



(12)发明专利申请

(10)申请公布号 CN 109136114 A

(43)申请公布日 2019.01.04

(21)申请号 201810953356.5

C12N 15/80(2006.01)

(22)申请日 2012.08.23

C12R 1/645(2006.01)

(30)优先权数据

61/526,809 2011.08.24 US

(62)分案原申请数据

201280052277.8 2012.08.23

(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 D·亚韦 金绮明

(74)专利代理机构 北京尚诚知识产权代理有限公司 11322

代理人 顾小曼 周琴

(51)Int.Cl.

C12N 1/15(2006.01)

C12N 15/56(2006.01)

权利要求书8页 说明书81页

序列表74页 附图14页

(54)发明名称

在丝状真菌宿主细胞中产生多种重组多肽的方法

(57)摘要

本发明涉及构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法。本发明还涉及在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法。本发明还涉及表达具有生物活性的多种重组多肽的丝状真菌菌株。

1. 一种构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,其包括:

(a) 通过向丝状真菌菌株中导入第一串联构建体,通过靶向整合取代内源性第一基因,所述第一串联构建体包含 (i) 所述第一基因的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个第一可选择性标记物, (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸, (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和 (v) 所述第一基因的同源3' 区、其同源侧翼区或其组合;

(b) 通过向所述丝状真菌菌株中导入第二串联构建体,通过靶向整合取代内源性第二基因,所述第二串联构建体包含 (i) 所述第二基因的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个第二可选择性标记物, (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和 (v) 所述第二基因的同源3' 区、其同源侧翼区或其组合;或

(c) (a) 和 (b) 的组合。

2. 根据权利要求1所述的方法,其中所述第一基因是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因: (i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I, (ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I, (iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I,和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

并且其中所述第二基因是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因: (i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II, (ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II, (iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II,和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

3. 根据权利要求1或2所述的方法,其中一个或多个串联构建体还包含一个或多个可选

择性标记物的5'侧翼的第一同源重复和所述一个或多个可选择性标记物的3'侧翼的第二同源重复,其中所述第一同源重复和所述第二同源重复进行同源重组,以切除所述一个或多个可选择性标记物。

4. 一种丝状真菌菌株,其包含:

(a) 通过靶向整合用第一串联构建体取代的内源性第一基因,所述第一串联构建体包含(i)所述第一基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)所述第一基因的同源3'区、其同源侧翼区或其组合;

(b) 通过靶向整合用第二串联构建体取代的内源性第二基因,所述第二串联构建体包含(i)所述第二基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个可选择性标记物,(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)所述第二基因的同源3'区、其同源侧翼区或其组合;或

(c) (a)和(b)的组合。

5. 根据权利要求4所述的丝状真菌菌株,其中所述第一基因是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因:(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I,(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I,(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I,和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

并且其中所述第二基因是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II,(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II,(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II,和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

6. 根据权利要求4或5所述的丝状真菌菌株, 其中一个或多个串联构建体还包含所述一个或多个可选择性标记物的5'侧翼的第一同源重复和所述一个或多个可选择性标记物的3'侧翼的第二同源重复, 其中所述第一同源重复和所述第二同源重复进行同源重组, 以切除所述一个或多个可选择性标记物。

7. 一种在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法, 其包括: 在有益于产生多肽的条件下培养丝状真菌宿主细胞, 其中所述丝状真菌宿主细胞包含:

(a) 通过靶向整合用第一串联构建体取代的内源性第一基因, 所述第一串联构建体包含 (i) 所述第一基因的同源5'区、其同源侧翼区或其组合, (ii) 一个或多个可选择性标记物, (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸, (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸, 和 (v) 所述第一基因的同源3'区、其同源侧翼区或其组合;

(b) 通过靶向整合用第二串联构建体取代的内源性第二基因, 所述第二串联构建体包含 (i) 所述第二基因的同源5'区、其同源侧翼区或其组合, (ii) 一个或多个可选择性标记物, (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 所述第二基因的同源3'区、其同源侧翼区或其组合; 或

(c) (a) 和 (b) 的组合。

8. 根据权利要求7所述的方法, 其还包括回收所述多种重组多肽。

9. 根据权利要求7或8所述的方法, 其中所述第一基因是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因: (i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I, (ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I, (iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I, 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

并且其中所述第二基因是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因: (i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II, (ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II, (iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸

所编码的纤维二糖水解酶II,和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

10. 根据权利要求7~9中任一项所述的方法,其中一个或多个串联构建体还包含所述一个或多个可选择性标记物的5'侧翼的第一同源重复和所述一个或多个可选择性标记物的3'侧翼的第二同源重复,其中所述第一同源重复和所述第二同源重复进行同源重组,以切除所述一个或多个可选择性标记物。

11. 一种串联构建体,其包含:(i)基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个第一可选择性标记物;(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v)所述基因的同源3'区、其同源侧翼区或其组合。

12. 根据权利要求11所述的串联构建体,其中所述基因是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因:(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I,(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I,(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I,和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

或其中所述基因是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II,(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II,(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II,和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

13. 一种构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,其包括:

(a) 通过靶向整合向内源性第一基因座中插入第一串联构建体,所述第一串联构建体包含(i)所述第一基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽

的第一多核苷酸, (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸, 和 (v) 所述第一基因座的同源3' 区、其同源侧翼区或其组合;

(b) 通过靶向整合向内源性第二基因座中插入第二串联构建体, 所述第二串联构建体包含 (i) 所述第二基因座的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个第二可选择性标记物, (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 所述第二基因座的同源3' 区、其同源侧翼区或其组合; 或

(c) (a) 和 (b) 的组合。

14. 根据权利要求13所述的方法, 其中所述第一基因座是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因: (i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I, (ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I, (iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I, 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

并且其中所述第二基因座是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因: (i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II, (ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II, (iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II, 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

15. 根据权利要求13或14所述的方法, 其中一个或多个串联构建体还包含一个或多个可选择性标记物的5' 侧翼的第一同源重复和所述一个或多个可选择性标记物的3' 侧翼的第二同源重复, 其中所述第一同源重复和所述第二同源重复进行同源重组, 以切除所述一个或多个可选择性标记物。

16. 一种丝状真菌菌株, 其包含:

(a) 通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座, 所述第一串联构建体包含 (i) 所述第一基因座的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个可选

择性标记物, (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸, (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸, 和 (v) 所述第一基因座的同源3' 区、其同源侧翼区或其组合;

(b) 通过靶向整合经插入第二串联构建体而修饰的内源性第二基因座, 所述第二串联构建体包含 (i) 所述第二基因座的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个可选择性标记物; (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 所述第二基因座的同源3' 区、其同源侧翼区或其组合; 或

(c) (a) 和 (b) 的组合。

17. 根据权利要求16所述的丝状真菌菌株, 其中所述第一基因座是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因: (i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I, (ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I, (iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I, 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

并且其中所述第二基因座是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因: (i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II, (ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II, (iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II, 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

18. 根据权利要求16或17所述的丝状真菌菌株, 其中一个或多个串联构建体还包含所述一个或多个可选择性标记物的5' 侧翼的第一同源重复和所述一个或多个可选择性标记物的3' 侧翼的第二同源重复, 其中所述第一同源重复和所述第二同源重复进行同源重组, 以切除所述一个或多个可选择性标记物。

19. 一种在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法, 其包括: 在有益于产生多肽的条件下培养丝状真菌宿主细胞, 其中所述丝状真菌宿主细胞包含:

(a) 通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座,所述第一串联构建体包含(i)所述第一基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)所述第一基因座的同源3'区、其同源侧翼区或其组合;

(b) 通过靶向整合经插入第二串联构建体而修饰的内源性第二基因座,所述第二串联构建体包含(i)所述第二基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个可选择性标记物;(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)所述第二基因座的同源3'区、其同源侧翼区或其组合;或

(c) (a)和(b)的组合。

20. 根据权利要求19所述的方法,其还包括回收所述多种重组多肽。

21. 根据权利要求19或20所述的方法,其中所述第一基因座是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因:(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I,(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I,(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I,和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

并且其中所述第二基因座是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II,(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II,(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II,和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

22. 根据权利要求19~21中任一项所述的方法,其中一个或多个串联构建体还包含所述一个或多个可选择性标记物的5'侧翼的第一同源重复和所述一个或多个可选择性标记物的3'侧翼的第二同源重复,其中所述第一同源重复和所述第二同源重复进行同源重组,

以切除所述一个或多个可选择性标记物。

23. 一种串联构建体,其包含:(i) 基因座的同源5' 区、其同源侧翼区或其组合;(ii) 一个或多个第一可选择性标记物;(iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v) 所述基因座的同源3' 区、其同源侧翼区或其组合。

24. 根据权利要求23所述的串联构建体,所述基因座是编码选自以下的纤维二糖水解酶I的纤维二糖水解酶I基因:(i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I,(ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I,(iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I,和(iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I;

或其中所述基因座是编码选自以下的纤维二糖水解酶II的纤维二糖水解酶II基因:(i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II,(ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II,(iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II,和(iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

