性的物质的表达, 其中通过至少一个拷贝的导入的核酸构建体使芽孢杆菌属宿主细胞成为感受态, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的, 所述宿主细胞在导入核酸构建体之前是非感受态的; 和

(b) 回收所述具有生物活性的物质。

14. 权利要求 13 的方法, 其中编码 ComS 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComS 多肽包含与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, 或 SEQ ID NO: 7 优选具有至少 60% 同一性的氨基酸序列, 更优选至少 65% 同一性, 甚至更优选至少 70% 同一性, 甚至更优选至少 75% 同一性, 甚至更优选至少 80% 同一性, 甚至更优选至少 85% 同一性, 最优选至少 90% 同一性, 和甚至最优选至少 95% 同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 优选具有至少 60% 同一性的核苷酸序列, 更优选至少 65% 同一性, 甚至更优选至少 70% 同一性, 甚至更优选至少 75% 同一性, 甚至更优选至少 80% 同一性, 甚至更优选至少 85% 同一性, 最优选至少 90% 同一性, 和甚至最优选至少 95% 同一性; (iii) 多核苷酸, 其在优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 和最优选至少高严紧性条件下, 与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9, 或其全长互补链杂交; 和 (iv) 多核苷酸, 其编码 ComS 变体, 该 ComS 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10。

15. 权利要求 13 的方法, 其中 ComS 多肽包含或其组成为 SEQ ID NO: 2,

SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10; 或其保持 ComS 多肽活性的片段。

16. 权利要求 13-15 中任一项的方法, 其中感受态芽孢杆菌属宿主细胞进一步包含至少一个拷贝的导入的第二核酸构建体, 所述核酸构建体包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区, 赋予芽孢杆菌属宿主细胞更进一步的感受态。

17. 权利要求 16 的方法, 其中编码 ComK 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComK 多肽包含与 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 优选具有至少 60% 同一性的氨基酸序列, 更优选具有至少 65% 同一性, 甚至更优选至少 70% 同一性, 甚至更优选至少 75% 同一性, 甚至更优选至少 80% 同一性, 甚至更优选至少 85% 同一性, 最优选至少 90% 同一性, 和甚至最优选至少 95% 同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 优选具有至少 60% 同一性的核苷酸序列, 更优选至少 65% 同一性, 甚至更优选至少 70% 同一性, 甚至更优选至少 75% 同一性, 甚至更优选至少 80% 同一性, 甚至更优选至少 85% 同一性, 最优选至少 90% 同一性, 和甚至最优选至少 95% 同一性; (iii) 多核苷酸, 其在优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 和最优选至少高严紧性条件下, 与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 或其全长互补链杂交; 和(iv)多核苷酸, 其编码 ComK 变体, 该 ComK 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16,

SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50。

18. 权利要求 16 的方法, 其中 ComK 多肽包含或其组成为 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50; 或其保持 ComK 多肽活性的片段。

19. 感受态芽孢杆菌属宿主细胞, 其包含至少一个拷贝的导入的第一核酸构建体, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的, 所述细胞在导入第一核酸构建体之前是非感受态的。

20. 权利要求 19 的感受态芽孢杆菌属宿主细胞, 其中编码 ComS 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComS 多肽包含与 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 优选具有至少 60%同一性的氨基酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9 优选具有至少 60%同一性的核苷酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (iii) 多核苷酸, 其在优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 和最优选至少高严紧性条件下, 与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9, 或其全长互补链杂交; 和(iv)多核苷酸, 其编码 ComS 变体, 该 ComS 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10。

21. 权利要求 19 的感受态芽孢杆菌属宿主细胞, 其中 ComS 多肽包含或

其组成为 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10; 或其保持 ComS 多肽活性的片段。

22. 权利要求 19-21 中任一项的感受态芽孢杆菌属宿主细胞, 其进一步包含至少一个拷贝的导入的第二核酸构建体, 所述核酸构建体包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区, 赋予芽孢杆菌属宿主细胞更进一步的感受态。

23. 权利要求 22 的感受态芽孢杆菌属宿主细胞, 其中编码 ComK 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComK 多肽包含与 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 优选具有至少 60% 同一性的氨基酸序列, 更优选至少 65% 同一性, 甚至更优选至少 70% 同一性, 甚至更优选至少 75% 同一性, 甚至更优选至少 80% 同一性, 甚至更优选至少 85% 同一性, 最优选至少 90% 同一性, 和甚至最优选至少 95% 同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 优选具有至少 60% 同一性的核苷酸序列, 更优选至少 65% 同一性, 甚至更优选至少 70% 同一性, 甚至更优选至少 75% 同一性, 甚至更优选至少 80% 同一性, 甚至更优选至少 85% 同一性, 最优选至少 90% 同一性, 和甚至最优选至少 95% 同一性; (iii) 多核苷酸, 其在优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 和最优选至少高严紧性条件下, 与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 或其全长互补链杂交; 和(iv)多核苷酸, 其编码 ComK 变体, 该 ComK 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 12, SEQ ID

NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50。

24. 权利要求 22 的感受态芽孢杆菌属宿主细胞, 其中 ComK 多肽包含或其组成为 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50; 或其保持 ComK 多肽活性的片段。

25. 权利要求 19-24 中任一项的感受态芽孢杆菌属宿主细胞, 其已经用外源 DNA 转化。

26. 通过权利要求 1-6 中任一项的方法获得的芽孢杆菌属转化体。

27. 通过权利要求 13-18 中任一项的方法获得的生物物质。

28. 产生亲本芽孢杆菌属细胞的突变体的方法, 其包括:

(a) 向亲本芽孢杆菌属细胞中导入包含核酸的外源 DNA, 以修饰亲本芽孢杆菌属细胞中编码多肽的基因, 这导致在相同条件下培养时与亲本细胞相比产生较少的所述多肽的突变细胞, 其中通过至少一个拷贝的导入的第一核酸构建体使亲本芽孢杆菌属细胞成为感受态, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于亲本芽孢杆菌属细胞是外源的, 所述细胞在导入第一核酸构建体之前是非感受态的; 和

(b) 分离所述突变细胞。

29. 权利要求 28 的方法, 其中编码 ComS 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComS 多肽包含与 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 优选具有至少 60% 同一性的氨基酸序列, 更优选至少 65% 同一性, 甚至更优选至少 70% 同一性, 甚至更优选至少 75% 同一性, 甚至更优选至少 80% 同一性, 甚至更优选至少 85% 同一性, 最优选至少 90% 同一性, 和甚至最优选至少 95% 同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ

ID NO: 9 优选具有至少 60%同一性的核苷酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (iii) 多核苷酸, 其在优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 和最优选至少高严紧性条件下, 与 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, 或 SEQ ID NO: 9, 或其全长互补链杂交; 和(iv)多核苷酸, 其编码 ComS 变体, 该 ComS 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10。

30. 权利要求 28 的方法, 其中 ComS 多肽包含或其组成为 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10; 或其保持 ComS 多肽活性的片段。

31. 权利要求 28-30 中任一项的方法, 其中亲本芽孢杆菌属细胞进一步包含至少一个拷贝的导入的第二核酸构建体, 所述核酸构建体包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区, 赋予芽孢杆菌属细胞更进一步的感受态。

32. 权利要求 31 的方法, 其中编码 ComK 多肽的多核苷酸选自下组: (i) 多核苷酸, 其编码的 ComK 多肽包含与 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 优选具有至少 60%同一性的氨基酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚至更优选至少 75%同一性, 甚至更优选至少 80%同一性, 甚至更优选至少 85%同一性, 最优选至少 90%同一性, 和甚至最优选至少 95%同一性; (ii) 多核苷酸, 其包含与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 优选具有至少 60%同一性的核苷酸序列, 更优选至少 65%同一性, 甚至更优选至少 70%同一性, 甚

至更优选至少 75%同一性,甚至更优选至少 80%同一性,甚至更优选至少 85%同一性,最优选至少 90%同一性,和甚至最优选至少 95%同一性; (iii) 多核苷酸,其在优选至少中严紧性条件下,更优选至少中-高严紧性条件下,和最优选至少高严紧性条件下,与 SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, SEQ ID NO: 45, SEQ ID NO: 47, 或 SEQ ID NO: 49 或其全长互补链杂交; 和(iv)多核苷酸,其编码 ComK 变体,该 ComK 变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50。

33. 权利要求 31 的方法,其中 ComK 多肽包含或其组成为 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50; 或其保持 ComK 多肽活性的片段。

34. 通过权利要求 28-33 中任一项的方法获得的突变芽孢杆菌属细胞。

在芽孢杆菌属细胞中获得遗传感受态的方法

对序列表的引用

本申请包含计算机可读格式的序列表。所述计算机可读格式通过提述并入本文。

发明背景

发明领域

本发明涉及在非感受态芽孢杆菌属(*Bacillus*)细胞中获得遗传感受态的方法。

相关领域描述

遗传感受态是这样的生理状态：外源 DNA 能够被内在化，产生转化事件 (Berka 等, 2002, *Mol. Microbiol.* 43: 1331-45)，但是与涉及电穿孔、原生质体和热休克或 CaCl_2 处理的人工转化不同。在革兰氏阳性和革兰氏阴性细菌菌种中都已经观察到了天然感受态 (Dubnau, 1999, *Annual Rev. Microbiol.* 53: 217-44)，并且过程需要十几种蛋白质，其表达经过精确地设计编排以满足每种生物的需要。

关于天然感受态的目的，已经提出了几种假设，这些假设可以总结为为食物的 DNA、为修复的 DNA 和为遗传多样性的 DNA (Dubnau, 1999, 见上文)。为食物的 DNA 假设得到如下观察结果的支持：感受态是稳定期现象，当细胞营养有限时出现，并且强大的非特异性核酸酶经常与转化特异性蛋白质共表达。第二种假设的证据来自下述事实：编码 DNA 修复酶的基因与编码 DNA 转运蛋白的那些基因协同表达。最后，用于遗传多样性的 DNA 假设提出，感受态是通过水平基因转移用于探索适应性景观 (for exploring the fitness landscape) 的机制。感受态受到细胞密度感受机制 (quorum-sensing mechanism) 的调节，并且其是一种双稳态状态 (Avery, 2005, *Trends Microbiol.* 13: 459-462)，这样的观察结果支持着这一假设。

公共数据库现在包含众多完整的细菌基因组，包括来自同一菌种不同菌

株的几个基因组。近来的分析表明,使用配对全基因组比对,同一菌种的不同菌株在基因内容上可能有很大差异。例如,大肠杆菌菌株 CFT073、EDL933 和 MG1655 的基因组比较表明,它们的联合蛋白质(基因产物)集合中只有 39.2%对于所有三株菌都是常见的,突出表明了同一菌种的菌株之间令人惊讶的多样性(Blattner 等, 1997, *Science* 277: 1453-74; Hayashi 等, 2006, *Mol. Syst. Biol.* doi:10.1038/msb4100049; Perna 等, 2001, *Nature* 409: 529-33; Welch 等, 2002, *Proc. Natl. Acad. Sci. USA* 99: 17020-17024)。此外,大肠杆菌菌株 CFT073 的基因组序列表明有 1,623 个菌株特异基因(21.2%)。根据这类比较,清楚地表明细菌基因组分成常见的保守骨架和菌株特异性序列。通常,菌种中指定菌株的基因组按照在所有菌株之间都保守的保守“骨架”基因和可能通过水平转移获得的非保守基因的分布而呈现镶嵌结构(Brzuszkiewicz 等, 2006, *Proc. Natl. Acad. Sci. USA* 103: 12879-12884; Welch 等, 2002, 见上文)。

按照实际应用,藉由天然感受态的转化是构建细菌菌株如芽孢杆菌的极为有用的工具,所述菌株中可以包含染色体基因的已改变的等位基因,或者通过重组 DNA 方法装配的质粒。尽管可以通过上述人工方法(例如,电穿孔、原生质体,和热休克或 CaCl_2 处理)实现用质粒和染色体 DNA 转化某些菌种,但是通过天然感受态导入 DNA 可以在简单、方便、速度和效率方面提供明显的优势。

在枯草芽孢杆菌(*Bacillus subtilis*)中,藉由涉及细胞密度感受、信号转导和基因表达级联的过程,群体中只有 5-10%的细胞分化成感受态状态(称为 *K*-状态)(Avery, 2005, 见上文)。已知至少 50 个基因与感受态直接有关,并且多达 165 个基因受到中央转录因子 ComK(直接或间接)的调节(Berka 等, 2002, 见上文)。枯草芽孢杆菌中的感受态级联由两个调节模块组成,所述模块由分子开关(图 1)间隔,所述分子开关涉及 ComS 与衔接分子 MecA 的结合,从而通过 ClpC/ClpP 蛋白酶干扰转录因子 ComK 的降解(Turgay 等, 1998, *EMBO J.* 17: 6730-6738)。

在密切相关的菌种地衣芽孢杆菌(*Bacillus licheniformis*)中,感受态相比而言知之甚少。Thorne 和同事(Gwinn 和 Thorne, 1964, 见上文; Leonard 等, 1964, *J. Bacteriol.* 88: 220-225; Thorne 和 Stull, 1966, *J. Bacteriol.* 91: 1012-1020)在 20 世纪 60 年代发表了一系列文章,描述通过天然感受态对源自地衣芽孢杆菌 ATCC 9945A 的三株营养缺陷型突变体的转化。仅在三株特定的营养缺陷

型突变体, 9945A-M28 (*gly*⁻)、-M30 (未表征的营养缺陷型), 和-M33 (*pur*⁻) 中观察到了天然感受态。源自相同亲本菌株(ATCC 9945A)的很多其他营养缺陷型不产生转化体, 包括需要硫胺素、赖氨酸、精氨酸、甲硫氨酸、色氨酸、组氨酸、尿嘧啶、腺嘌呤或次黄嘌呤, 和 13 种其他未表征的营养缺陷型需要 (Gwinn 和 Thorne, 1964, 见上文)的那些。此外, 这些研究者不能证明地衣芽孢杆菌 ATCC 10716 中通过天然感受态进行的转化(Gwinn 和 Thorne, 1964, 见上文)。如 Thorne 和同事的早期工作所建议的, 大多数地衣芽孢杆菌分离株不出现天然感受态, 并且近年来, 仅通过电穿孔(Tangney 等, 1994, *Biotechnol. Techniques* 8: 463-466)、接合(Herzog-Velikonja 等, 1994, *Plasmid* 31: 201-206)或原生质体(protoplasting) (Pragai 等, 1994, *Microbiol (Reading)* 140: 305-310), 已经实现了很多地衣芽孢杆菌分离株的转化。地衣芽孢杆菌中明显缺少感受态状态的原因尚不可知。

Ashikaga 等, 2000, *Journal of Bacteriology* 182: 2411-2415, 描述了枯草芽孢杆菌纳豆亚种(*Bacillus subtilis* subsp. *natto*)发展遗传感受态和表达并入外源 DNA 所需的后期感受态基因(late competence gene)的能力。Liu 等, 1996, *Journal of Bacteriology* 178: 5144-5152, 描述了通过多拷贝表达 *comS* 而提高野生型枯草芽孢杆菌中感受基因转录和转化效率。Tortosa 等, 2000, *Molecular Microbiology* 35: 1110-1119, 证明了破坏 *ylbF* 基因会引起 *comK* 表达的减少, 并且 *comS* 的过量表达足以避开 *ylbF* 突变的感受态表型。

因为地衣芽孢杆菌是具有工业重要性的菌种, 所以极为期望表现天然感受态的工程菌株用于构建新的改进型生产菌株。提供用于在转化性差的地衣芽孢杆菌菌株中诱导感受态的交钥匙方法(turn-key method)可以提高导入染色体标记/等位基因和表达质粒的速度与效率。如本文所述, 术语转化性差和非感受态可互换使用, 这些术语指, 当如前所述用于在枯草芽孢杆菌或土芽孢杆菌中感受态介导的转化时, 每微克 DNA 中的转化体数目小于自发突变效率的两倍(Anagnostopoulos 和 Spizizen, 1961, *J. Bacteriol.* 81: 741-746; Thorne 和 Stull, 1966, *J. Bacteriol.* 91: 1012-1020; Gwinn 和 Thorne, 1964, 见上文)。

本发明涉及在非感受态芽孢杆菌属细胞中获得遗传感受态的方法。

发明概述

本发明涉及获得感受态芽孢杆菌属宿主细胞的方法, 包含:

(a) 将至少一个拷贝的第一核酸构建体导入非感受态芽孢杆菌属宿主细胞, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的; 和

(b) 分离包含编码 ComS 多肽的多核苷酸的感受态芽孢杆菌属宿主细胞。
本发明还涉及获得芽孢杆菌属转化体的方法, 包含:

(a) 将外源 DNA 转化进芽孢杆菌属宿主细胞, 所述细胞通过至少一个拷贝的导入的第一核酸构建体而成为感受态, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的; 和

(b) 分离包含所述外源 DNA 的芽孢杆菌属宿主细胞转化体。

本发明还涉及产生生物物质的方法, 包含:

(a) 在有利于产生所述物质的条件下培养用外源 DNA 转化的芽孢杆菌属宿主细胞, 所述 DNA 编码或参与具有生物活性的物质的表达, 其中通过至少一个拷贝的导入的核酸构建体而使芽孢杆菌属宿主细胞成为感受态, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的, 并且所述细胞在导入该核酸构建体之前是非感受态的; 和

(b) 回收所述具有生物活性的物质。

本发明还涉及包含至少一个拷贝的导入的第一核酸构建体的感受态芽孢杆菌属宿主细胞, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的, 并且所述细胞在导入该核酸构建体之前是非感受态的。

本发明还涉及产生亲本芽孢杆菌属细胞的突变体的方法, 其包括:

(a) 用包含核酸构建体的外源 DNA 转化亲本芽孢杆菌属细胞, 以修饰亲本芽孢杆菌属细胞中编码多肽的基因, 这产生在相同条件下培养时与亲本细胞相比产生的多肽较少或产生的多肽生物活性较低的突变细胞, 其中通过至少一个拷贝的导入的第一核酸构建体而使芽孢杆菌属细胞成为感受态, 所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区, 其中编码 ComS 多肽的多核苷酸对于亲本芽孢杆菌属细胞是外源的, 并且所述细胞在导入该核酸构建体之前是非感受态的; 和

(b) 分离所述突变细胞。

在优选的方面，上述成为感受态的芽孢杆菌属细胞进一步包含至少一个拷贝的导入的第二核酸构建体，所述构建体包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区，赋予芽孢杆菌属宿主细胞更进一步的感受态。

附图简述

图 1 表明枯草芽孢杆菌的感受态调控级联。模块 1 包括感受态信息素 CSF 的检测和通过磷酸中继机制(phosphorelay mechanism)进行的信号传导，其引起 ComS 肽的合成。ComS 通过与 MecA 结合而干扰转录因子 ComK 的蛋白质分解降解，MecA 活化编码后期感受态功能的模块 2，后期感受态功能编码 DNA 转运机制。

图 2A 和 2B 表示地衣芽孢杆菌 DNA 甲基转移酶的基因组 DNA 序列和推导的氨基酸序列(分别为 SEQ ID NO: 51 和 52)。

图 3A、3B 和 3C 表示地衣芽孢杆菌 Bli1904II 限制-修饰系统的基因组 DNA 序列，其包含编码 Bli1904II 限制性内切核酸酶和 M.Bli1904II DNA 甲基转移酶的基因(SEQ ID NO: 53)。Bli1904II 限制性内切核酸酶编码区的反向补体(reverse complement)用双下划线表示，而 M.Bli1904II DNA 甲基转移酶编码区用单下划线表示。

图 4 显示 pMDT138 的限制图谱。

图 5 显示 pKK223-3 的限制图谱。

图 6 显示 pNBT51 的限制图谱。

图 7 显示 pNBT52 的限制图谱。

图 8 显示 pNBT53 的限制图谱。

图 9 显示 pNBT54 的限制图谱。

图 10 显示 pNBT35 的限制图谱。

图 11 显示 pNBT30 的限制图谱。

图 12 显示 pNBT31 的限制图谱。

图 13 显示 pNBT36 的限制图谱。

图 14 显示 pMDT100 的限制图谱。

图 15 显示枯草芽孢杆菌和地衣芽孢杆菌基因组编码的 ComS 蛋白质的氨基酸序列比对。

图 16 显示 pMRT098 的限制图谱。

图 17 显示 pMRT098/*comK* 的限制图谱。

图 18 显示 pMRT098/*comK/amyL3*'的限制图谱。

图 19 显示 pMRT098/*comK/amyL#24* 的限制图谱。

图 20 显示 pMMar2 的限制图谱。

图 21 显示枯草芽孢杆菌 *comS* 和地衣芽孢杆菌 *comK* 共表达的示意图。

定义

感受态: 术语“感受态”在本文中定义为天然生理状态，在这种状态中外源(胞外)DNA 能够内在化至芽孢杆菌属宿主细胞中，产生转化事件(Berka 等, 2002, *Mol. Microbiol.* 43: 1331-45)。感受态不同于涉及电穿孔、原生质体、热休克或 CaCl₂ 处理的人工转化。

感受态机制(级联): 术语“感受态机制”和“感受态级联”在本文可以互换使用，指细胞分化过程，所述过程将芽孢杆菌属细胞转化成天然可转化的细胞，所述可转化细胞能使用后期感受态基因编码的特定转运蛋白质摄取和合并外源(胞外)DNA，所述基因包含 *comC*、*comE*、*comF* 和 *comG* 操纵子。

非感受态: 如本文所述，术语“非感受态”和“可转化性差”可互换使用，并且这些术语指使用如前所述在枯草芽孢杆菌或地衣芽孢杆菌中由感受态介导的转化方法时(Anagnostopoulos 和 Spizizen, 1961, *J. Bacteriol.* 81: 741-746; Thorne 和 Stull, 1966, *J. Bacteriol.* 91: 1012-1020; Gwinn 和 Thorne, 1964, 见上文)，每微克 DNA 中转化体的数目少于自发突变频率的两倍。

ComS 多肽: 术语“ComS 多肽”在本文中定义为 *comS* 基因的产物，其参与遗传感受态的调控。ComS 是感受态信号转导途径中其他调控组分之间的装配接头(Ogura 等, 1999, *Mol. Microbiol.* 32: 799-812; Liu and Zuber, 1998, *J. Bacteriol.* 180: 4243-4251)。

ComK 多肽: 术语“ComK 多肽”在本文中定义为 *comK* 基因的产物；在感受态发展之前作为最后的自身调控控制开关起作用的转录因子；参与后期感受态基因表达的活化，所述基因参与 DNA-结合和摄取以及重组(Liu 和 Zuber, 1998, 见上文; Hamoen 等, 1998, *Genes Dev.* 12:1539-1550)。

外源多核苷酸: 术语“外源多核苷酸”及其变体(variation)用于本文中指对于芽孢杆菌属细胞非天然的多核苷酸，或者对于芽孢杆菌属细胞是天然的，但是已经通过使用对于芽孢杆菌属细胞非天然的遗传元件修饰的多核苷酸，

或者已经通过使用天然元件修饰的多核苷酸, 该天然元件已经经过操作, 以芽孢杆菌属细胞中通常不存在的方式起作用。

外源 DNA: 术语“外源 DNA”在本文表示在芽孢杆菌属细胞外部的 DNA。

同一性: 参数“同一性”描述两个氨基酸序列之间或两个核苷酸序列之间的相关性。

就本发明而言, 两个氨基酸序列之间的同一性程度使用 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, *J. Mol. Biol.* 48: 443-453)来测定, 如在 EMBOSS 软件包 (EMBOSS: The European Molecular Biology Open Software Suite, Rice 等, 2000, *Trends in Genetics* 16: 276-277)的 Needle 程序中所执行的, 优选 3.0.0 版或以后的版本。使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 EBLOSUM62 (BLOSUM62 的 EMBOSS 版)取代矩阵。使用 Needle 标记为“最长同一性”的输出结果(使用-nobrief选项获得)作为百分比同一性, 并如下计算:

$(\text{相同残基} \times 100) / (\text{比对长度} - \text{比对中缺口的总数})$

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使用 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, 见上文)测定, 如 EMBOSS 软件包 (EMBOSS: The European Molecular Biology Open Software Suite, Rice 等, 2000, 见上文)的 Needle 程序所执行的, 优选 3.0.0 版或以后的版本。使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 Needle 标记为“最长同一性”的输出结果(使用-nobrief选项获得)作为百分比同一性, 并如下计算:

$(\text{相同脱氧核糖核苷酸} \times 100) / (\text{比对长度} - \text{比对中缺口的总数})$

肽片段: 术语“肽片段”在本文中定义为从 ComS 多肽或 ComK 多肽的氨基和/或羧基末端缺失一个或多个氨基酸的 ComS 多肽或 ComK 多肽, 其中所述片段分别具有 ComS 或 ComK 活性。在优选的方面, SEQ ID NO: 2、4、6、8 或 10 的 ComS 片段, 或其同源物, 含有至少 30 个氨基酸残基, 更优选至少 35 个氨基酸残基, 并且最优选至少 40 个氨基酸残基。在另一个优选的方面, SEQ ID NO: 12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、42、44、46、48 或 50 的 ComK 片段, 或其同源物, 含有至少 400 个氨基酸残基, 更优选至少 420 个氨基酸残基, 并且最优选至少 440 个氨基酸残基。

亚序列: 术语“亚序列(subsequence)”在本文中定义为编码 ComS 多肽或 ComK 多肽的多核苷酸, 其从所述多核苷酸的 5'和/或 3'端缺失一个或多个核苷酸, 其中所述亚序列编码具有 ComS 或 ComK 活性的肽片段。在优选的方面, SEQ ID NO: 1、3、5、7, 或 9 的 *comS* 亚序列, 或其同源物, 含有至少 90 个核苷酸, 更优选至少 105 个核苷酸, 并且最优选至少 120 个核苷酸。在另一个优选的方面, SEQ ID NO: 11、13、15、17、19、21、23、25、27、29、31、33、35、37、39、41、43、45、47 或 49 的 *comK* 亚序列, 或其同源物, 含有至少 1200 个核苷酸, 更优选至少 1260 个核苷酸, 并且最优选至少 1320 个核苷酸。

等位变体(allelic variant): 术语“等位变体”在本文中占据相同染色体基因座的基因的任何两种或两种以上可选形式。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。

分离的多核苷酸: 术语“分离的多核苷酸”用于本文中指从来源分离的多核苷酸。在优选的方面, 所述多核苷酸如通过琼脂糖电泳测定为至少 1%纯, 优选至少 5%纯, 更优选至少 10%纯, 更优选至少 20%纯, 更优选至少 40%纯, 更优选至少 60%纯, 甚至更优选至少 80%纯, 并且最优选至少 90%纯。

基本上纯的多核苷酸: 术语“基本上纯的(substantially pure)多核苷酸”用于本文指不含其它外来的或不期望的核苷酸的多核苷酸制备物, 并且所述多核苷酸制备物处于适合于在遗传工程的蛋白质生产体系中使用的形式。因此, 基本上纯的多核苷酸按重量计含有至多 10%, 优选至多 8%, 更优选至多 6%, 更优选至多 5%, 更优选至多 4%, 更优选至多 3%, 甚至更优选至多 2%, 最优选至多 1%, 并且甚至最优选至多 0.5%的与其天然或重组结合的其它多核苷酸材料。然而, 基本上纯的多核苷酸可以包括天然存在的 5'和 3'非翻译区, 例如启动子和终止子。优选基本上纯的多核苷酸是按重量计至少 90%纯, 优选至少 92%纯, 更优选至少 94%纯, 更优选至少 95%纯, 更优选至少 96%纯, 更优选至少 97%纯, 甚至更优选至少 98%纯, 最优选至少 99%, 并且甚至最优选至少 99.5%纯。本发明所述多核苷酸优选为基本上纯的形式, 即, 所述多核苷酸制备物基本上不含与其天然或重组结合的其它多核苷酸材料。所述多核苷酸可以是基因组、cDNA、RNA、半合成、合成来源的, 或它们的任

何组合。

分离的多肽: 术语“分离的多肽”用于本文中指由其来源分离的多肽。在优选的方面,所述多肽如通过 SDS-PAGE 测定是至少 1%纯,优选至少 5%纯,更优选至少 10%,更优选至少 20%纯,更优选至少 40%纯,更优选至少 60%纯,甚至更优选至少 80%纯,和最优选至少 90%纯的多肽。

基本上纯的多肽: 术语“基本上纯的多肽”在本文表示多肽制备物,所述多肽制备物按重量计含有至多 10%,优选至多 8%,更优选至多 6%,更优选至多 5%,更优选至多 4%,更优选至多 3%,甚至更优选至多 2%,最优选至多 1%,并且甚至最优选至多 0.5%的与其天然或重组结合的(associated)的其它多肽材料。因此,优选所述基本上纯的多肽是按存在于制备物中的全部多肽材料的重量计至少 92%纯,优选至少 94%纯,更优选至少 95%纯,更优选至少 96%纯,更优选至少 96%纯,更优选至少 97%纯,更优选至少 98%纯,甚至更优选至少 99%纯,最优选至少 99.5%纯,并且甚至最优选 100%纯。本发明的多肽优选是基本上纯的形式,即,所述多肽制备物基本上(essentially)不含与其天然或重组结合的其它多肽材料。这能够通过以下方法实现,例如,通过公知的重组方法或由经典纯化方法制备多肽。

核酸构建体: 术语“核酸构建体”用于本文指单链或双链的核酸分子,其分离自天然存在的基因,或将其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。当所述核酸构建体含有表达编码序列所需的调控序列时,术语核酸构建体与术语“表达盒”同义。

调控序列(control sequence): 术语“调控序列”在本文定义为包括对于表达编码本发明多肽的多核苷酸是必需的或有利的的所有组分。各种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,或各种调控序列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、信号肽序列和转录终止子。最少的情况,调控序列包括启动子和转录和翻译的终止信号。调控序列可与用于导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核苷酸序列编码区的连接。

启动子: 术语“启动子”在本文定义为与 RNA 聚合酶结合并指导聚合酶到达多核苷酸正确的下游转录起始位点以启动转录的 DNA 序列,所述多核苷酸编码具有生物活性的多肽。RNA 聚合酶有效地催化与编码区合适 DNA 链互

补的信使 RNA 的装配。术语“启动子”还应理解为包括 5'非编码区(在启动子和翻译起点之间), 用于转录成 mRNA 后的翻译, 顺式作用转录控制元件如增强子, 和/或其他能与转录因子相互作用的核苷酸序列。启动子可以是野生型、变体、杂合或共有(consensus)启动子。

启动子区: 术语“启动子区”在本文定义为包含一个或多个(几个)启动子序列的核苷酸序列, 例如, 三联启动子(tandem triple promoter)。

启动子变体: 术语“启动子变体”在本文定义为这样的启动子, 其具有的核苷酸序列包含对亲本启动子的一个或多个(几个)核苷酸的取代、缺失和/或插入, 其中突变启动子具有比相应的亲本启动子更多或更少的启动子活性。术语“启动子变体”还包含天然变体和使用本领域已知方法获得的体外产生的变体, 所述方法如经典诱变、定点诱变, 和 DNA 改组。

串联启动子: 术语“串联启动子”在本文定义为两个或更多的启动子序列, 每个都与编码序列可操作地连接, 并介导编码序列转录成 mRNA。

杂合启动子: 术语“杂合启动子”在本文定义为两个或更多启动子的部分, 其融合在一起产生为所述两个或多个启动子的融合体的序列, 与编码具有生物活性的多肽的多核苷酸编码序列可操作连接时介导该编码序列转录成 mRNA。

可操作地连接: 术语“可操作地连接”在本文表示这样的构型, 其中将调控序列置于相对于多核苷酸序列的编码序列的合适位置, 使得调控序列指导多肽的编码序列的表达。

编码序列: 当用于本文时术语“编码序列”的意思是直接指定其蛋白产物的氨基酸序列的核苷酸序列。编码序列的边界通常由开放阅读框确定, 所述开放阅读框通常以 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 DNA、cDNA、合成的或重组的核苷酸序列。

表达: 术语“表达”包括涉及感兴趣的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。

表达载体: 术语“表达载体”在本文定义为线性的或环状的 DNA 分子, 其包含编码感兴趣的多肽的多核苷酸, 并且与提供用于其表达的额外核苷酸可操作地连接。

宿主细胞: 如本文中所使用的术语“宿主细胞”包括任何细胞类型, 所述

细胞类型对于使用核酸构建体或表达载体的转化、转染、转导、接合等是易感的(susceptible)。

转化: 术语“转化”在本文定义为将外源 DNA 导入芽孢杆菌属细胞,使所述 DNA 作为染色体整合体或作为自复制的染色体外载体保留。

转染: 术语“转染”在本文定义为用病毒核酸转化芽孢杆菌属宿主细胞。

转导: 术语“转导”在本文定义为将来自第一芽孢杆菌属细胞的 DNA 包装入病毒颗粒,并通过用病毒颗粒感染第二芽孢杆菌属细胞,将细菌 DNA 转入第二细胞。

接合: 术语“接合”在本文定义为通过细胞与细胞的接触,将 DNA 直接从一个芽孢杆菌属细胞转移至另一个芽孢杆菌属细胞。

转化体: 术语“转化体”在本文定义为通常包含任何芽孢杆菌属宿主细胞,其中已经通过转化将外源 DNA 导入了所述细胞。术语“转化体”不包括由人工方法如电穿孔、原生质体、热休克或 CaCl_2 处理产生的转染体、接合体和转化体。

修饰: 术语“修饰”在本文的意思是,对 ComS 多肽或 ComK 多肽的任何化学修饰,以及对编码如 ComS 多肽或 ComK 多肽的 DNA 的遗传操作。所述修饰可以是一个或多个氨基酸的取代、缺失和/或插入,以及一个或多个氨基酸侧链的置换。

人工变体: 当用在本文时,术语“ComS 人工变体”是指由表达修饰的 ComS 编码序列的生物产生的 ComS 多肽。术语“ComK 人工变体”是指由表达修饰的 ComK 编码序列的生物产生的 ComK 多肽。所述修饰的核苷酸序列通过人为干预(human intervention),通过修饰亲本 ComS 或亲本 ComK 编码序列来获得。亲本序列可以是野生型序列、合成序列、突变序列等。