在丝状真菌宿主细胞中产生多种重组多肽的方法

[0001] 本申请是2014年4月24日提交到中华人民共和国国家知识产权局的发明名称为“在丝状真菌宿主细胞中产生多种重组多肽的方法”、申请号为“201280052277.8”的发明专利申请的分案申请。

[0002] 序列表的引用

[0003] 本申请含有计算机可读形式的序列表,其通过引用的方式合并入本文。

发明领域

[0004] 本发明涉及在丝状真菌宿主细胞中产生多种重组多肽的方法。

背景技术

[0005] 在丝状真菌宿主细胞中重组产生多肽可以为生产商业相关量的多肽提供更为合意的载体。多肽的重组产生通常通过构建表达盒来实现,其中在表达盒中,将编码多肽的DNA置于来自调控基因的启动子的表达控制下。通常通过质粒介导的转化将表达盒导入到宿主细胞中。然后通过表达盒中含有的启动子适当发挥作用所必需的诱导性条件下培养所转化的宿主细胞来实现多肽的生产。

[0006] 可以通过涉及原生质体形成、原生质体的转化和细胞壁的再生的方法以本身已知的方式用载体转化丝状真菌细胞。用两种或更多种载体单独或一起(共转化)转化丝状真菌宿主细胞的效率是非常低的,并受到可用的可选择性标记物的可得性的限制。

[0007] 本领域需要一些方法,通过将串联表达构建体定向至一个或多个(例如几个)特定基因组基因座来构建能够表达多种重组多肽的丝状真菌菌株,从而实现所需表达量的所有目标多肽。

[0008] 本发明提供在丝状真菌宿主细胞中产生多种重组多肽的改进方法。

发明内容

[0009] 本发明涉及构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,包括:

[0010] (a) 通过向丝状真菌菌株中导入第一串联构建体,通过靶向整合(targeted integration)来取代内源性第一基因,该第一串联构建体包含:(i) 第一基因的同源5'区、其同源侧翼区或其组合;(ii) 一个或多个(例如,几个)第一可选择性标记物;(iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v) 第一基因的同源3'区、其同源侧翼区或其组合;

[0011] (b) 通过向丝状真菌菌株中导入第二串联构建体,通过靶向整合来取代内源性第二基因,该第二串联构建体包含:(i) 第二基因的同源5'区、其同源侧翼区或其组合;(ii) 一个或多个(例如,几个)第二可选择性标记物;(iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸;(iv) 与第四启动子和第四终止子可操

作连接的编码具有生物活性的第四多肽的第四多核苷酸;和(v)第二基因的同源3'区、其同源侧翼区或其组合;或

[0012] (c) (a)和(b)的组合。

[0013] 本发明还涉及丝状真菌菌株,其包含:

[0014] (a)通过向丝状真菌菌株中导入第一串联构建体经靶向整合被取代的内源性第一基因,该第一串联构建体包含:(i)第一基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个(例如,几个)第一可选择性标记物;(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v)第一基因的同源3'区、其同源侧翼区或其组合;

[0015] (b)通过向丝状真菌菌株中导入第二串联构建体经靶向整合被取代的内源性第二基因,该第二串联构建体包含:(i)第二基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个(例如,几个)第二可选择性标记物;(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸;(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸;和(v)第二基因的同源3'区、其同源侧翼区或其组合;或

[0016] (c) (a)和(b)的组合。

[0017] 本发明还涉及在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法,其包括:

[0018] (A)在有益于产生多肽的条件下培养丝状真菌宿主细胞,其中该丝状真菌宿主细胞包含:(a)通过靶向整合用第一串联构建体取代的内源性第一基因,该第一串联构建体包含(i)第一基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个(例如,几个)第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)第一基因的同源3'区、其同源侧翼区或其组合;(b)通过靶向整合用第二串联构建体取代的内源性第二基因,该第二串联构建体包含(i)第二基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个(例如,几个)第二可选择性标记物,(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)第二基因的同源3'区、其同源侧翼区或其组合;或(c) (a)和(b)的组合;以及任选地

[0019] (B)回收多种重组多肽。

[0020] 本发明还涉及串联构建体,其包含:(i)基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个(例如,几个)可选择性标记物;(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v)基因的同源3'区、其同源侧翼区或其组合。

[0021] 本发明还涉及构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,其包括:

[0022] (a) 通过靶向整合向内源性第一基因座中插入第一串联构建体, 该第一串联构建体包含: (i) 第一基因座的同源5' 区、其同源侧翼区或其组合; (ii) 一个或多个第一可选择性标记物; (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸; (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸; 和 (v) 第一基因座的同源3' 区、其同源侧翼区或其组合;

[0023] (b) 通过靶向整合向内源性第二基因座中插入第二串联构建体, 该第二串联构建体包含: (i) 第二基因座的同源5' 区、其同源侧翼区或其组合; (ii) 一个或多个第二可选择性标记物; (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸; (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸; 和 (v) 第二基因座的同源3' 区、其同源侧翼区或其组合; 或

[0024] (c) (a) 和 (b) 的组合。

[0025] 本发明还涉及丝状真菌菌株, 其包含: (a) 通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座, 该第一串联构建体包含 (i) 第一基因座的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个第一可选择性标记物, (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸, (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸, 和 (v) 第一基因座的同源3' 区、其同源侧翼区或其组合; (b) 通过靶向整合经插入第二串联构建体而修饰的内源性第二基因座, 该第二串联构建体包含 (i) 第二基因座的同源5' 区、其同源侧翼区或其组合; (ii) 一个或多个第二可选择性标记物; (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 第二基因座的同源3' 区、其同源侧翼区或其组合; 或 (c) (a) 和 (b) 的组合。

[0026] 本发明还涉及在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法, 其包括:

[0027] (A) 在有益于产生多肽的条件下培养丝状真菌宿主细胞, 其中该丝状真菌宿主细胞包含: (a) 通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座, 该第一串联构建体包含 (i) 第一基因座的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个第一可选择性标记物, (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸, (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸, 和 (v) 第一基因座的同源3' 区、其同源侧翼区或其组合; (b) 通过靶向整合经插入第二串联构建体而修饰的内源性第二基因座, 该第二串联构建体包含 (i) 第二基因座的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个第二可选择性标记物, (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 第二基因座的同源3' 区、其同源侧翼区或其组合; 或 (c) (a) 和 (b) 的组合; 以及

[0028] (B) 回收多种重组多肽。

[0029] 本发明还涉及串联构建体, 其包含: (i) 基因座的同源5' 区、其同源侧翼区或其组合; (ii) 一个或多个可选择性标记物; (iii) 与第一启动子和第一终止子可操作连接的编码

具有生物活性的第一多肽的第一多核苷酸；(iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸；和 (v) 基因座的同源3' 区、其同源侧翼区或其组合。

附图说明

- [0030] 图1示出质粒pAG43的限制性酶切图谱。
- [0031] 图2示出质粒pSMai214的限制性酶切图谱。
- [0032] 图3示出质粒pDM287的限制性酶切图谱。
- [0033] 图4示出阳性转化体对于 β -葡萄糖苷酶活性的比较：pDM287的45 个转化子与pEJG107+pSMai214的45个转化子之间。
- [0034] 图5示出质粒pDM286的限制性酶切图谱。
- [0035] 图6示出质粒pDM290的限制性酶切图谱。
- [0036] 图7示出质粒pJfyS142的限制性酶切图谱。
- [0037] 图8示出质粒pJfyS144的限制性酶切图谱。
- [0038] 图9示出质粒pJfyS139的限制性酶切图谱。
- [0039] 图10示出质粒pQM18的限制性酶切图谱。
- [0040] 图11示出质粒pQM21的限制性酶切图谱。
- [0041] 图12示出质粒pAG121的限制性酶切图谱。
- [0042] 图13示出质粒pRRAB01的限制性酶切图谱。
- [0043] 图14示出质粒pDFng113-3的限制性酶切图谱。
- [0044] 图15示出质粒pAmFs074的限制性酶切图谱。
- [0045] 图16示出质粒pQM22的限制性酶切图谱。

具体实施方式

[0046] 定义

[0047] 乙酰木聚糖酯酶：术语“乙酰木聚糖酯酶”是指催化乙酰基从聚合木聚糖、乙酰化木糖、乙酰化葡萄糖、乙酸 α -萘酯和乙酸对硝基苯酯水解的羧酸酯酶 (E.C.3.1.1.72)。出于本发明的目的，乙酰木聚糖酯酶的活性是使用在含0.01% TWEENTM 20 (聚氧乙烯失水山梨醇单月桂酸酯) 的50mM乙酸钠pH 5.0中的0.5mM乙酸对硝基苯酯作为底物而确定的。一个单位的乙酰木聚糖酯酶被定义为能够在pH 5、25℃下每分钟释放1微摩尔(μ mole) 对硝基苯酚阴离子的酶量。