发明详述

本发明涉及获得感受态芽孢杆菌属宿主细胞的方法,其包括:(a)将至少一个拷贝的第一核酸构建体导入非感受态芽孢杆菌属宿主细胞,所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区,其中编码 ComS 多肽的多核苷酸对于所述芽孢杆菌属宿主细胞是外源的;和(b)分离包含编码 ComS 多肽的多核苷酸的感受态芽孢杆菌属宿主细胞。

本发明还涉及产生获得芽孢杆菌属转化体的方法,其包括:(a)将外源 DNA

转化进芽孢杆菌属宿主细胞,所述细胞通过至少一个拷贝的导入的第一核酸构建体而成为感受态,所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区,其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的,所述细胞在导入第一核酸构建体之前是非感受态的;和(b)分离包含所述 DNA 的芽孢杆菌属宿主细胞的转化体。

与非感受态芽孢杆菌属细胞相比,本发明的方法将获得的转化体的数目增加至少 10 倍,优选至少 100 倍,更优选至少 1000 倍,甚至更优选至少 10,000 倍,并且最优选至少 100,000 倍。

芽孢杆菌属宿主细胞

在本发明的方法中,芽孢杆菌属宿主细胞可以是任何非感受态或可转化性差的芽孢杆菌属细胞。如本文所述,术语非感受态或可转化性差指在枯草芽孢杆菌或地衣芽孢杆菌中在使用感受态介导的转化方法时,每微克 DNA 中的转化体的数目小于自发突变频率的两倍。在本文,术语非感受态和可转化性差可以互换使用。应理解的是,本文中的术语“芽孢杆菌属”还涵盖芽孢杆菌属的同义词和曾被分类为芽孢杆菌的属如土芽孢杆菌属(*Geobacillus*)和类芽孢杆菌属(*Paenibacillus*)。在本发明的实践中有用的非感受态芽孢杆菌属宿主细胞包括,但不限于,嗜碱芽孢杆菌(*Bacillus alkalophilus*)、解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*)、萎缩芽孢杆菌(*Bacillus atrophaeus*)、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*)、坚固芽孢杆菌(*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*)、莫海威芽孢杆菌(*Bacillus mojavenensis*)、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*)、苏云金芽孢杆菌(*Bacillus thuringiensis*)、和花域芽孢杆菌(*Bacillus vallismortis*)细胞。

在优选的方面,非感受态芽孢杆菌属宿主细胞是解淀粉芽孢杆菌、迟缓芽孢杆菌、地衣芽孢杆菌、嗜热脂肪芽孢杆菌或枯草芽孢杆菌细胞。在更优选的方面,非感受态芽孢杆菌属宿主细胞是解淀粉芽孢杆菌细胞。在另外的更优选方面,非感受态芽孢杆菌属宿主细胞是克劳氏芽孢杆菌细胞。在另外的更优选方面,非感受态芽孢杆菌属宿主细胞是地衣芽孢杆菌细胞。在另外的更优选方

面，非感受态芽孢杆菌属宿主细胞是枯草芽孢杆菌细胞。在最优选的方面，非感受态芽孢杆菌属宿主细胞是地衣芽孢杆菌细胞。

在本发明的另外的方面，芽孢杆菌属宿主细胞可以另外包含一个或多个(几个)修饰，例如，对其他基因的缺失或破坏，所述其它基因可能不利于感兴趣的多肽或生物化学物质(biochemical)的产生、回收或应用。在优选的方面，芽孢杆菌属宿主细胞是蛋白酶缺陷型细胞。在更优选的方面，芽孢杆菌属宿主细胞包含对 *aprE* 和 *nprE* 的破坏或缺失。在另外的优选方面，芽孢杆菌属宿主细胞不产生孢子。在另外的更优选的方面，芽孢杆菌属宿主细胞包含对 *spoIIAC* 的破坏或缺失。在另外的优选方面，芽孢杆菌属宿主细胞包含破坏或缺失与表面活性肽的生物合成有关的基因之一，所述基因例如 *srfA*、*srfB*、*srfC* 和 *srfD*。参见，例如，美国专利 5,958,728。也可以破坏或缺失不利于感兴趣的多肽或生物物质产生、回收或应用的其他基因，例如，*amyE* 基因。

本发明还涉及感受态芽孢杆菌属宿主细胞，其包含至少一个拷贝的导入的第一核酸构建体，所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区，其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是外源的，所述细胞在导入第一核酸构建体之前是非感受态的。

在优选的方面，上述成为感受态的芽孢杆菌属宿主细胞进一步包含至少一个拷贝的导入的第二核酸构建体，所述核酸构建体包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区，赋予芽孢杆菌属宿主细胞高于通过表达 ComS 多肽获得的感受态的更进一步的感受态。通过进一步将获得的转化体数目与通过表达 ComS 多肽获得感受态的芽孢杆菌属细胞相比，增加至少 2 倍，优选至少 5 倍，更优选至少 10 倍，更优选至少 100 倍，甚至更优选至少 1000 倍，最优选至少 10,000 倍，和甚至最优选至少 100,000 倍，由此赋予芽孢杆菌属宿主细胞更进一步的感受态。

本发明还涉及这样的感受态芽孢杆菌属宿主细胞，所述细胞包含核酸构建体或重组表达载体，其包含编码或参与生物物质表达的感兴趣的 DNA。

ComS 多肽和 ComK 多肽和它们的多核苷酸

在本发明的方法中，可以使用适于赋予非感受态芽孢杆菌属细胞遗传感受态的编码 ComS 多肽的任何分离的多核苷酸。此外，可以使用适于赋予感受态芽孢杆菌属细胞在遗传上更强的感受态的编码 ComK 多肽的任何分离的多核

苷酸。

分离的多核苷酸可以是基因组、cDNA、半合成、合成来源，或它们的任何组合。

编码 ComS 多肽的多核苷酸可以获得自，例如，解淀粉芽孢杆菌(登录号 Q70KJ5)、枯草芽孢杆菌(登录号 P80355 和 Q83WC2)，或地衣芽孢杆菌。

编码 ComK 多肽的多核苷酸可以获得自，例如，枯草芽孢杆菌 168(登录号 P40396)、地衣芽孢杆菌(DSM 13/ATCC 14580) (登录号 Q65LN7)、地衣芽孢杆菌(登录号 Q8VQ66)、芽孢杆菌菌种 Bt 24 (登录号 Q2HQ42)、韦氏芽孢杆菌(*Bacillus weihenstephanensis*) KBAB4 (登录号 Q2AUN4)、苏云金芽孢杆菌 *konkukian* 亚种(登录号 Q6HM51)、蜡状芽孢杆菌(*Bacillus cereus*) ATCC 10987 (登录号 Q73C31)、蜡状芽孢杆菌菌株 ZK/E33L (登录号 Q63EM6)、蜡状芽孢杆菌 G9241 (登录号 Q4MPH9)、炭疽芽孢杆菌(*Bacillus anthracis*) (登录号 Q81TW5)、蜡状芽孢杆菌(ATCC 14579/DSM 31; 登录号 Q81GQ3)、蜡状芽孢杆菌 *cytotoxis* 亚种 NVH 391-98 (登录号 Q2E900)、芽孢杆菌菌种 NRRL B-14911(登录号 Q2B9A0)、芽孢杆菌菌种 Ob 20(登录号 Q2HQ30)、芽孢杆菌菌种 Bt 26(登录号 Q2HQ36)、芽孢杆菌菌种 Ob 07(登录号 Q2HQ38)、芽孢杆菌菌种 Bt 30(登录号 Q2HQ39)、芽孢杆菌菌种 Bt 35(登录号 Q2HQ35)、芽孢杆菌菌种 Ob 12b(登录号 Q2HQ37)，和苏云金芽孢杆菌以色列亚种(*Bacillus thuringiensis* subsp. *israelensis*) (ATCC 35646; 登录号 Q3EYL1)。

在第一方面，编码 ComS 多肽的分离的多核苷酸包含与 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 具有如下同一性程度的氨基酸序列，优选至少 60%，更优选具有至少 65%，更优选至少 70%，更优选至少 75%，更优选至少 80%，更优选至少 85%，甚至更优选至少 90%，最优选至少 95%，并且甚至最优选至少 96%，至少 97%，至少 98%，或至少 99% (下文中的“同源 ComS 多肽”或“ComS 同源物”)。在优选的方面，同源的 ComS 多肽包含这样的氨基酸序列，其与 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 有十个氨基酸不同，优选五个氨基酸不同，更优选四个氨基酸不同，甚至更优选三个氨基酸不同，最优选两个氨基酸不同，并且甚至最优选一个氨基酸不同。

分离的多核苷酸优选编码包含 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 的氨基酸序列的 ComS 多肽，或其等位

变体；或其保持 ComS 多肽活性的片段。在优选的方面，ComS 多肽包含 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 的氨基酸序列。在另外的优选方面，ComS 多肽由 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 的氨基酸序列，或其等位变体；或其具有 ComS 活性的片段组成。在另外优选的方面，ComS 多肽由 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, 或 SEQ ID NO: 10 的氨基酸序列组成。

在另外的第一方面，编码 ComK 多肽的分离的多核苷酸包含与 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 具有如下同一性程度的氨基酸序列，优选至少 60%，更优选至少 65%，更优选至少 70%，更优选至少 75%，更优选至少 80%，更优选至少 85%，甚至更优选至少 90%，最优选至少 95%，并且甚至最优选至少 96%，至少 97%，至少 98%，或至少 99%(下文中的“同源 ComK 多肽”或“ComK 同源物”)。在优选的方面，同源的 ComK 多肽包含氨基酸序列，所述氨基酸序列与 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 的氨基酸有十个氨基酸不同，优选五个氨基酸不同，更优选四个氨基酸不同，甚至更优选三个氨基酸不同，最优选两个氨基酸不同，并且甚至最优选一个氨基酸不同。

分离的多核苷酸优选编码包含 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 的 ComK 多肽，或其等位变体；或其保持 ComK 多肽活性的片段。在优选的方面，ComK 多肽包含 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18,

SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 的氨基酸序列。在另外的优选方面, ComK 多肽由 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 的氨基酸序列, 或其等位变体; 或其具有 ComK 活性的片段组成。在另外优选的方面, ComK 多肽由 SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 44, SEQ ID NO: 46, SEQ ID NO: 48, 或 SEQ ID NO: 50 的氨基酸序列组成。

在第二方面, 编码 ComS 多肽的分离的多核苷酸优选在至少非常低严紧性条件下, 更优选至少低严紧性条件下, 更优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 甚至更优选至少高严紧性条件下, 并且最优选至少非常高严紧性条件下, 与 SEQ ID NO: 1、SEQ ID NO: 3、SEQ ID NO: 5、SEQ ID NO: 7 或 SEQ ID NO: 9 杂交: (ii) (i)的亚序列, 或(iii) (i)或(ii)的全长互补链(J. Sambrook, E.F. Fritsch 和 T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 第 2 版, Cold Spring Harbor, New York)。SEQ ID NO: 1、SEQ ID NO: 3、SEQ ID NO: 5、SEQ ID NO: 7 或 SEQ ID NO: 9 的亚序列含有至少 90 个连续的核苷酸或优选至少 120 个连续的核苷酸。此外, 所述亚序列可编码具有 ComS 活性的多肽片段。

在另一个第二方面, 编码 ComK 多肽的分离的多核苷酸在优选至少非常低严紧性条件下, 更优选至少低严紧性条件下, 更优选至少中严紧性条件下, 更优选至少中-高严紧性条件下, 甚至更优选至少高严紧性条件下, 并且最优选至少非常高严紧性条件下, 与 SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 15、SEQ ID NO: 17、SEQ ID NO: 19、SEQ ID NO: 21、SEQ ID NO: 23、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29、SEQ ID NO: 31、SEQ ID NO: 33、

SEQ ID NO: 35、SEQ ID NO: 37、SEQ ID NO: 39、SEQ ID NO: 41、SEQ ID NO: 43、SEQ ID NO: 45、SEQ ID NO: 47 或 SEQ ID NO: 49 杂交, (ii) (i)的亚序列, 或(iii) (i)或(ii)的全长互补链(J. Sambrook, E.F. Fritsch 和 T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 第2版, Cold Spring Harbor, New York)。SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 15、SEQ ID NO: 17、SEQ ID NO: 19、SEQ ID NO: 21、SEQ ID NO: 23、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29、SEQ ID NO: 31、SEQ ID NO: 33、SEQ ID NO: 35、SEQ ID NO: 37、SEQ ID NO: 39、SEQ ID NO: 41、SEQ ID NO: 43、SEQ ID NO: 45、SEQ ID NO: 47 或 SEQ ID NO: 49 的亚序列含有至少 100 个连续的核苷酸或优选至少 200 个连续的核苷酸。此外, 所述亚序列可编码具有 ComK 活性的多肽片段。

上述核苷酸序列或其亚序列, 以及上述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和种的菌株鉴定和克隆编码 ComS 多肽和 ComK 多肽的 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 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, 但长度应为至少 14, 优选至少 17, 更优选至少 20, 并且最优选至少 50 个核苷酸。然而, 优选所述核酸探针长度为至少 60 个核苷酸。例如, 所述核酸探针的长度可以是至少 100 个核苷酸。DNA 和 RNA 两种探针均可使用。通常将探针标记以探测相应的基因(例如, 用 ^{32}P 、 ^3H 、 ^{35}S 、生物素或抗生物素蛋白(avidin)标记)。这些探针包含于本发明中。

因而, 可从制备自这些其它生物体的基因组 DNA 文库中筛选与上述探针杂交并且编码 ComS 多肽或 ComK 多肽的 DNA。可以通过琼脂糖或聚丙烯酰胺凝胶电泳, 或其它分离技术分离来自这些其它生物体的基因组 DNA。可以将来自文库的 DNA 或分离的 DNA 转移至硝酸纤维素(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、SEQ ID NO: 3、SEQ ID NO: 5、SEQ ID NO: 7 或 SEQ ID NO: 9 或其亚序列, 或 SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 15、SEQ ID NO: 17、SEQ ID NO: 19、SEQ ID NO: 21、SEQ ID NO: 23、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29、SEQ ID NO: 31、SEQ ID NO: 33、SEQ ID NO: 35、SEQ ID NO: 37、SEQ ID NO: 39、SEQ ID NO: 41、SEQ ID NO: 43、SEQ ID NO: 45、SEQ ID NO: 47 或 SEQ ID

NO: 49, 其全长互补链, 或其亚序列同源的克隆或 DNA, 将所述载体材料用在 Southern 印迹中。

就本发明而言, 杂交表示核苷酸序列在非常低至非常高严谨性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、SEQ ID NO: 3、SEQ ID NO: 5、SEQ ID NO: 7 或 SEQ ID NO: 9 中所示的核苷酸序列, 它的全长互补链, 或它们的亚序列, 或 SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 15、SEQ ID NO: 17、SEQ ID NO: 19、SEQ ID NO: 21、SEQ ID NO: 23、SEQ ID NO: 25、SEQ ID NO: 27、SEQ ID NO: 29、SEQ ID NO: 31、SEQ ID NO: 33、SEQ ID NO: 35、SEQ ID NO: 37、SEQ ID NO: 39、SEQ ID NO: 41、SEQ ID NO: 43、SEQ ID NO: 45、SEQ ID NO: 47 或 SEQ ID NO: 49, 它的全长互补链, 或它们的亚序列。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。

在优选的方面, 核酸探针是多核苷酸, 其编码 SEQ ID NO: 2 的 ComS 多肽, 或其亚序列。在另一个优选方面, 核酸探针是 SEQ ID NO: 1 或其全长互补链。

在另一个优选的方面, 核酸探针是多核苷酸, 其编码 SEQ ID NO: 4 的 ComS 多肽, 或其亚序列。在另一个优选方面, 核酸探针是 SEQ ID NO: 3 或其全长互补链。

在另一个优选的方面, 核酸探针是多核苷酸, 其编码 SEQ ID NO: 6 的 ComS 多肽, 或其亚序列。在另一个优选方面, 核酸探针是 SEQ ID NO: 5 或其全长互补链。

在另一个优选的方面, 核酸探针是多核苷酸, 其编码 SEQ ID NO: 8 的 ComS 多肽, 或其亚序列。在另一个优选方面, 核酸探针是 SEQ ID NO: 7 或其全长互补链。

在另一个优选的方面, 核酸探针是多核苷酸, 其编码 SEQ ID NO: 10 的 ComS 多肽, 或其亚序列。在另一个优选方面, 核酸探针是 SEQ ID NO: 9 或其全长互补链。

在另一个优选的方面, 核酸探针是多核苷酸, 其编码 SEQ ID NO: 12 的 ComK 多肽, 或其亚序列。在另一个优选方面, 核酸探针是 SEQ ID NO: 11 或其全长互补链。

在另一个优选的方面, 核酸探针是多核苷酸, 其编码 SEQ ID NO: 14 的

ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 13 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 16 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 15 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 18 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 17 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 20 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 19 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 22 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 21 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 24 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 23 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 26 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 25 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 28 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 27 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 30 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 29 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 32 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 31 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 34 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 33

或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 36 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 35 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 38 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 37 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 40 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 39 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 42 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 41 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 44 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 43 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 46 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 45 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 48 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 47 或其全长互补链。

在另一个优选的方面，核酸探针是多核苷酸，其编码 SEQ ID NO: 50 的 ComK 多肽，或其亚序列。在另一个优选方面，核酸探针是 SEQ ID NO: 49 或其全长互补链。

对于长度至少 100 个核苷酸的长探针，将非常低至非常高严紧性条件定义为在 42°C，在 5X SSPE、0.3% SDS、200 µg/ml 已剪切并且变性的鲑精 DNA，并且对于非常低和低严紧性为 25% 的甲酰胺、对于中和中-高严紧性为 35% 的甲酰胺、或对于高和非常高严紧性为 50% 的甲酰胺中，根据标准的 Southern 印迹法进行预杂交和杂交最佳 12 至 24 小时。

对于长度为至少 100 个核苷酸的长探针，使用 2X SSC、0.2% SDS 优选

至少在 45°C (非常低严谨性), 更优选至少在 50°C (低严谨性), 更优选至少在 55°C (中严谨性), 更优选至少在 60°C (中-高严谨性), 甚至更优选至少在 65°C (高严谨性), 并且最优选至少在 70°C (非常高严谨性)将载体材料最终洗涤三次, 每次 15 分钟。

对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针, 将严谨性条件定义为在比使用 Bolton 和 McCarthy 的算法(1962, *Proceedings of the National Academy of Sciences USA* 48:1390)得出的 T_m 低大约 5°C 至大约 10°C, 在 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1×Denhardt 溶液, 1 mM 焦磷酸钠(sodium pyrophosphate), 1 mM 磷酸二氢钠(sodium monobasic phosphate), 0.1 mM ATP 和 0.2 mg/ml 的酵母 RNA 中, 根据标准的 Southern 印迹步骤进行预杂交、杂交和杂交后洗涤最佳 12 至 24 小时。

对于长度大约 15 个核苷酸至大约 70 个核苷酸的短探针, 将载体材料在 6×SSC 加 0.1% SDS 中洗涤一次 15 分钟, 并用 6×SSC 在比计算的 T_m 低 5°C 至 10°C 的温度下洗涤两次, 每次 15 分钟。

在第三个方面, 分离的多核苷酸编码 ComS 多肽的人工变体, 所述人工变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:6、SEQ ID NO:8 或 SEQ ID NO:10, 或其同源序列; 或其成熟多肽。

在另外的第三方面, 分离的多核苷酸编码 ComK 多肽的人工变体, 所述人工变体包含取代、缺失和/或插入一个或多个氨基酸的 SEQ ID NO:12、SEQ ID NO:14、SEQ ID NO:16、SEQ ID NO:18、SEQ ID NO:20、SEQ ID NO:22、SEQ ID NO:24、SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:30、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:36、SEQ ID NO:38、SEQ ID NO:40、SEQ ID NO:42、SEQ ID NO:44、SEQ ID NO:46、SEQ ID NO:48 或 SEQ ID NO:50, 或其同源序列; 或其成熟多肽。

优选地, 氨基酸改变对性质是较不重要的(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; 小缺失, 通常缺失 1 至大约 30 个氨基酸; 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, 例如多组氨酸序列(poly histidine tract)、抗原表位(antigenic epitope)或结合域(binding domain)。

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)、疏水性氨基酸组(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, *In, The Proteins*, Academic Press, New York 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。

除了 20 个基本氨基酸，非基本氨基酸(例如 4-羟脯氨酸、6-N-甲基赖氨酸、2-氨基异丁酸、异缬氨酸和 α -甲基丝氨酸)可以取代野生型多肽的氨基酸残基。有限数量的非保守氨基酸、不由遗传密码编码的氨基酸和非天然氨基酸可以取代氨基酸残基。“非天然氨基酸”在蛋白质合成后已经过修饰，和/或在它们的侧链具有不同于基本氨基酸的化学结构。非天然氨基酸能够以化学方法合成，并且优选是商业上可获得的，包括六氢吡啶羧酸(pipecolic acid)、噻唑烷羧酸(thiazolidine carboxylic acid)、脱氢脯氨酸、3-和 4-甲基脯氨酸，和 3,3-二甲基脯氨酸。

可选地，氨基酸改变具有这样的性质以使 ComS 多肽或 ComK 多肽的物理化学性质改变。例如，氨基酸改变可改进 ComS 或 ComK 对于 MecA 的结合亲和力和/或结合动力学，或 ComK 与基因组中它的 DNA 序列靶标的结合亲和力等。

能够根据本领域已知的方法，例如定点诱变或丙氨酸分区诱变法(alanine-scanning mutagenesis) (Cunningham 和 Wells, 1989, *Science* 244: 1081-1085)来鉴定亲本 ComS 或 ComK 多肽中的必需氨基酸。在最后一技术中，将单一丙氨酸突变导入到分子中的每个残基，并且测试所得突变分子的生物活性(即，限制性内切核酸酶活性)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物理分析结构来测定，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测定。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

氨基酸的身份(identity)也能够通过分析与多肽的同一性来推断,所述多肽与根据本发明的多肽相关。

能够使用已知的诱变、重组和/或改组方法,继之以有关的筛选方法,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; WO 95/17413; 或 WO 95/22625 公开的那些方法,来进行并测试单个或多个氨基酸取代。能够使用的其它方法包括易错 PCR、噬菌体展示(例如, Lowman 等, 1991, *Biochem.* 30: 10832-10837; 美国专利 5,223,409; WO 92/06204)和区域定向的诱变(region directed-mutagenesis) (Derbyshire 等, 1986, *Gene* 46: 145; Ner 等, 1988, *DNA* 7: 127)。

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,并且使用本领域内标准方法快速测序。这些方法允许快速确定感兴趣的多肽中单个氨基酸残基的重要性,并且能够应用于未知结构的多肽。

氨基酸取代、缺失和/或插入的总数优选是 10,更优选 9,更优选 8,更优选 7,更优选至多 6,更优选 5,更优选 4,甚至更优选 3,最优选 2,并且甚至最优选 1。

ComS 和 ComK 多核苷酸的表达

可以用许多方式操作编码 ComS 多肽或 ComK 多肽的多核苷酸,以用于多核苷酸在芽孢杆菌属宿主细胞中的表达。依赖于核酸构建体或载体或芽孢杆菌属宿主细胞,在将多核苷酸的序列插入核酸构建体或载体之前对其进行操作可能是理想的或必需的。使用克隆方法修饰核苷酸序列的技术是本领域熟知的。

包含编码 ComS 多肽或 ComK 多肽的多核苷酸的核酸构建体可以与一个或多个调控序列可操作地连接,所述调控序列在芽孢杆菌属宿主细胞中在与该调控序列相容的条件下能够指导编码序列的表达。

每个调控序列对于编码 ComS 多肽或 ComK 多肽的核苷酸序列可以是天然或外源的。这样的调控序列包括,但不限于,前导序列、启动子、信号序列和转录终止子。最少的情况,调控序列包括启动子和转录和翻译的终止信号。调控序列可以与用于导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 ComS 多肽或 ComK 多肽的核苷酸序列编码区

的连接。

调控序列可以是合适的启动子序列，其是由用于表达编码 ComS 多肽或 ComK 多肽的多核苷酸的芽孢杆菌属宿主细胞识别的核苷酸序列。启动子序列含有介导 ComS 多肽或 ComK 多肽表达的转录调控序列。启动子区可以是在所选芽孢杆菌属宿主细胞中显示转录活性的任何核苷酸序列，并可获得自指导与芽孢杆菌属宿主细胞同源或异源的具有生物活性的胞外或胞内多肽合成的基因。

启动子区可以包含单一启动子或启动子的组合。当启动子区包含启动子的组合时，启动子优选串联(in tandem)。启动子区的启动子可以是能启动编码具有生物活性的多肽的多核苷酸在感兴趣的芽孢杆菌属宿主细胞中转录的任何启动子。启动子对于编码具有生物活性的多肽的核苷酸序列可以是天然的、外源的或其组合。这样的启动子能获得自指导与芽孢杆菌属宿主细胞同源或异源的具有生物活性的胞外或胞内多肽合成的基因。

在优选的方面，启动子区包含获得自细菌来源的启动子。在更优选的方面，启动子区包含获得自革兰氏阳性细菌的启动子。在另一个更优选的方面，启动子区包含获得自革兰氏阴性细菌的启动子。革兰氏阳性细菌包括，但不限于，芽孢杆菌属、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*)、土芽孢杆菌属和海洋芽孢杆菌属(*Oceanobacillus*)。革兰氏阴性细菌包括，但不限于，大肠杆菌、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*)、奈瑟氏球菌属(*Neisseria*)和尿枝原体属(*Ureaplasma*)。

在最优选的方面，启动子区包含获得自芽孢杆菌属菌株的启动子，例如，*Bacillus agaradherens*、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、坚固芽孢杆菌、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、枯草芽孢杆菌或苏云金芽孢杆菌；或获得自链霉菌属菌株，例如，浅青紫链霉菌(*Streptomyces lividans*)或鼠灰链霉菌(*Streptomyces murinus*)。

用于指导编码本发明方法中具有生物活性的多肽的多核苷酸转录的合适启动子的实例是从下述获得的启动子：大肠杆菌 *lac* 操纵子、天蓝色链霉菌

(*Streptomyces coelicolor*)琼脂糖酶基因(*dagA*)、迟缓芽孢杆菌或克劳氏芽孢杆菌碱性蛋白酶基因(*aprH*)，地衣芽孢杆菌碱性蛋白酶基因(枯草蛋白酶 Carlsberg 基因(*subtilisin Carlsberg gene*))，枯草芽孢杆菌果聚糖蔗糖酶基因(*sacB*)、枯草芽孢杆菌 α -淀粉酶基因(*amyE*)、地衣芽孢杆菌 α -淀粉酶基因(*amyL*)、嗜热脂肪芽孢杆菌产麦芽糖淀粉酶基因(*amyM*)、解淀粉芽孢杆菌 α -淀粉酶基因(*amyQ*)、地衣芽孢杆菌青霉素酶基因(*penP*)、枯草芽孢杆菌 *xylA* 和 *xylB* 基因，苏云金芽孢杆菌拟步行甲亚种(*Bacillus thuringiensis* subsp. *tenebrionis*) CryIII A 基因(*cryIII A*)或其部分，原核 β -内酰胺酶基因(Villa-Kamaroff 等, 1978, *Proceedings of the National Academy of Sciences USA* 75: 3727-3731)，和巨大芽孢杆菌 *xylA* 基因 (Rygus and Hillen, 1992, *J. Bacteriol.* 174: 3049-3055; Kim 等, 1996, *Gene* 181: 71-76)，以及 *tac* 启动子 (DeBoer 等, 1983, *Proceedings of the National Academy of Sciences USA* 80: 21-25)，质粒 pUB110 的 *orf β* 启动子(Tortosa 等, 2000, *Mol. Microbiol.* 35: 1110-1119)，和 *spac* 启动子(Henner, 1990, *Methods Enzymol.* 185: 223-228)。其他实例是 *spoI* 细菌噬菌体启动子和 *tac* 启动子的启动子(DeBoer 等, 1983, *Proceedings of the National Academy of Sciences USA* 80:21-25)。在"Useful proteins from recombinant bacteria"于 *Scientific American*, 1980, 242: 74-94 中；和 Sambrook, Fritsch, 和 Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 第 2 版, Cold Spring Harbor, New York 中描述了其他启动子。

在另一个优选的方面，启动子区包含启动子，其为“共有”启动子，在“-35”区具有序列 TTGACA，并且在“-10”区具有 TATAAT。共有启动子可以获得自能在芽孢杆菌属宿主细胞中起作用的任何启动子。可以通过下述方法完成“共有”启动子的构建：使用本领域公知的方法进行定点诱变产生启动子，其更完美地符合枯草芽孢杆菌营养型“ σ A-型”启动子的“-10”和“-35”区已确定的共有序列(Voskuil 等, 1995, *Molecular Microbiology* 17: 271-279)。

在另一个优选的方面，启动子区包含“共有”启动子，该“共有”启动子获得自从下述获得的启动子：大肠杆菌 *lac* 操纵子、天蓝色链霉菌琼脂糖酶基因(*dagA*)、克劳氏芽孢杆菌或迟缓芽孢杆菌碱性蛋白酶基因(*aprH*)，地衣芽孢杆菌碱性蛋白酶基因(枯草蛋白酶 Carlsberg 基因)，枯草芽孢杆菌果聚糖蔗糖酶基因(*sacB*)、枯草芽孢杆菌 α -淀粉酶基因(*amyE*)、地衣芽孢杆菌 α -淀粉酶基因(*amyL*)、嗜热脂肪芽孢杆菌产麦芽糖淀粉酶基因(*amyM*)、解淀粉芽孢杆

菌 α -淀粉酶基因(*amyQ*)、地衣芽孢杆菌青霉素酶基因(*penP*)、枯草芽孢杆菌 *xylA* 和 *xylB* 基因, 苏云金芽孢杆菌拟步行甲亚种 *CryIII A* 基因(*cryIII A*)或其部分, 或原核 β -内酰胺酶基因 *spoI* 细菌噬菌体启动子。

在更优选的方面, 启动子区包含获得自解淀粉芽孢杆菌 α -淀粉酶基因(*amyQ*)的“共有”启动子。

在另一个优选的方面, 启动子区包含启动子, 其为杂合启动子。

在另一个优选的方面, 启动子区包含启动子, 其为变体启动子。参见, 例如, WO 05/098-16, 美国专利 5,698,415 和美国专利 6,100,063。在优选的方面, 变体启动子是 $P_{amyL4199}$, 其中 P=启动子。

在另一个优选的方面, 启动子区包含启动子, 其为串联启动子。参见, 例如, WO 99/043835 和 WO 05/098016。在优选的方面, 串联启动子是 $P_{amyQ}-P_{cryIII A}-cryIII A$ mRNA 加工/稳定序列 (mRNA processing/stablizing sequence)。在另一个优选的方面, 串联启动子是 $P_{amyL4199}-P_{amyQ}-P_{cryIII A}-cryIII A$ mRNA 加工/稳定序列。

在本发明的方法中, 杂合或串联启动子应理解为对于编码具有生物活性的多肽的多核苷酸是异源的, 即使其野生型启动子对于所述多核苷酸序列是天然的。例如, 在一个由至少两个启动子组成的串联启动子中, 一个启动子可以是编码生物物质的多核苷酸的野生型启动子。

调控序列也可以是合适的转录终止子序列, 其是由芽孢杆菌属宿主细胞识别以终止转录的序列。所述终止子序列与编码 ComS 多肽或 ComK 多肽的核苷酸序列的 3'末端可操作地连接。在所选芽孢杆菌属宿主细胞中有功能的任何终止子都可用于本发明中。

调控序列还可以是合适的前导序列, 其是对于芽孢杆菌属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指导具有生物活性的多肽合成的核苷酸序列的 5'-末端。在所选芽孢杆菌属宿主细胞中有功能的任何前导序列都可用于本发明中。

调控序列还可以是 mRNA 稳定序列。术语“mRNA 稳定序列”在本文中定义为位于启动子区下游和编码 ComS 多肽或 ComK 多肽的多核苷酸编码序列上游的序列, 启动子区与其可操作地连接, 从而使所有从启动子区合成的 mRNA 可以被加工以产生在转录物的 5'末端包含稳定物序列 (stabilizer sequence) 的 mRNA 转录物。在 mRNA 转录物的 5'末端存在这样的稳定物序

列可以增加其半衰期(Agaisse 和 Lereclus, 1994, 见上文, Hue 等, 1995, *Journal of Bacteriology* 177: 3465-3471)。mRNA 加工/稳定序列与细菌 16S 核糖体 RNA 的 3' 末端互补。在优选的方面, mRNA 加工/稳定序列基本产生单一大小的转录物, 其在 5' 末端包含稳定序列。mRNA 加工/稳定序列优选为一个与细菌 16S 核糖体 RNA 的 3' 末端互补的序列。参见, 美国专利 6,255,076 和 5,955,310。

对于芽孢杆菌属宿主细胞有效的 mRNA 加工/稳定序列是 WO 94/25612 中公开的苏云金芽孢杆菌 *cryIIIA* mRNA 加工/稳定序列, 或其保持 mRNA 加工/稳定功能的部分, 或 Hue 等, 1995, *Journal of Bacteriology* 177: 3465-3471 中公开的枯草芽孢杆菌 SP82 mRNA 加工/稳定序列, 或其保持 mRNA 加工/稳定功能的部分。

然后使用本领域已知的方法或本文所述用于表达 ComS 多肽或 ComK 多肽的方法, 将核酸构建体导入芽孢杆菌属宿主细胞。

还可以与如上所述相似地构建包含感兴趣的 DNA 的核酸构建体, 所述 DNA 编码或参与具有生物活性的物质的表达。

为了获得导入 DNA 的产物的分泌, 调控序列还可以是信号肽编码区, 其编码与多肽的氨基末端相连的氨基酸序列, 其可指导表达的多肽进入细胞的分泌途径。信号肽编码区对于所述多肽可为天然的或可从外部来源获得。核苷酸序列的编码序列的 5' 端可固有地包含信号肽编码区, 其与编码分泌多肽的编码区片段一起天然地连接在翻译阅读框中。可选地, 编码序列 5' 端可含有信号肽编码区, 其对于编码分泌多肽的编码序列的部分是外源的。外源信号肽编码区在编码序列不正常含有信号肽编码区时可能是必需的。或者, 外源信号肽编码区可以简单地取代天然信号肽编码区以相对于与编码序列正常结合的天然信号肽编码区获得增强的多肽分泌。信号肽编码区可以获得自芽孢杆菌属菌种的淀粉酶或蛋白酶基因。然而, 能够指导表达的多肽进入所选芽孢杆菌属宿主细胞的分泌途径的任何信号肽编码区可用于本发明中。

对于芽孢杆菌属宿主细胞有效的信号肽编码区是从如下获得的信号肽编码区: 芽孢杆菌属 NCIB 11837 产麦芽糖淀粉酶基因、嗜热脂肪芽孢杆菌 α -淀粉酶基因、地衣芽孢杆菌枯草蛋白酶基因、地衣芽孢杆菌 β -内酰胺酶基因、嗜热脂肪芽孢杆菌中性蛋白酶基因(*nprT*, *nprS*, *nprM*)和枯草芽孢杆菌 *prsA* 基因。另外的信号肽由 Simonen 和 Palva, 1993, *Microbiological Reviews* 57: 109-137 描述。

重组表达载体

在本发明的方法中,可以使用重组表达载体重组产生 ComS 多肽或 ComK 多肽,所述重组表达载体包含编码 ComS 多肽或 ComK 多肽的多核苷酸,启动子,和转录和翻译终止信号。上述各种核酸和调控序列可以结合在一起以产生重组表达载体,所述载体可以包括一个或多个方便的限制位点以允许在这些位点插入或取代指导 ComS 多肽或 ComK 多肽合成的多核苷酸。可选地,可通过将多核苷酸或包含所述多核苷酸的核酸构建体插入合适的用于表达的载体中来表达多核苷酸。在制备表达载体的过程中,将编码序列置于载体中,使得该编码序列与用于表达和可能的分泌的合适的调控序列可操作地连接。

重组表达载体可以是任何载体,其能够方便地进行重组 DNA 步骤,并且能够产生核苷酸序列的表达。载体的选择将通常依赖于载体与将导入该载体的芽孢杆菌属宿主细胞的相容性。载体可以是线状或闭合环状质粒。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导入芽孢杆菌属宿主细胞中时,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。载体系统可以是单独的载体或质粒或两个或更多的载体或质粒,其共同含有待导入芽孢杆菌属细胞基因组的完整 DNA (total DNA),或者是转座子(transposon)。

当导入芽孢杆菌属宿主细胞时,载体可以整合入基因组。为了整合,载体可依赖于指导 ComS 多肽或 ComK 多肽合成的核苷酸序列,或通过同源重组将载体稳定整合入基因组的任何其它载体元件。或者,载体可以含有额外的核苷酸序列,用于指导通过同源重组整合入芽孢杆菌属宿主细胞的基因组。所述额外的核苷酸序列使载体能够整合入芽孢杆菌属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,整合元件应该优选含有足够数目的核酸,如 100 至 1,500 碱基对,优选 400 至 1,500 碱基对,并且最优选 800 至 1,500 碱基对,其与相应的靶序列高度同源以增强同源重组的概率。整合元件可以是任何序列,其与芽孢杆菌属宿主细胞基因组中的靶序列同源。此外,整合元件可以是非编码或编码的核苷酸序列。

为了自主复制,载体可以进一步包含复制起点,其使载体能够在所述的

芽孢杆菌属宿主细胞中自主地复制。细菌复制起点的实例是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 pAMB1 的复制起点。复制起点可以是具有突变以使其在芽孢杆菌属宿主细胞中的功能对温度敏感的复制起点(参见，例如，Ehrlich, 1978, *Proceedings of the National Academy of Sciences USA* 75:1433-1436)。

可以将多于一个拷贝的指导具有生物活性的多肽，或 ComS 多肽或 ComK 多肽合成的核苷酸序列导入芽孢杆菌属宿主细胞以扩增核苷酸序列的表达。核苷酸序列的稳定扩增可通过如下方式获得：使用本领域公知的方法将至少一个额外拷贝的序列整合入芽孢杆菌属宿主细胞基因组并选择转化体。WO 94/14968 中描述了用于实现基因组 DNA 序列扩增的方便的方法。

载体优选地含有一个或多个选择性标记，其允许简单选择经转化的细胞。选择性标记是基因，其产物提供杀生物剂抗性、对重金属的抗性、对营养缺陷型的原养性(protothrophy to auxotrophs)等。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，所述抗生素抗性例如氨苄青霉素、卡那霉素、红霉素、氯霉素或四环素抗性。此外，可以通过共转化完成选择，例如，如 WO 91/09129 中所述，其中选择性标记在单独的载体上。

用于连接上述元件以构建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)。

也可以与如上所述相似地构建包含感兴趣的 DNA 的重组表达载体，所述 DNA 编码或参与具有生物活性的物质的表达。

将载体导入芽孢杆菌属细胞可，例如，通过如下实现：原生质体转化(参见，例如，Chang 和 Cohen, 1979, *Molecular General Genetics* 168: 111-115)，使用感受态细胞(参见，例如，Young 和 Spizizen, 1961, *Journal of Bacteriology* 81: 823-829 或 Dubnau 和 Davidoff-Abelson, 1971, *Journal of Molecular Biology* 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, *Journal of Bacteriology* 169: 5771-5278)。

DNA

在本发明的方法中，根据本发明的方法获得的，导入感受态芽孢杆菌属细胞的外源 DNA 可以是任何感兴趣的 DNA。DNA 可以是基因组、cDNA、半合成、合成来源，或其任意组合。DNA 可以编码具有感兴趣的生物活性的任何物质(下文中的“生物物质”)或者可以是参与所述生物物质表达的 DNA，例如，启动子。

具有生物活性的物质可以是任何感兴趣的多肽。多肽对于感兴趣的芽孢杆菌属宿主细胞可以是天然的或异源(外源)的。术语“异源多肽”在本文中定义为对宿主细胞不是天然的多肽；天然多肽，其中进行了结构修饰以使天然多肽改变，例如，使天然多肽的蛋白质序列改变；或作为通过重组 DNA 技术对编码多肽的 DNA 操作的结果，例如更强的启动子，而使其表达量改变的天然多肽。多肽可以是下述多肽和杂合多肽的天然存在的等位变体和工程改造的变体。

术语“多肽”在本文并不指特定长度的编码产物，因此，包括肽、寡肽和蛋白质。术语“多肽”还包括杂合多肽，其包含获得自至少两个不同多肽的部分或全部多肽序列的组合，其中一个或多个多肽对芽孢杆菌属细胞可以是异源的。多肽进一步包括多肽的天然存在的等位变体和工程改造的变体。

在优选的方面，多肽是抗体、抗原、抗微生物肽、酶、生长因子、激素、免疫调节剂(immunodilator)、神经递质、受体、报告蛋白质、结构蛋白质和转录因子。

在更优选的方面，多肽是氧化还原酶、转移酶、水解酶、裂合酶、异构酶或连接酶。在最优选的方面，多肽是 α -葡糖苷酶、氨肽酶、淀粉酶、糖酶、羧肽酶、过氧化氢酶、纤维素酶、几丁质酶、角质酶、环式糊精糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、葡糖脑苷脂酶、 α -葡糖苷酶、 β -葡糖苷酶、转化酶、漆酶、脂肪酶、甘露糖苷酶、变构酶(mutanase)、氧化酶、果胶分解酶、过氧化物酶、磷脂酶、肌醇六磷酸酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶、尿激酶或木聚糖酶。

在另一个优选方面，多肽是清蛋白、胶原、原弹性蛋白、弹性蛋白或明胶。

在另一个优选方面，多肽是杂合多肽，其包含获得自至少两个不同多肽的部分或完整多肽序列的组合，其中一个或多个多肽对于芽孢杆菌属宿主细胞可以是异源的。

在另一个优选的方面，多肽是融合的多肽，其中将另外的多肽融合到所述多肽或其片段的 N 末端或 C 末端。通过将编码一种多肽的核苷酸序列(或其部分)与编码另一种多肽的核苷酸序列(或其部分)融合而产生融合的多肽。产生融合多肽的技术是本领域已知的，且包括连接编码多肽的编码序列以使它们在阅读框中，并且使融合的多肽的表达在相同启动子和终止子的调控下。

编码感兴趣的多肽的 DNA 可以获得自任何原核、真核或其他来源。就本发明而言，用于本文与给定的来源有关的术语“获得自”，意思应为多肽由所述来源产生，或由其中插入了来自所述来源的基因的细胞产生。

用于分离或克隆编码感兴趣的多肽的 DNA 的技术是本领域内已知的，且包括从基因组 DNA 分离，从 cDNA 制备，或它们的组合。可通过例如使用熟知的聚合酶链式反应(PCR)实现从这种基因组 DNA 克隆感兴趣的多核苷酸。参见，例如，Innis 等，1990, *PCR Protocols: A Guide to Methods and Application*, Academic Press, New York。克隆步骤可以涉及包含编码多肽的核酸序列的期望核酸片段的切除与分离，向载体分子中插入该片段，并将重组载体并入突变芽孢杆菌属细胞，其中将复制多个拷贝或克隆的所述核酸序列。DNA 可以是基因组、cDNA、RNA、半合成、合成来源的，或它们的任意组合。

可以用许多方式操作编码感兴趣的多肽的 DNA 以提供 DNA 在合适的芽孢杆菌属宿主细胞中的表达。用于编码感兴趣的多肽的 DNA 的核酸构建体和重组表达载体的构建可以如本文 ComS 多肽或 ComK 多肽的表达中所述进行。

DNA 还可以是调控序列，例如，启动子，用于操作感兴趣的基因的表达。调控序列的非限定性实例在本文中描述。

DNA 还可以是用于将芽孢杆菌属细胞中感兴趣的基因失活的核酸构建体。

DNA 的范围不限定于上述公开的具体实例，因为这些实例意欲作为对本发明几个方面的说明。

产生方法

本发明还涉及产生生物物质的方法，其包括：(a)在有益于产生生物物质的条件下培养芽孢杆菌属宿主细胞，所述细胞用编码或参与具有生物活性的物质的表达的外源 DNA 转化，其中通过至少一个拷贝导入的核酸构建体使芽孢杆

菌属宿主细胞成为感受态，所述核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区，其中编码 ComS 多肽的多核苷酸对于芽孢杆菌属宿主细胞是异源的，所述细胞在导入该核酸构建体之前是非感受态的；和(b)回收具有生物活性的物质。

在优选的方面，上述成为感受态的芽孢杆菌属宿主细胞进一步包含至少一个拷贝的导入的第二核酸构建体，所述第二核酸构建体包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区，赋予芽孢杆菌属宿主细胞更进一步的感受态。

使用本领域已知的方法在适合于产生感兴趣的多肽的营养培养基中培养芽孢杆菌属宿主细胞。例如，可以通过在合适培养基中和允许表达和/或分离感兴趣的多肽的条件下进行的摇瓶培养，和在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布的组成制备(例如，在美国典型培养物保藏中心的目录中)。分泌的感兴趣的物质能够从培养基中直接回收。

感兴趣的生物物质，例如多肽，可以使用本领域已知的特定用于该物质的方法来检测。这些检测方法可包括特异性抗体的使用、高效液相层析、毛细管层析、酶产物的形成、酶底物的消失或 SDS-PAGE。例如，酶试验(enzyme assay)可用于测定具有酶活性的多肽的活性。对于很多酶，用于测定酶活性的方法在本领域中已知(参见，例如，D. Schomburg 和 M. Salzmänn (编), *Enzyme Handbook*, Springer-Verlag, New York, 1990)。

所得感兴趣的生物物质，例如，多肽，可以用本领域已知的方法分离。例如，感兴趣的多肽可以通过常规方法从培养基中分离，所述常规方法包括但不限于离心、过滤、提取、喷雾干燥、蒸发或沉淀。然后分离的感兴趣的生物物质可以通过多种本领域已知的方法进一步纯化，所述方法包括但不限于层析(例如，离子交换、亲和、疏水、层析聚焦和大小排阻)、电泳方法(例如，制备型(preparative)等电聚焦(IEF))、差示溶解度(例如，硫酸铵沉淀)或提取(参见，例如，*Protein Purification*, J.-C. Janson 和 Lars Ryden 编, VCH Publishers, New York, 1989)。

基因的修饰

本发明还涉及产生亲本芽孢杆菌属细胞突变体的方法，其包括(a)将包含核酸的外源 DNA 转化入亲本芽孢杆菌属细胞，以修饰亲本芽孢杆菌属细胞中编码多肽的基因，这产生在相同条件下培养时与亲本细胞相比产生较少的所述多肽或产生的多肽生物活性较低的突变细胞；其中亲本芽孢杆菌属细胞通过至少一个拷贝导入的第一核酸构建体成为感受态，所述第一核酸构建体包含与编码 ComS 多肽的多核苷酸可操作连接的启动子区，其中编码 ComS 多肽的多核苷酸对于亲本芽孢杆菌属细胞是外源的，所述细胞在导入第一核酸构建体之前是非感受态的；和(b)分离突变细胞。

在优选的方面，修饰是将使其产物的产生消失的基因失活。

在另一个优选的方面，上述成为感受态的芽孢杆菌属细胞进一步包含至少一个拷贝导入的第二核酸构建体，其包含与编码 ComK 多肽的多核苷酸可操作连接的启动子区，赋予芽孢杆菌属细胞更进一步的感受态。

可以使用本领域公知的方法构建包含修饰基因的突变细胞，例如，通过插入、破坏、替代或缺失。待修饰的基因可以是，例如，编码区或其对于活性而言关键的部分，或编码区表达所需的调节元件。这样的调节或调控序列的实例可以是启动子序列或其功能部分，即，足以影响基因表达的部分。其他用于可能的修饰的调控序列包括，但不限于，前导序列、聚腺苷酸化序列、前肽序列、信号肽序列、转录终止子和转录激活因子。

可以通过在基因或其转录或翻译所需的调节元件中导入、取代或去除一个或多个核苷酸而完成基因的修饰。例如，可以插入或删除核苷酸，导致终止密码子的导入，起始密码子的去除，或开放阅读框的改变。

修饰基因的方便方法的实例是基于基因置换、基因缺失或基因破坏的技术。例如，在基因破坏方法中，将对应于内源核苷酸序列的核酸序列在体外突变，产生缺陷型核酸序列，然后将所述缺陷型核酸序列导入亲本细胞，产生缺陷型基因。通过同源重组，缺陷核酸序列替代内源核苷酸序列。可能更理想的是，缺陷型核苷酸序列还编码标记，其可用于选择核苷酸序列已被修饰或破坏的转化体。在特别优选的方面，用选择性标记如本文所述的那些破坏核苷酸序列。

这样构建的芽孢杆菌属突变细胞作为宿主细胞用于表达对细胞是天然或外源的多肽特别有用。因此，本发明进一步涉及产生天然或外源多肽的方法，

其包括：(a)在有益于多肽产生的条件下培养突变细胞；和(b)回收所述多肽。术语“外源多肽”在本文定义为对于宿主细胞不是天然的多肽，其中经过修饰而使天然序列改变的天然蛋白质，或作为通过重组DNA技术操作宿主细胞的结果而表达量发生改变的天然蛋白质。

能在这样的突变体中表达的多肽的实例在本文描述。

用于培养和纯化感兴趣的产物的方法可以通过本领域已知和本文所述的方法进行。

通过下述实施例进一步描述本发明，但不应将下述实施例理解为对本发明范围的限制。

实施例

DNA 测序

使用 Applied Biosystems Model 3130X Genetic Analyzer (3130X 型遗传分析仪) (Applied Biosystems, Foster City, CA, USA)，利用染料终止子化学(dye terminator chemistry) (Giesecke 等, 1992, *Journal of Virol. Methods* 38: 47-60)进行 DNA 测序。使用 phred/phrap/consed (University of Washington, Seattle, WA, USA)用测序特定引物组装序列。

大肠杆菌菌株

使用 ONE SHOT® TOP10 化学感受态大肠杆菌细胞 (Invitrogen Corporation, Carlsbad, CA, USA)，SURE®感受态大肠杆菌细胞(Stratagene, La Jolla, CA, USA)，XL1-Blue 感受态大肠杆菌细胞(Stratagene, La Jolla, CA, USA)，和 SOLOPACK® Gold 超感受态大肠杆菌细胞(Stratagene, La Jolla, CA, USA)用于常规的质粒构建与增殖。

芽孢杆菌属菌株

枯草芽孢杆菌 168Δ4 源自枯草芽孢杆菌典型菌株 168 (BGSC 1A1, Bacillus Genetic Stock Center, Columbus, OH, USA)，并在 *spoIIAC*、*aprE*、*nprE* 和 *amyE* 基因中有缺失。基本如对枯草芽孢杆菌 A164Δ5 所述进行这四个基因的缺失，过程如美国专利 5,891,701 中详细所述。向枯草芽孢杆菌 168Δ4 的培养物中补充 50 μg/ml 色氨酸。

在枯草芽孢杆菌 168 Δ 4 (枯草芽孢杆菌 168 Δ sigF Δ aprE Δ nprE Δ amyE) 中构建了所有温度敏感型质粒。使用枯草芽孢杆菌 A164 Δ 5 (枯草芽孢杆菌 A164 Δ spoIIAC, Δ aprE, Δ nprE, Δ amyE, Δ srfAC) 作为宿主评价地衣芽孢杆菌 *comK* 过表达对枯草芽孢杆菌转化效率的影响。枯草芽孢杆菌菌株 MDT101, 如本文所述, 表达地衣芽孢杆菌 SJ1904 限制-修饰系统(restriction-modification system) 的 DNA 甲基转移酶成分, 将该菌株用于在转化实验前修饰质粒 DNA。地衣芽孢杆菌 SJ1904 (美国专利 5,733,753) 用作宿主用于表达枯草芽孢杆菌 *comS* 基因, 用于增加地衣芽孢杆菌 *comK* 基因的表达, 并且用于后续在地衣芽孢杆菌中诱导感受态。

根据 Anagnostopolous 和 Spizizen, 1961, *J. Bacteriol.* 81: 741-746 的方法转化枯草芽孢杆菌。根据 Susanna 等, 2004, *J. Bacteriol.* 186: 1120-1128 的方法, 通过电穿孔转化地衣芽孢杆菌菌株 SJ1904。用已经甲基化的质粒 DNA 转化限制-健全型(restriction-proficient)地衣芽孢杆菌菌株, 赋予其对地衣芽孢杆菌中限制(restriction)的抗性。为了提供正确的甲基化, 从枯草芽孢杆菌 MDT101 的先前转化体分离了 DNA。

培养基

2X YT 平板由每升 16 g 胰蛋白胨、10 g 酵母提取物, 5 g NaCl, 和 15 g 细菌用琼脂(bacto agar)组成。

2X YT 氨苄青霉素平板由每升 16 g 胰蛋白胨、10 g 酵母提取物, 5 g NaCl, 和 15 g 细菌用琼脂, 补充有 100 μ g/ml 氨苄青霉素组成。

TBAB 由 Tryptose Blood Agar Base (胰蛋白胨血琼脂基底) (BD Diagnostics, Franklin Lakes, NJ, USA)组成。

LB 培养基由每升 10 g 胰蛋白胨、5 g 酵母提取物和 5 g NaCl 组成。

LB 平板由 LB 培养基与每升 15 g 细菌用琼脂组成。

LB 红霉素培养基由包含 5 μ g/ml 红霉素的 LB 培养基组成。

LB 红霉素/林可霉素平板由 LB 培养基与每毫升 1 μ g 红霉素和 25 μ g 林可霉素组成。

LB 氯霉素平板由 LB 培养基与每毫升 5 μ g 氯霉素组成。

LB 红霉素/氯霉素平板由 LB 培养基与每毫升 1 μ g 红霉素和 5 μ g 氯霉素组成。

VY 培养基由每升 25 g 小牛肉浸出物(veal infusion) (BD Diagnostics, Franklin Lakes, NJ, USA)和 5 g 酵母提取物组成。

Spizizen I培养基由 1X Spizizen 盐、0.5%葡萄糖、0.1%酵母提取物和 0.02%酪蛋白水解物组成。这个培养基在本文也称作基本培养基。

1X Spizizen 盐由每升 6 g KH_2PO_4 , 14 g K_2HPO_4 , 2 g $(\text{NH}_4)_2\text{SO}_4$, 1 g 柠檬酸钠和 0.2 g of MgSO_4 组成, pH 7.0。

Spizizen II培养基由 Spizizen I培养基补充 0.5 mM CaCl_2 和 2.5 mM MgCl_2 组成。

TBAB 红霉素/林可霉素平板由 TBAB 培养基和每毫升 1 μg 红霉素和 25 μg 林可霉素组成。

实施例 1: 测定地衣芽孢杆菌菌株 SJ1904 的基因组序列

从使用 454 DNA 测序技术(Margulies 等, 2005, *Nature* 437: 376-380)产生的重叠群(contig), 使用 Sanger 测序技术的随机配对读数(random paired reads), 和为了关闭缺口和解析重复序列的来自基因组 DNA 的 PCR 片段的读数, 确定地衣芽孢杆菌菌株 SJ1904 完整染色体的基因组序列。使用 Phrap 组装测序数据, 并在 Consed 中编辑和查看。使用 Glimmer (Delcher 等, 1999, *Nucleic Acids Research* 27: 4636-4641)由基因组 DNA 序列预测基因模型。使用 E-值阈值为 1×10^{-5} 的 BLASTP, 通过与无冗余数据库 PIR-NREF (Wu 等, 2002, *Nucleic Acids Research* 30: 35-37)的比较, 对基因模型进行机器注解。

实施例 2: 地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶基因的鉴定

使用 BLASTP (Altschul 等, 1997, *Nucleic Acids Research* 25: 3389-3402)将地衣芽孢杆菌菌株 SJ1904 基因模型的推定的氨基酸序列与来自 REBASE (Roberts, R.J., Macelis, M., Rebase. 2005)的蛋白质序列进行比较。因为 DNA 甲基转移酶具有中等水平的序列保守性, 所以这项分析鉴定了这个基因组中所有推定的 DNA 甲基转移酶。使用通过 InterProScan v3.3 版执行的 Prints-S 16 版, 鉴定了 M.Bli1904II 中的胞嘧啶特异性 DNA 甲基转移酶特征(signature)。此外, 发现存在于胞嘧啶特异性 DNA 甲基转移酶中的六个高度保守的基序(motif)在地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶中也是保守的。

实施例 3: 地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶基因的表征

地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶基因的核苷酸序列(SEQ ID NO: 51)和推定的氨基酸序列(SEQ ID NO: 52)如图 2A 和 2B 中所示。编码序列为 1014 bp, 其包括终止密码子。编码区为 36.1% G+C。编码的预测蛋白质为 337 个氨基酸, 分子量为 38.5 kDa。

使用 Needleman-Wunsch 算法(Needleman 和 Wunsch, 1970, 见上文), 如 EMBOSS 的 Needle 程序中所执行的, 缺口开放罚分为 10, 缺口延伸罚分为 0.5, 使用 EBLOSUM 62 矩阵, 确定了氨基酸序列的比较性配对全局比对(comparative pairwise global alignment)。比对显示, 地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶的推定氨基酸序列与韦氏芽孢杆菌 C-5 胞嘧啶特异性 DNA 甲基转移酶前体(UniRef100_Q2AVE0)共享 64% 的同一性, 并且与 *Oceanobacillus iheyensis* 的胞嘧啶特异性 DNA 甲基转移酶(UniRef100_Q8EL98)共享 47% 的同一性。当使用 Needle 标记为“最长同一性”的输出结果作为百分比同一性并如下计算时:

$$(\text{相同的残基} \times 100) / (\text{比对长度} - \text{比对中缺口数目})$$

地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶的推定氨基酸序列与韦氏芽孢杆菌 C-5 胞嘧啶特异性 DNA 甲基转移酶前体(UniRef100_Q2AVE0)共享 68.5% 的同一性, 并且与 *Oceanobacillus iheyensis* 的胞嘧啶特异性 DNA 甲基转移酶(UniRef100_Q8EL98)共享 55.9% 的同一性。

实施例 4: 地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶基因的克隆

为在枯草芽孢杆菌中表达而通过 PCR 克隆了地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶基因。

根据 Pitcher 等, 1989, *Lett. Appl. Microbiol.* 8: 151-156 的方法从地衣芽孢杆菌 SJ1904 分离了基因组 DNA。图 3 显示了包含编码 Bli1904II 限制性内切核酸酶和 M.Bli1904II DNA 甲基转移酶的基因的地衣芽孢杆菌染色体区域。使用如下所示的引物 999611 和 999612, 通过 PCR 从地衣芽孢杆菌 SJ1904 基因组 DNA 扩增了地衣芽孢杆菌 SJ1904 染色体中约 1043 bp 的片段, 其包括 M.Bli1904II DNA 甲基转移酶基因的核糖体结合位点和编码区, 包含 SEQ ID NO: 53 的核苷酸 2019-3049 (图 3A、3B 和 3C)。引物 999611 并入了 *SacI* 限制性位点, 而引物 999612 并入了 *MluI* 限制性位点。

引物 999611:

5'-GAGCTCTGCAAGGAGGTATAATTTTG-3' (SEQ ID NO: 54)

引物 999612:

5'-ACGCGTTTATTCAGCTATTGCATATTC-3' (SEQ ID NO: 55)

使用 *Pfx* PLATINUM® DNA 聚合酶(Invitrogen, Carlsbad, CA, USA)进行 PCR。扩增反应(50 μ l)由下述组成: 1X *Pfx* 扩增缓冲液(Invitrogen, Carlsbad, CA, USA), 1 mM MgSO₄, 300 μ M 每种 dNTP, 0.3 μ M 每条引物, 1.25 单位 PLATINUM® *Pfx* DNA 聚合酶, 和约 200 ng 模板 DNA。使用 ROBOCYCLER®40 温度循环仪(Stratagene Corporation, La Jolla, CA, USA)进行反应, 程序为 95°C 2 分钟的 1 个循环; 95°C 1 分钟、55°C 1 分钟和 68°C 1 分钟的 30 个循环; 和 68°C 3 分钟的 1 个循环。

使用用于测序的 ZERO BLUNT® TOPO® PCR 克隆试剂盒(Invitrogen, Carlsbad, CA, USA)将获得的约 1043 bp PCR 产物克隆入载体 pCR4Blunt, 并根据制造商说明转化入 ONE SHOT® TOP10 化学感受态大肠杆菌细胞中。使用 Plasmid Midi 试剂盒(QIAGEN Inc., Valencia, CA, USA)从一个转化体分离质粒 DNA, 并通过用 *Eco*RI、*Nco*I 和 *Sna*BI 消化然后在 TBE (50 mM Tris 碱 -50 mM 硼酸-1 mM EDTA 二钠)缓冲液中的 0.8%琼脂糖电泳进行验证, 用 *Eco*RI 消化获得了 3939 bp 和 1061 bp 的期望片段; *Nco*I 为 3217 bp 和 1783 bp; 而 *Sna*BI 为 4165 bp 和 835 bp。通过 DNA 测序确认了克隆的 PCR 片段的 DNA 序列。将这个质粒命名为 pMDT138 (图 4)。

根据制造商说明, 将质粒 pMDT138 转化入大肠杆菌 XL1-Blue 细胞 (Stratagene Corporation, La Jolla, CA, USA), 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。将一个转化体命名为 MDT45, 并按照布达佩斯条约的条款于 2006 年 9 月 7 日将其保藏在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), 1815 University Street, Peoria, Illinois, 61604, 并且给予登录号 NRRL B-41967。

实施例 5: pMDT100 的构建

质粒 pMDT100 是大肠杆菌复制子, 其包含 P_{amyL4199}/P_{短共有 amyQ}/P_{cryIIIa}/cryIIIastab 三联启动子, 其驱动克劳氏芽孢杆菌碱性蛋白酶基因(*aprH*)的表达。

这个 *aprH* 表达盒和 pC194 的 *cat* 基因(Horinouchi 和 Weisblum, 1982, *J. Bacteriol.* 150: 804-814)两侧侧翼均为枯草芽孢杆菌 α -淀粉酶(*amyE*)基因的片段, 允许通过双同源重组藉由两个 *amyE* 片段在枯草芽孢杆菌染色体的 *amyE* 基因座插入 *aprH* 表达盒和 *cat* 基因。用另一个基因替代 pMDT100 中的 *aprH* 基因允许将所述基因插入枯草芽孢杆菌染色体并在枯草芽孢杆菌中表达。pMDT100 的构建如下所述。

质粒 pNBT51。根据制造商的说明, 使用 QIAGEN®质粒试剂盒(QIAGEN Inc., Valencia, CA, USA)从大肠杆菌 DH5 α 宿主分离了质粒 pNBT10 (pDG268MCS-Pr_{cryIII A}/cryIII A_{stab}/SAV; 美国专利 No. 6,255,076), 并用 *Cla*I 和 *Sca*I 消化。裂解发生在 *aprH* 编码序列大约在密码子 326 处的 *Cla*I 位点, 而不是大约在密码子 23 处的 *Cla*I 位点, 后者通过大肠杆菌 Dam DNA 甲基转移酶导致的甲基化而被阻断。使用 Klenow 片段(New England Biolabs, Inc., Beverly, MA, USA)和 dNTP, 根据制造商的说明将 *Cla*I 末端平端化。通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析了消化的质粒, 并使用 QIAQUICK®凝胶提取试剂盒(QIAGEN Inc., Valencia, CA, USA)纯化了约 6615 bp 的载体片段。用 *Sal*I 和 *Sca*I 消化质粒 pOS4301 (Bacillus Genetic Stock Center, Ohio State University, Columbus, OH, USA), 并使用 Klenow 片段和 dNTP 将 *Sal*I 末端平端化, 如上所述。通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析了消化的质粒, 并使用 QIAQUICK®凝胶提取试剂盒纯化了携带大肠杆菌 *rrnB* 转录终止子的约 840 bp 的片段。可以从载体 pKK223-3 (GE Healthcare, Piscataway, NJ, USA) (图 5)分离同样的 840 bp *Sal*I/*Sca*I 片段。根据制造商的说明, 用 T4 DNA 连接酶(Roche Diagnostics Corporation, Indianapolis, IN, USA)将 pNBT10 载体片段和携带终止子的片段连接在一起, 并根据制造商的说明, 用所述连接物转化了大肠杆菌 DH5 α (Gibco BRL, Gaithersburg, MD, USA), 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。将得到的质粒命名为 pNBT51 (pDG268- P_{cryIII A}/cryIII A_{stab}/SAV Δ) (图 6)。

质粒 pNBT52。用 *Sfi*I 消化质粒 pNBT51, 在 11°C 与 T4 DNA 聚合酶(Roche Diagnostics Corporation, Indianapolis, IN, USA)和 25 μ M 每种 dNTP 一起温育 20 分钟, 将末端平端化, 然后 75°C 温育 10 分钟, 将聚合酶热失活。然后用 *Dra*III 消化末端平端化的质粒, 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析, 使用 QIAQUICK®凝胶提取试剂盒纯化了约 5920 bp 的载体片段。用 *Dra*III

和 *Ecl136II* 消化质粒 pNBT20 (pDG268MCS-P_{短共有 amyQ}/SAV; 美国专利 No. 6,255,076), 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 1641 bp 的携带短共有 *amyQ* 启动子(P_{短共有 amyQ})的片段。如上所述连接 pNBT51 载体片段和 P_{短共有 amyQ} 片段, 并如上所述用所述连接物转化了大肠杆菌 DH5 α , 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。使用 QIAPREP® 8 小量制备试剂盒(QIAGEN, Valencia, CA, USA)从几个转化体分离了质粒 DNA, 用 *SphI* 消化, 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析。将具有期望的约 4873 bp 和 2688 bp 的限制性片段的一个质粒命名为 pNBT52 (pDG268-P_{短共有 amyQ}/P_{cryIII A}/cryIII A_{stab}/SAV Δ) (图 7)。

质粒 pNBT53。用 *SfiI* 和 *SacI* 消化质粒 pNBT6 (pHP13amp-SAV; 美国专利 No. 6,255,076), 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析, 使用 QIAQUICK®凝胶提取试剂盒纯化了约 6438 bp 的载体片段。用 *SfiI* 和 *SacI* 消化质粒 pNBT52, 并通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析, 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 727 bp 的携带 P_{短共有 amyQ}/P_{cryIII A}/cryIII A_{stab} 串联启动子的片段。如上所述连接 pNBT6 载体片段和 P_{短共有 amyQ}/P_{cryIII A}/cryIII A_{stab} 片段, 并如上所述用所述连接物转化了大肠杆菌 DH5 α 细胞, 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。使用 QIAPREP® 8 小量制备试剂盒(QIAGEN, Valencia, CA, USA), 从几个转化体分离了质粒 DNA, 用 *PvuII* 消化, 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析。将具有期望的约 4903 bp、1320 bp 和 942 bp 的限制性片段的一个质粒命名为 pNBT53 (pHP13amp-P_{短共有 amyQ}/P_{cryIII A}/cryIII A_{stab}/SAV) (图 8)。

质粒 pNBT54。用 *SfiI* 和 *BamHI* 消化质粒 pNBT1 (pDG268MCS; 美国专利 No. 6,255,076), 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析, 使用 QIAQUICK®凝胶提取试剂盒纯化了约 6040 bp 的载体片段。用 *SfiI* 和 *BamHI* 消化质粒 pNBT53, 并通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析, 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 1953 bp 的携带 P_{短共有 amyQ}/P_{cryIII A}/cryIII A_{stab}/SAV 表达盒的片段。如上所述连接 pNBT1 载体片段和 P_{短共有 amyQ}/P_{cryIII A}/cryIII A_{stab}/SAV 片段, 并如上所述用所述连接物转化了大肠杆菌 DH5 α 细胞, 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。使用 QIAPREP® 8 小量制备试剂盒, 从几个转化体分离了质粒 DNA, 并通过用 *SfiI* 和 *BamHI* 同时消化, 然后是 TBE 缓冲液中的 0.8%琼脂糖凝胶电泳来分析。