[0048] 等位变体 (allelic variant)：术语“等位变体”是指占据同一染色体基因座的基因的两种或更多种 (例如，几种) 可选形式中的任意种。等位变化通过突变而自然发生，并可能引起种群内的多态性。基因突变可以是沉默的 (在所编码的多肽中没有变化) 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所编码的多肽。

[0049] α -L-阿拉伯呋喃糖苷酶：术语“ α -L-阿拉伯呋喃糖苷酶”是指催化 α -L-阿拉伯糖苷中末端非还原性 α -L-阿拉伯呋喃糖苷残基的水解的 α -L-阿拉伯呋喃糖苷阿拉伯呋喃水解酶 (alpha-L-arabinofuranoside arabinofuranohydrolase) (E.C.3.2.1.55)。该酶作用于 α -L-阿拉伯呋喃糖苷、含有 (1,3)-和/或 (1,5)-键的 α -L-阿拉伯聚糖、阿拉伯木聚糖、和阿

拉伯半乳聚糖。 α -L-阿拉伯呋喃糖苷酶也称为阿拉伯糖苷酶、 α -阿拉伯糖苷酶、 α -L-阿拉伯糖苷酶、 α -阿拉伯呋喃糖苷酶、多糖 α -L-阿拉伯呋喃糖苷酶、 α -L-阿拉伯呋喃糖苷水解酶、L-阿拉伯糖苷酶、或 α -L-阿拉伯聚糖酶。出于本发明的目的， α -L-阿拉伯呋喃糖苷酶的活性是使用总体积200 μ l的100mM乙酸钠pH 5中的每ml 5mg中等粘度小麦阿拉伯木聚糖(Megazyme International Ireland, Ltd., Bray, Co. Wicklow, 爱尔兰)在40℃下30分钟,随后通过AMINEX® HPX-87H柱层析(Bio-Rad Laboratories, Inc., Hercules, CA, 美国)的阿拉伯糖分析而确定的。

[0050] α -葡糖醛酸糖苷酶:术语“ α -葡糖醛酸糖苷酶”是指催化 α -D-葡糖苷酸水解为D-葡糖醛酸和醇的 α -D-葡糖苷酸葡糖醛酸水解酶(α -D-glucosiduronate glucuronohydrolase)(E.C.3.2.1.139)。出于本发明的目的， α -葡糖醛酸糖苷酶活性是根据de Vries, 1998, J. Bacteriol. 180:243-249而确定的。一个单位的 α -葡糖醛酸糖苷酶等于能够在pH 5、40℃下每分钟释放1微摩尔的葡糖醛酸或4-O-甲基葡糖醛酸的酶量。

[0051] 天冬氨酸蛋白酶:术语“天冬氨酸蛋白酶”是指使用天冬氨酸残基来催化肽和蛋白中肽键的水解的蛋白酶。天冬氨酸蛋白酶是使用天冬氨酸残基来催化它们的肽底物的水解的蛋白酶家族。一般情况下，它们在活性位点中具有两个高度保守的天冬氨酸盐(酯)，且在酸性pH下活性最佳(Szecsni, 1992, Scand. J. Clin. Lab. Invest. Suppl. 210:5-22)。出于本发明的目的，天冬氨酸蛋白酶活性根据Aikawa等人, 2001, J. Biochem. 129:791-794描述的过程来确定。

[0052] β -葡糖苷酶:术语“ β -葡糖苷酶”是指催化末端非还原性 β -D-葡萄糖残基的水解并释放 β -D-葡萄糖的 β -D-葡糖苷葡萄糖水解酶(β -D-glucoside glucohydrolase)(E.C.3.2.1.21)。出于本发明的目的， β -葡糖苷酶活性是根据Venturi等人, 2002, Extracellular β -D-glucosidase from Chaetomium thermophilum var. coprophilum: production, purification and some biochemical properties, J. Basic Microbiol. 42:55-66的过程使用对硝基苯基- β -D-吡喃葡萄糖苷作为底物而确定的。一个单位的 β -葡糖苷酶被定义为在25℃、pH 4.8下在含0.01% TWEEN® 20的50mM柠檬酸钠中从作为底物的1mM对硝基苯基- β -D-吡喃葡萄糖苷中每分钟生产1.0微摩尔对硝基苯酚阴离子。

[0053] β -木糖苷酶:术语“ β -木糖苷酶”是指催化短 β -(4)-低聚木糖的外水解以从非还原性末端去除连续的D-木糖残基的 β -D-木糖苷木糖水解酶(EC 3.2.1.37)。出于本发明的目的，一个单位的 β -木糖苷酶被定义为在40℃、pH 5下在含0.01% TWEEN® 20的100mM柠檬酸钠中从作为底物的1mM对硝基苯基- β -D-木糖苷中每分钟生产1.0微摩尔对硝基苯酚阴离子。

[0054] cDNA:术语“cDNA”是指可以通过反转录从成熟的、剪接的从真核或原核细胞获得的mRNA分子制备而来的DNA分子。cDNA缺少可以存在于对应基因组DNA中的内含子序列。最初的初级RNA转录体是在呈现为成熟剪接的mRNA前的经由一系列步骤包括剪接而加工的mRNA的前体。

[0055] 纤维二糖水解酶:术语“纤维二糖水解酶”是指催化纤维素、纤维寡糖、或任何含有 β -1,4-连接的葡萄糖的聚合物中1,4- β -D-糖苷键的水解以从链的还原末端或非还原末端释放纤维二糖的1,4- β -D-葡聚糖纤维二糖水解酶(E.C.3.2.1.91和E.C.3.2.1.176)

(Teeri,1997,Crystalline cellulose degradation:New insight into the function of cellobiohydrolases, Trends in Biotechnology 15:160-167;Teeri等人,1998, Trichoderma reesei cellobiohydrolases:why so efficient on crystalline cellulose? Biochem.Soc.Trans.26:173-178)。纤维二糖水解酶活性根据Lever等人, 1972,Anal.Biochem.47:273-279;van Tilbeurgh等人,1982,FEBS Letters, 149:152-156;van Tilbeurgh and Claeysens,1985,FEBS Letters,187: 283-288;以及Tomme等人,1988,Eur.J.Biochem.170:575-581描述的过程来确定。在本发明中,Tomme等人的方法可被用于确定纤维二糖水解酶的活性。

[0056] 纤维分解酶或纤维素酶:术语“纤维分解酶”或“纤维素酶”是指一种或多种(例如,几种)水解纤维素材料的酶。这样的酶包括内切葡聚糖酶、纤维二糖水解酶、 β -葡萄糖苷酶或其组合。测量纤维分解活性的两种基本方法包括:(1)测量总纤维分解活性,和(2)测量单独的纤维分解活性(内切葡聚糖酶、纤维二糖水解酶和 β -葡萄糖苷酶),如 Zhang等人,Outlook for cellulase improvement:Screening and selection strategies,2006, Biotechnology Advances 24:452-481所综述的。总纤维分解活性通常是使用不溶性底物来测量的,该底物包括WhatmanNo.1 滤纸、微晶纤维素、细菌纤维素、藻类纤维素、棉花、经预处理的木质纤维素等。最常见的总纤维分解活性检定是使用WhatmanNo.1滤纸作为底物的滤纸检定。该检定是由国际纯粹与应用化学联合会(International Union of Pure and Applied Chemistry,IUPAC)确立的(Ghose,1987,Measurement of cellulase activities,Pure Appl.Chem.59: 257-68)。

[0057] 出于本发明的目的,纤维分解酶活性通过测量在下述条件下由纤维素分解酶进行的纤维素材料的水解的增加而加以确定:1~50mg纤维分解酶蛋白/g PCS中纤维素(或其他经预处理的纤维素材料),在合适的温度例如50℃、55℃或60℃下3~7天,与未添加纤维分解酶蛋白的对照水解相比较。通常条件为:1ml反应物,经洗涤或未洗涤的PCS,5%不溶性固体,50mM乙酸钠pH 5,1mM $MnSO_4$,50℃、55℃或 60℃,72小时,通过AMINEX®HPX-87H柱(Bio-Rad Laboratories,Inc., Hercules,CA,美国)进行糖分析。

[0058] 纤维素材料:术语“纤维素材料”是指包含纤维素的任何材料。在生物质的初生细胞壁中的主要多糖是纤维素,其次丰富的是半纤维素,而第三是果胶。在细胞停止生长后产生的次生细胞壁也含有多糖,且次生细胞壁通过与半纤维素共价交联的聚合木质素而进行加强。纤维素是脱水纤维二糖的均聚物,并且因此是直链 β -(1-4)-D-葡聚糖;而半纤维素包括多种化合物,例如木聚糖、木葡聚糖(xyloglucan)、阿拉伯木聚糖和甘露聚糖,处于具有系列取代基的复杂分支结构。尽管通常是多晶形的,存在于植物组织中的纤维素主要是平行葡聚糖链的不溶晶体基质。半纤维素通常与纤维素以及其他半纤维素以氢键相连,这帮助稳定细胞壁基质。