将具有期望的约 6040 bp 和 1953 bp 的限制性片段的一个质粒命名为 pNBT54 (pDG268MCS-P_{短共有 amyQ}/P_{cryIII A}/cryIII Astab/SAV) (图 9)。

质粒 pNBT35。用 *Sfi*I 和 *Bam*HI 消化质粒 pNBT2 (pDG268MCSΔ-Pr_{cryIII A}/cryIII Astab/SAV; 美国专利 No. 6,255,076), 并通过 TBE 缓冲液中的 0.8%琼脂糖凝胶电泳分析, 使用 QIAQUICK®凝胶提取试剂盒纯化了约 5394 bp 的载体片段。用 *Sfi*I 和 *Bam*HI 消化质粒 pNBT54, 并通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析, 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 1953 bp 的携带 P_{短共有 amyQ}/P_{cryIII A}/cryIII Astab/SAV 表达盒的片段。如上所述连接 pNBT2 载体片段和 P_{短共有 amyQ}/P_{cryIII A}/cryIII Astab/SAV 片段, 并如上所述用所述连接物转化了大肠杆菌 DH5α 细胞, 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。使用 QIAPREP® 8 小量制备试剂盒, 从几个转化体分离了质粒 DNA, 用 *Nco*I 消化, 并通过 TBE 缓冲液中的 0.8%琼脂糖凝胶电泳分析。将具有期望的约 5492 bp 和 1855 bp 的限制性片段的一个质粒命名为 pNBT35 (pDG268MCSΔ-P_{短共有 amyQ}/P_{cryIII A}/cryIII Astab/SAV) (图 10)。

质粒 pNBT30。构建了质粒 pNBT30, 其包含 *amyL* 基因启动子的 *amyL*4199 变体的 PCR 克隆(美国专利 No. 6,100,063)。根据 Pitcher 等, 1989, 见上文的方法分离了地衣芽孢杆菌 SJ1904 的基因组 DNA。使用如下所示的引物 950872 和 991151, 通过 PCR 从地衣芽孢杆菌 SJ1904 的基因组 DNA 扩增了 *amyL*4199 启动子(P_{amyL4199})基因。引物 950872 并入了 *Sfi*I 限制性位点, 而引物 991151 并入了 *Sac*I 限制性位点和 P_{amyL4199} 的变体核苷酸。

引物 950872:

5'-CCAGGCCTTAAGGGCCGCATGCGTCCTTCTTTGTGCT-3' (SEQ ID NO: 56)

引物 991151:

5'-GAGCTCCTTTCAATGTGATACATATGA-3' (SEQ ID NO: 57)

使用 AMPLITAQ® Gold DNA 聚合酶(Applied Biosystems, Foster City, CA, USA)根据制造商的推荐进行了 PCR, 只是 MgCl₂ 浓度是 3 mM, 而不是标准的 1.5 mM。扩增反应(50 μl)由下述组成: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 3.0 mM MgCl₂, 200 μM 每种 dNTP, 0.5 μM 每条引物, 0.25 单位 AMPLITAQ® Gold DNA 聚合酶, 和约 200 ng 模板 DNA。在 ROBOCYCLER® 40 温度循环仪中进行 PCR, 程序为 95°C 9 分钟的 1 个循环; 95°C 1 分钟、55°C 1 分钟和

72°C 1 分钟的 30 个循环; 和 72°C 3 分钟的 1 个循环。

使用 TOPO® TA 克隆试剂盒(Invitrogen, Carlsbad, CA, USA)将得到的约 625 bp 的 PCR 产物克隆入载体 pCR2.1, 并根据制造商的说明转化入 ONE SHOT® TOP10 化学感受态大肠杆菌细胞(Invitrogen, Carlsbad, CA, USA)。使用 QIAPREP® 8 小量制备试剂盒从几个转化体分离了质粒 DNA, 并通过用 *EcoRI* 消化, 然后是 TBE 缓冲液中 0.8%的琼脂糖电泳来分析克隆的 PCR 片段的存在。将一个具有预期的约 3913 bp 和 640 bp 的限制性片段的质粒命名为 pNBT30 (pCR2.1-amyL4199) (图 11)。通过 DNA 测序确认了克隆的 PCR 片段的 DNA 序列。

质粒 pNBT31。用 *SfiI* 和 *SacI* 消化质粒 pNBT3 (pDG268MCS Δ neo-Pr_{cryIII}A/cryIII_Astab/SAV; 美国专利 No. 6,255,076), 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析, 使用 QIAQUICK®凝胶提取试剂盒纯化了约 7931 bp 的载体片段。用 *SfiI* 和 *SacI* 消化质粒 pNBT30, 并通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析, 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 612 bp 的携带 P_{amyL4199} 的片段。如上所述连接 pNBT3 载体片段和 P_{amyL4199} 片段, 并根据制造商的说明用所述连接物转化了大肠杆菌 XL1-Blue 细胞(Stratagene Corporation, La Jolla, CA, USA), 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。使用 QIAPREP® 8 小量制备试剂盒从几个转化体分离了质粒 DNA, 用 *NcoI* 消化, 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析。将具有预期的约 6802 bp 和 1741 bp 的限制性片段的一个质粒命名为 pNBT31 (图 12)。

质粒 pNBT36。用 *SfiI* 消化质粒 pNBT35, 并使用 T4 DNA 聚合酶和 dNTP 将末端平端化, 如上所述。然后将末端平端化的质粒用 *DraIII* 消化, 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析。使用 QIAQUICK®凝胶提取试剂盒纯化了约 5808 bp 的载体片段。用 *DraIII* 和 *Ecl136II* 消化质粒 pNBT31, 并通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析, 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 2150 bp 携带 P_{amyL4199} 的片段。如上所述连接 pNBT35 载体片段和 P_{amyL4199} 片段, 并根据制造商的说明用所述连接物转化了大肠杆菌 SURE®细胞 (Stratagene Corporation, La Jolla, CA, USA), 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。使用 QIAPREP® 8 小量制备试剂盒从几个转化体分离了质粒 DNA, 用 *NcoI* 消化, 并通过 TBE 缓冲液中的 0.8%琼脂糖电泳分析。将具有预期的约 5492 bp 和 2466 bp 的限制性片段的一个质粒命名为 pNBT36

(图 13)。

质粒 pMDT100。用 *Dra*III 和 *Sac*I 消化质粒 pNBT13 (pDG268 Δ neo-P_{amyL}/P_{cryIII_A}/cryIII_Astab/SAV;美国专利 No. 6,255,076), 并使用 QIAQUICK®凝胶提取试剂盒纯化了约 6395 bp 的载体片段。用 *Dra*III 和 *Sac*I 消化质粒 pNBT36, 并通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析, 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 2873 bp 的携带 P_{amyL4199}/P_{短共有 amyQ}/P_{cryIII_A} 三联启动子的片段。如上所述连接 pNBT13 载体片段和 P_{amyL4199}/P_{短共有 amyQ}/P_{cryIII_A} 片段, 并如上所述用所述连接物转化了大肠杆菌 SURE®细胞, 在 37°C 在 2X YT 氨苄青霉素平板上选择氨苄青霉素抗性。使用 QIAPREP® 8 小量制备试剂盒从几个转化体分离了质粒 DNA, 用 *Apa*I 消化, 并通过 TBE 缓冲液中的 0.8%琼脂糖分析。将具有预期的约 4974 bp 和 4294 bp 的限制性片段的一个质粒命名为 pMDT100 (图 14)。

实施例 6: 地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶基因在枯草芽孢杆菌中的表达

将地衣芽孢杆菌 M.Bli1904II DNA 甲基转移酶基因插入枯草芽孢杆菌的染色体中, 以在该宿主中表达甲基转移酶, 从而允许枯草芽孢杆菌中 DNA 的甲基化。

用 *Sac*I 和 *Mlu*I 消化质粒 pMDT100 并通过 TBE 缓冲液中 0.8%的琼脂糖电泳分析, 并使用 QIAQUICK®凝胶提取试剂盒纯化了约 8100 bp 的载体片段。用 *Sac*I 和 *Mlu*I 消化质粒 pMDT138, 并且使用 QIAQUICK®凝胶提取试剂盒纯化了约 1033 bp 的携带 M.Bli1904II 基因的片段。如上所述连接 pMDT100 载体片段和 M.Bli1904II 基因片段。此连接将 M.Bli1904II 基因置于 P_{amyL4199}/P_{短共有 amyQ}/P_{cryIII_A}/cryIII_Astab 启动子的下游和 *aprH* 转录终止子的上游。根据 Anagnostopoulos 和 Spizizen, 1961, *J. Bacteriol.* 81: 741-746 的方法用所述连接物转化枯草芽孢杆菌 168 Δ 4, 在 37°C 在 TBAB 氯霉素平板上选择氯霉素抗性的转化体。在 37°C 在 TBAB 新霉素平板上筛选新霉素敏感的抗氯霉素转化体, 以确定是否已将 DNA 通过双交换插入到枯草芽孢杆菌染色体的 *amyE* 基因中。

使用如下所示的引物 994112 和 999592 (其分别结合在三联启动子和 M.Bli1904II DNA 甲基转移酶基因内)和如下所示的引物 999611 和 960456 (其分别结合在 M.Bli1904II DNA 甲基转移酶基因和 *amyE* 基因内), 通过 PCR 确

认了在 *amyE* 基因座存在 M.Bli1904II DNA 甲基转移酶表达盒。将在 *amyE* 基因座包含 *cat* 基因和 M.Bli1904II DNA 甲基转移酶表达盒的一个这样的转化体，命名为枯草芽孢杆菌 MDT101。

引物 994112:

5'-GCGGCCGCTCGCTTTCCAATCTGA-3' (SEQ ID NO: 58)

引物 999592:

5'-ATCGATCAGCTTGGATAAACCCCTA-3' (SEQ ID NO: 59)

引物 999611:

5'-GAGCTCTGCAAGGAGGTATAATTTTG-3' (SEQ ID NO: 60)

引物 960456:

5'-CGTCGACGCCTTTGCGGTAGTGGTGCTT-3' (SEQ ID NO: 61)

使用 *Taq* DNA 聚合酶(New England Biolabs, Inc., Beverly, MA, USA)根据制造商的说明进行了 PCR。扩增反应(50 μ l)由下述组成: 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 3.0 mM MgCl₂, 200 μ M 每种 dNTP, 0.5 μ M 每条引物, 0.25 单位 *Taq* DNA 聚合酶, 和约 200 ng 基因组 DNA。在 ROBOCYCLER® 40 温度循环仪中进行 PCR, 程序为 95°C 2 分钟的 1 个循环; 95°C 2 分钟、55°C 2 分钟和 72°C 2 分钟的 30 个循环; 和 72°C 3 分钟的 1 个循环。

实施例 7: 地衣芽孢杆菌感受态基因的编目

在查询指令中使用术语“感受态”为关键字搜索枯草芽孢杆菌数据库 (Subtilist; Moszer 等, 2002, *Nucleic Acids Res.* 30: 62-65)产生了 50 个基因的列表, 这些基因在该物种的感受态发展中起作用(表 1)。使用 BLAST (McGinnis 和 Madden, 2004, *Nucleic Acids Res.* 32: W20-5), 用 1×10^{-10} 的最小期望分值, 在地衣芽孢杆菌 ATCC 14580 的基因组序列(Rey 等, 2004, *Genome Biol.* 5: R77) 中鉴定了枯草芽孢杆菌感受态基因的直系同源基因。

表 1. 由枯草芽孢杆菌和地衣芽孢杆菌基因组编码的感受态基因的比较

枯草芽孢杆菌基因	SwissProt	功能	地衣芽孢杆菌中是否存在
<i>abrB</i>	P08874	转换状态基因的多效转录调节子	+
<i>addA</i>	P23478	ATP 依赖型脱氧核糖核酸酶(A 亚基)	+

<i>addB</i>	P23477	ATP 依赖型脱氧核糖核酸酶(B 亚基)	+
<i>bdbC</i>		巯基-二硫化物氧化还原酶	+
<i>cinA</i>	P46323	感受态损伤诱导型蛋白质	+
<i>clpC</i>	P37571	III 类压力响应相关 ATPase	+
<i>clpP</i>	P80244	ATP 依赖型 Clp 蛋白酶蛋白水解亚基(III 类热休克蛋白质)	+
<i>clpX</i>	P50866	ATP 依赖型 Clp 蛋白酶 ATP 结合亚基(III 类热休克蛋白质)	+
<i>comA</i>	P14204	后期感受态基因的双组分响应调节子/表面活性肽产生	+
<i>comC</i>	P15378	ComGC, ComGD, ComGE, ComGG 的加工和易位所需的后期感受态蛋白质	+
<i>comEA</i>	P39694	外源 DNA 结合蛋白	+
<i>comEB</i>	P32393	DNA 结合和摄取所需的后期感受态操纵子	+
<i>comEC</i>	P39695	DNA 结合和摄取所需的后期感受态操纵子	+
<i>comER</i>	P39696	感受态的非必需基因	+
<i>comFA</i>	P39145	DNA 摄取所需的后期感受态蛋白质	+
<i>comFB</i>	P39146	后期感受态基因	+
<i>comFC</i>	P39147	后期感受态基因	+
<i>comGA</i>	P25953	后期感受态基因	+
<i>comGB</i>	P25954	DNA 转运机制	+
<i>comGC</i>	P25955	外源 DNA 结合	+
<i>comGD</i>	P25956	DNA 转运机制	+
<i>comGE</i>	P25957	DNA 转运机制	+
<i>comGF</i>	P25958	DNA 转运机制	+
<i>comGG</i>	P25959	DNA 转运机制	+
<i>comK</i>	P40396	感受态转录因子 (CTF)	+
<i>comP</i>	Q99027	与早期感受态有关的双组分传感器组氨酸激酶	IS3BII 插入
<i>comQ</i>	P33690	后期感受态操纵子(<i>comG</i>)和表面活性肽表达(<i>srfA</i>)的转录调节子	+
<i>comS</i>	P80355	感受态信号转导通路的调节组分之间的装配连接	-
<i>comX</i>	P45453	感受态信息素前体	+
<i>comZ</i>		后期感受态基因	+

<i>degS</i>	P13799	与降解性酶和感受态调节有关的双组分传感器组氨酸激酶	+
<i>degU</i>	P13800	与降解性酶和感受态调节有关的双组分传感器组氨酸激酶	+
<i>lspA</i>	Q45479	信号肽酶 II	+
<i>mecA</i>	P37958	感受态的负调节子	+
<i>med</i>		comK 的正调节子	+
<i>nin</i>	P12669	NucA 的 DNA 降解活性的抑制子	+
<i>nucA</i>	P12667	膜相关核酸酶	+
<i>oppA</i>	P24141	寡肽 ABC 转运蛋白(结合蛋白)(启动孢子形成、感受态发展)	+
<i>oppB</i>	P24138	寡肽 ABC 转运蛋白(通透酶)(启动孢子形成、感受态发展)	+
<i>oppC</i>	P24139	寡肽 ABC 转运蛋白(通透酶)(启动孢子形成、感受态发展)	+
<i>oppD</i>	P24136	寡肽 ABC 转运蛋白(ATP-结合蛋白)(启动孢子形成、感受态发展)	+
<i>oppF</i>	P24137	寡肽 ABC 转运蛋白(ATP-结合蛋白)(启动孢子形成、感受态发展)	+
<i>phrC</i>		磷酸酶(RapC)调节子/感受态和孢子形成刺激因子(CSF)	+
<i>pnpA</i>	P50849	多核苷酸磷酸化酶(PNPase)	+
<i>rapE</i>	P45943	响应调节子天冬氨酸磷酸酶	+
<i>recA</i>	P16971	参与同源重组和 DNA 修复的多功能蛋白质(LexA-自裂解)	+
<i>sinR</i>	P06533	指数期后基因的转录调节子	+
<i>slr</i>		感受态发展和孢子形成基因的转录活化因子	+
<i>smf</i>	P39813	DNA 加工 Smf 蛋白质类似物	+
<i>spo0F</i>	P06628	参与孢子形成启动的双组分响应调节子	+

如表 1 所示,地衣芽孢杆菌 ATCC 14580 基因组表现出携带除 *comP* 基因和 *comS* 基因之外的感受态发展所需全部基因,其中 *comP* 基因已经通过插入序列 IS3Bl1 (Lapidus 等, 2002, *FEMS Microbiol. Lett.* 209: 23-30)破坏,而 *comS* 或者不存在,或者与枯草芽孢杆菌中的相应基因本质上不同。没有活性 *comP* 基因产物,感受态信号转导级联的早期部分不能在地衣芽孢杆菌中正确工作。然而,能通过增加中央转录因子 ComK 的表达而绕过感受态级联的早期部分,ComK 诱导编码 DNA 结合和摄取机制的后期感受态基因的转录(Susanna 等, 2004, *J. Bacteriol.* 186: 1120-8)。然而,如果 MecA 蛋白的水平高到足以结合

并失活所有 ComK 蛋白, 那么有可能仅增加 *comK* 基因的表达不足以诱导感受态。取而代之的, 需要增加 *comS* 基因的表达来克服 MecA 的活性, 并且从而释放 ComK 来活化后期感受态基因的转录。

在枯草芽孢杆菌中, 将 *comS* 基因嵌入 *srfA* 基因第四个氨基酸活化域的编码区中。因此, 扫描地衣芽孢杆菌中相应的区域(*srfA* 直系同源基因)来定位可能的 ComS 样序列。图 15 中的比较比对显示, 最接近的预测的地衣芽孢杆菌直系同源物与枯草芽孢杆菌中已知的 ComS 基因产物略有不同, 并且多个已知对生物活性重要的残基在地衣芽孢杆菌中出现歧化。还不了解地衣芽孢杆菌中推定的 ComS 直系同源物是否有功能。

进行了两种实验方法。第一种方法涉及增加 *comK* 的表达以绕过感受态级联的早期部分, 而第二种方法涉及增加 *comS* 的表达以避免 ComK 由 MecA/ClpCP 复合物降解(参见图 1)。

实施例 8: pMRT098 的构建

使用如下所示的引物 992129 和 992130, 通过 PCR 从质粒 pAX01 (Härtl 等, 2001, *J. Bact.* 183: 2696-2699)扩增了 *xylA* 启动子和 *xylR* 基因。

引物 992129

5'-GAGCTCGGATCCCATTTCC-3' (SEQ ID NO: 62)

引物 992130

5'-ATCTCTGAGCTCGCGATGATTAATTAATTCAGAACGCTCGGTTGC
CGCCGGGCGTTTTTTTATGCAGCAATGGCAAGAACGTCCTCCGGTTAGCTCC
-3' (SEQ ID NO: 63)

PCR 扩增在 50 μ l 反应中进行, 所述反应由下述组成: 10 ng pAX01 DNA, 0.4 μ M 每种引物, dATP、dCTP、dGTP 和 dTTP 各 200 μ M, 包含 2.5 mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AMPLITAQ GOLD®酶(Applied Biosystems, Inc., Foster City, CA, USA)。在 ROBOCYCLER® 40 温度循环仪中进行反应, 程序为 95°C 10 分钟的 1 个循环; 95°C 1 分钟、53°C 1 分钟和 72°C 1.5 分钟的 25 个循环; 和 72°C 7 分钟的 1 个循环。通过 0.5X TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳使 PCR 产物可见。期望的片段长约 1500 bp。

使用 TA-TOPO®克隆试剂盒(Invitrogen, Carlsbad, CA, USA)将 PCR 片段克隆入载体 pCR2.1, 并根据制造商的说明转化入大肠杆菌 ONE SHOT®感受

态细胞。在补充有 100 µg/ml 氨苄青霉素的 2X YT 琼脂平板上筛选转化体，在 37°C 温育 16 小时。根据制造商的说明，使用 BIOROBOT® 9600 (QIAGEN Inc., Valencia, CA, USA) 从这些转化体中的几个纯化了质粒 DNA，并使用 M13(-20) 正向和 M13 反向引物 (Invitrogen, Inc, Carlsbad, CA, USA)，通过 DNA 测序确认了插入物的 DNA 序列。将携带正确 PCR 片段的质粒命名为 pMRT091。

用 *Bam*HI 和 *Sac*I 消化质粒 pMRT091 和 pUC18 (Yanisch-Perron 等, 1985, *Gene* 33: 103-119)。通过 0.5X TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳解析消化的产物，使用 QIAQUICK® DNA 提取试剂盒，根据制造商的说明将来自 pUC18 的较大载体片段和来自 pMRT091 的较小片段凝胶纯化。使用快速 DNA 连接试剂盒 (Roche Applied Science, Indianapolis, IN, USA)，根据制造商的说明将两个纯化的片段连接在一起，并将连接混合物转化进大肠杆菌 XL1 SE 感受态细胞 (Stratagene, Inc., La Jolla, CA, USA)。在补充有 100 µg/ml 氨苄青霉素的 2X YT 琼脂平板上筛选转化体。

根据制造商的说明，使用 BIOROBOT® 9600 从几个转化体中纯化了质粒 DNA，并通过 *Bam*HI 和 *Sac*I 消化，然后由 0.5X TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳分析。通过存在约 700 bp 的 *Ava*I/*Bam* HI pMRT091 片段而鉴定正确的质粒，并命名为 pMRT096。

此外，用引物 992131 和 992132，然后用引物 992129 和 992131，通过 SOE PCR (Horton 等, 1989, *Gene* 77: 61-68)，缺失 pMRT096 的 *xylR* 基因中存在的 *Hind*III 和 *Eco*RI 位点。

引物 992131

5'-CTTCTCGAGAATAATATTTTCCTTCTAAGTCGGTTAGGATTCCG-3'

(SEQ ID NO: 64)

引物 992132

5'-CAAGCATCAAAAAACACCAACTTAGTTCGGTGGATAAACAAG
GAGTGGTTATTA TTCAAATTGCAGATCAGGCTTTAG-3' (SEQ ID NO: 65)

PCR 扩增在 50 µl 反应中进行，所述反应由下述组成：10 ng pAX01 DNA，0.4 µM 每种引物，200 µM 每种 dATP、dCTP、dGTP 和 dTTP，包含 2.5 mM MgCl₂ 的 1X PCR Buffer II 和 2.5 单位 AMPLITAQ GOLD® 酶。在 ROBOCYCLER® 40 温度循环仪中进行反应，程序为 95°C 10 分钟的 1 个循环；95°C 1 分钟、55°C

1 分钟和 72°C 1 分钟的 25 个循环; 和 72°C 7 分钟的 1 个循环。通过 0.5X TBE 缓冲液中的 0.8%琼脂糖凝胶电泳使 PCR 产物可见。期望的片段长约 700 bp。

使用 TA-TOPO®克隆试剂盒将 PCR 片段克隆入 pCR2.1, 并根据制造商的说明转化入大肠杆菌 ONE SHOT®感受态细胞。在补充有 100 µg/ml 氨苄青霉素的 2X YT 琼脂平板上筛选转化体, 在 37°C 温育 16 小时。根据制造商的说明, 使用 BIOROBOT® 9600 (QIAGEN Inc., Valencia, CA, USA)从这些转化体中的几个纯化了质粒 DNA, 并使用 M13(-20)正向和 M13 反向引物 (Invitrogen, Inc, Carlsbad, CA, USA), 通过 DNA 测序确认了插入物的 DNA 序列。将携带正确 PCR 片段的质粒命名为 pMRT092。

用 *Bam*HI 和 *Ava*I 消化质粒 pMRT096 和 pMRT092。通过 0.5X TBE 缓冲液中的 0.8%琼脂糖凝胶电泳解析消化的产物, 使用 QIAQUICK® DNA 提取试剂盒, 根据制造商的说明将来自 pMRT096 的较大载体片段和来自 pMRT092 的较小片段凝胶纯化。使用快速 DNA 连接试剂盒, 根据制造商的说明将两个纯化的片段连接在一起, 并将连接混合物转化进大肠杆菌 XL1 SE 感受态细胞 (Stratagene, Inc., La Jolla, CA, USA)。在补充有 100 µg/ml 氨苄青霉素的 2X YT 琼脂平板上筛选转化体。

根据制造商的说明, 使用 BIOROBOT® 9600 从几个转化体中纯化了质粒 DNA, 并通过 *Eco*RI 和 *Hind*III 消化, 然后由 0.5X TBE 缓冲液中的 0.8%琼脂糖凝胶电泳分析。通过用 *Eco*RI 或 *Hind*III 消化时存在单一的 4200 bp 片段而鉴定了正确的质粒。将这个构建体命名为 pMRT098 (图 16)。

实施例 9: pΔComS 的构建

如下构建质粒 pΔComS: 用 *Bam*HI 加 *Hind*III 消化 pBD2528 (也称作 pComS, Hahn 等, 1996, *Mol. Microbiol.* 21: 763-75), 用 DNA 聚合酶 I (Klenow 片段)处理以产生平末端, 并且通过用 T4 DNA 连接酶将载体再次连接而产生对照质粒, 其除了缺少 *comS* 基因外均与 pBD2528 相同。为了确保正确的甲基化以进一步转化入地衣芽孢杆菌 SJ1904, 根据 Anagnostopoulos 和 Spizizen, 1961, 见上文的过程, 将质粒 pΔcomS 和 pComS 转化入本文所述的枯草芽孢杆菌 MDT101。在补充有 20 µg/ml 卡那霉素的 TBAB 培养基上筛选转化体。

实施例 10: 扩增地衣芽孢杆菌 SJ1904 *comK* 基因并将其克隆进大肠杆菌

载体 pMRT098

设计了下述 PCR 引物从地衣芽孢杆菌 SJ1904 扩增编码 ComK 的 DNA。加入限制酶位点 *Bam*HI 和 *Pst*I (下划线)以便将 *comK* 基因片段克隆进 pMRT098。

引物 999722:

5'-GTGGATCCgattaggaggatcaaaatg-3' (SEQ ID NO: 66)

*Bam*HI

引物 999723:

5'-CAGTACTGCAGtcaatagcgcttttcagctccctgaggatAaattcgatatc-3' (SEQ ID NO: 67)

*Pst*I

使用 Expand 高保真 PCR 系统(Roche Applied Science, Indianapolis, IN, USA), 通过 PCR 扩增了 *comK* 基因片段。根据 Pitcher 等, 1989, 见上文的过程, 从地衣芽孢杆菌 SJ1904 分离了基因组 DNA。PCR 扩增反应混合物包含 1 μ l 145 ng/ μ l 的地衣芽孢杆菌 SJ1904 基因组 DNA, 1 μ l 引物 999722 (50 pmol/ μ l), 1 μ l 引物 999723 (50 pmol/ μ l), 包含 15 mM MgCl₂ 的 5 μ l 10X PCR 缓冲液, 1 μ l dNTP 混合物(每种 10 mM), 37.25 μ l 水, 和 0.75 μ l (3.5 单位/ μ l)DNA 聚合酶混合物。使用 EPPENDORF® MASTERCYCLER® 5333 (Hamburg, Germany)扩增片段, 程序为 94°C 2 分钟的 1 个循环; 94°C 15 秒、60°C 30 秒和 72°C 1 分钟的 10 个循环; 94°C 15 秒、60°C 30 秒和 72°C 1 分钟的 15 个循环, 并在每个连续的循环中加上 5 秒延长的 72°C; 和 72°C 7 分钟的 1 个循环; 和 4°C 保温。

使用 NANOSEP® 30K OMEGA™离心设备, 根据制造商的说明(Pall Life Science, Inc., Ann Arbor, MI, USA)纯化了 579 bp PCR 产物。然后用 *Bam*HI 和 *Pst*I 消化 579 bp PCR 产物和载体 pMRT098。使用快速 DNA 连接试剂盒, 根据制造商的说明将所述片段连接在一起。使用两微升反应物, 根据制造商的说明转化大肠杆菌 SURE®细胞。

由大肠杆菌转化体制备质粒 DNA, 并使用 1 μ l 质粒模板、1.6 ng 引物 999722 或引物 999723 (如上所述)并加水至 6 μ l 来测序。用 Applied Biosystems Model 377 Sequencer XL, 使用染料终止子化学进行 DNA 测序。将鉴定为具有正确序列的所得质粒命名为 pMRT098/*comK* (图 17)。

实施例 11: 构建包含在木糖诱导型启动子调控下的地衣芽孢杆菌 SJ1904 *comK* 基因的大肠杆菌质粒, 其中基因和启动子由 *amyL* 整合臂从侧翼包围

设计了下述引物从 pMRT074 (美国公开申请 2003/175902) 扩增编码 3'-*amyL* 整合臂的 DNA:

引物 999726:

5'-ctgaacaacaaaaacggctttac-3' (SEQ ID NO: 68)

引物 999727:

5'-ACTGAAGCTTggttgccggtcagcgggatcg-3' (SEQ ID NO: 69)

*Hind*III

因为 3'-*amyL* 整合臂具有天然的 *Pst*I 位点, 所以为将 3'-*amyL* 整合臂以 *Pst*I-*Hind*III 片段克隆进 pMRT098/*comK* 而增加 *Hind*III 切割位点。使用 Expand 高保真 PCR 系统, 通过 PCR 扩增了目的片段。PCR 扩增反应混合物包含约 10 ng pMRT074 质粒 DNA, 1 μ l 引物 999726 (50 pmol/ μ l), 1 μ l 引物 999727 (50 pmol/ μ l), 包含 15 mM MgCl₂ 的 5 μ l 10X PCR 缓冲液, 1 μ l dNTP 混合物(每种 10 mM), 37.25 μ l 水, 和 0.75 μ l (3.5 单位/ μ l) DNA 聚合酶混合物。使用 EPPENDORF® MASTERCYCLER® 5333 扩增片段, 程序为 94°C 2 分钟的 1 个循环; 94°C 15 秒、60°C 30 秒和 72°C 1 分钟的 10 个循环; 94°C 15 秒、60°C 30 秒和 72°C 1 分钟的 15 个循环, 并在每个连续的循环中加上 5 秒延长的 72°C; 和 72°C 7 分钟的 1 个循环; 和 4°C 保温。

使用 NANOSEP® 30K OMEGA™ 离心设备, 根据制造商的说明纯化了 450 bp PCR 产物。然后用 *Hind*III 和 *Pst*I 消化纯化的 PCR 产物和载体 pMRT098/*comK*, 由 TBE 缓冲液中的 1% 琼脂糖凝胶电泳分析, 并使用 QIAQUICK® 凝胶提取试剂盒纯化两个片段。使用快速 DNA 连接试剂盒, 根据制造商的说明连接片段。使用两微升反应物, 根据制造商的说明转化大肠杆菌 SURE® 细胞。由大肠杆菌转化体制备质粒 DNA, 并用 *Hind*III 和 *Pst*I 消化, 然后由 TBE 缓冲液中的 1% 琼脂糖凝胶电泳分析。将鉴定为具有正确限制图谱的所得质粒命名为 pMRT098/*comK*/*amyL*3' (图 18)。

设计了下述引物从 pMRT074 扩增编码 5'-*amyL* 整合臂的 DNA:

引物 999724:

*Eco*RI

5'-AGTCgaattcgactggaagcagagc-3' (SEQ ID NO: 70)

引物 999756:

Sac I

5'-TCAGGAGCTCagtaccattttccctata-3' (SEQ ID NO: 71)

加入 *Eco*RI 和 *Sac*I 限制性位点以便将 5'-*amyL* 整合臂克隆进 pMRT098/*comK/amyL3'*(如上所述)。使用上述条件, 通过 PCR 扩增目的片段。

使用 NANOSEP® 30K OMEGA™ 离心设备, 根据制造商的说明纯化了 523 bp PCR 产物。然后用 *Eco*RI 和 *Sac*I 消化 523bp 的 PCR 产物和载体 pMRT098/*comK/amyL3'*, 由 TBE 缓冲液中的 1%琼脂糖凝胶电泳分析, 并使用 QIAQUICK® 凝胶提取试剂盒纯化两个片段。使用快速 DNA 连接试剂盒, 根据制造商的说明连接片段。使用连接物的 2 µl 等分试样, 根据制造商的说明转化大肠杆菌 SURE® 细胞。由大肠杆菌转化体制备质粒 DNA, 并用 *Eco*RI 和 *Sac*I 消化, 然后由 1X TBE 缓冲液中的 1%琼脂糖凝胶电泳分析。将鉴定为具有正确限制图谱的所得质粒命名为 pMRT098/*comK/amyL#24*(图 19)。

实施例 12: 地衣芽孢杆菌 SJ1904 *comK* 表达载体 pMMar2 的构建

用 *Eco*RI, *Sca*I 和 *Hind*III 消化质粒 pMRT098/*comK/amyL#24*, 并由 TAE 缓冲液(每升 4.84 g Tris 碱, 1.14 ml 冰醋酸, 和 2 ml 0.5 M EDTA pH 8.0)中的 0.7%琼脂糖凝胶电泳, 连同 QIAQUICK® 凝胶提取试剂盒, 纯化了 3178 bp 片段。通过用 *Eco*RI 和 *Hind*III 消化从 pMRT077 (WO 2003/054163)产生了载体片段, 并由 TAE 缓冲液中的 0.7%琼脂糖凝胶电泳, 连同 QIAQUICK® 凝胶提取试剂盒, 纯化了 4340 bp 片段。然后使用 T4 DNA 连接酶, 在 16°C 用 16 小时以大致相当的摩尔浓度连接 3178 bp 和 4340 bp 片段。根据 Anagnostopoulos 和 Spizizen, 1961, 见上文的过程, 使用全部连接混合物转化枯草芽孢杆菌 168Δ4 感受态细胞。在 TBAB 红霉素/林可霉素平板上筛选转化体。

根据 Pitcher 等, 1989, 见上文所述的过程, 从几个转化体制备了枯草芽孢杆菌基因组 DNA。使用 PCR 扩增, 使用 Expand 高保真 PCR 系统确认了质粒构建。50 µl PCR 扩增反应混合物包含约 100 ng 基因组 DNA, 1 µl 引物 999722 (50 pmol/µl), 1 µl 引物 999727 (50 pmol/µl), 包含 15 mM MgCl₂ 的 5 µl 10X PCR 缓冲液, 1 µl dNTP 混合物(每种 10 mM), 37.25 µl 水, 和 0.75 µl (3.5 单位/µl) DNA 聚合酶混合物。使用 Eppendorf Mastercycler 5333 扩增片段, 程序为 94°C 2 分钟的 1 个循环; 94°C 15 秒、60°C 30 秒和 72°C 1 分钟的 10 个循环; 94°C 15