[0059] 纤维素通常见于例如植物的茎、叶、壳、皮和穗轴,或树的叶、枝和木材。纤维素材料可以是,但不限于,农业残余物、草本材料(包括能量作物)、城市固体废物、纸浆与造纸厂残余物、废纸和木材(包括林业残余物)(参见,例如,Wiselogel等,1995,于Handbook on Bioethanol(Charles E.Wyman编辑),pp.105-118,Taylor&Francis, Washington D.C.; Wyman,1994,Bioresource Technology 50:3-16;Lynd, 1990,Applied Biochemistry and Biotechnology 24/25:695-719;Mosier等人,1999,Recent Progress in Bioconversion

of Lignocellulosics, 于 Advances in Biochemical Engineering/Biotechnology, T.Scheper 主编, Volume 65, pp.23-40, Springer-Verlag, New York)。在本文中应理解的是,纤维素可以是木质纤维素的形式,即在混合基质中包含木质素、纤维素和半纤维素的植物细胞壁材料。在优选的方面,纤维素材料是任何生物质材料。在另一个优选的方面,纤维素材料是木质纤维素,其包含纤维素、半纤维素和木质素。

[0060] 在一个方面,纤维素材料是农业残余物。在另一个方面,纤维素材料是草本材料(包括能量作物)。在另一个方面,纤维素材料是城市固体废物。在另一个方面,纤维素材料是纸浆和造纸厂残余物。在另一个方面,纤维素材料是废纸。在另一个方面,纤维素材料是木材(包括林业残余物)。

[0061] 在另一个方面,纤维素材料是芦竹。在另一个方面,纤维素材料是甘蔗渣。在另一个方面,纤维素材料是竹。在另一个方面,纤维素材料是玉米穗轴。在另一个方面,纤维素材料是玉米纤维。在另一个方面,纤维素材料是玉米秸秆。在另一个方面,纤维素材料是芒草。在另一个方面,纤维素材料是橘皮。在另一个方面,纤维素材料是稻秆。在另一个方面,纤维素材料是柳枝稷。在另一个方面,纤维素材料是麦秆。

[0062] 在另一个方面,纤维素材料是山杨。在另一个方面,纤维素材料是桉树。在另一个方面,纤维素材料是冷杉。在另一个方面,纤维素材料是松树。在另一个方面,纤维素材料是白杨。在另一个方面,纤维素材料是云杉。在另一个方面,纤维素材料是柳树。

[0063] 在另一个方面,纤维素材料是藻类纤维素。在另一个方面,纤维素材料是细菌纤维素。在另一个方面,纤维素材料是棉绒。在另一个方面,纤维素材料是滤纸。在另一个方面,纤维素材料是微晶纤维素。在另一个方面,纤维素材料是磷酸处理的纤维素。

[0064] 在另一个方面,纤维素材料是水生生物质。如本文所使用的,术语“水生生物质”是指在水生环境中由光合作用过程产生的生物质。水生生物质可以为藻类、挺水植物、浮叶植物或沉水植物。

[0065] 如本文所述,纤维素材料可以使用本领域已知的常规方法按原样使用或进行预处理。在优选的方面,纤维素材料被预处理。

[0066] 编码序列:术语“编码序列”是指直接指定多肽的氨基酸序列的多核苷酸。编码序列的边界一般由开放阅读框确定,开放阅读框以起始密码子如ATG、GTG或TTG开始并以终止密码子如TAA、TAG或 TGA结束。编码序列可以是基因组DNA、cDNA、合成DNA或其组合。

[0067] 控制序列:术语“控制序列”是指编码多肽的多核苷酸在表达时所必需的核酸序列。各个控制序列对于编码多肽的多核苷酸可以是内生的(即,来自相同基因)或外来的(即,来自不同基因)或彼此为内生或外来的。这些控制序列包括但不限于前导序列、多腺苷酸化序列、前肽序列、启动子、信号肽序列和转录终止子。在最低限度,控制序列包括启动子、以及转录和翻译终止信号。控制序列可以设有接头,以用于导入特定的促进控制序列与编码多肽的多核苷酸编码区进行连接的限制性酶切位点。

[0068] 异位整合:术语“异位整合”是指向微生物的基因组在非靶位点或除其通常的染色体基因座以外的位点插入核酸,即,随机整合。

[0069] 内切葡聚糖酶:术语“内切葡聚糖酶”是指催化纤维素、纤维素衍生物(例如羧甲基纤维素和羟乙基纤维素)、地衣多糖中的1,4- β -D-糖苷键的内水解(endohydrolysis),催化混合的 β -1,3-葡聚糖例如谷类 β -D-葡聚糖或木葡聚糖以及含有纤维素组分的其他植物

材料中的 β -1,4 键的内水解的内切-1,4-(1,3;1,4)- β -D-葡聚糖4-葡聚糖水解酶 (endo-1,4-(1,3;1,4)-beta-D-glucan 4-glucanohydrolase) (E.C.3.2.1.4)。内切葡聚糖酶的活性可以通过测量底物粘度的减少或由还原糖检定 (Zhang等人,2006,Biotechnology Advances 24:452-481) 确定的还原端的增加来确定。出于本发明的目的,内切葡聚糖酶的活性根据Ghose, 1987,Pure and Appl.Chem.59:257-268的方法,在pH 5、40℃下使用羧甲基纤维素 (CMC) 作为底物而确定。

[0070] 表达:术语“表达”包括在多肽生产中涉及的任何步骤,包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

[0071] 表达载体:术语“表达载体”是指包含编码多肽的多核苷酸并与提供用于其表达的控制序列可操作地连接的线性或环状DNA分子。

[0072] 糖苷水解酶61家族:术语“糖苷水解酶61家族”或“GH61家族”或“GH61”是指根据Henrissat B.,1991,A classification of glycosyl hydrolases based on amino-acid sequence similarities,Biochem.J.280: 309-316及Henrissat B.和Bairoch A.,1996, Updating the sequence-based classification of glycosyl hydrolases, Biochem.J.316:695-696而落入糖苷水解酶61家族的多肽。原先,该家族中的酶基于一个家族成员中测量的非常弱的内切-1,4- β -D-葡聚糖酶活性而归类为糖苷水解酶家族。这些酶的结构和作用模式是非公认的,且它们无法被视为真正的糖苷酶。然而,基于它们在与纤维素酶或纤维素酶的混合物一同使用时增强木质纤维素分解的能力,它们被保留在CAZy分类中。

[0073] 阿魏酸酯酶:术语“阿魏酸酯酶”是指催化4-羟基-3-甲氧基肉桂酰(阿魏酰)基团从酯化的糖(在天然生物质底物中通常为阿拉伯糖)中水解以产生阿魏酸酯(盐)(4-羟基-3-甲氧基肉桂酸酯(盐))的4-羟基-3-甲氧基肉桂酰-糖水解酶(EC 3.1.1.73)。阿魏酸酯酶也被称为阿魏酸酯酶、羟基肉桂酰基酯酶、FAE-III、肉桂酸酯水解酶、FAEA、cinnAE、FAE-I或FAE-II。出于本发明的目的,阿魏酸酯酶的活性是使用50mM乙酸钠pH 5.0中的0.5mM阿魏酸对硝基苯酯作为底物而确定的。一个单位的阿魏酸酯酶等于能够在pH 5、25℃下每分钟释放1 微摩尔对硝基苯酚阴离子的酶量。

[0074] 侧翼:术语“侧翼”是指在特定DNA序列、基因座或基因的任一侧延伸的DNA序列。侧翼DNA紧邻将要整合到丝状真菌细胞的基因组中的另一DNA序列、基因座或基因。

[0075] 片段:术语“片段”是指在成熟多肽域的氨基端和/或羧基端缺少一个或多个(如,几个)氨基酸的多肽;其中该片段具有酶活性。在一方面,片段含有酶的成熟多肽的至少85%,例如至少90%或至少95%的氨基酸残基。

[0076] 半纤维素分解酶或半纤维素酶:术语“半纤维素分解酶”或“半纤维素酶”是指一种或多种(例如,几种)水解半纤维素材料的酶。参见,例如Shallom,D.和Shoham,Y.Microbial hemicellulases.Current Opinion In Microbiology,2003,6(3):219-228)。半纤维素酶是植物生物质降解中的关键组分。半纤维素酶的实例包括,但不限于,乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、香豆酸酯酶、阿魏酸酯酶、半乳糖苷酶、葡糖苷酸酶、葡糖醛酸酯酶、甘露聚糖酶、甘露糖苷酶、木聚糖酶和木糖苷酶。这些酶的底物,半纤维素,是支链和直链多糖的混杂群,它们通过氢键键合于植物细胞壁中的纤维素微纤维,将其交联为稳定网路。半纤维素亦共价地附于木质素,与纤维素一同形成高度复杂