秒、60°C 30 秒和 72°C 1 分钟的 15 个循环，并在每个连续的循环中加上 5 秒延长的 72°C；和 72°C 7 分钟的 1 个循环；和 4°C 保温。

包含期望的 1029 bp 扩增片段的转化体，其由 TBE 缓冲液中 0.8% 琼脂糖凝胶电泳所确定，命名为 pMMar2 (图 20)。此外，由枯草芽孢杆菌 168Δ4/pMMar2 制备了质粒 DNA，然后进行限制性酶消化，通过凝胶电泳分析，得到期望大小的片段。为了确保正确的甲基化以进一步转化进地衣芽孢杆菌 SJ1904，根据 Anagnostopoulos 和 Spizizen, 1961, 见上文的过程，将质粒 pMMar2 转化进本文所述的枯草芽孢杆菌 MDT101。在 TBAB 红霉素/林可霉素平板上筛选转化体。

实施例 13: 将 pMMar2 转化进地衣芽孢杆菌 SJ1904 的 *amyL* 基因座

将包含木糖诱导型 *xylA* 启动子(Kim 等, 1996, *Gene* 181: 71-6)调控下的地衣芽孢杆菌 *comK* 基因的表达盒，通过染色体整合和温度敏感质粒 pMMar2 的切除合并入地衣芽孢杆菌 SJ1904 的基因组 DNA。将包含质粒 pMMar2 的地衣芽孢杆菌转化体在 45°C 涂布于 TBAB 红霉素/林可霉素平板上，迫使载体整合。根据在 45°C 在 TBAB 红霉素/林可霉素平板上生长的能力选择期望的整合体。然后不加选择在 34°C 在 VY 培养基中培养整合体，诱导切除整合的质粒。将细胞涂布在 LB 平板或基本培养基平板上，并筛选对红霉素敏感的菌落。筛选对红霉素敏感的克隆通过 PCR 进行的基因转变，检测整合的 *xylA::comK* 表达盒。获得的菌株包含整合于 *amyL* 基因座、驱动地衣芽孢杆菌 *comK* 表达的 *xylA* 启动子，将所述菌株命名为地衣芽孢杆菌 SJ1904 *xylA::comK*。

实施例 14: 用 pMMar2、pComS 或 pΔComS 转化地衣芽孢杆菌 SJ1904 和 SJ1904 *xylA::comK*

使用质粒 Midi 试剂盒，从枯草芽孢杆菌 MDT101 分离了质粒 pMMar2、pComS 和 pΔComS。用 pMMar2、pComS 和 pΔComS 质粒 DNA 转化了地衣芽孢杆菌菌株 SJ1904，并如上所述通过电穿孔用 pComS 质粒 DNA 转化了地衣芽孢杆菌 *xylA::comK*。将获得的地衣芽孢杆菌转化体分别命名为 SJ1904 (pMMar2)、SJ1904 (pComS)、SJ1904 (pΔComS) 和 SJ1904 *xylA::comK* (pComS)。

实施例 15: 地衣芽孢杆菌 *comK* 基因在枯草芽孢杆菌 A164Δ5 和地衣芽孢

杆菌 SJ1904 中的表达

首先如上所述通过转化将携带在木糖诱导型 *xylA* 启动子转录调控下的地衣芽孢杆菌 *comK* 基因的质粒 pMMar2 导入枯草芽孢杆菌 164Δ5。pMMar2 载体还携带赋予红霉素抗性的基因。然后将命名为枯草芽孢杆菌 164Δ5/pMMar2 的红霉素抗性转化体，在包含葡萄糖(抑制 *xylA* 启动子)或葡萄糖加木糖(部分抑制 *xylA* 启动子)或木糖(脱抑制 *xylA* 启动子)的培养基中测试感受态发展。使用上述方法，通过在包含 1%木糖和/或 0.5%葡萄糖的 Spizizen I 培养基中培养而制备枯草芽孢杆菌 164Δ5/pMMar2 和枯草芽孢杆菌 164Δ5 感受态细胞。使用前在 -80°C 冻存细胞。为了转化，将细胞混合物在 37°C 水浴中快速解冻。向每种转化混合物加入一微克 pGME086 质粒 DNA 和包含 0.5%葡萄糖或 1%木糖和 0.2 μg/ml 氯霉素的 LB 培养基。质粒 pGME086 是 pE194 (Gryczan 等, 1982, *J. Bacteriol.* 152: 722-735) 的衍生物，携带来自 pC194 (Horinouchi 等, 1982, *J. Bacteriol.* 150: 815-825) 的氯霉素抗性标记。转化混合物在 34°C 在振荡培养箱中培养 1 小时。1 小时后，将反应混合物涂布在 LB 氯霉素/红霉素平板上。在 34°C 培养平板 24 小时。次日计数菌落，确定转化效率。

表 2 显示了在以木糖作为唯一碳源的培养基中受体菌株生长后，转化体的数目约为在葡萄糖或葡萄糖加木糖中生长后获得的数目的 200 倍。这些结果证明，异源地衣芽孢杆菌 *comK* 基因不仅由 *xylA* 启动子转录，而且地衣芽孢杆菌 ComK 蛋白质有效地诱导了枯草芽孢杆菌中的感受态状态。

表 2. 使用来自地衣芽孢杆菌的 *comK* 基因在枯草芽孢杆菌中的感受态诱导

生长培养基	兼具氯霉素和红霉素抗性的菌落总数
对照培养基†	81 (102)
包含葡萄糖的 Spizizen I 培养基	69 (156)
包含葡萄糖和木糖的 Spizizen I 培养基	180 (149)
包含木糖的 Spizizen I 培养基	36,600 (34,400)‡

†对照培养基是标准的枯草芽孢杆菌感受态培养基(Anagnostopolous 和 Spizizen, 1961, 见上文)。

‡这些数字是从转化反应的 1:50 稀释测定的。括号中的数字来自重复的实验。

实施例 16: DNA 微阵列分析

使用 DNA 微阵列比较葡萄糖培养基(*comK*抑制型)上和木糖培养基(*comK*诱导型)上生长的地衣芽孢杆菌菌株 SJ1904 *xylA::comK* 的全局转录谱。

通过点印 CDS 特异性寡核苷酸(50mer)而制备 DNA 微阵列, 所述寡核苷酸选自地衣芽孢杆菌 ATCC 14580 基因组中的蛋白质编码基因, 如 Genbank 中所保存的(登录号 CP000002)。寡核苷酸购自 MWG-Biotech, Inc., Highpoint, NC, USA。用于微阵列点印、杂交和分析的方法如 Berka 等, 2003, *Proc. Natl. Acad. Sci. USA* 100: 5682-5687 所述进行。

在包含 0.5%葡萄糖(抑制培养基)或 1%木糖(诱导培养基)的 Spizizen I 培养基中培养地衣芽孢杆菌 SJ1904 *xylA::comK* 细胞。在接种后 1、3 和 5 小时收获细胞, 并使用 Berka 等, 2003, 见上文中所述方法分离总细胞 RNA。通过 25 μ g 总 RNA 的反转录制备荧光探针, 根据 Berka 等, 2003, 见上文的过程在第一链 cDNA 中并入氨基烯丙基-dUTP。然后根据 Berka 等, 2003, 见上文的过程, 通过直接偶联至 Cy3 或 Cy5 单功能反应活性染料(Amersham Pharmacia Biotech, Arlington Heights, IL, USA)而标记氨基-cDNA 产物。用 Cy3 标记来自葡萄糖培养基中生长的细胞的探针, 并且用 Cy5 标记来自木糖培养基中生长的细胞的探针。杂交和洗涤条件与 Berka 等, 2003, 见上文中所述相同。

使用 GENEPIX® 4000B 扫描仪(Axon Instruments, Union City, CA, USA)使微阵列片成像。使用 GENEPIX®软件(Axon Instruments)将微阵列点的荧光强度值量化(包括背景减除), 并且使用 S+ARRAYANALYZER™软件(Insightful Corporation, Seattle, WA, USA)中提供的 Lowess 功能将获得的数字标准化。根据 Cy5/Cy3 比例 ≥ 2.0 来指定受到 *xylA::comK* 表达单元的表达诱导的基因。

使用 DNA 微阵列比较葡萄糖培养基(*comK*抑制型)上和木糖培养基(*comK*诱导型)上生长的地衣芽孢杆菌菌株 SJ1904 *xylA::comK* 的全局转录谱。这个分析的结果显示, 与葡萄糖培养基相比, 木糖培养基上生长的细胞中 *comK* 转录水平有实质性增加(10 至 30 倍)。然而, 在这个实验中, 后期感受态基因(*comE*, *comF* 和 *comG* 操纵子)的转录未出现相应的增加。先前在枯草芽孢杆菌中的研究(Brzuszkiewicz 等, 2006, *Proc. Natl. Acad. Sci. USA* 103: 12879-84)表明, *comK* 转录的增加引起了后期感受态基因的转录增加。然而, 这种关系未在地衣芽孢杆菌中观察到。

实施例 17: pMDT131 的构建

构建了质粒 pMDT131, 以创建赋予氯霉素抗性的温度敏感型质粒。用 *EcoRI* 消化质粒 pMRT074 (美国公开申请 2003/0175902), 然后用 T4 DNA 聚合酶加 dNTP 处理以产生平末端, 如实施例 5 中所述。然后用 *NotI* 消化质粒, 由 TBE 缓冲液中 0.8% 琼脂糖凝胶电泳分析, 使用 QIAQUICK® 凝胶提取试剂盒纯化了约 4355 bp 的载体片段。用 *Eco47III* 和 *NotI* 消化质粒 pNBT1, 由 TBE 缓冲液中的 0.8% 琼脂糖凝胶电泳分析, 并使用 QIAQUICK® 凝胶提取试剂盒纯化携带 *cat* 基因和多克隆位点的约 1222 bp 的片段。使用 T4 DNA 连接酶, 如上所述连接 pMRT074 载体片段和 pNBT1 *cat* 片段, 并根据 Anagnostopoulos 和 Spizizen, 1961, 见上文的过程, 用所述连接物转化了枯草芽孢杆菌 168Δ4, 在 34°C 在 TBAB 氯霉素平板上筛选氯霉素抗性。使用质粒 Midi 试剂盒, 从一株转化体分离了质粒 DNA, 并通过用 *BamHI* 消化, 然后由 TBE 缓冲液中 0.8% 琼脂糖凝胶电泳来确认, 电泳得到了约 3779 bp 和 1802 bp 的期望片段。将获得的质粒命名为 pMDT131 (图 21)。

实施例 18: 枯草芽孢杆菌 *comS* 和地衣芽孢杆菌 *comK* 基因在地衣芽孢杆菌中的共表达-

有两种可能的方法测试下述假说: MecA/ClpCP 复合物以意想不到的高活性阻止 ComK 诱导地衣芽孢杆菌中的后期感受态基因。第一种方法涉及 *mecA* 基因的破坏。然而, 先前的研究已经表明, *mecA* 通常可以作为衔接分子, 靶向受调控降解的蛋白质 (Persuh 等, 1999, *Mol. Microbiol.* 33: 886-94), 并且因此, 感受态之外的处理可能对 *mecA*-缺陷型细胞具有负面影响。第二种方法涉及增加 ComS 的表达, 其表面上可以从 MecA/ClpCP 复合物释放 ComK, 保护它不被降解并且从而使后期感受态基因能够被诱导。为了使用这种方法, 通过用质粒和染色体 DNA 转化, 测试地衣芽孢杆菌菌株 SJ1904 *xylA::comK* + pComS (实施例 14) 的感受态发展。地衣芽孢杆菌菌株 SJ1904 *xylA::comK* + pComS (a) 在 *amyL* 基因座包含 *xylA::comK* 转录单元的拷贝, 并且 (b) 携带包含枯草芽孢杆菌 *comS* 基因拷贝的质粒 (图 22)。作为对照, 在相同的试验中测试了很多其他地衣芽孢杆菌菌株, 包括 SJ1904 背景菌株, 和仅携带 *xylA::comK* 表达单元、pComS 载体或 pΔComS 对照质粒的菌株 (实施例 14)。

将下述地衣芽孢杆菌转化宿主从冷冻甘油储液涂布至合适的选择培养基上, 在过夜培养后获得连片生长: SJ1904 *xylA::comK*, SJ1904 *xylA::comK* + pComS, SJ1904 + pComS, SJ1904 + p Δ ComS 和 SJ1904。向 500 ml 侧口瓶 (side-arm flask) 中加入包含 2% 木糖的五十毫升 Spizizen I 培养基。向培养平板上加入另外 5 ml 包含 2% 木糖的 Spizizen I 培养基, 通过用无菌涂抹棒刮抹而收集细胞并转移至无菌管中。向侧口瓶中加入每 5 ml 培养物中的五百微升, 获得 Klett 读数 30。在 37°C, 250 rpm 温育培养物 11 小时。向 Falcon 2059 管加入来自每个 11 小时培养物的二百五十微升加包含 2% 木糖和 2 mM EGTA 的 250 μ l Spizizen II 培养基。向每个管加入一微克转化 DNA, 或者质粒或者染色体 DNA, 使用 10 mM Tris-0.1 mM EDTA (TE) 缓冲液作为阴性对照。对于染色体 DNA, 将试管在 37°C、250 rpm 温育 1 小时, 对于质粒 DNA, 将试管在 34°C、250 rpm 温育 1 小时。将包含染色体 DNA 的转化反应物涂布在包含 100 μ g/ml 壮观霉素的 TBAB 平板上。将包含质粒 DNA 的转化反应物涂布在 TBAB 红霉素/林可霉素平板上。对于壮观霉素选择在 37°C 温育平板, 而对于在含有红霉素的平板上的选择在 34°C 温育。在次日计数菌落, 确定转化效率。

表 4 中的结果表明, 在本实验所用条件下, 含质粒携带的枯草芽孢杆菌 *comS* 基因的地衣芽孢杆菌受体菌株(样品 2 和 3)用染色体 DNA 每个平板得到了约 20 至 45 个转化体, 而使用质粒 DNA 时每个平板得到了 3 至 7 个转化体。相反, 不含枯草芽孢杆菌 *comS* 基因的菌株仅得到背景水平的壮观霉素抗性菌落, 没有红霉素抗性菌落。*comS* 和 *comK* 基因表达提高的组合大概使地衣芽孢杆菌中的转化效率翻倍, 尽管只有 *comK* 表达提高不诱导感受态。对源自地衣芽孢杆菌 SJ1904 *xylA::comK* + pComS 菌株的 pMDT131 转化的三株独立的红霉素和卡那霉素抗性菌落的 PCR 分析表明, 这些菌株全部包含整合的 *xylA::comK* 表达盒、pComS 质粒和基于 pE194 的质粒, 证明它们是真正的转化体。

表 3. 用质粒和染色体 DNA 对地衣芽孢杆菌 SJ1904 衍生物进行的感受态介导转化

每平板的菌落					
样品	受体	MDT232 染色体 DNA*	无染色体 DNA 对照***	pMDT131 质粒 DNA**	无 pMDT131 质粒 DNA 对照***
1	<i>xylA::comK</i>	1	3	0	0

2	<i>xylA::comK</i> + pComS	45	1	7	0
3	pComS	19	0	3	0
4	pΔComS	2	1	0	0
5	SJ1904	2	1	0	0

* 三块板的平均值

** 四块板的平均值

*** 将无 DNA 对照涂布在包含壮观霉素的培养基上[用壮观霉素筛选可能产生低水平的自发壮观霉素抗性突变体(Kimura 等, 1973, *Mol. Gen. Genet.* 124: 107-115)]。

本文描述的和要求保护的发明并不局限于本文所公开的具体实施方案的范围内, 因为这些实施方案意欲作为本发明几个方面的说明。任何等同的实施方案意欲在本发明的范围内。事实上, 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技术人员将是显而易见的。这些修改也意欲落入所附权利要求的范围内。在冲突的情况下, 以包括定义的本公开为准。

本文引用了许多参考文献, 其公开的内容通过提述以其整体并入。

- <110> 诺维信股份有限公司(Novozymes, Inc.)
- <120> 在芽孢杆菌属细胞中获得遗传感受态的方法
- <130> 11025.204-W0
- <150> 60/877,053
- <151> 2006-12-21
- <160> 71
- <170> PatentIn version 3.5
- <210> 1
- <211> 141
- <212> DNA
- <213> 枯草芽孢杆菌(*Bacillus subtilis*)
- <400> 1
- ttgaaccgat caggcaagca tcttatcagc agcattatcc tgtatccccg gccacgcgga 60
- gaatgtatat cctcaatcag cttggacaag caaacacaag ctacaacgtc cccgctgtac 120
- ttctgctgga gggagaagta g 141
- <210> 2
- <211> 46
- <212> PRT
- <213> 枯草芽孢杆菌(*Bacillus subtilis*)
- <400> 2
- Met Asn Arg Ser Gly Lys His Leu Ile Ser Ser Ile Ile Leu Tyr Pro
1 5 10 15
- Arg Pro Ser Gly Glu Cys Ile Ser Ser Ile Ser Leu Asp Lys Gln Thr
20 25 30
- Gln Ala Thr Thr Ser Pro Leu Tyr Phe Cys Trp Arg Glu Lys
35 40 45
- <210> 3
- <211> 141
- <212> DNA
- <213> 枯草芽孢杆菌纳豆变种(*Bacillus subtilis* var. *natto*)
- <400> 3
- ttgaaccgat caggcaagca tcttatcagc tgcattatcc tgtatccccg gccacgcgga 60
- gaatgtatat cctcaatcag cttggacaag caaacacaag ctacaacgtc cccgctgtac 120
- ttctgctgga gggagaagta g 141
- <210> 4
- <211> 46
- <212> PRT
- <213> 解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*)
- <400> 4
- Met Asn Arg Ser Asp Lys Arg Leu Ile Arg Ser Ile Ile Leu Phe Pro
1 5 10 15
- Gln His Ser Ala Gly Cys Ile Ser Leu Ile Ser Ser Asp Arg Pro Ala
20 25 30
- Arg Ala Thr Thr Ser Leu Leu Tyr Phe Cys Trp Lys Gly Gln
35 40 45

<210> 5
 <211> 141
 <212> DNA
 <213> 解淀粉芽孢杆菌(*Bacillus amyloliquefaciens*)

 <400> 5
 ttgaaccgat ccgacaagcg ccttatcaga agcattatcc tgtttcctca gcacagcgca 60
 ggatgtatat ccttaatcag ctcggacagg ccagcacgag ctacaacgtc cctgctgtac 120
 ttttgctgga agggtcagta g 141

<210> 6
 <211> 46
 <212> PRT
 <213> 枯草芽孢杆菌纳豆亚种(*Bacillus subtilis* subsp. *natto*)

<400> 6
 Met Asn Arg Ser Gly Lys His Leu Ile Ser Cys Ile Ile Leu Tyr Pro
 1 5 10 15
 Arg Pro Ser Gly Glu Cys Ile Ser Ser Ile Ser Leu Asp Lys Gln Thr
 20 25 30
 Gln Ala Thr Thr Ser Pro Leu Tyr Phe Cys Trp Arg Glu Lys
 35 40 45

<210> 7
 <211> 171
 <212> DNA
 <213> 地衣芽孢杆菌(*Bacillus licheniformis*)

 <400> 7
 atggacaggc agaacaaagc gggattcagc ctgccgaaaa acgcgactgg tatcccgttt 60
 catccgctca gcaacggatg tatgccctcc atcatattga aaagaacgga acgggctaca 120
 acatgccgtc tgttctcatg ctggaaggcg tacttgatag ggatcgctta a 171

<210> 8
 <211> 56
 <212> PRT
 <213> 地衣芽孢杆菌(*Bacillus licheniformis*)

<400> 8
 Met Asp Arg Gln Asn Lys Ala Gly Phe Ser Leu Pro Lys Asn Ala Thr
 1 5 10 15
 Gly Ile Pro Phe His Pro Leu Ser Asn Gly Cys Met Pro Ser Ile Ile
 20 25 30
 Leu Lys Arg Thr Glu Arg Ala Thr Thr Cys Arg Leu Phe Ser Cys Trp
 35 40 45
 Lys Ala Tyr Leu Ile Arg Ile Ala
 50 55

<210> 9
 <211> 171
 <212> DNA
 <213> 地衣芽孢杆菌(*Bacillus licheniformis*)

 <400> 9
 atggacaggc agaacaaagc gggattcagc ctgccgaaaa acgcgactgg tatcccgttt 60

catccgctca gcaacggatg tatgccctcc atcatattga aaagaacgga acgggctaca 120
 acatgccgctc tggttctcatg ctggaaggcg tacttgatac ggatcgetta a 171

<210> 10
 <211> 56
 <212> PRT
 <213> 地衣芽孢杆菌(Bacillus licheniformis)

<400> 10

Met Asp Arg Gln Asn Lys Ala Gly Phe Ser Leu Pro Lys Asn Ala Thr
 1 5 10 15

Gly Ile Pro Phe His Pro Leu Ser Asn Gly Cys Met Pro Ser Ile Ile
 20 25 30

Leu Lys Arg Thr Glu Arg Ala Thr Thr Cys Arg Leu Phe Ser Cys Trp
 35 40 45

Lys Ala Tyr Leu Ile Arg Ile Ala
 50 55

<210> 11
 <211> 576
 <212> DNA
 <213> 枯草芽孢杆菌(Bacillus subtilis)

<400> 11

atgagccaga agaccgacgc cccctggag agctacgagg tgaacggcgc caccatgcc 60
 gtgctgcccg aggagatcga cggcaagatc tgcagcaaga tcatcgagaa ggactgctg 120
 ttctacgtga acatgaagcc cctgcagatc gtggacagaa gctgcagatt ctteggcagc 180
 agctacgccg gcagaaaggc cggcacctac gaggigacca agatcagcca caagcccc 240
 atcatggcgg accccagcaa ccagatctc ctgttcccc cctgagcag caccagacc 300
 cagtggcgtt ggatcagcca cgtgcacgtg aaggagtca agcccaccga gttcagcagc 360
 accgaggtga cctcagcaa cggcaagacc atggagctgc ccatcageta caacagcttc 420
 gagaaccagg tglacagaac cgcttgctg agaaccaagt tccaggacag aatcgaccac 480
 agagtgccca agagacagga gttcatgctg tacccaagg aggagagaac caagatgatc 540
 tacgacttca tctgagaga gctggcggag agatac 576

<210> 12
 <211> 192
 <212> PRT
 <213> 枯草芽孢杆菌(Bacillus subtilis)

<400> 12

Met Ser Gln Lys Thr Asp Ala Pro Leu Glu Ser Tyr Glu Val Asn Gly
 1 5 10 15

Ala Thr Ile Ala Val Leu Pro Glu Glu Ile Asp Gly Lys Ile Cys Ser
 20 25 30

Lys Ile Ile Glu Lys Asp Cys Val Phe Tyr Val Asn Met Lys Pro Leu
 35 40 45

Gln Ile Val Asp Arg Ser Cys Arg Phe Phe Gly Ser Ser Tyr Ala Gly
 50 55 60

Arg Lys Ala Gly Thr Tyr Glu Val Thr Lys Ile Ser His Lys Pro Pro
 65 70 75 80

Ile Met Val Asp Pro Ser Asn Gln Ile Phe Leu Phe Pro Thr Leu Ser
85 90 95
Ser Thr Arg Pro Gln Cys Gly Trp Ile Ser His Val His Val Lys Glu
100 105 110
Phe Lys Ala Thr Glu Phe Asp Asp Thr Glu Val Thr Phe Ser Asn Gly
115 120 125
Lys Thr Met Glu Leu Pro Ile Ser Tyr Asn Ser Phe Glu Asn Gln Val
130 135 140
Tyr Arg Thr Ala Trp Leu Arg Thr Lys Phe Gln Asp Arg Ile Asp His
145 150 155 160
Arg Val Pro Lys Arg Gln Glu Phe Met Leu Tyr Pro Lys Glu Glu Arg
165 170 175
Thr Lys Met Ile Tyr Asp Phe Ile Leu Arg Glu Leu Gly Glu Arg Tyr
180 185 190

<210> 13
<211> 579
<212> DNA
<213> 韦氏芽孢杆菌(*Bacillus weihenstephanensis*)

<400> 13
atggagagca aggtggagag atactgtggag aactactgtg tgagcaagaa caccatggcc 60
ctgcigcccc tgggtgctggg cgagaagaag gtggtgacca gaatcgtgga gatggaggac 120
agctttcttc tgttcagaa gccctggac atcatcgaga gaagctgcag aaagcacggc 180
agcagcttct tggcagaaa ggaggcacc aaggagctga ccagaatcac ccacaaggcc 240
cccctgcca tcagccccac cgaccagctg tactttcttc ccacctacag ctacagcaga 300
aaggagtgcg cctggctgag ccacttccac atcgaggaca acaaggagct gaaggacggc 360
aacctgatca tcagattcat caacggcttc gccgtgaagc tggagatgag caagagcagc 420
ttcgagaacc agcagaacag aaccgccaag ctgagaaccg agtacgagga cagaaagaag 480
aagcagggca acccctgctt caaggagatc gacaagaacg aggagagcaa gctgaccccc 540
gcctacgaga agtgttactt cgtgaaggag ggcgagtg 579

<210> 14
<211> 192
<212> PRT
<213> 地衣芽孢杆菌(*Bacillus licheniformis*)

<400> 14
Met Ser Thr Glu Asp Met Thr Lys Asp Thr Tyr Glu Val Asn Ser Ser
1 5 10 15
Thr Met Ala Val Leu Pro Leu Gly Glu Gly Glu Lys Pro Ala Ser Lys
20 25 30
Ile Leu Glu Thr Asp Arg Thr Phe Arg Val Asn Met Lys Pro Phe Gln
35 40 45
Ile Ile Glu Arg Ser Cys Arg Tyr Phe Gly Ser Ser Tyr Ala Gly Arg
50 55 60
Lys Ala Gly Thr Tyr Glu Val Ile Lys Val Ser His Lys Pro Pro Ile
65 70 75 80
Met Val Asp His Ser Asn Asn Ile Phe Leu Phe Pro Thr Phe Ser Ser

	85		90		95
Thr Arg Pro Gln Cys Gly Trp Leu Ser His Ala His Val His Glu Phe					
	100		105		110
Cys Ala Ala Lys Tyr Asp Asn Thr Phe Val Thr Phe Val Asn Gly Glu					
	115		120		125
Thr Leu Glu Leu Pro Val Ser Ile Ser Ser Phe Glu Asn Gln Val Tyr					
	130		135		140
Arg Thr Ala Trp Leu Arg Thr Lys Phe Ile Asp Arg Ile Glu Gly Asn					
	145		150		160
Pro Met Gln Lys Lys Gln Glu Phe Met Leu Tyr Pro Lys Glu Asp Arg					
	165		170		175
Asn Gln Leu Ile Tyr Glu Phe Ile Leu Arg Glu Leu Lys Lys Arg Tyr					
	180		185		190

<210> 15
 <211> 579
 <212> DNA
 <213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 15	
atgaaccacc tggacatgca caagaagaga ctggtggagg agtacgagat caaccccagc	60
accatgatca tcttgcccca gatctacggc aagaagatct acagcagaat cttegaggtg	120
gaggacgagt tcttgagccc cttcaagctg ttcgacatcg tgaagaagag ctgctggctac	180
ttcggcagca gctacgaggg cagaaaggac gccaccaagg acatcatcgg cgtgacccac	240
aaggtgccc tctgatcga ccccaccaac ctgctgtact tcttccccac caccagcccc	300
aacaaccccg actgcatctg gatcagctac gagcacatcg ccgcccacca cagaaccgac	360
cccagccaca ccaaggtggt gttcggcaac aagcacaccc tgatcctgcc cgtgagcagc	420
agcagcttcg agaaccagct gctgagaacc gcccacctga gaaccaagct gcaccagaga	480
atcgagggcc acggcagaaa gatgltctac ttcaccgaca accagagact gaacagagcc	540
agcgagcccc tgggcagcta cagaaacaga aagctggag	579

<210> 16
 <211> 192
 <212> PRT
 <213> 地衣芽孢杆菌(Bacillus licheniformis)

<400> 16	
Met Ser Thr Glu Asp Met Thr Lys Asp Thr Tyr Glu Val Asn Ser Ser	
1	5 10 15
Thr Met Ala Val Leu Pro Leu Gly Glu Gly Glu Lys Pro Ala Ser Lys	
	20 25 30
Ile Leu Glu Thr Asp Arg Thr Phe Arg Val Asn Met Lys Pro Phe Gln	
	35 40 45
Ile Ile Glu Arg Ser Cys Arg Tyr Phe Gly Ser Ser Tyr Ala Gly Arg	
	50 55 60
Lys Ala Gly Thr Tyr Glu Val Ile Lys Val Ser His Lys Pro Pro Ile	
	65 70 75 80
Met Val Asp His Ser Asn Asn Ile Phe Leu Phe Pro Thr Phe Ser Ser	
	85 90 95

Thr Arg Pro Gln Cys Gly Trp Leu Ser His Ala His Val His Glu Phe
 100 105 110
 Cys Ala Ala Lys Tyr Asp Asn Thr Phe Val Thr Phe Val Asn Gly Glu
 115 120 125
 Thr Leu Glu Leu Pro Val Ser Ile Ser Ser Phe Glu Asn Gln Val Tyr
 130 135 140
 Arg Thr Ala Trp Leu Thr Thr Lys Phe Ile Asp Arg Ile Glu Gly Asn
 145 150 155 160
 Pro Met Gln Lys Lys Gln Glu Phe Met Leu Tyr Pro Lys Glu Asp Arg
 165 170 175
 Asn Gln Leu Ile Tyr Glu Phe Ile Leu Arg Glu Leu Lys Lys Arg Tyr
 180 185 190

<210> 17
 <211> 588
 <212> DNA
 <213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 17
 atggacggca gagccgagag atacgtggag aactacgtga tcaacaagaa gaccatggcc 60
 ctgctgcccg tgatcctggg cgagaagaac gtgatcacca gaggatcga ggtggaggac 120
 agcttcttca tgttccagaa gcccttgac atcgtggaga gaagctgcag aaagcacggc 180
 agcagcttcc tgggcagaaa ggagggcacc aaggagctga ccagaatcac ccacaaggcc 240
 cccatgcgca tcagccccac cgaccagctg tacttcttcc ccacctacag ctacagcaga 300
 aaggagtgcg cctggctgag ccacttccac atcggcagca acaaggagct ggccgacggc 360
 aacctgatca tcagattcat caacggcttc gccgtgaagc tggagatgag caagagcagc 420
 ttcgagaacc agcagaacag aaccgccaag ctgagaaccg agtacgagga cagaaaggac 480
 aagcagggca acctgcagtt caagcccgtg gcccaaggagg ccatcagcac cctgagacc 540
 gcctacgaga aggtgtacct ggtgaaggag gaggacatcg agggcgag 588

<210> 18
 <211> 134
 <212> PRT
 <213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 18
 Gly Glu Lys Pro Ala Ser Lys Ile Leu Glu Thr Asp Arg Thr Phe Arg
 1 5 10 15
 Val Asn Met Lys Pro Phe Gln Ile Ile Glu Arg Ser Cys Arg Tyr Phe
 20 25 30
 Gly Ser Ser Tyr Ala Gly Arg Lys Ala Gly Thr Tyr Glu Val Ile Lys
 35 40 45
 Val Ser His Lys Pro Pro Ile Met Val Asp His Ser Asn Asn Ile Phe
 50 55 60
 Leu Phe Pro Thr Phe Ser Ser Thr Arg Pro Gln Cys Gly Trp Leu Ser
 65 70 75 80
 His Ala His Val His Glu Phe Cys Ala Ala Lys Tyr Gly Asn Thr Phe
 85 90 95
 Val Thr Phe Val Asn Gly Glu Thr Leu Glu Leu Pro Val Ser Ile Ser
 100 105 110

Ser Phe Glu Asn Gln Val Tyr Arg Thr Ala Trp Leu Arg Thr Lys Phe
115 120 125

Ile Asp Arg Ile Glu Gly
130

<210> 19
<211> 423
<212> DNA
<213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 19
aactacgtgg tgaacaagaa caccatggcc ctgctgccca tcatcctgag cgagaagaga 60
atcgtgacca gagggtgga gatggaggac agcttcttcg tgitccagaa gccctggac 120
atcatcgaga gaagctgcag aaagcacggc agcagcttcc tgggcagaaa ggagggcacc 180
aaggagctga cccacatcac ccacaaggcc cccatcgcca tcagccccac cgaccagctg 240
tacttcttcc ccacctacag ctacagcaga aaggagtgcg cctggctgag ccacttctac 300
atcgagagca acaaggagag caaggacggc aacgtgatcg tgagattcat caacggcttc 360
gccgtgaagc tggagatcag caagagcagc ttcgagaacc agctgaacag aaccgccaag 420
ctg 423

<210> 20
<211> 193
<212> PRT
<213> 韦氏芽孢杆菌(Bacillus weihenstephanensis)

<400> 20

Met Glu Ser Lys Val Glu Arg Tyr Val Glu Asn Tyr Val Val Ser Lys
1 5 10 15
Asn Thr Met Ala Leu Leu Pro Val Val Leu Gly Glu Lys Lys Val Val
20 25 30
Thr Arg Ile Val Glu Met Glu Asp Ser Phe Phe Val Phe Gln Lys Pro
35 40 45
Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Phe
50 55 60
Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr Arg Ile Thr His Lys Ala
65 70 75 80
Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
85 90 95
Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe His Ile Glu
100 105 110
Asp Asn Lys Glu Leu Lys Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
115 120 125
Gly Phe Ala Val Lys Leu Glu Met Ser Lys Ser Ser Phe Glu Asn Gln
130 135 140
Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Lys Lys
145 150 155 160
Lys Gln Gly Asn Pro Cys Phe Lys Glu Ile Asp Lys Asn Glu Glu Ser
165 170 175
Lys Leu Thr Pro Ala Tyr Glu Lys Val Tyr Phe Val Lys Glu Gly Glu
180 185 190

Val

<210> 21
 <211> 423
 <212> DNA
 <213> 芽孢杆菌属菌种(Bacillus sp.)

 <400> 21
 aactacgtgg tgaccaagaa caccatggcc ctgctgccg tgatcctgag cgagaagaag 60
 atcgccacca gagggtgga gatgaacgac agcttcttcg tgtccagaa gccctggac 120
 atcatcgaga gaagctgcag aaagcacggc agcagcttcc tgggcagaaa ggagggcacc 180
 aaggagctga cccacatcac ccacaaggcc cccatcgcca tcagccccgc cgaccagctg 240
 tacttcttcc ccacctacag ctacagcaga aaggagtgcg cctggtgag ccacttctac 300
 atcgagagca acaaggagct gaaggacggc aacctgatca tcagattcat caacggcttc 360
 gccgtgaagc tggagatcag caagaccagc ttcgagaacc agcagaacag aaccgccaag 420
 ctg 423

<210> 22
 <211> 193
 <212> PRT
 <213> 苏云金芽孢杆菌(Bacillus thuringiensis)

<400> 22

Met Glu Asn Lys Val Glu Arg Tyr Val Glu Asn Tyr Val Val Asn Lys
 1 5 10 15
 Asn Thr Met Ala Leu Leu Pro Val Ile Leu Ser Glu Lys Lys Ile Val
 20 25 30
 Thr Arg Val Val Glu Val Gln Asp Ser Phe Phe Val Phe Gln Lys Pro
 35 40 45
 Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Leu
 50 55 60
 Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr His Ile Thr His Lys Ala
 65 70 75 80
 Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
 85 90 95
 Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe Tyr Ile Glu
 100 105 110
 Ser Asn Lys Glu Leu Lys Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
 115 120 125
 Gly Phe Ala Val Lys Leu Glu Ile Ser Lys Thr Ser Phe Glu Asn Gln
 130 135 140
 Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Arg Lys
 145 150 155 160
 Lys Gln Gly Asn Pro Cys Phe Lys Glu Val Asp Gln Arg Asp Glu Ser
 165 170 175
 Thr Leu Arg Pro Ala Tyr Glu Arg Val Tyr Val Val Arg Glu Glu Asp
 180 185 190

Val

<210> 23
 <211> 423
 <212> DNA
 <213> 芽孢杆菌属菌种(Bacillus sp.)

 <400> 23
 aactacgtgg tgaacaagaa caccatggcc ctgctgccc tgatcctgag cgagaagaag 60
 atcgtgacca gagggtgga gatggcgac agcttcttcg tgttccagga gcccttgac 120
 atcatcgaga gaagctgcag aaagcacggc agcagcttc tgggcagaaa ggagggcacc 180
 aaggagctga cccacatcac ccacaaggcc cccatcgcca tcagccccac cgaccagctg 240
 tacttcttc ccacctacag ctacagcaga aaggagtgcg cctggctgag ccacttctac 300
 atcgagagca acaaggagag caaggacggc aacgtgatca tcagattcat caacggcttc 360
 gccgtgaagc tggagatcag caagagcagc ttcgagaacc agctgaacag aaccgccaag 420
 ctg 423

<210> 24
 <211> 193
 <212> PRT
 <213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 24
 Met Glu Ser Lys Val Glu Arg Tyr Val Glu Asn Tyr Val Val Asn Lys
 1 5 10 15
 Asn Thr Met Ala Leu Leu Pro Val Ile Leu Ser Glu Lys Lys Ile Val
 20 25 30
 Thr Arg Val Val Glu Val Gln Asp Ser Phe Phe Val Phe Gln Lys Pro
 35 40 45
 Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Leu
 50 55 60
 Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr His Ile Thr His Lys Ala
 65 70 75 80
 Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
 85 90 95
 Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe Tyr Ile Glu
 100 105 110
 Ser Asn Lys Glu Leu Lys Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
 115 120 125
 Gly Phe Ala Val Lys Leu Glu Ile Ser Lys Thr Ser Phe Glu Asn Gln
 130 135 140
 Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Arg Lys
 145 150 155 160
 Lys Gln Gly Asn Pro Cys Phe Lys Glu Val Asp Gln Arg Asp Glu Ser
 165 170 175
 Thr Leu Arg Pro Ala Tyr Glu Lys Val Tyr Phe Val Arg Glu Glu Asp
 180 185 190
 Leu

<210> 25

<211> 423
 <212> DNA
 <213> 芽孢杆菌属菌种(Bacillus sp.)
 <400> 25
 aactacgtgg tgaccaagaa caccatggcc ctgctgcccg tgatcctgag cgagaagaag 60
 atcgtgacca gagggtgga gatgaacgac agcttcttcg tgttccagaa gcccttgac 120
 atcatcgaga gaagctgcag aaagaacggc agcagcttcc tgggcagaaa ggagggcacc 180
 aaggagctga cccacatcac ccacaaggcc cccatcgcca tcagccccgc cgaccagctg 240
 tactttcttc ccacctacag ctacagcaga aaggagtgcg cctggctgag ccacttctac 300
 atcgagagca acaaggagct gaaggacggc aacctgatca tcagattcat caacggcttc 360
 gccgtgaagc tggagatcag caagaccagc ttcgagaacc agcagaacag aaccgccaag 420
 ctg 423

<210> 26
 <211> 193
 <212> PRT
 <213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 26
 Met Glu Asn Lys Val Glu Arg Tyr Val Glu Asn Tyr Val Val Asn Lys
 1 5 10 15
 Asn Thr Met Ala Leu Leu Pro Val Ile Leu Ser Glu Lys Lys Ile Val
 20 25 30
 Thr Arg Val Val Glu Val Gln Asp Ser Phe Phe Val Phe Gln Lys Pro
 35 40 45
 Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Leu
 50 55 60
 Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr His Ile Thr His Lys Ala
 65 70 75 80
 Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
 85 90 95
 Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe Tyr Ile Glu
 100 105 110
 Ser Asn Lys Glu Leu Lys Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
 115 120 125
 Gly Phe Ala Val Lys Leu Glu Ile Ser Lys Thr Ser Phe Glu Asn Gln
 130 135 140
 Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Arg Lys
 145 150 155 160
 Lys Gln Gly Asn Pro Cys Phe Lys Glu Val Asp Lys Lys Glu Glu Ser
 165 170 175
 Arg Leu Lys Pro Ala Tyr Glu Ser Val Tyr Phe Val Lys Glu Glu Glu
 180 185 190
 Val

<210> 27
 <211> 423
 <212> DNA
 <213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 27
 aactacgtgg tgaacaagaa caccatggcc ctgctgccc tgatcctgag cgagaagaag 60
 atcgtgacca gagggtgga gatggaggac agcttcttcg tgttccagaa gcccttgac 120
 atcatcgaga gaagctgcag aaagcacggc agcagcttcc tgggcagaaa ggagggcacc 180
 aaggagctga cccacatcac ccacaaggcc cccatcgcca tcagccccgc cgaccagttc 240
 tacttcttcc ccacctacag ctacagcaga aaggagtgcg cctggctgag ccacttctac 300
 atcgagagca acaaggagct gaaggacggc aacgtgatcg tgagattcat caacggcttc 360
 gccgtgaagc tggagatcag caagagcagc agcgagaacc agctgaacag aaccgccaag 420
 ctg 423

<210> 28
 <211> 193
 <212> PRT
 <213> 蜡状芽孢杆菌(*Bacillus cereus*)

<400> 28
 Met Glu Ser Lys Val Glu Arg Tyr Val Glu Asn Tyr Val Val Asn Lys
 1 5 10 15
 Asn Thr Met Ala Leu Leu Pro Ile Ile Leu Ser Glu Lys Lys Ile Val
 20 25 30
 Thr Arg Val Val Glu Val Gln Asp Ser Phe Phe Val Phe Gln Lys Pro
 35 40 45
 Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Leu
 50 55 60
 Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr His Ile Thr His Lys Ala
 65 70 75 80
 Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
 85 90 95
 Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe Tyr Ile Glu
 100 105 110
 Asn Asn Lys Glu Leu Lys Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
 115 120 125
 Gly Phe Ala Val Lys Leu Glu Ile Ser Lys Thr Ser Phe Glu Asn Gln
 130 135 140
 Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Arg Lys
 145 150 155 160
 Lys Gln Gly Asn Pro Cys Phe Lys Glu Ile Asp Lys Ser Glu Glu Ser
 165 170 175
 Thr Leu Arg Pro Ala Tyr Glu Lys Val Tyr Val Val Arg Glu Glu Glu
 180 185 190
 Val

<210> 29
 <211> 423
 <212> DNA
 <213> 芽孢杆菌属菌种(*Bacillus* sp.)

<400> 29
 aactacgtgg tgaacaagaa caccatggcc ctgctgagcg tgatcctgag cgagaagaag 60

atcgtgacca gagggtgga gatggcgac agcttcttcg tgtccagaa gccctggac 120
atcatcgaga gaagctgcag aaagcacggc agcagcttcc tgggcagaaa ggagggcacc 180
aaggagctga cccacatcac ccacaaggcc cccatcgcca tcagceccac cgaccagctg 240
tacttcttcc ccacctacag ctacagcaga aaggagtgcg cctggtgag ccacttctac 300
atcgagagca acaaggagag caaggacggc aacgtgatca tcagattcat caacggcttc 360
gccgtgaagc tggagatcag caagagcagc ttcgagaacc agctgaacag aaccgccaag 420
ctg 423

<210> 30
<211> 193
<212> PRT
<213> 蜡状芽孢杆菌(*Bacillus cereus*)

<400> 30

Met Glu Asn Lys Val Glu Arg Tyr Val Glu Asn Tyr Val Val Asn Lys
1 5 10 15
Asn Thr Met Ala Leu Leu Pro Val Ile Leu Ser Glu Lys Lys Ile Val
20 25 30
Thr Arg Val Val Glu Val Gln Asp Ser Phe Phe Val Phe Gln Lys Pro
35 40 45
Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Leu
50 55 60
Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr His Ile Thr His Lys Ala
65 70 75 80
Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
85 90 95
Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe Tyr Ile Glu
100 105 110
Ser Asn Lys Glu Leu Lys Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
115 120 125
Gly Phe Ala Val Lys Leu Glu Ile Ser Lys Thr Ser Phe Glu Asn Gln
130 135 140
Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Arg Lys
145 150 155 160
Lys Gln Gly Asn Pro Cys Phe Lys Glu Val Asp Lys Asn Glu Glu Ser
165 170 175
Arg Leu Lys Pro Ala Tyr Glu Ser Val Tyr Phe Val Lys Glu Glu Glu
180 185 190
Val

<210> 31
<211> 402
<212> DNA
<213> 芽孢杆菌属菌种(*Bacillus sp.*)

<400> 31

ggcgagaagc ccgccagcaa gatcctggag accgacagaa ccttcagagt gaacatgaag 60
cccttcaga tcatcgagag aagctgcaga tacttcggca gcagctacgc cggcagaaag 120

gccggcacct acgaggtgat caaggtgagc cacaagcccc ccatcatggt ggaccacagc 180
 aacaacatct tcctgttccc caccttcagc agcaccagac cccagtgcgg ctggctgagc 240
 cagccccag tgcacgagtt ctgcgcccc aagtacgca acaccttcgt gaccttcgtg 300
 aacggcgaga ccctggagct gcccgtgagc atcagcagct tcgagaacca ggtgtacaga 360
 accgcctggc tgagaaccaa gttcatcgac agaatcgagg gc 402

<210> 32
 <211> 193
 <212> PRT
 <213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 32

Met Glu Ser Lys Val Glu Arg Tyr Val Glu Asn Tyr Val Val Thr Lys
 1 5 10 15

Asn Thr Met Ala Leu Leu Pro Val Ile Leu Ser Glu Lys Lys Ile Val
 20 25 30

Thr Arg Val Val Glu Met Asn Asp Ser Phe Phe Val Phe Gln Lys Pro
 35 40 45

Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Leu
 50 55 60

Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr His Ile Thr His Lys Ala
 65 70 75 80

Pro Ile Ala Ile Ser Pro Ala Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
 85 90 95

Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe Tyr Ile Glu
 100 105 110

Ser Asn Lys Glu Leu Lys Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
 115 120 125

Gly Phe Ala Val Lys Leu Glu Ile Ser Lys Thr Ser Phe Glu Asn Gln
 130 135 140

Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Arg Lys
 145 150 155 160

Lys Gln Gly Asn Pro Cys Phe Lys Glu Val Asp Lys Lys Glu Glu Ser
 165 170 175

Thr Leu Arg Pro Ala Tyr Glu Ser Val Tyr Phe Val Lys Glu Gly Glu
 180 185 190

Val

<210> 33
 <211> 504
 <212> DNA
 <213> 苏云金芽孢杆菌(Bacillus thuringiensis)

<400> 33

atgaacgacg agaacaacat catcatcagc agcagcacca tgatgctggt gcctacaac 60

caccctact acagaaccaa gatcatcgac agcagcggca acgacctgtg cagctgccag 120

accgcctgc agctgatcaa gaagagctgc ctgaccaga tccacagcac ctaccagggc 180

agaagaaacg cegtgcagac caacttcaga ttcaagcaga acgtgcccac ccccatcaac 240

cacagagagt acatctgcac ctccccacc gagagcccca gcagcccaaa ctgcatctgg 300

ctgttctaca accacatcca cgacatcgag ttctgccaga agatcaagga ggccaagatc 360
 cacttcagca acggcaccac caccaccate agcatcagcc cccacaagct gcagcagcag 420
 ctgctgaagg cgggtacat cctgagcaga atgaacatgc aggacagcct gcagttcaag 480
 aaccccctgc tgcacctgct gcac 504

<210> 34
 <211> 196
 <212> PRT
 <213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 34

Met Asp Gly Arg Ala Glu Arg Tyr Val Glu Asn Tyr Val Ile Asn Lys
 1 5 10 15
 Lys Thr Met Ala Leu Leu Pro Val Ile Leu Gly Glu Lys Asn Val Ile
 20 25 30
 Thr Arg Val Ile Glu Val Glu Asp Ser Phe Phe Met Phe Gln Lys Pro
 35 40 45
 Leu Asp Ile Val Glu Arg Ser Cys Arg Lys His Gly Ser Ser Phe Leu
 50 55 60
 Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr Arg Ile Thr His Lys Ala
 65 70 75 80
 Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu Tyr Phe Phe Pro Thr Tyr
 85 90 95
 Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu Ser His Phe His Ile Ala
 100 105 110
 Ser Asn Lys Glu Leu Ala Asp Gly Asn Leu Ile Ile Arg Phe Ile Asn
 115 120 125
 Gly Phe Ala Val Lys Leu Glu Met Ser Lys Ser Ser Phe Glu Asn Gln
 130 135 140
 Gln Asn Arg Thr Ala Lys Leu Arg Thr Glu Tyr Glu Asp Arg Lys Asp
 145 150 155 160
 Lys Gln Gly Asn Leu Gln Phe Lys Pro Val Ala Lys Glu Ala Ile Ser
 165 170 175
 Thr Leu Arg Pro Ala Tyr Glu Lys Val Tyr Leu Val Lys Glu Glu Asp
 180 185 190
 Ile Glu Gly Glu
 195

<210> 35
 <211> 579
 <212> DNA
 <213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 35

atggagagca aggtggagag atactgtggag aactactgtg tgaacaagaa caccatggcc 60
 ctgctgccca tcatactgag cgagaagaag atcgtgacca gactgtgtgga ggtgcaggac 120
 agcttctctg tgttcagaa gccctggac atcatcgaga gaagctgcag aaagcacggc 180
 agcagcttcc tgggcagaaa ggagggcacc aaggagctga cccacatcac ccacaaggcc 240
 cccatgcca tcagcccccac cgaccagctg tacttcttcc ccacctacag ctacagcaga 300

aaggagtgcg cctggctgag ccacttctac atcgagaaca acaaggagct gaaggacggc 360
aacctgatca tcagattcat caacggcttc gccgtgaagc tggagatcag caagaccagc 420
ttcgagaacc agcagaacag aaccgccaag ctgagaaccg agtacgagga cagaagaaag 480
aagcagggca acccctgctt caaggagatc gacaagagcg aggagagcac cctgagaccc 540
gcctacgaga aggtgtacgt ggtgagagag gaggaggtg 579

<210> 36
<211> 193
<212> PRT
<213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 36

Met Asn His Leu Asp Met His Lys Lys Arg Leu Val Glu Glu Tyr Glu
1 5 10 15
Ile Asn Pro Ser Thr Met Ile Ile Leu Pro Gln Ile Tyr Gly Lys Lys
20 25 30
Ile Tyr Ser Arg Ile Phe Glu Val Glu Asp Glu Phe Leu Ser Pro Phe
35 40 45
Lys Leu Phe Asp Ile Val Lys Lys Ser Cys Gly Tyr Phe Gly Ser Ser
50 55 60
Tyr Glu Gly Arg Lys Asp Ala Thr Lys Asp Ile Ile Gly Val Thr His
65 70 75 80
Lys Val Pro Ile Val Ile Asp Pro Thr Asn Leu Leu Tyr Phe Phe Pro
85 90 95
Thr Thr Ser Pro Asn Asn Pro Asp Cys Ile Trp Ile Ser Tyr Glu His
100 105 110
Ile Ala Ala His His Arg Thr Asp Pro Ser His Thr Lys Val Val Phe
115 120 125
Gly Asn Lys His Thr Leu Ile Leu Pro Val Ser Ser Ser Ser Phe Glu
130 135 140
Asn Gln Leu Leu Arg Thr Ala His Leu Arg Thr Lys Leu His Gln Arg
145 150 155 160
Ile Glu Gly His Gly Arg Lys Met Phe Tyr Phe Thr Asp Asn Gln Arg
165 170 175
Leu Asn Arg Ala Ser Glu Pro Val Gly Ser Tyr Arg Asn Arg Lys Leu
180 185 190
Glu

<210> 37
<211> 579
<212> DNA
<213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 37

atggagaaca aggtggagag atacgtggag aactacgtgg tgaacaagaa caccatggcc 60
ctgctgcccc tgatcctgag cgagaagaag atcgtgacca gagtgggtgga ggtgcaggac 120
agctttcttcg tgttccagaa gcccctggac atcatcgaga gaagctgcag aaagcacggc 180
agcagcttcc tgggcagaaa ggagggcacc aaggagctga cccacatcac ccacaaggcc 240
cccategcca tcageccccc cgaccagctg tacttcttcc ccacctacag ctacagcaga 300

aaggagtgcg cctggctgag ccacttctac atcgagagca acaaggagct gaaggacggc 360
aacctgatca tcagattcat caacggcttc gccgtgaagc tggagatcag caagaccagc 420
ttcgagaacc agcagaacag aaccgccaag ctgagaaccg agtacgagga cagaagaaag 480
aagcagggca acccctgctt caaggaggtg gacaagaagg aggagagcag actgaagccc 540
gcctacgaga gcgtgtactt cgtgaaggag gaggaggtg 579

<210> 38
<211> 141
<212> PRT
<213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 38

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

<210> 39
<211> 576
<212> DNA
<213> 地衣芽孢杆菌(Bacillus licheniformis)

<400> 39

atgagcaccg aggacatgac caaggacacc tacgaggtga acagcagcac catggccgtg 60
ctgcccttgg gcgagggcga gaagcccgcc agcaaatcc tggagaccga cagaaccttc 120
agagtgaaca tgaagccctt ccagatcatc gagagaagct gcagatactt cggcagcagc 180
tacgccggca gaaaggccgg cacctacgag gtgatcaagg tgagccacaa gccccccatc 240
atggtgacc acagcaacaa catcttctg ttccccacct tcagcagcac cagaccccag 300
tgcggctggc tgagccacgc ccacgtgcac gatttctgag ccgccaagta cgacaacacc 360
ttcgtgacct tcgtgaacgg cgagaccctg gagctgcccg tgagcatcag cagcttcgag 420
aaccaggtgt acagaaccgc ctggctgaga accaagtcca tcgacagaat cgagggeaac 480
cccatgcaga agaagcagga gttcatgctg taccccaagg aggacagaaa ccagctgatc 540
tacgagttca tcctgagaga gctgaagaag agatac 576

<210> 40
 <211> 141
 <212> PRT
 <213> 芽孢杆菌属菌种(Bacillus sp.)

 <400> 40
 Asn Tyr Val Val Asn Lys Asn Thr Met Ala Leu Leu Pro Val Ile Leu
 1 5 10 15
 Ser Glu Lys Lys Ile Val Thr Arg Val Val Glu Met Gly Asp Ser Phe
 20 25 30
 Phe Val Phe Gln Glu Pro Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys
 35 40 45
 His Gly Ser Ser Phe Leu Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr
 50 55 60
 His Ile Thr His Lys Ala Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu
 65 70 75 80
 Tyr Phe Phe Pro Thr Tyr Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu
 85 90 95
 Ser His Phe Tyr Ile Glu Ser Asn Lys Glu Ser Lys Asp Gly Asn Val
 100 105 110
 Ile Ile Arg Phe Ile Asn Gly Phe Ala Val Lys Leu Glu Ile Ser Lys
 115 120 125
 Ser Ser Phe Glu Asn Gln Leu Asn Arg Thr Ala Lys Leu
 130 135 140

<210> 41
 <211> 579
 <212> DNA
 <213> 苏云金芽孢杆菌(Bacillus thuringiensis)

 <400> 41
 atggagaaca aggtggagag atactgtggag aactactgtg tgaacaagaa caccatggcc 60
 ctgctgcccc tgatcctgag cgagaagaag atcgtgacca gactggtgga ggtgcaggac 120
 agctttcttcg tgttcagaa gccctggac atcatcgaga gaagctgcag aaagcacggc 180
 agcagcttcc tgggcagaaa ggaggcacc aaggagctga cccacatcac ccacaaggcc 240
 cccatcgcca tcagccccac cgaccagctg tactttcttc ccacctacag ctacagcaga 300
 aaggagtgcg cctggctgag ccacttctac atcgagagca acaaggagct gaaggacggc 360
 aacctgatca tcagattcat caacggcttc gccgtgaagc tggagatcag caagaccagc 420
 ttcgagaacc agcagaacag aaccgccaag ctgagaaccg agtacgagga cagaagaaag 480
 aagcagggca acccctgett caaggagtg gaccagagag acgagagcac cctgagaccc 540
 gcctacgaga gactgtactg ggtgagagag gaggacgtg 579

<210> 42
 <211> 141
 <212> PRT
 <213> 芽孢杆菌属菌种(Bacillus sp.)

 <400> 42
 Asn Tyr Val Val Asn Lys Asn Thr Met Ala Leu Leu Pro Val Ile Leu
 1 5 10 15

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

<210> 43
 <211> 579
 <212> DNA
 <213> 蜡状芽孢杆菌(Bacillus cereus)

<400> 43
 atggagagca aggtggagag atactggag aactactggg tgaacaagaa caccatggcc 60
 ctgctgcccc tgatcctgag cgagaagaag atcgtgacca gagtgggtga ggtgcaggac 120
 agcttcttcg tgttcagaa gccctggac atcatcgaga gaagctgcag aaagcacggc 180
 agcagcttcc tgggcagaaa ggagggcacc aaggagctga cccacatcac ccacaaggcc 240
 cccatcgcca tcagccccac cgaccagctg tacttcttcc ccacctacag ctacagcaga 300
 aaggagtgcg cctggctgag ccacttctac atcgagagca acaaggagct gaaggacggc 360
 aacctgatca tcagattcat caacggcttc gccgtgaagc tggagatcag caagaccagc 420
 ttcgagaacc agcagaacag aaccgccaag ctgagaaccg agtacgagga cagaagaaag 480
 aagcagggca acccctgctt caaggagtg gaccagagag acgagagcac cctgagacct 540
 gcctacgaga aggtgtactt cgtgagagag gaggacctg 579

<210> 44
 <211> 141
 <212> PRT
 <213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 44
 Asn Tyr Val Val Asn Lys Asn Thr Met Ala Leu Leu Ser Val Ile Leu
 1 5 10 15
 Ser Glu Lys Lys Ile Val Thr Arg Val Val Glu Met Gly Asp Ser Phe
 20 25 30
 Phe Val Phe Gln Lys Pro Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys
 35 40 45
 His Gly Ser Ser Phe Leu Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr
 50 55 60
 His Ile Thr His Lys Ala Pro Ile Ala Ile Ser Pro Thr Asp Gln Leu
 65 70 75 80

130 135 140

<210> 47
 <211> 579
 <212> DNA
 <213> 炭疽芽孢杆菌(Bacillus anthracis)

<400> 47
 atggagaaca aggtggagag atacgtggag aactacgtgg tgaacaagaa caccatggcc 60
 ctgctgcccc tgatcctgag cgagaagaag atcgtgacca gagtgggtgga ggtgcaggac 120
 agctttcttcg tgttccagaa gcccttgac atcatcgaga gaagctgcag aaagcacggc 180
 agcagcttcc tgggcagaaa ggagggcacc aaggagctga cccacatcac ccacaaggcc 240
 cccatcgcca tcagccccac cgaccagctg tactttcttc ccacctacag ctacagcaga 300
 aaggagtgcg cctggctgag ccacttctac atcgagagca acaaggagct gaaggacggc 360
 aacctgatca tcagattcat caacggcttc gccgtgaagc tggagatcag caagaccagc 420
 ttcgagaacc agcagaacag aaccgccaag ctgagaaccg agtacgagga cagaagaaag 480
 aagcagggca acccctgctt caaggaggtg gacaagaacg aggagagcag actgaagccc 540
 gcctacgaga gctgttactt cgtgaaggag gaggaggtg 579

<210> 48
 <211> 141
 <212> PRT
 <213> 芽孢杆菌属菌种(Bacillus sp.)

<400> 48

Asn Tyr Val Val Thr Lys Asn Thr Met Ala Leu Leu Pro Val Ile Leu
 1 5 10 15

Ser Glu Lys Lys Ile Val Thr Arg Val Val Glu Met Asn Asp Ser Phe
 20 25 30

Phe Val Phe Gln Lys Pro Leu Asp Ile Ile Glu Arg Ser Cys Arg Lys
 35 40 45

Asn Gly Ser Ser Phe Leu Gly Arg Lys Glu Gly Thr Lys Glu Leu Thr
 50 55 60

His Ile Thr His Lys Ala Pro Ile Ala Ile Ser Pro Ala Asp Gln Leu
 65 70 75 80

Tyr Phe Phe Pro Thr Tyr Ser Tyr Ser Arg Lys Glu Cys Ala Trp Leu
 85 90 95

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

Ile Ile Arg Phe Ile Asn Gly Phe Ala Val Lys Leu Glu Ile Ser Lys
 115 120 125

Thr Ser Phe Glu Asn Gln Gln Asn Arg Thr Ala Lys Leu
 130 135 140

<210> 49
 <211> 576
 <212> DNA
 <213> 地衣芽孢杆菌(Bacillus licheniformis)

<400> 49
 atgagcaccg aggacatgac caaggacacc tacgaggtga acagcagcac catggccgtg 60

ctgcccctgg gcgagggcga gaagcccgcc agcaagatcc tggagaccga cagaaccttc 120
 agagtgaaca tgaagccctt ccagatcatc gagagaagct gcagatactt cggcagcagc 180
 tacgccggca gaaaggccgg cacctacgag gtgatcaagg tgagccacaa gccccccatc 240
 atggtggacc acagcaacaa catcttctg ttccccacct tcagcagcac cagaccccag 300
 tgcggctgge tgagccacgc ccacgtgac gagttctgcg ccgccaagta cgacaacacc 360
 ttcgtgacct tegtgaacgg cgagaccctg gagctgcccg tgagcatcag cagcttcgag 420
 aaccaggtgt acagaaccgc ctggtgac accaagtcca tcgacagaat cgagggcaac 480
 cccatgcaga agaagcagga gttcatgctg taccccaagg aggacagaaa ccagctgatc 540
 tacgagttca tcctgagaga gctgaagaag agatac 576

<210> 50
 <211> 168
 <212> PRT
 <213> 苏云金芽孢杆菌(Bacillus thuringiensis)

<400> 50

Met Asn Asp Glu Asn Asn Ile Ile Ile Ser Ser Ser Thr Met Met Leu
 1 5 10 15
 Val Pro Tyr Asn His Pro Tyr Tyr Arg Thr Lys Ile Ile Asp Ser Ser
 20 25 30
 Gly Asn Asp Leu Cys Ser Cys Gln Thr Ala Leu Gln Leu Ile Lys Lys
 35 40 45
 Ser Cys Leu Thr Gln Ile His Ser Thr Tyr Gln Gly Arg Arg Asn Ala
 50 55 60
 Val Gln Thr Asn Phe Arg Phe Lys Gln Asn Val Pro Ile Pro Ile Asn
 65 70 75 80
 His Arg Glu Tyr Ile Cys Thr Phe Pro Thr Glu Ser Pro Ser Ser Pro
 85 90 95
 Asn Cys Ile Trp Leu Phe Tyr Asn His Ile His Asp Ile Glu Phe Cys
 100 105 110
 Gln Lys Ile Lys Glu Ala Lys Ile His Phe Ser Asn Gly Thr Thr Thr
 115 120 125
 Thr Ile Ser Ile Ser Pro His Lys Leu Gln Gln Gln Leu Leu Lys Ala
 130 135 140
 Gly Tyr Ile Leu Ser Arg Met Asn Met Gln Asp Ser Leu Gln Phe Lys
 145 150 155 160
 Asn Pro Leu Leu His Leu Leu His
 165

<210> 51
 <211> 1143
 <212> DNA
 <213> 地衣芽孢杆菌(Bacillus licheniformis)

<400> 51

atgttctata ctaatcaacc agccatcaac tgcactacat acaaacaaat gtcctcgtca 60
 actggttcgc tatccaattt gttctctgaa agtgactcgc cttatttggt ctcaaggaat 120
 gtggaaaatg ctttttgta agcatttga gctgaaaact tggggaggtc agactgttct 180
 gctgacgctt cattaatcg tgtcggaatt ggtattaaga cttttcttca tgtaatggt 240

catactcttc aaaaagtagc tgaattcaat aaagactcag acttgtatcg tgggaaatct 300
 ccaaaaagagc taataaacac ggttgcttct ctccgtaacg agagaattga atttactaaa 360
 agaacataatg gtattgattc aatgatatac cactgtgtaa caagaaagcc agggaaaatt 420
 cttatTTTTg aagagccaat ggacttggtt gaaatctcct caattacaaa tgtgaaagta 480
 agtaacaaca gaaatacaat cacctttgaa gacggctac acgaatacag ctttaatgct 540
 actaagagca ccctttataa gcgttttatac actgataaac ctattgaaga aattaatggt 600
 gaaatcttag aaaatcctta tcatgaattg gctaaactat ttggctttga aattccaaaa 660
 attccagcac caactgtcaa tccttttgaa aaccttgagc acgttattct tccactcttt 720
 tcagaccgtg gctcaaaagc tcatgtacca gaaaaaagcg gtctaaacca atggaatgct 780
 ttaggtcgac cagcaaaccc taacgagatt tatataccaa ttccaaaatg gattcataat 840
 gtattcccaa catttttccc agctcgtgat aaaccttttc agttacgctt gccagacaaa 900
 tegcttttat cagccaaggt atgccaagac aatagtaaag cacttatgct taatccaaat 960
 agtgcctttg gagaatggct actaagacaa gttatgaact tagaggaaaa agaacttcta 1020
 acctatgaaa tgcgtgaaag actaaatatt gactcagtaa ttgtttataa acacagcgaa 1080
 caacattact ccattgattt ttgtgaaatg gttcttatg atgaatttga aatgaaaac 1140
 aaa 1143

<210> 52
 <211> 381
 <212> PRT
 <213> 地衣芽孢杆菌(Bacillus licheniformis)

<400> 52

Met Phe Tyr Thr Asn Gln Pro Ala Ile Asn Cys Thr Thr Tyr Lys Gln
 1 5 10 15
 Met Leu Arg Ser Thr Gly Ser Leu Ser Asn Leu Phe Ser Glu Ser Asp
 20 25 30
 Ser Pro Tyr Leu Val Ser Arg Asn Val Glu Asn Ala Phe Cys Glu Ala
 35 40 45
 Phe Gly Ala Glu Asn Leu Gly Arg Ser Asp Cys Ser Ala Asp Ala Ser
 50 55 60
 Leu Asn Arg Val Gly Ile Gly Ile Lys Thr Phe Leu His Gly Asn Gly
 65 70 75 80
 His Thr Leu Gln Lys Val Ala Glu Phe Asn Lys Asp Ser Asp Leu Tyr
 85 90 95
 Arg Gly Lys Ser Pro Lys Glu Leu Ile Asn Thr Val Ala Ser Leu Arg
 100 105 110
 Asn Glu Arg Ile Glu Phe Thr Lys Arg Thr Tyr Gly Ile Asp Ser Met
 115 120 125
 Ile Tyr His Cys Val Thr Arg Lys Pro Gly Lys Ile Leu Ile Phe Glu
 130 135 140
 Glu Pro Met Asp Leu Val Glu Ile Ser Ser Ile Thr Asn Val Lys Val
 145 150 155 160
 Ser Asn Asn Arg Asn Thr Ile Thr Phe Glu Asp Gly Leu His Glu Tyr
 165 170 175

Ser Phe Asn Val Thr Lys Ser Thr Leu Tyr Lys Arg Phe Ile Thr Asp
180 185 190

Lys Pro Ile Glu Glu Ile Asn Val Glu Ile Leu Glu Asn Pro Tyr His
195 200 205

Glu Leu Ala Lys Leu Phe Gly Phe Glu Ile Pro Lys Ile Pro Ala Pro
210 215 220

Thr Val Asn Pro Phe Glu Asn Leu Glu His Val Ile Leu Pro Leu Phe
225 230 235 240

Ser Asp Arg Gly Ser Lys Arg His Val Pro Glu Lys Ser Gly Leu Asn
245 250 255

Gln Trp Asn Ala Leu Gly Arg Pro Arg Asn Pro Asn Glu Ile Tyr Ile
260 265 270

Pro Ile Pro Lys Trp Ile His Asn Val Phe Pro Thr Phe Phe Pro Ala
275 280 285

Arg Asp Lys Pro Phe Gln Leu Arg Leu Pro Asp Lys Ser Leu Leu Ser
290 295 300

Ala Lys Val Cys Gln Asp Asn Ser Lys Ala Leu Met Ser Asn Pro Asn
305 310 315 320

Ser Ala Leu Gly Glu Trp Leu Leu Arg Gln Val Met Asn Leu Glu Glu
325 330 335

Lys Glu Leu Leu Thr Tyr Glu Met Leu Glu Arg Leu Asn Ile Asp Ser
340 345 350

Val Ile Val Tyr Lys His Ser Glu Gln His Tyr Ser Ile Asp Phe Cys
355 360 365

Glu Met Gly Ser Tyr Asp Glu Phe Glu Asn Glu Asn Lys
370 375 380

<210> 53

<211> 3049

<212> DNA

<213> 地衣芽孢杆菌(Bacillus licheniformis)