的结构。半纤维素的可变结构和组织需要许多酶的协同作用来使其完全降解。半纤维素酶的催化单元为水解糖苷键的糖苷水解酶 (GH), 或水解乙酸酯或阿魏酸侧基的酯键的糖酯酶 (CE)。这些催化单元, 基于其一级序列的同源性, 可分类到GH和CE家族。具有总体上类似的折叠的一些家族还可以归类为宗族, 按字母顺序标记 (例如, GH-A)。这些和其他糖类活性酶的最具信息性和最新的分类可以在糖类活性酶 (CAZy) 数据库中获得。半纤维素分解酶活性可以根据Ghose和Bisaria, 1987, Pure&Appl. Chem. 59:1739-1752在合适的温度例如50℃、55℃或60℃以及pH例如5.0 或5.5下进行测量。

[0077] 高等严格条件: 术语“高等严格条件”是指至少100个核苷酸长度的探针于42℃在5×SSPE、0.3% SDS、200微克/ml剪切并变性的鲑鱼精子DNA和50%甲酰胺中预杂交和杂交, 遵循标准DNA印迹程序进行12至24小时。载体材料最后使用2×SSC、0.2% SDS在65℃洗涤三次, 每次15分钟。

[0078] 同源3' 或5' 区: 术语“同源3' 区”是指以下DNA片段, 其与基因组中区域的序列相同或者序列同一性为至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%, 并且当与同源5' 区结合时可以通过同源重组将一条DNA靶向整合至基因组中的特定位点。术语“同源5' 区”是指以下DNA片段, 其与基因组中区域的序列相同, 并且当与同源3' 区结合时可以通过同源重组将一条DNA靶向整合至基因组中的特定位点。同源5' 和3' 区必须连接在基因组中, 这意味着它们处于同一染色体中, 彼此在至少200kb内。

[0079] 同源侧翼区: 术语“同源侧翼区”是指以下DNA片段, 其与基因组中区域的序列相同或者序列同一性为至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%, 并且紧邻基因组中的细胞外DNA靶向整合入的特定位点的上游或下游。

[0080] 同源重复: 术语“同源重复”是指在导入宿主细胞的重组DNA中重复至少两次且可以通过同源重组而促进DNA (即, 插在两个同源重复之间的可选择性标记物) 的丢失的DNA片段。

[0081] 宿主细胞: 术语“宿主细胞”是指对具有含编码多肽的多核苷酸的核酸构建体或表达载体的转化、转染、转导等敏感的任何细胞类型。术语“宿主细胞”包括由于复制过程中发生的突变而不与亲本细胞相同的任何亲本细胞后代。

[0082] 分离的: 术语“分离的”是指, 处于非天然发生的形式或环境中的物质。分离的物质的非限制性实例包括 (1) 任何非天然发生的物质; (2) 包括但不限于任何酶、变体、核酸、蛋白、肽或辅助因子的任何物质, 其至少部分地从与其天然相关联的一种或多种或所有天然发生成分中移出; (3) 相对于天然发现的物质, 经人工修饰的任何物质; 或者 (4) 通过增加物质相对于与其天然相关的其他组分的含量而修饰的任意物质 (例如, 宿主细胞中的重组生产; 多个拷贝的物质编码基因; 相比于与编码物质的基因天然相关的启动子, 使用更强的启动子)。

[0083] 低等严格条件: 术语“低等严格条件”是指, 长度为至少100个核苷酸的探针, 在42℃下, 在5×SSPE、0.3% SDS、200微克/ml剪切并变性的鲑鱼精子DNA和25%甲酰胺中预杂交

并杂交,遵循标准DNA 印迹过程进行12~24个小时。载体材料最终使用2×SSC、0.2%SDS在50℃下洗涤三次,每次15分钟。

[0084] 成熟多肽:术语“成熟多肽”是指翻译和任何翻译后修饰例如N-端加工、C-端截短、糖基化、磷酸化等之后的最终形式的多肽。在一方面,基于预测SEQ ID NO:2的氨基酸1~17是信号肽的SignalP程序 (Nielsen等人,1997,Protein Engineering 10:1-6),里氏木霉(T.Reesei) 纤维二糖水解酶I的成熟多肽是SEQ ID NO:2的氨基酸18~514。另一方面,基于预测SEQ ID NO:4的氨基酸1~18是信号肽的SignalP程序,里氏木霉纤维二糖水解酶II的成熟多肽是SEQ ID NO:4的氨基酸 19~471。另一方面,基于预测SEQ ID NO:6的氨基酸1~22是信号肽的SignalP程序,里氏木霉内切葡聚糖酶I的成熟多肽是SEQ ID NO:6的氨基酸23~459。另一方面,基于预测SEQ ID NO:8的氨基酸1~21 是信号肽的SignalP程序,里氏木霉内切葡聚糖酶II的成熟多肽是SEQ ID NO:8的氨基酸22~418。另一方面,基于预测SEQ ID NO:10的氨基酸1~19是信号肽的SignalP程序,里氏木霉β-葡糖苷酶的成熟多肽是SEQ ID NO:10的氨基酸20~744。另一方面,基于预测SEQ ID NO: 12的氨基酸1~19是信号肽的SignalP程序,里氏木霉木聚糖酶I的成熟多肽是SEQ ID NO:12的氨基酸20~229。另一方面,基于预测SEQ ID NO:14的氨基酸1~19是信号肽的SignalP程序,里氏木霉木聚糖酶II 的成熟多肽是SEQ ID NO:14的氨基酸20~223。另一方面,基于预测 SEQ ID NO:16的氨基酸1~16是信号肽的SignalP程序,里氏木霉木聚糖酶III的成熟多肽是SEQ ID NO: 16的氨基酸17~347。另一方面,基于预测SEQ ID NO:18的氨基酸1~20是信号肽的SignalP程序,里氏木霉β-木糖苷酶的成熟多肽是SEQ ID NO:18的氨基酸21~797。另一方面,基于预测SEQ ID NO:20的氨基酸1~18是信号肽的SignalP程序,里氏木霉膨胀因子的成熟多肽是SEQ ID NO:20的氨基酸19~493。另一方面,基于预测SEQ ID NO:22的氨基酸1~19是信号肽的SignalP 程序,里氏木霉的枯草杆菌蛋白酶样丝氨酸蛋白酶的成熟多肽是SEQ ID NO:22的氨基酸20~882。另一方面,基于预测SEQ ID NO:24的氨基酸1~20是信号肽的SignalP程序,里氏木霉天冬氨酸蛋白酶的成熟多肽是SEQ ID NO:24的氨基酸21~407。另一方面,基于预测SEQ ID NO:26的氨基酸1~19是信号肽的SignalP程序,里氏木霉胰蛋白酶样丝氨酸蛋白酶的成熟多肽是SEQ ID NO:26的氨基酸20~259。另一方面,基于预测SEQ ID NO:28的氨基酸1~15是信号肽的SignalP程序,另一种里氏木霉枯草杆菌蛋白酶样丝氨酸蛋白酶的成熟多肽是SEQ ID NO:28的氨基酸16~540。另一方面,基于预测SEQ ID NO:30的氨基酸1~17是信号肽的SignalP程序,另一种里氏木霉天冬氨酸蛋白酶的成熟多肽是SEQ ID NO:30的氨基酸18~395。在本领域已知宿主细胞可以产生由同一多核苷酸表达的两种或更多种不同成熟多肽(例如,具有不同的C-端和/或N-端氨基酸)的混合物。

[0085] 成熟多肽编码序列:术语“成熟多肽编码序列”是指编码具有酶活性的成熟多肽的多核苷酸。在一方面,基于预测SEQ ID NO:1的核苷酸1~51编码信号肽的SignalP程序,里氏木霉纤维二糖水解酶I的成熟多肽编码序列是SEQ ID NO:1的核苷酸52~1545或其cDNA序列。另一方面,基于预测SEQ ID NO:3的核苷酸1~54编码信号肽的SignalP 程序,里氏木霉纤维二糖水解酶II的成熟多肽编码序列是SEQ ID NO: 3的核苷酸55~1608或其cDNA序列。另一方面,基于预测SEQ ID NO: 5的核苷酸1~66编码信号肽的SignalP程序,里氏木霉内切葡聚糖酶I 的成熟多肽编码序列是SEQ ID NO:5的核苷酸67~1374或其cDNA序列。另一方面,基于预测SEQ ID NO:7的核苷酸1~63编码信号肽的 SignalP程序,里氏木霉内切