<400> 53

tcatgttccc atattctttt aatgttccaa tccttttctc cgtaatatatt attaacttcc 60

ttatctcttt ttttatttct itcgagtttt ttctcccaat attccgtatt actttttggt 120

atattcccgl gtttttcaca cgcattgccag aaacaagaat caatgaatat gactatttta 180

tattttctgta ttactataac tggactaccg tataatttct taacattttt tcggaatctt 240

attccacggt gccatagttc tttagtaacc ttatcttcta attttgaacg agatttgatt 300

gcttgcattgt tttttcttct ttgttctttt gaaaccgtgt cagtcataga agagtcctcc 360

aaagccacaa taattgtatt ctataaacga ggaagcaagc cctcaagctt acccctctt 420

agttcctttt ttgctacttt atttatttgt ttccattttc aaattcatca taagaacca 480

tttcacaaaa atcaatggag taatgttggt cgtgtgtttt ataaacaatt actgagtcaa 540

tatttagtct ttccagcatt tcalaggttt gaagttcttt ttctcttaag ttcataactt 600

gtcttagtag ccattctcca agagcactat ttggattaga cataagtget ttactattgt 660

cttggcatac cttggctgat aaaagcgatt tgtctggcaa gcgtaactga aaagttttat 720

cacgagctgg gaaaaatggt gggaatacat tatgaatcca ttttgaatt ggtatataaa 780

tctcgtagg gtttcgtggt cgacctaaag cattccattg gtttagaccg cttttttctg	840
gtacatgacg ctttgagcca cggctctgaaa agagtggaag aataacgtgc tcaagggttt	900
caaaaggatt gacagttggt gctggaattt ttggaatttc aaagccaaat agtttagcca	960
attcatgata aggattttct aagatttcaa cattaatttc ttcaataggt ttatcagtga	1020
taaaacgctt ataaagggtg ctcttagtga cattaagct gtattcgtgt agaccgtctt	1080
caaaggatgat tgtattttctg ttgttactta ctttcacatt tglaattgag gagatttcaa	1140
ccaagtccat tggctcttca aaaataagaa ttttccttgg ctttcttggt acacagtgg	1200
atatcattga atcaatacca tatgttcttt tagtaaattc aattctctcg ttacggagag	1260
aagcaaccgt gtttattagc tcttttgag atttcccacg atacaagtct gagtctttat	1320
tgaattcagc tactttttga agagtatgac cattaccatg aagaaaagtc ttaataccaa	1380
ttccgacacg atttaatgaa gcgtcagcag aacagtctga cctcccaag ttttcagctc	1440
caaagtcttc acaaaaagca ttttccat tcccttgagac caaataaggc gagtcacttt	1500
cagagaacaa attggatagc gaaccagttg agcggagcat ttgtttgtat gtagtgcagt	1560
tgatggctgg ttgattagta tagaacatta ttttctctc tcttttatgc ttgtcatttc	1620
ttctttcaga cccaaaaggt agtcagctga tacgttcaat gtttcagcta ttcttttgaa	1680
agtgaccaat gatggagttc tattttact ttcatatagt gaccaagtgc ttctagtgac	1740
cccactttt tcagegattt ggetgggtaa taacctacga gcttctctg cattttgaat	1800
acgalltcca aggaaaggta tcatttttgc acctccaaga ttgtttggtt tcagagtatc	1860
accagaacc cggaaaatag tccaaagta gtaacagca acaataaaa aataaataag	1920
ttgtttactc ttagcaaact tgttactaaa atttgataaa gttattcatt taatccagct	1980
cttatgctaa aattgcatta gcggacaagc ttaatgtttg caaggaggta taattttgac	2040
ttatcgagta ggtagtatgt ttgctgggat aggtggaact tgtttaggt ttatccaagc	2100
tggegttagg altgtctggg caaatgaaat agacaaaaat gcttgtatta cttatagaaa	2160
ttattttggg gatgcttact tacaagaggg tgacattaac ctaatagata aaaactccat	2220
acctgaactg gacattttga ttggaggttt tccctgcca gccttctcta tagctggcta	2280
tcgtaaaggg ttigaagatg aaaggggaaa cgtgttcttt caaatattag aggtattgga	2340
agcacaaaga aatgtttatg gacacttacc ccaagcaata atgcttgaga atgtaaagaa	2400
cttatttaca catgatagag gtaatacgtg cagagtaata aaagaggctt tggaagcctt	2460
tggttatacc gtaaaagctg aggttcttaa ttcaatggaa tacggtaacg tgccacaaaa	2520
cagagagcgg atttatattg taggttttca agatgagagc caagctgaaa ggtttagctt	2580
tccagaccca attcctttaa caaatcaact taatgatgta attgaccgaa ctcgagagat	2640
tgataaaaga tattattatg atgaaacctc tcaatattat gatatgttgc gagaagccat	2700
ggacagtaca gatacaactt atcaataag acgtatatat gttcagaaa atagaagcaa	2760
tgtttgcct aactgacag cgaatatggg aactggaggg cataatgttc ctattgtatt	2820
agactttgaa aataataata gaaaactaac accagaagaa tgcttactat tgcaaggttt	2880
cccagctgac tatcatttcc cagaaggcat ggcaaacact cacaaatata aacaagctgg	2940
taactctggt acggtgccag ttataagaag aattgccact aatattatta gcgtattgaa	3000

cattggaatg aatataaatc aagaacatga atatgcaata gctgaataa	3049
<210> 54	
<211> 26	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 54	
gagctctgca aggagtata attttg	26
<210> 55	
<211> 27	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 55	
acgcgtttat tcagctattg catattc	27
<210> 56	
<211> 37	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 56	
ccaggcctta agggccgcat gcgtccttct ttgtgct	37
<210> 57	
<211> 27	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 57	
gagctccttt caatgtgata catatga	27
<210> 58	
<211> 24	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 58	
gcgcccgctc gctttccaat ctga	24
<210> 59	
<211> 24	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 59	
atcgatcagc ttgataaac ccta	24
<210> 60	
<211> 26	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 60	
gagctctgca aggagtata attttg	26
<210> 61	
<211> 28	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 61	

cgtcgacgcc ttgcggtag tggtgctt	28
<210> 62	
<211> 19	
<212> DNA	
<213> 巨大芽孢杆菌(<i>Bacillus megaterium</i>)	
<400> 62	
gagctcgat cccatttc	19
<210> 63	
<211> 94	
<212> DNA	
<213> 巨大芽孢杆菌(<i>Bacillus megaterium</i>)	
<400> 63	
atctctgagc tcgcatgat taattaattc agaacgctcg gttgccgccc ggcgtttttt	60
atgcagcaat ggcaagaac tcccggttag ctcc	94
<210> 64	
<211> 43	
<212> DNA	
<213> 巨大芽孢杆菌(<i>Bacillus megaterium</i>)	
<400> 64	
cttctcgaga ataatatttc cttctaagtc gtttaggatt ccg	43
<210> 65	
<211> 79	
<212> DNA	
<213> 巨大芽孢杆菌(<i>Bacillus megaterium</i>)	
<400> 65	
caagcatcaa aaaacaccaa cttagttcgg tggataaaca aaggagtggc tattattcaa	60
attgcagatc aggcttttag	79
<210> 66	
<211> 27	
<212> DNA	
<213> 地衣芽孢杆菌(<i>Bacillus licheniformis</i>)	
<400> 66	
gtggatccga ttagggat caaatg	27
<210> 67	
<211> 54	
<212> DNA	
<213> 地衣芽孢杆菌(<i>Bacillus licheniformis</i>)	
<400> 67	
cagtactgca gtcaatagcg ctttttcagc tcctgagga taaattcgta tacc	54
<210> 68	
<211> 24	
<212> DNA	
<213> 地衣芽孢杆菌(<i>Bacillus licheniformis</i>)	
<400> 68	
ctgaaacaac aaaaacggct ttac	24
<210> 69	
<211> 30	
<212> DNA	

<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 69	
actgaagctt ggttgcggtc agcgggatcg	30
<210> 70	
<211> 25	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 70	
agtcgaattc gactggaagc agagc	25
<210> 71	
<211> 28	
<212> DNA	
<213> 地衣芽孢杆菌(Bacillus licheniformis)	
<400> 71	
tcaggagctc agtaccattt tccctata	28

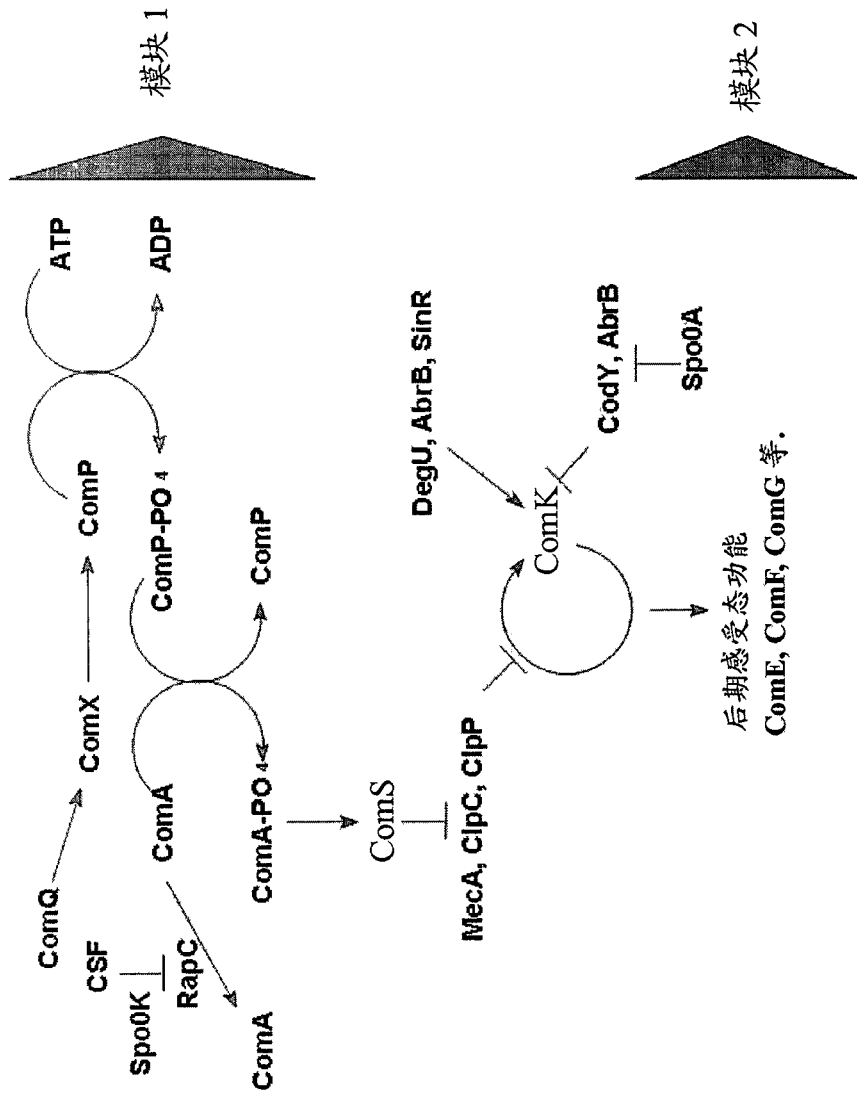


图 1

M F Y T N Q P A I N C T T Y K Q M L R S
ATGTTCTATACTAATCAACCAGCCACTCAACTGCACTACATACAAAACAAATGCTCCGCTCA60

T G S L S N L F S E S D S P Y L V S R N
ACTGGTTCGCTATCCAATTTGTTCTCTGAAAGTGAAGTCACTCGCCCTTATTTGGTCTCAAGGAAT120

V E N A F C E A F G A E N L G R S D C S
GTGAAAATGCTTTTGTGAAGCATTGGAGCTGAAAACCTGGGAGGTCAGACTGTTCT180

A D A S L N R V G I G I K T F L H G N G
GCTGACGCTTCATTAATAATCGTTCGGAATGGTATTAAGACTTTTCTTCAATGGTAATGGT240

H T L Q K V A E F N K D S D L Y R G K S
CATACTCTCAAAAAGTAGCTGAATTCAATAAAGACTCAGACTTGTATCGTGGAAAATCT300

P K E L I N T V A S L R N E R I E F T K
CCAAAAGAGCTAATAAACACGGTTGCTCTCCGTAACGAGAGAAATGAAATTTACTAAA360

R T Y G I D S M I Y H C V T R K P G K I
AGAACATATGGTATGATTCAATGATATACCCTGTGTAACAAGAAAGCCAGGAAAAT420

L I F E E P M D L V E I S S I T N V K V
CTTATTTTGAAGAGCCAAATGGACTTGGTTGAAATCTCCTCAATTACAAAATGTGAAAGTA480

S N N R N T I T F E D G L H E Y S F N V
AGTAACAACAGAAAATACAATCACCTTTGAAGACGGTCTACACGAATACAGCTTTAATGTC540

T K S T L Y K R F I T D K P I E E I N V
ACTAAGAGCACCCCTTTATAAGCGTTTTATCAGCTGATAAACCTTATTGAAGAAAATTAATGTT600

图 2A

E I L E N P Y H E L A K L F G F E I P K
 GAAATCTTAGAAAAATCCTTATCATGAATGGCTAAACTATTTGGCTTTGAAATTCCAAAA660
 I P A P T V N P F E N L E H V I L P L F
 ATTCCAGCACCAACTGTCAATCCCTTTGAAAACCTTGAGCACGTTATTTCCACTCTTT720
 S D R G S K R H V P E K S G L N Q W N A
 TCAGACCGTGGCTCAAAGCGTCAATGACCAGAAAAAGCGGTCTAAACCAATGGAATGCT780
 L G R P R N P N E I Y I P I P K W I H N
 TTAGGTCGACCACGAAACCCCTAACGAGATTTATATACCAATTCCAAAATGGATTCATAAT840
 V F P T F F P A R D K P F Q L R L P D K
 GTATTCCCAACATTTTCCCAGCTCGTGATAAACCTTTTCAGTTACGCTTGCCAGACAAA900
 S L L S A K V C Q D N S K A L M S N P N
 TCGCTTTTATCAGCCAAGGTATGCCAAGACAATAGTAAAGCACTTATGTCTAATCCAAAT960
 S A L G E W L L R Q V M N L E E K E L L
 AGTGCTCTTGGAGAAATGGCTACTAAGACAAGTTATGAACTTAGAGGAAAAGAACTTCTA1020
 T Y E M L E R L N I D S V I V Y K H S E
 ACCATGAAATGCTGGAAGACTAAAATATTGACTCAGTAAATGTTTATAAACACAGCGAA1080
 Q H Y S I D F C E M G S Y D E F E N E N
 CAACATTACTCCAATTGATTTTGTGAAATGGTCTTATGATGAATTTGAAAATGAAAAC1140

K *
 AAATAA

图 2B

1146

TCAATGTTCCCAATATCTTTTAAATGTTCCAATCCCTTTCCCTCGTAAATATTTATTAACCTTCC60
 TTATCTCTTTTATTTTTCGAGTTTTTCTCCCAATATCCCGATTTACTTTTTTGGT120
 ATATTCCCCTGTTTTTTCACACGGCATGCCAGAAAACAAGAATCAAATGAATATGACIATTTTA180
 TATTTCTGTATTACTATATCTGGACTACCGTATAATTTCTTAACATTTTTTCGGAAATCTT240
 ATCCACGGTGCATAGTTCTTTAGTAACCTTATCTCTAAATTTTGAACGAGATTTGATTT300
 GCCTGCATGTTTTTCTTTCTTTTGAACCGTGTGAGTCTAGAAAGAGTCCCTCC360
 AAAGCCACAATAATTGATTTCTATAAACCAGGAAGCAAGCCCTCAAGCTTACCCCTCTT420
 AGTTCCTTTTTGGCTACTTATTTAATTTGTTTTTCAATTTTCAAAATTCATCAATAAGAACC480
 TTTCACAAAAATCAAATGGAGTAAATGTTGTTCCGCTGTGTTTTATAAACAATTTACTGAGTCAA540
 TATTTAGTCTTTCCAGCATTTCCATAGGTTAGAAGTCTTTTTTCCCTAAGTTCATAACTTT600
 GTCTTAGTAGCCATTTCCCAAGAGCACTAATTTGGATTAGACATAAGTGCTTTTACTATTGT660
 CTTGGCATACTTGGCTGATAAAAAGCGAATTTGTCTGGCAAGCGTAACCTGAAAAGGTTTAT720
 CACGAGCTGGGAAAAATGTTGGGAATACATATATGAATCCATTTTGGAAATGGTATATAAA780
 TCTCGTTAGGGTTTCGTGGTCCGACCTAAAGCATCCATTTGGTTTAGACCGCTTTTTTCTG840
 GTACATGACGCTTTGAGCCACGGTCTGAAAAGAGAGTGAAGAAATAACGTGCTCAAGGTTT900
 CAAAAGGATTGACAGTTGGTCTGGAATTTTGGAAATTTCAAAGCCAAATAGTTTAGCCA960
 ATTCAATGATAAGGATTTTCTAAGATTTCAACATTAATTTCTCAATAGGTTTATCAGTGA1020
 TAAAACGCTTATAAAGGGTCTTTAGTGACATTAAGCTGTAATTCGTGTAGACCGTCTT1080

图 3A

CAAAGGIGATTGATTTCTGTTGTTACTTACTTTCACATTTGTAATGAGGAGATTTCAA1140
CCAAAGTCCATTGGCTCTTCAAAAAATAAGAATTTCCCTGGCTTTCTGTTACACAGTGGT1200
ATATCATTTGAAATCAATACCATAATGTTCTTTTAGTAAATTCAAATTCCTCGTTACGGAGAG1260
AAGCAACCGTGTATTATAGCTCTTTGGAGATTTCCCACGATACAAGTCTGAGTCTTTAT1320
TGAATTCAGCTACTTTTGAAGAGTAGACCAATTACCATGAAGAAAAAGTCTTAATACCAA1380
TTCCGACACGATTTAATGAAGCGTCAGCAGAACAGTCTGACCCTCCCCAAGTTTTCAGCTC1440
CAAATGCTTCACAAAAAGCAATTTCCACATTTCCATTCCCTTGAGACCAAATAAGCGGAGTCACTTT1500
CAGAGAACAAAATTGGATAGCGAACCAAGTTGAGCGGAGCATTTGTTGATGTAGTGCAGT1560
TGATGGCTGGTTGATTTAGTATAGAACATTTATTTTCCCTCCTCTTTTATGCTTGTCAATTT1620
TTCTTTCAGACCCAAAAGGTAGTCAGCTGATACGTTCAATGTTTCAGCTATTTCTTTGAA1680
AGTGTCCAATGATGGAGTCTATTTTCACTTTCATATAGTGACCAAGTGCCTTAGTGAC1740
CCCGACTTTTTCAGCGAATTTGGCTGGGTAATAACCTACGAGCTTCTCTTGCATTTTGAAT1800
ACGATTTCCAAGGAAAGGTATCATTTTGCACCTCCAAGATTTGTTGTTTTCAGAGTATC1860
ACCAGAACCCCGAAAAATAGTCCAAAAGTTAGTAAACAGCAACAAAATAAAAATAAATAAG1920
TTGTTACTCTTAGCAAACTTGTACTAAAAATTTGATAAAGTTATTCATTTAATCCAGCT1980
CTTATGCTAAAAATTGCATTAGCGGACAAGCTTAATGTTTGC AAGGAGGTATAATTTTGAC2040
TTATCGAGTAGGTAGTATGTTTGTCTGGGATAGGTGGAACTTGTTTAGGGTTTATCCAAGC2100
TGGCGTAGGATTGTCTGGGCAAAATGAAATAGACAAAAATGCTTGTATTACTTATAGAAA2160
TTATTTGGGGATGCTTACTTACAAGAGGGTGACATTAACCTAATAGATAAAAACTCCAT2220

图 3B

ACCTGAACTGGACATTTTGAATGGAGGTTTTCCCTTGCCAAAGCCTTCTCTATAGCTGGCTA2280
TCGTAAAGGGTTTGAAGATGAAAGGGGAAACGTGTTCTTTCAAATATTAGAGGTAATTGGA2340
AGCACAAGAAAATGTTTATGGACACTTACCCCAAAGCAATAATGCTTGAGAAATGTAAGAA2400
CTTATTACACATGATAGAGGTAATACGTACAGAGTAATAAAAGAGGCTTTGGAAAGCCTT2460
TGGTTATACCGTAAAAGCTGAGGTTCTTAATTCAATGGAAATACGGTAAACGTGCCACAAAAA2520
CAGAGCGGGATTTATATTGTAGGTTTTCAAGATGAGAGCCCAAGCTGAAAGGTTTAGCTT2580
TCCAGACCCCAATTCCTTTAAACAAAATCAACTTAAATGATGTAAATTGACCCGAACTCGGAGAGT2640
TGATAAAAGATATTATTATGATGAAACCCTCTCAAATATTATGATAATGTTGCCGAGAAGCCAT2700
GGACAGTACAGATACAACCTTATCAAATAAGACGTAATATATGTTCCGAGAAAATAGAAAGCAA2760
TGTTTGTCCCTACACTGACAGCGAATATGGAACTGGAGGCAATAATGTTCCCTATTGTATT2820
AGACTTTGAAAAATAATAAGAAAACTAACACCCAGAAGAAATGCTTACTATTGCAAGGTTT2880
CCCAGCTGACTATCATTTTCCAGAAGGCAATGGCAAACACTCACAAAATAATAACAAGCTGG2940
TAACTCTGTTACGGTGCCAGTTATAAGAAGAATTGCCCACTAATATTATTAGCGTATTGAA3000
CATTGGAATGAATAATAAATCAAGAACAATGAATAATGCAATAGCTGAATAA