葡聚糖酶II的成熟多肽编码序列是SEQ ID NO:7的核苷酸64~1254或其cDNA序列。另一方面,基于预测SEQ ID NO:9的核苷酸1~57编码信号肽的SignalP程序,里氏木霉 β -葡糖苷酶的成熟多肽编码序列是SEQ ID NO:9的核苷酸58~2612或其cDNA 序列。另一方面,基于预测SEQ ID NO:11的核苷酸1~57编码信号肽的SignalP程序,里氏木霉木聚糖酶I的成熟多肽编码序列是SEQ ID NO:11的核苷酸58~749或其cDNA序列。另一方面,基于预测SEQ ID NO:13的核苷酸1~57编码信号肽的SignalP程序,里氏木霉木聚糖酶 II的成熟多肽编码序列是SEQ ID NO:13的核苷酸58~778或其cDNA 序列。另一方面,基于预测SEQ ID NO:15的核苷酸1~48编码信号肽的SignalP程序,里氏木霉木聚糖酶III的成熟多肽编码序列是SEQ ID NO:15的核苷酸49~1349或其cDNA序列。另一方面,基于预测SEQ ID NO:17的核苷酸1~60编码信号肽的SignalP程序,里氏木霉 β -木糖苷酶的成熟多肽编码序列是SEQ ID NO:17的核苷酸61~2391或其 cDNA序列。另一方面,基于预测SEQ ID NO:19的核苷酸1~54编码信号肽的SignalP程序,里氏木霉膨胀因子的成熟多肽编码序列是SEQ ID NO:19的核苷酸55~2776或其cDNA序列。另一方面,基于预测 SEQ ID NO:21的核苷酸1~57编码信号肽的SignalP程序,里氏木霉枯草杆菌蛋白酶样丝氨酸蛋白酶的成熟多肽编码序列是SEQ ID NO:21 的核苷酸58~2774。另一方面,基于预测SEQ ID NO:23的核苷酸1~60 编码信号肽的SignalP程序,里氏木霉天冬氨酸蛋白酶的成熟多肽编码序列是SEQ ID NO:23的核苷酸61~1299。另一方面,基于预测SEQ ID NO:25的核苷酸1~57编码信号肽的SignalP程序,里氏木霉胰蛋白酶样蛋白酶的成熟多肽编码序列是SEQ ID NO:25的核苷酸58~930。另一方面,基于预测SEQ ID NO:27的核苷酸1~45编码信号肽的SignalP 程序,另一种里氏木霉枯草杆菌蛋白酶样丝氨酸蛋白酶的成熟多肽编码序列是SEQ ID NO:27的核苷酸46~1681。另一方面,基于预测SEQ ID NO:29的核苷酸1~51编码信号肽的SignalP程序,另一种里氏木霉天冬氨酸蛋白酶的成熟多肽编码序列是SEQ ID NO:29的核苷酸 52~1339。

[0086] 中等严格条件:术语“中等严格条件”是指,长度为至少100个核苷酸的探针,在42℃下,在5×SSPE、0.3%SDS、200微克/ml剪切并变性的鲑鱼精子DNA和35%甲酰胺中预杂交并杂交,按照标准DNA 印迹过程进行12~24个小时。载体材料最终使用2×SSC、0.2%SDS在55℃下洗涤三次,每次15分钟。

[0087] 中到高等严格条件:术语“中到高等严格条件”是指,长度为至少100个核苷酸的探针,在42℃下,在5×SSPE、0.3%SDS、200微克 /ml剪切并变性的鲑鱼精子DNA和35%甲酰胺中预杂交并杂交,按照标准DNA印迹过程进行12~24个小时。载体材料最终使用2×SSC、0.2%SDS在60℃下洗涤三次,每次15分钟。

[0088] 核酸构建体:术语“核酸构建体”是指单链或双链的核酸分子,其从天然存在的基因中分离出或以自然中不会出现或是合成的方式修饰成包含核酸片段,其包含一个或多个(例如,几个)控制序列。

[0089] 可操作地连接:术语“可操作地连接”是指一种构造,其中控制序列位于相对于多核苷酸的编码序列为合适的位置,从而使控制序列引导编码序列的表达。

[0090] 具有纤维素分解增强活性的多肽:术语“具有纤维素分解增强活性的多肽”是指通过具有纤维素分解活性的酶来催化纤维素材料水解的增强的GH61多肽。出于本发明的目的,通过测量由纤维素分解酶在下述条件下水解纤维素材料而来的还原糖的增加或纤维二糖与葡萄糖总量的增加来确定纤维素分解增强活性:PCS中的1~50mg总蛋白/g 纤维素,其

中总蛋白包含50~99.5%w/w的纤维素分解酶蛋白和 0.5~50%w/w的具有纤维素分解增强活性的GH61多肽的蛋白,在合适的温度例如50℃、55℃或60℃以及pH例如5.0或5.5下1~7天,与用等量的无纤维素分解增强活性的总蛋白加载量(PCS中的1~50mg纤维素分解蛋白/g纤维素)所进行的对照水解相比。在优选的方面,在 2~3%总蛋白质重量的米曲霉(*Aspergillus oryzae*) β -葡糖苷酶(根据 W0 02/095014在米曲霉中重组产生)或2~3%总蛋白质重量的烟曲霉(*Aspergillus fumigatus*) β -葡糖苷酶(根据W0 2002/095014在米曲霉中重组产生)的纤维素酶蛋白质加载量的存在下将CELLUCLAST® 1.5L (Novozymes A/S, Bagsværd, Denmark)的混合物用作纤维素分解活性的来源。

[0091] 具有纤维素分解增强活性的GH61多肽通过降低达到相同水解程度所需的纤维素分解酶的量来增强由具有纤维素分解活性的酶催化的纤维素材料的水解,优选降低至少1.01倍,例如至少1.05倍、至少1.10 倍、至少1.25倍、至少1.5倍、至少2倍、至少3倍、至少4倍、至少5倍、至少10倍或至少20倍。

[0092] 经预处理的玉米秸秆:术语“PCS”或“经预处理的玉米秸秆”是指通过用热和稀硫酸处理、碱预处理或中性预处理而得自玉米秸秆的纤维素材料。

[0093] 序列同一性:两个氨基酸序列之间或两个核苷酸序列之间的关联性通过参数“序列同一性”来描述。

[0094] 出于本发明的目的,使用在EMBOSS软件包(EMBOSS:The European Molecular Biology Open Software Suite(欧洲分子生物学开放软件包),Rice等人,2000,Trends Genet.16:276-277)(优选5.0.0或更新版本)的Needle程序中执行的Needleman-Wunsch算法(Needleman and Wunsch,1970,J.Mol.Biol.48:443-453),来确定两个氨基酸序列之间的序列同一性。所使用的参数是,空位开放罚分为10,空位扩展罚分为0.5,以及EBLOSUM62(BLOSUM62的EMBOSS版本)替换矩阵。Needle标记的“最长同一性”的输出(使用-nobrief选项得到的)被用作百分比同一性并计算如下:

[0095] $(\text{相同残基} \times 100) / (\text{比对长度} - \text{比对中的空位总数})$ 。

[0096] 出于本发明的目的,使用在EMBOSS软件包(EMBOSS:The European Molecular Biology Open Software Suite(欧洲分子生物学开放软件包),Rice等人,2000,同上)(优选5.0.0或更新版本)的Needle 程序中执行的Needleman-Wunsch算法(Needleman and Wunsch,1970,同上),来确定两个脱氧核糖核苷酸序列之间的序列同一性。所使用的参数是,空位开放罚分为10,空位扩展罚分为0.5,以及EDNAFULL (NCBI NUC4.4的EMBOSS版本)替换矩阵。Needle标记的“最长同一性”的输出(使用-nobrief选项得到的)被用作百分比同一性并计算如下:

[0097] $(\text{相同的脱氧核糖核苷酸} \times 100) / (\text{比对长度} - \text{比对中的空位总数})$ 。

[0098] 子序列:术语“子序列”是指从成熟多肽编码序列的5'和/或3'端缺失一个或多个(例如,几个)核苷酸的多核苷酸;其中子序列编码具有酶活性的片段。在一个方面,子序列含有酶的成熟多肽编码序列的至少85%,例如,至少90%或至少95%的核苷酸。

[0099] 枯草杆菌蛋白酶样丝氨酸蛋白酶:术语“枯草杆菌蛋白酶样丝氨酸蛋白酶”是指底物特异性类似于枯草杆菌蛋白酶的蛋白酶,枯草杆菌蛋白酶使用丝氨酸残基来催化肽和蛋白中肽键的水解。枯草杆菌蛋白酶样蛋白酶(枯草杆菌酶)为丝氨酸蛋白酶,其特征三个氨基酸天冬氨酸、组氨酸和丝氨酸的催化三联体(catalytic triad)。这些催化残基的排

布为来自地衣芽孢杆菌 (*Bacillus licheniformis*) 的原型枯草杆菌蛋白酶所共有 (Siezen 和 Leunissen, 1997, *Protein Science* 6:501-523)。枯草杆菌蛋白酶样丝氨酸蛋白酶的活性可以使用合成的底物 N-琥珀酰-L-Ala-L-Ala-L-Pro-L-Phe-对硝基酰苯胺 (AAPF) (Bachem AG, Bubendorf, 瑞士) 于 100mM NaCl-100mM MOPS pH 7.0 中在 50℃ 下 3 小时并接着测量在 405nm 处吸光度而确定。

[0100] 靶向整合: 术语“靶向整合”是指细胞外 DNA 在指定的基因组基因座处的稳定整合。

[0101] 转化体: 术语“转化体”是指已接纳胞外 DNA (外来的、人工的或修饰的) 并表达包含其中的基因的细胞。

[0102] 转化: 术语“转化”是指将胞外 DNA 导入细胞, 即, 由从其周围直接摄入、整合并表达外来遗传物质 (外来 DNA) 以及经细胞膜接收而引起的细胞的遗传改变。

[0103] 转化效率: 术语“转化效率”是指细胞能够摄入胞外 DNA 并表达其中所含的基因的效率, 转化效率通过将表达基因的阳性转化体的数目除以转化过程中使用的 DNA 量来计算。

[0104] 胰蛋白酶样丝氨酸蛋白酶: 术语“胰蛋白酶样丝氨酸蛋白酶”是指底物特异性类似于胰蛋白酶的蛋白酶, 胰蛋白酶使用丝氨酸残基来催化肽和蛋白中肽键的水解。出于本发明的目的, 胰蛋白酶样丝氨酸蛋白酶的活性是根据 Dienes 等人, 2007, *Enzyme and Microbial Technology* 40:1087-1094 所述的方法而确定的。

[0105] 变体: 术语“变体”是指, 在一个或多个 (例如, 几个) 位置上包含改变 (即替换、插入和/或缺失) 的具有酶活性的多肽。替换是指用不同的氨基酸取代占据位置的氨基酸; 缺失是指去除占据位置的氨基酸; 而插入是指将氨基酸添加至紧邻占据位置的氨基酸。

[0106] 极高等严格条件: 术语“极高等严格条件”是指, 长度为至少 100 个核苷酸的探针, 在 42℃ 下, 在 5×SSPE、0.3% SDS、200 微克/ml 剪切并变性的鲑鱼精子 DNA 和 50% 甲酰胺中预杂交并杂交, 按照标准 DNA 印迹过程进行 12 至 24 个小时。载体材料最终使用 2×SSC、0.2% SDS 在 70℃ 下洗涤三次, 每次 15 分钟。

[0107] 极低等严格条件: 术语“极低等严格条件”是指, 长度为至少 100 个核苷酸的探针, 在 42℃ 下, 在 5×SSPE、0.3% SDS、200 微克/ml 剪切并变性的鲑鱼精子 DNA 和 25% 甲酰胺中预杂交并杂交, 按照标准 DNA 印迹过程进行 12 至 24 个小时。载体材料最终使用 2×SSC、0.2% SDS 在 45℃ 下洗涤三次, 每次 15 分钟。

[0108] 含木聚糖的材料: 术语“含木聚糖的材料”是指任何包含植物细胞壁多糖的材料, 植物细胞壁多糖含有 β-(1-4)-连接的木糖残基的骨架。陆生植物的木聚糖是具有 β-(1-4)-D-吡喃木糖骨架的杂聚物, 其由短的糖链分支。它们包含 D-葡萄糖醛酸或其 4-O-甲基醚、L-阿拉伯糖和/或多种由 D-木糖、L-阿拉伯糖、D-或 L-半乳糖和 D-葡萄糖构成的寡糖。木聚糖类的多糖可以分为均木聚糖 (homoxylan) 和杂木聚糖 (heteroxylan), 包括葡萄糖醛酸木聚糖、(阿拉伯) 葡萄糖醛酸木聚糖、(葡萄糖醛酸) 阿拉伯木聚糖、阿拉伯木聚糖和复合杂木聚糖。参见, 例如 Ebringerova 等人, 2005, *Adv. Polym. Sci.* 186:1-67。

[0109] 在本发明的方法中, 可以使用任何含有木聚糖的材料。在优选的方面, 含木聚糖的材料为木质纤维素。

[0110] 木聚糖降解活性或木聚糖分解活性: 术语“木聚糖降解活性”或“木聚糖分解活性”是指水解含木聚糖的材料的生物学活性。两种基础的测量木聚糖分解活性的方法包括: (1) 测量总木聚糖分解活性, 和 (2) 测量单个的木聚糖分解活性 (例如, 内切木聚糖酶、β-木糖苷

酶、阿拉伯呋喃糖苷酶、 α -葡萄糖醛酸酶、乙酰木聚糖酯酶、阿魏酸酯酶和 α -葡萄糖醛酸酯酶(α -glucuronoyl esterase))。木聚糖分解酶检定的最新进展被总结于几个公开文献中,包括Biely and Puchard,Recent progress in the assays of xylanolytic enzymes,2006,Journal of the Science of Food and Agriculture 86(11):1636-1647;Spanikova and Biely,2006, Glucuronoyl esterase-Novel carbohydrate esterase produced by Schizophyllum commune,FEBS Letters 580(19):4597-4601;Herrmann, Vrsanska, Jurickova, Hirsch, Biely and Kubicek,1997,The beta-D-xylosidase of Trichoderma reesei is a multifunctional beta-D-xylan xylohydrolase, Biochemical Journal 321:375-381。

[0111] 总木聚糖降解活性可以通过确定由多种类型的木聚糖形成的还原糖来测量,木聚糖包括,例如,燕麦(oat spelt)、山毛榉材和落叶松材木聚糖,或可以通过光度法确定从多种共价染色的木聚糖中释放出的染色木聚糖片段而测量。最常见的总木聚糖分解活性检定基于由多聚 4-O-甲基葡萄糖醛酸木聚糖生产还原糖,如Bailey, Biely, Poutanen,1992, Interlaboratory testing of methods for assay of xylanase activity,Journal of Biotechnology 23(3):257-270中所述。木聚糖酶活性亦可以用0.2% AZCL-阿拉伯木聚糖作为底物在37℃下于0.01% TRITON®X-100 (4-(1,1,3,3-四甲基丁基)苯基-聚乙二醇)和200mM磷酸钠缓冲液pH 6 中加以确定。一个单位的木聚糖酶活性定义为在37℃、pH 6 下在200 mM磷酸钠pH 6缓冲液中从作为底物的0.2%AZCL-阿拉伯木聚糖中每分钟产生1.0微摩尔可可碱醋酸钠(azurine)。

[0112] 出于本发明的目的,木聚糖降解活性是通过在以下通常条件下测量由木聚糖降解酶引起的白桦木聚糖(Sigma Chemical Co., Inc., St. Louis, MO, 美国)水解的增加而确定的:1ml反应,5mg/ml底物(总固体),5mg木聚糖分解蛋白/g底物,50mM乙酸钠pH 5,50℃,24小时,如Lever,1972,A new reaction for colorimetric determination of carbohydrates,Anal.Biochem 47:273-279所述的使用对羟基苯甲酸酰肼(PHBAH)检定的糖分析。

[0113] 木聚糖酶:术语“木聚糖酶”是指催化木聚糖中1,4- β -D-木糖苷键的内水解的1,4- β -D-木聚糖-木糖水解酶(1,4-beta-D-xylan-xylohydrolase)(E.C.3.2.1.8)。出于本发明的目的,木聚糖酶活性是用0.2%AZCL-阿拉伯木聚糖作为底物在37℃下于0.01% TRITON®X-100和200mM磷酸钠缓冲液pH 6中加以确定的。一个单位的木聚糖酶活性定义为在37℃、pH 6下于200mM磷酸钠pH 6缓冲液中从作为底物的0.2%AZCL-阿拉伯木聚糖中每分钟产生1.0微摩尔可可碱醋酸钠。

[0114] 发明详述

[0115] 本发明涉及构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,包括:(a)通过向丝状真菌菌株中导入第一串联构建体,通过靶向整合取代内源性第一基因,该第一串联构建体包含(i)第一基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个(例如,几个)第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)第一基因的同源3'区、其同源侧翼区或其组合;(b)通过向丝状真菌菌株中导入第二串联构建体,通过靶向整合取代内源性第

二基因,该第二串联构建体包含(i)第二基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个(例如,几个)第二可选择性标记物,(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)第二基因的同源3'区、其同源侧翼区或其组合;或(c)(a)和(b)的组合。

[0116] 本发明还涉及串联构建体,其包含:(i)基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个(例如,几个)可选择性标记物;(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v)基因的同源3'区、其同源侧翼区或其组合。

[0117] 本发明还涉及构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,其包括:(a)通过向丝状真菌菌株中导入第一串联构建体,通过靶向整合插入到内源性第一基因座中,该第一串联构建体包含(i)第一基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)第一基因座的同源3'区、其同源侧翼区或其组合;(b)通过向丝状真菌菌株中导入第二串联构建体,通过靶向整合插入到内源性第二基因座中,该第二串联构建体包含(i)第二基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个第二可选择性标记物,(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)第二基因座的同源3'区、其同源侧翼区或其组合;或(c)(a)和(b)的组合。

[0118] 本发明还涉及串联构建体,其包含:(i)基因座的同源5'区、其同源侧翼区或其组合;(ii)一个或多个第一可选择性标记物;(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v)基因座的同源3'区、其同源侧翼区或其组合。

[0119] 在本发明的方法中,各个串联构建体均通过双同源重组整合至丝状真菌菌株染色体中的目标位点。在一方面,第一基因或第一基因座的同源5'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。另一方面,第一基因或第一基因座的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。另一方面,第二基因或第二基因座的同源5'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。另一方面,第二基因或第二基因座的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0120] 本发明的方法还可以包括将一个或多个(例如,几个)其他内源性基因各自通过靶向整合用针对各个基因的对应串联构建体取代,串联构建体包含:(i)基因的同源5'区、其