3049

图 3C

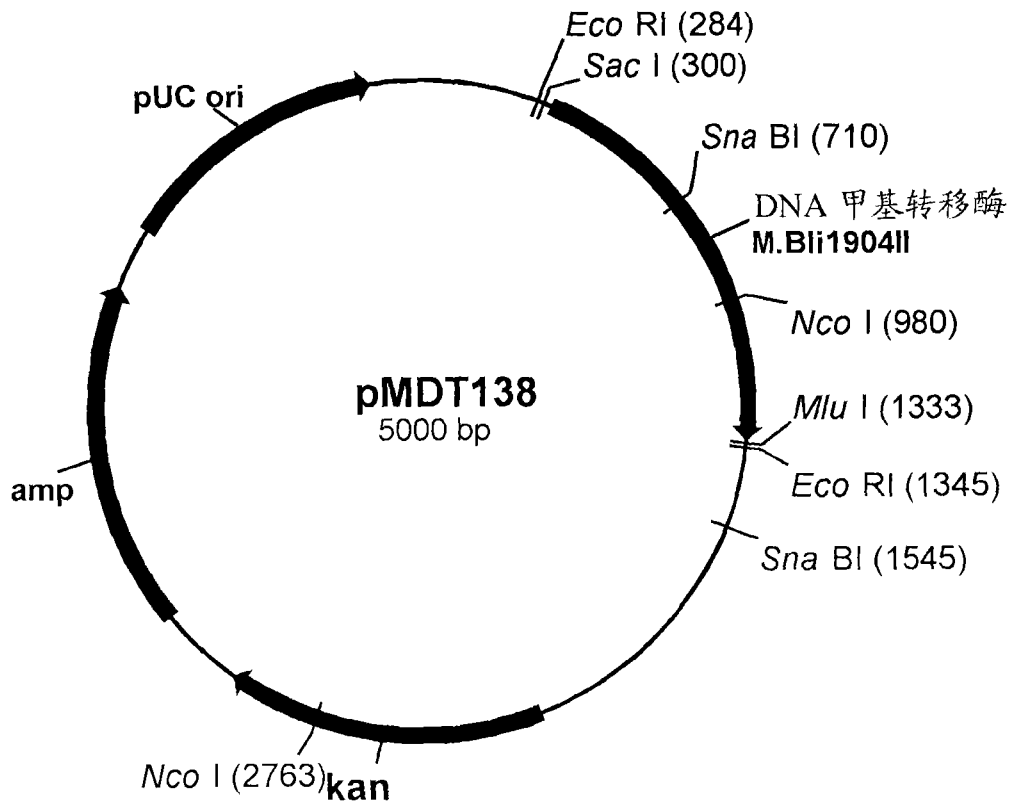


图 4

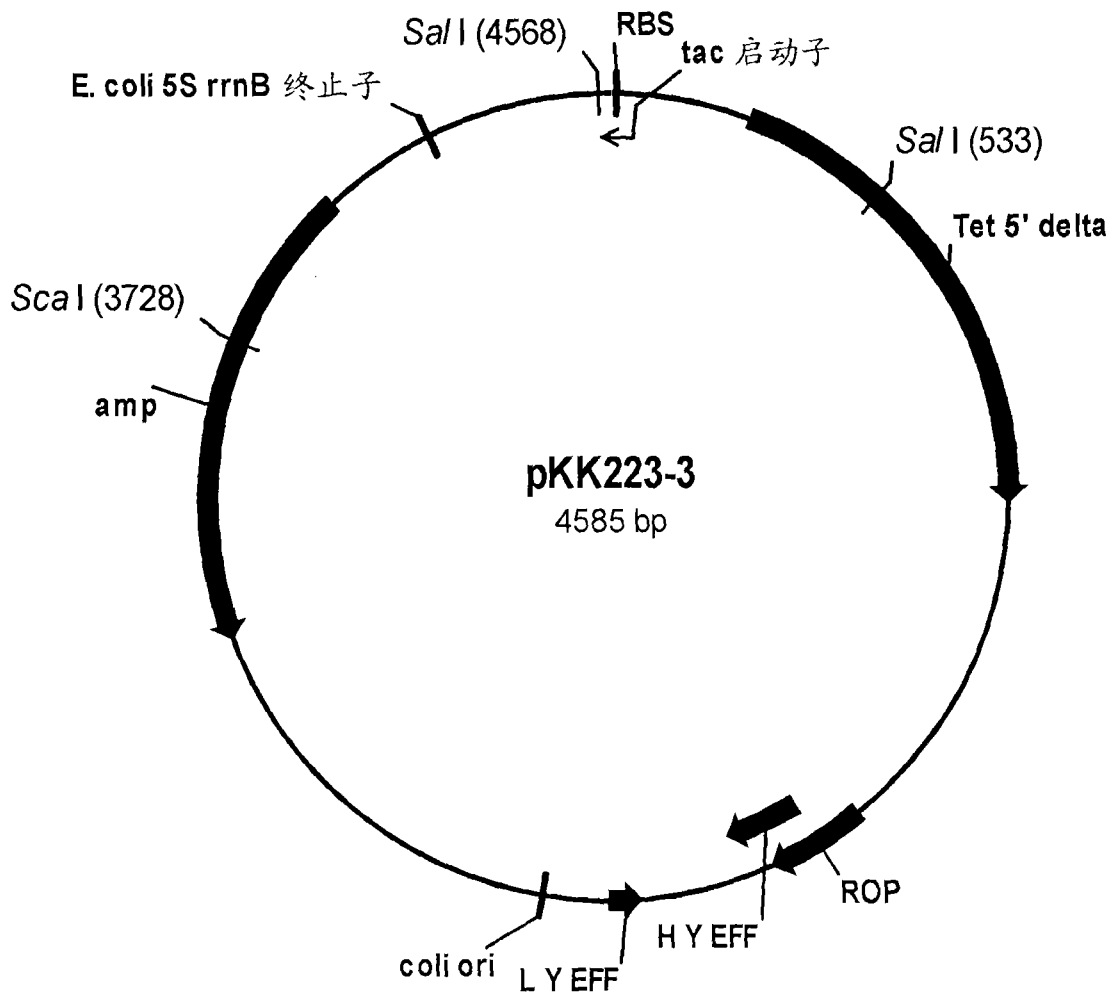


图 5

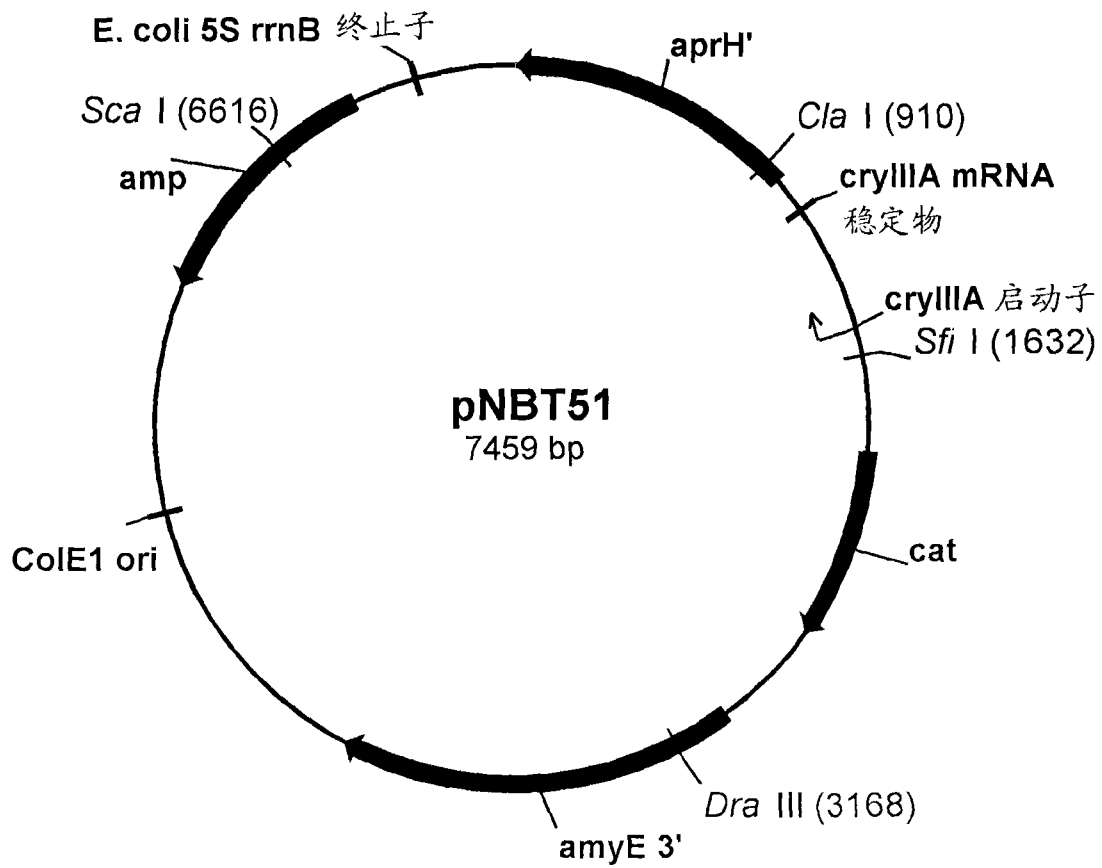


图 6

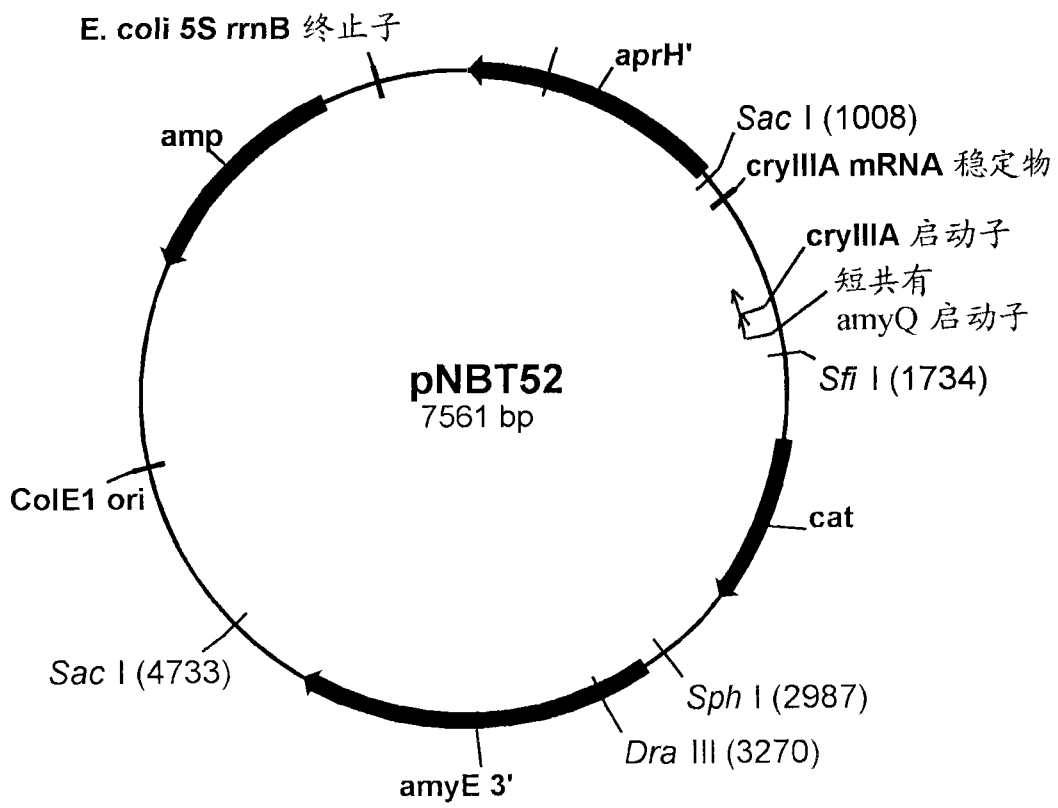


图 7

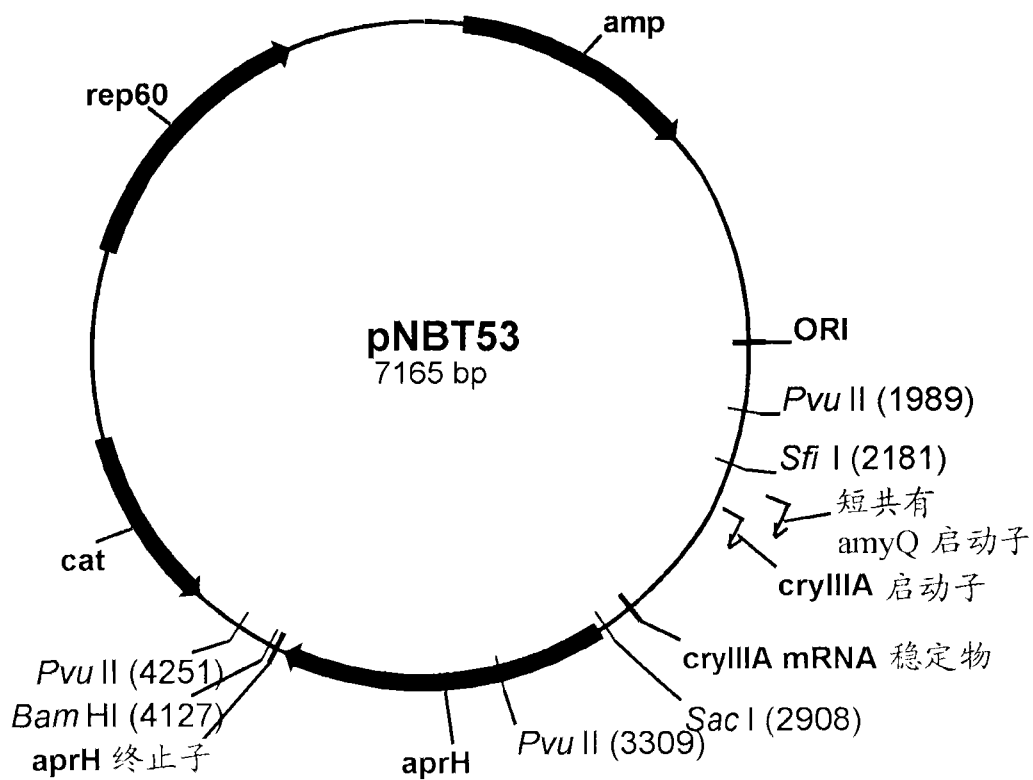


图 8

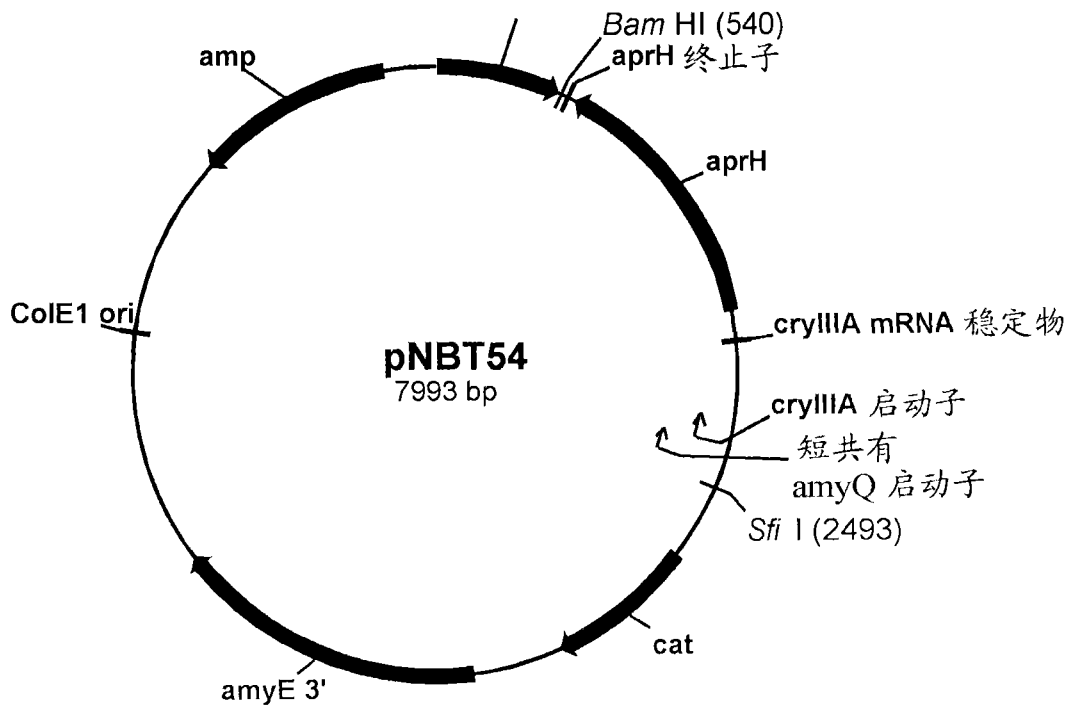


图 9

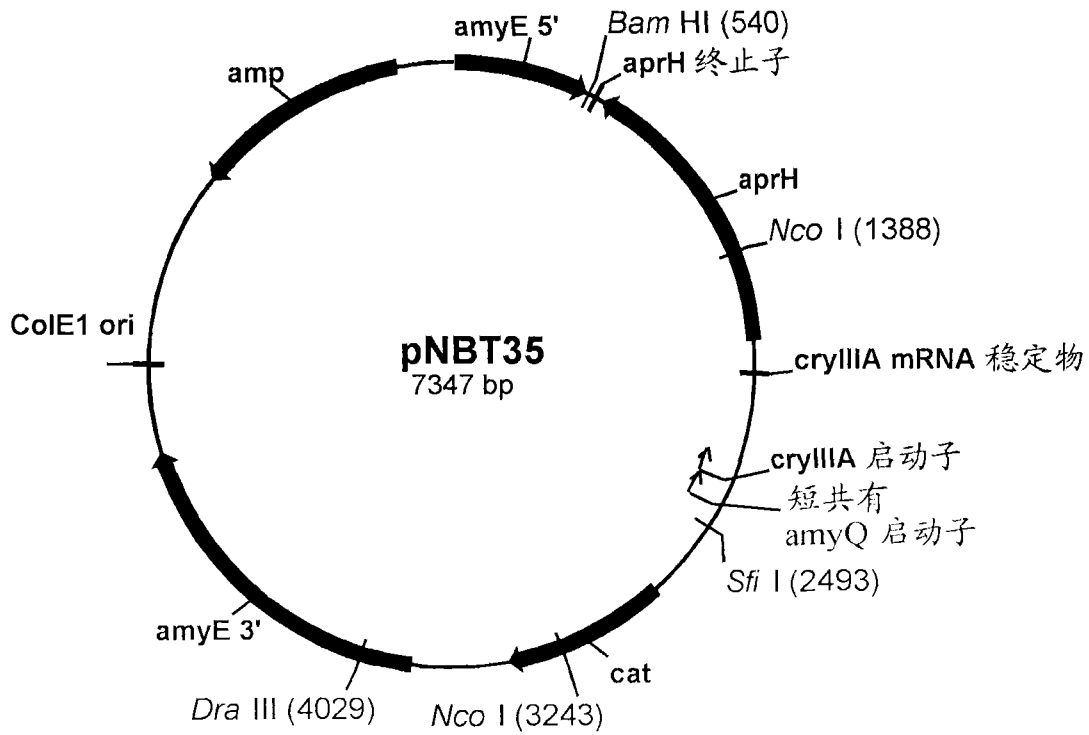


图 10

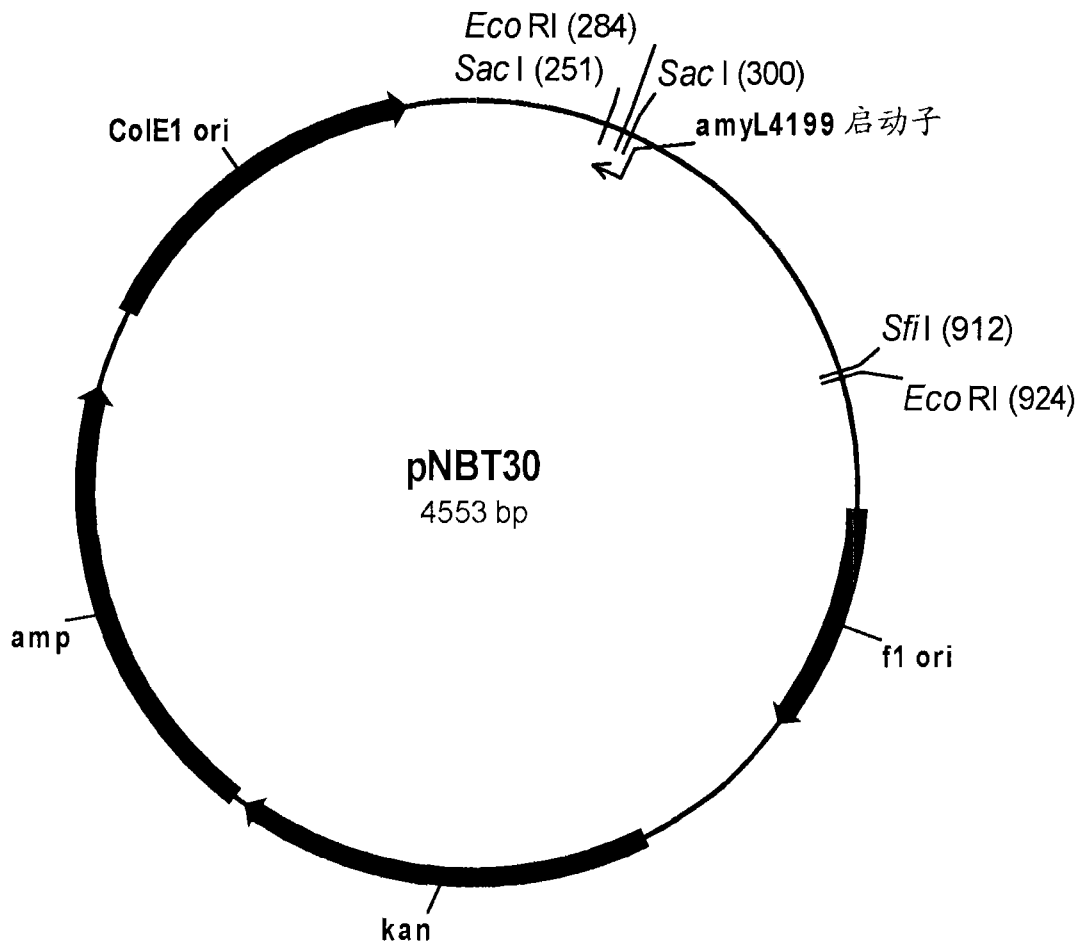


图 11

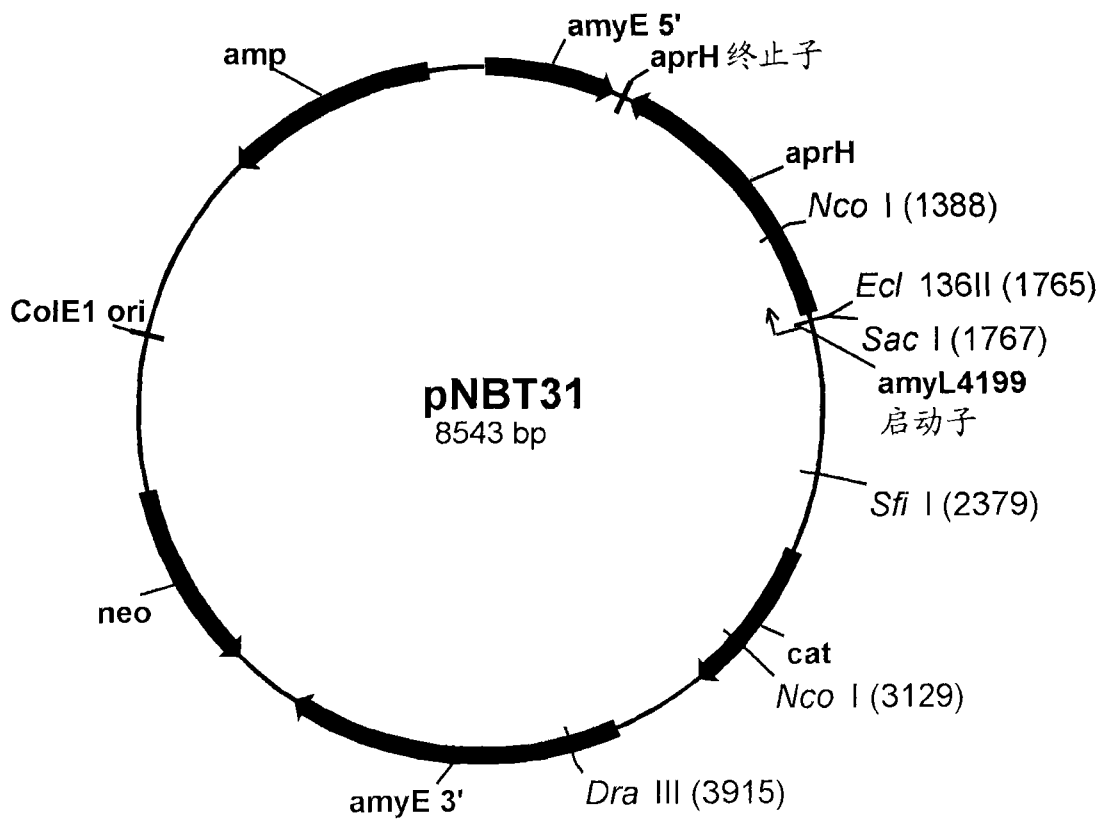


图 12

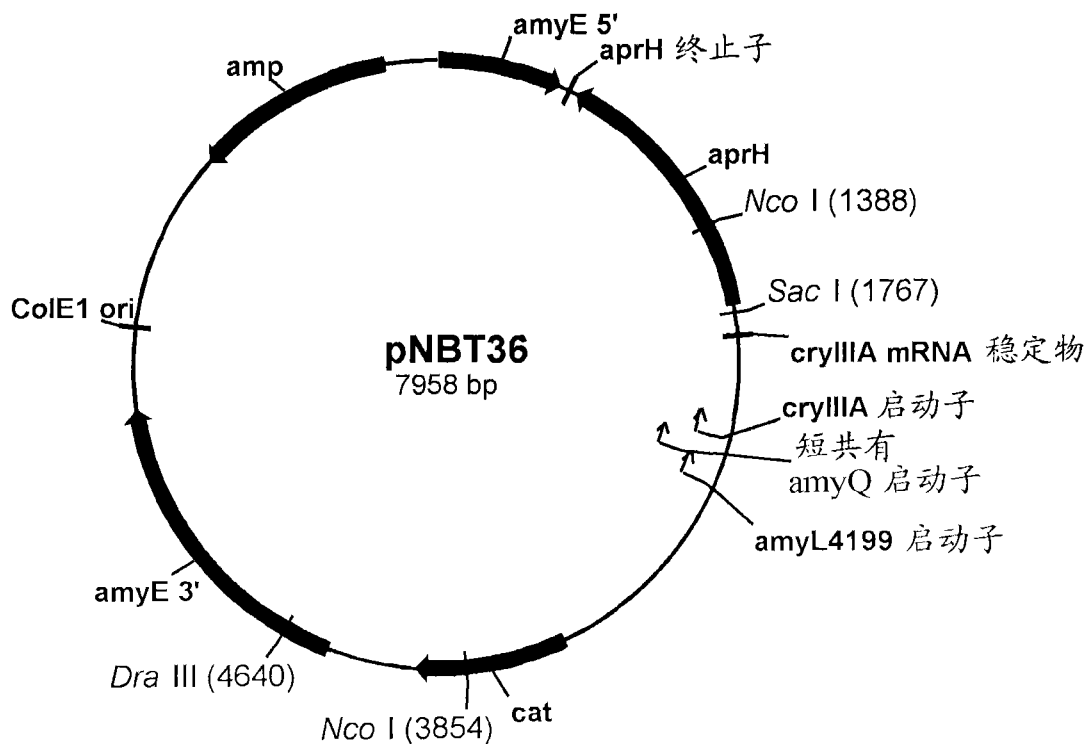


图 13

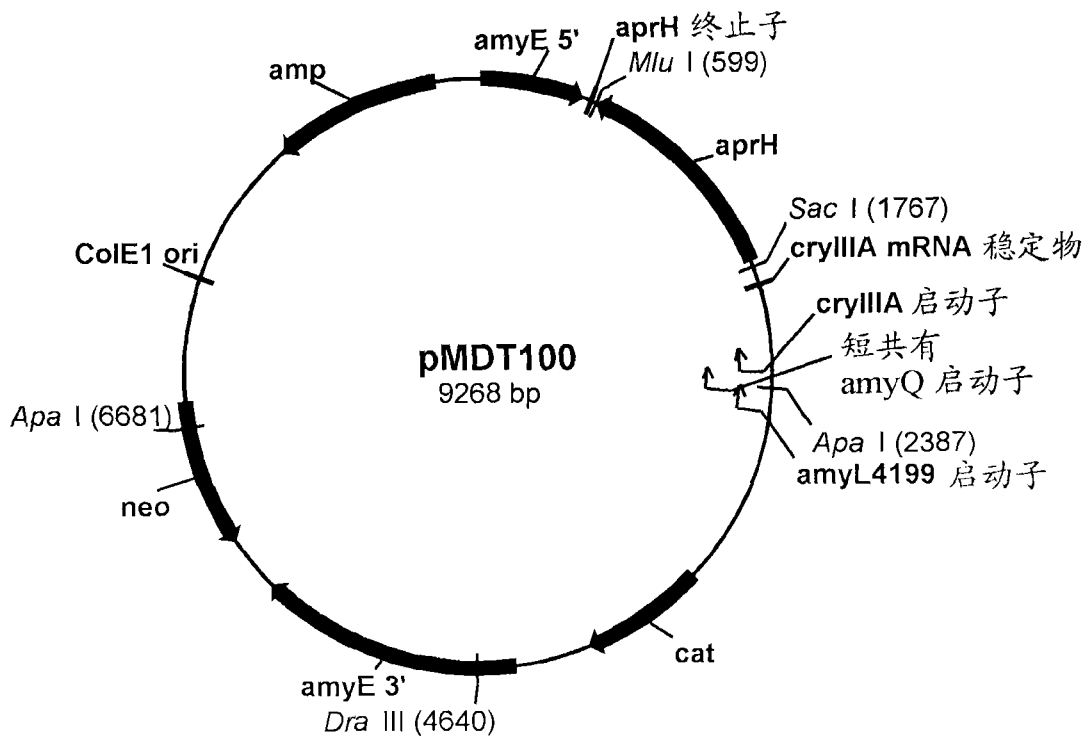


图 14

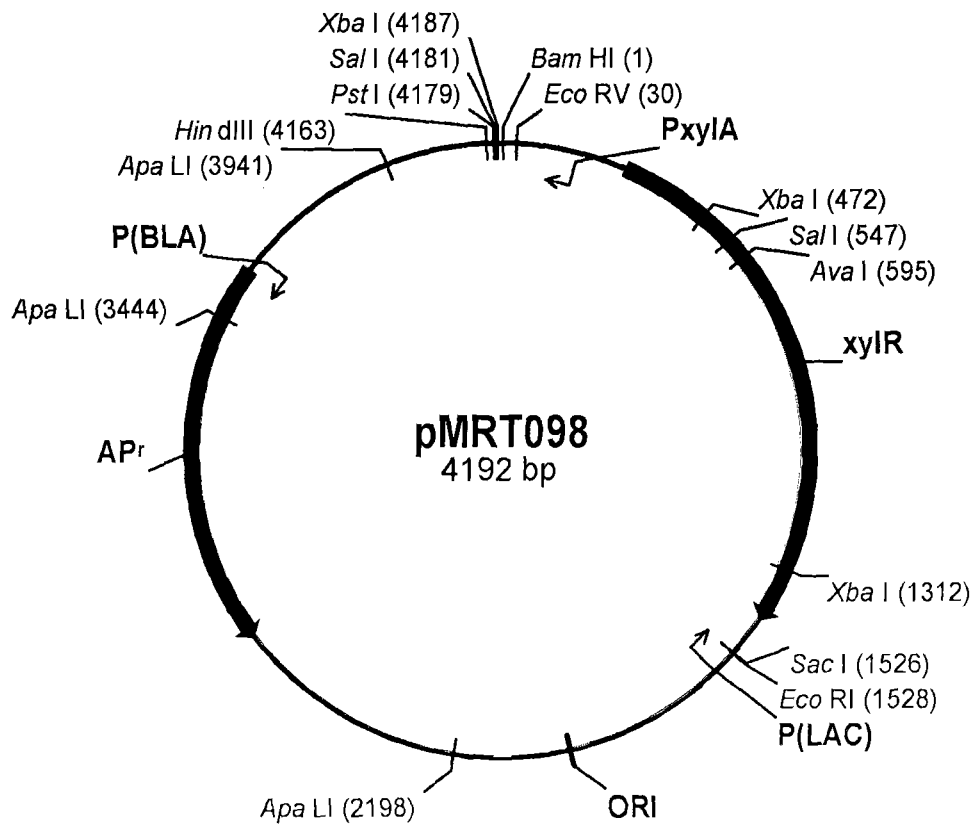


图 16

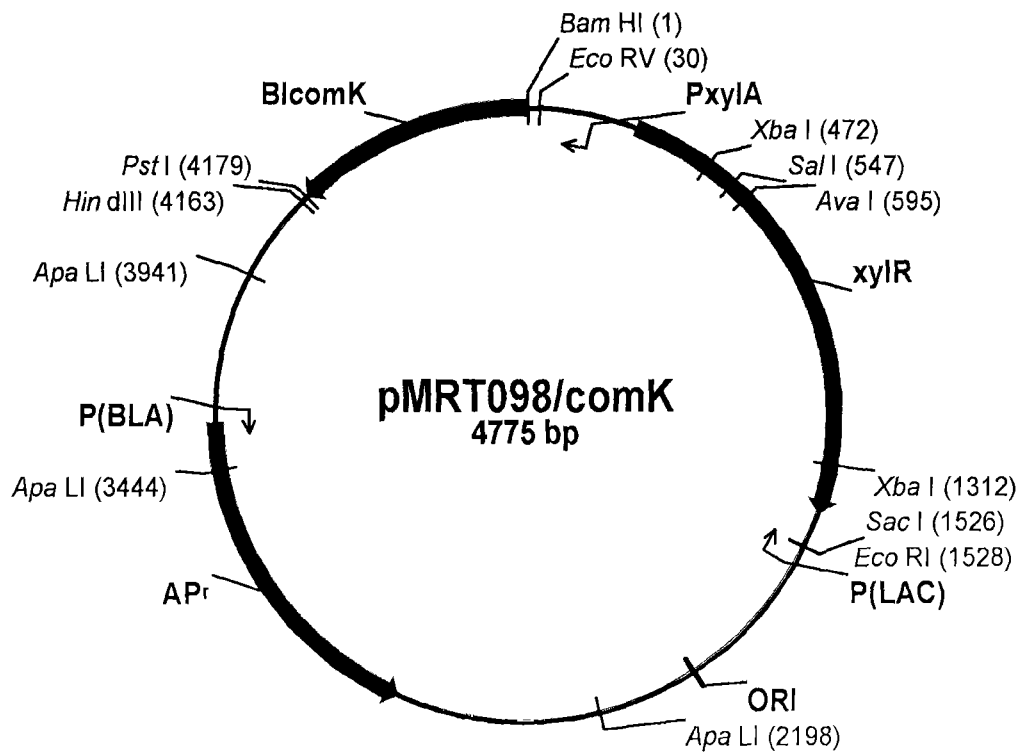


图 17

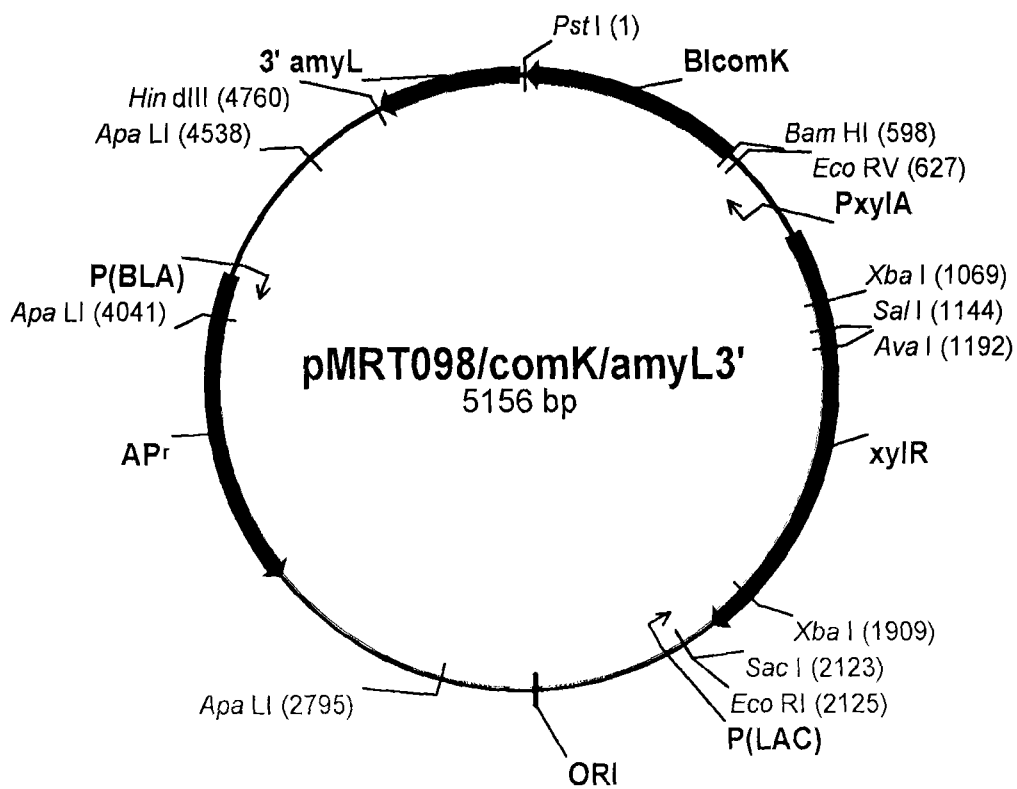


图 18

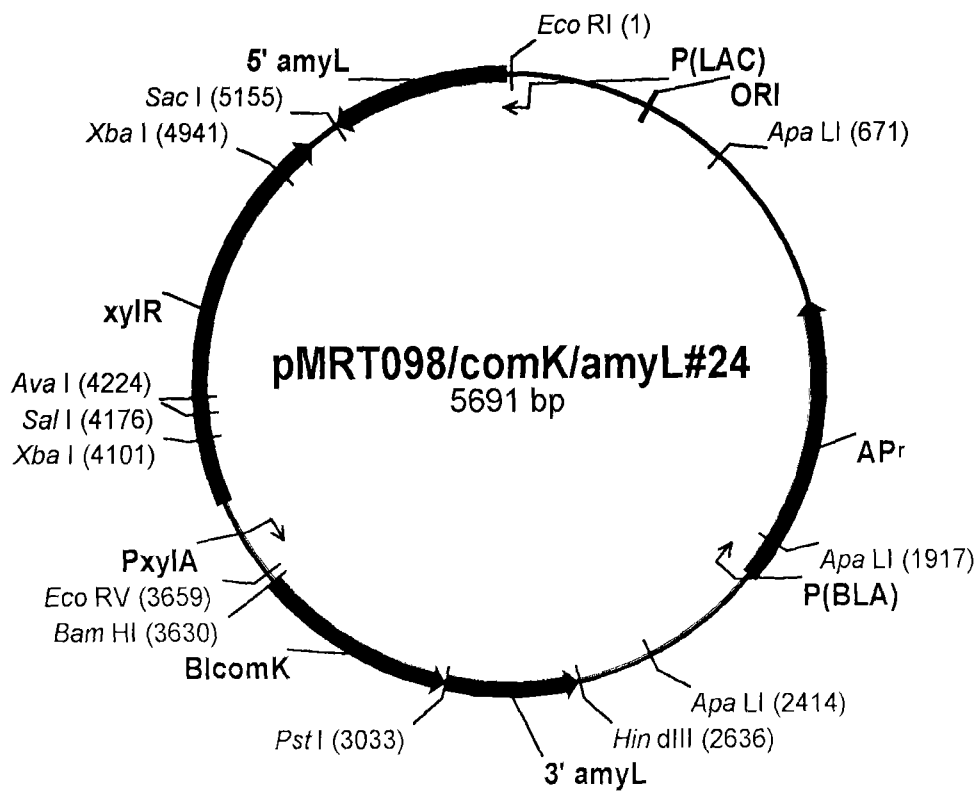


图 19

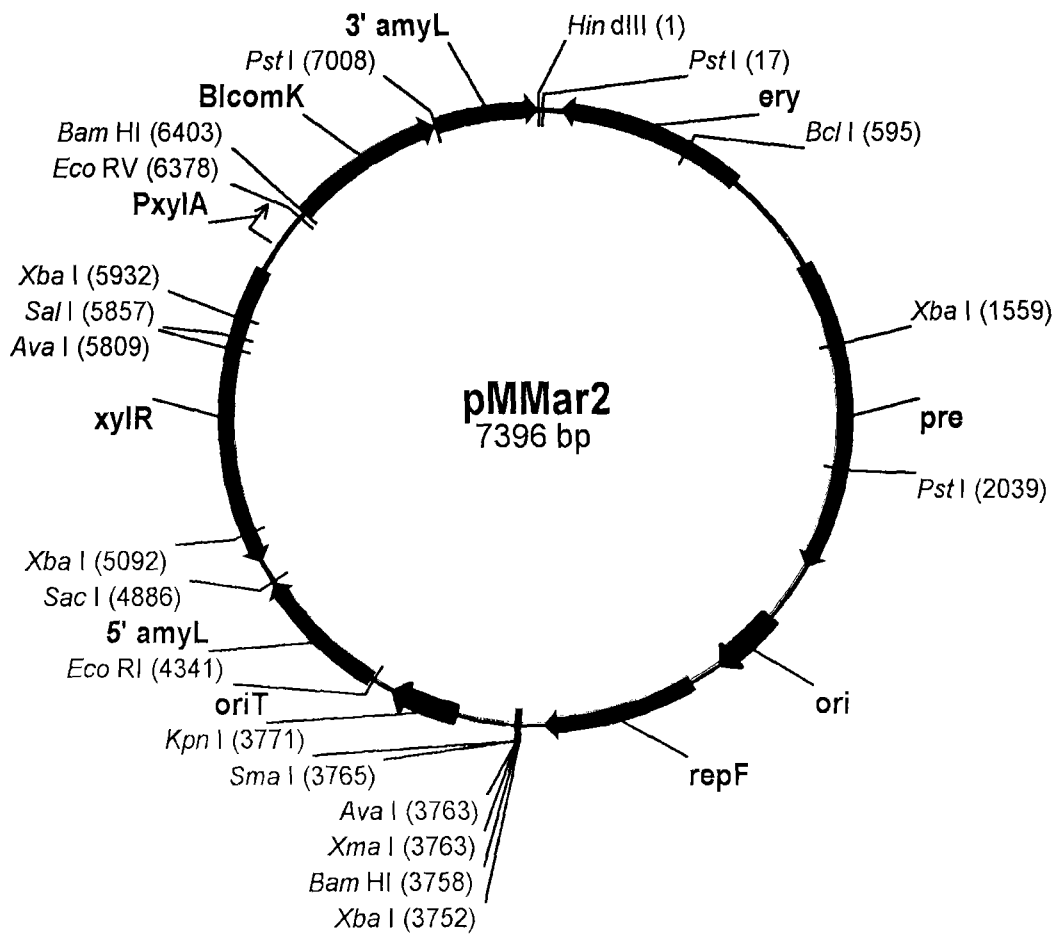


图 20

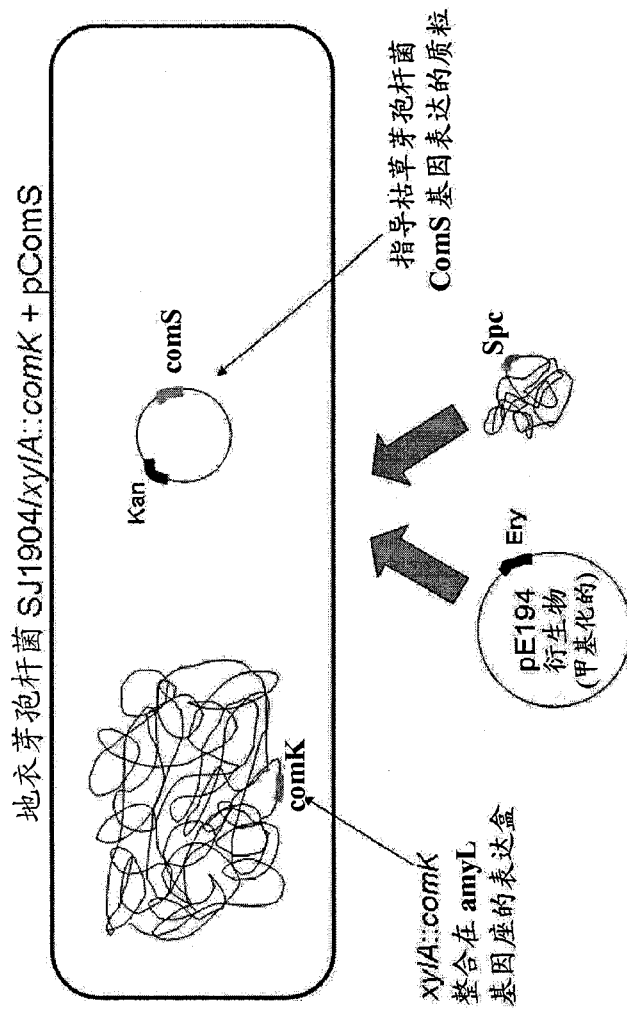


图 21