同源侧翼区或其组合；(ii) 一个或多个(例如,几个)可选择性标记物；(iii) 与启动子可操作连接的编码具有生物活性的多肽的多核苷酸；(iv) 与另一启动子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸；和(v) 基因的同源3' 区、其同源侧翼区或其组合。

[0121] 本发明的方法还可以包括向一个或多个(例如,几个)其他内源性基因座中各自通过靶向整合插入针对各个基因座的对应串联构建体,该串联构建体包含:(i) 基因座的同源5' 区、其同源侧翼区或其组合；(ii) 一个或多个(例如,几个)可选择性标记物；(iii) 与启动子和终止子可操作连接的编码具有生物活性的多肽的多核苷酸；(iv) 与另一启动子和另一终止子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸；和(v) 基因座的同源3' 区、其同源侧翼区或其组合。

[0122] 在本发明的方法中,丝状真菌宿主细胞还可以用串联构建体进行转化,该串联构建体包含:(i) 一个或多个(例如,几个)可选择性标记物；(ii) 与第五启动子和第五终止子可操作连接的编码具有生物活性的第五多肽的第五多核苷酸；和(iii) 与第六启动子和第六终止子可操作连接的编码具有生物活性的第六多肽的第六多核苷酸,其中串联构建体通过异位整合来整合。

[0123] 本发明提供几项优势,包括在丝状真菌宿主中产生多种重组多肽的改良方法；使得一个或多个导入到丝状真菌菌株中的重组多肽可以简单地得以取代和/或缺失的方法(如果一个或多个重组多肽在单个基因座处导入,该一个或多个重组多肽可以在单一步骤中被取代和/或缺失,而不是在多个步骤中)；以及使得丝状真菌宿主中存在的一个或多个重组多肽可以简单地得以修饰(如果一个或多个重组多肽在单个基因座处导入,一个或多个这些重组多肽的变体/突变体可以很容易地在单一步骤中加以修饰,而不是在多个步骤中)。本发明构建的丝状真菌菌株的其他优势是可以诱变菌株并选择收率提高的突变体。然后将这些突变体的串联构建体用表达新的多种重组多肽的新的串联构建体取代,从而利用作为宿主细胞的突变体的提高的收率生产率。

[0124] 内源性基因

[0125] 在本发明的方法中,任何对于丝状真菌菌株来说为内源性的的基因可以被取代,或者可以是基因座。基因对于丝状真菌菌株来说可以是内生的或外来的。术语“内源性基因”或其变型,例如,“内源性第一基因”或“内源性第二基因”将理解成包括一个或多个(例如,几个)基因。当多于一个基因被取代时,基因优选是邻接的。这样的多个基因可以是,例如,代谢通路或其一部分。

[0126] 内源性基因可以编码选自抗体、抗原、抗微生物肽、酶、生长因子、激素、免疫增强剂(immunodilator)、神经递质、受体、报告蛋白质、结构蛋白或转录因子的多肽。

[0127] 在一方面,酶选自氧化还原酶、转移酶、水解酶、裂解酶、异构酶和连接酶。另一方面,酶选自乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、氨基肽酶、 α -淀粉酶、 α -半乳糖苷酶、 α -葡萄糖苷酶、 α -1,6-葡萄糖苷转移酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、 β -半乳糖苷酶、 β -葡萄糖苷酶、 β -木糖苷酶、糖酶、羧肽酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、几丁质酶、香豆酸酯酶、环糊精糖基转移酶、角质酶、环糊精糖基转移酶、脱氧核糖核酸酶、内切葡聚糖酶、酯酶、阿魏酸酯酶、具有纤维素分解促进活性的GH61多肽、葡萄糖脑苷脂酶、葡萄糖氧化酶、葡萄糖醛酸酶、葡萄糖醛酸酯酶、卤过氧化物酶(haloperoxidase)、半纤维素酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、甘露聚糖酶、甘露糖苷酶、变聚糖酶(mutanase)、氧化酶、果胶酶、过氧

化物酶、磷脂酶、植酸酶、酚氧化酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶、尿激酶和木聚糖酶。

[0128] 另一方面,一个或多个(例如,几个)内源性蛋白酶基因是失活的。另一方面,一个或多个(例如,几个)内源性蛋白酶基因是枯草杆菌蛋白酶样丝氨酸蛋白酶、天冬氨酸蛋白酶和胰蛋白酶样丝氨酸蛋白酶基因,如WO 2011/075677中所述,其通过引用的方式全部并入本文。

[0129] 另一方面,酶是枯草杆菌蛋白酶样丝氨酸蛋白酶。另一方面,酶是天冬氨酸蛋白酶。另一方面,酶是胰蛋白酶样丝氨酸蛋白酶。

[0130] 另一方面,酶是内切葡聚糖酶。另一方面,酶是纤维二糖水解酶。另一方面,酶是 β -葡萄糖苷酶。另一方面,酶是具有纤维素分解增强活性的GH61多肽。另一方面,酶是木聚糖酶。另一方面,酶是 β -木糖苷酶。另一方面,酶是膨胀因子。另一方面,酶是乙酰木聚糖酯酶。另一方面,酶是阿魏酸酯酶。另一方面,酶是阿拉伯呋喃糖苷酶。另一方面,酶是葡萄糖醛酸酶。另一方面,酶是乙酰甘露聚糖酯酶。另一方面,酶是阿拉伯聚糖酶。另一方面,酶是香豆酸酯酶。另一方面,酶是半乳糖苷酶。另一方面,酶是葡萄糖醛酸酯酶。另一方面,酶是甘露聚糖酶。另一方面,酶是甘露糖苷酶。

[0131] 另一方面,内源性基因可以是纤维二糖水解酶I基因。另一方面,纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I:(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I;(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I;(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I;和(iv)由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0132] 另一方面,内源性基因可以是纤维二糖水解酶II基因。另一方面,纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II;(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II;(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II;和(iv)由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0133] 另一方面,内源性基因可以是内切葡聚糖酶I基因。另一方面,内切葡聚糖酶I基因编码选自以下的内切葡聚糖酶I: (i) 包含SEQ ID NO: 6的成熟多肽的内切葡聚糖酶I; (ii) 包含与SEQ ID NO:6的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的内切葡聚糖酶I; (iii) 由包含与SEQ ID NO:5的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的内切葡聚糖酶I; 和 (iv) 由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:5的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的内切葡聚糖酶 I。

[0134] 另一方面,内源性基因可以是内切葡聚糖酶II基因。另一方面,内切葡聚糖酶II基因编码选自以下的内切葡聚糖酶II: (i) 包含SEQ ID NO:8的成熟多肽的内切葡聚糖酶II; (ii) 包含与SEQ ID NO:8的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的内切葡聚糖酶II; (iii) 由包含与SEQ ID NO:7的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的内切葡聚糖酶II; 和 (iv) 由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:7的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的内切葡聚糖酶II。

[0135] 另一方面,内源性基因可以是 β -葡糖苷酶基因。另一方面, β -葡糖苷酶基因编码选自以下的 β -葡糖苷酶: (i) 包含SEQ ID NO:10的成熟多肽的 β -葡糖苷酶; (ii) 包含与SEQ ID NO:10的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的 β -葡糖苷酶; (iii) 由包含与SEQ ID NO:9的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的 β -葡糖苷酶; 和 (iv) 由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:9的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的 β -葡糖苷酶。

[0136] 另一方面,内源性基因可以是木聚糖酶I基因。另一方面,木聚糖酶I基因编码选自以下的木聚糖酶I: (i) 包含SEQ ID NO:12的成熟多肽的木聚糖酶I; (ii) 包含与SEQ ID NO:12的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少

92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的木聚糖酶I; (iii) 由包含与SEQ ID NO:11的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的木聚糖酶I; 和 (iv) 由在至少高等严格条件, 例如非常高等严格条件下与SEQ ID NO:11的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的木聚糖酶I。

[0137] 另一方面, 内源性基因可以是木聚糖酶II基因。另一方面, 木聚糖酶II基因编码选自以下的木聚糖酶II: (i) 包含SEQ ID NO:14的成熟多肽的木聚糖酶II; (ii) 包含与SEQ ID NO:14的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的木聚糖酶II; (iii) 由包含与SEQ ID NO:13的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的木聚糖酶II; 和 (iv) 由在至少高等严格条件, 例如非常高等严格条件下与SEQ ID NO:13的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的木聚糖酶II。

[0138] 另一方面, 内源性基因可以是木聚糖酶III基因。另一方面, 木聚糖酶III基因编码选自以下的木聚糖酶III: (i) 包含SEQ ID NO:16的成熟多肽的木聚糖酶III; (ii) 包含与SEQ ID NO:16的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的木聚糖酶III; (iii) 由包含与SEQ ID NO:15的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的木聚糖酶III; 和 (iv) 由在至少高等严格条件, 例如非常高等严格条件下与SEQ ID NO:15的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的木聚糖酶III。

[0139] 另一方面, 内源性基因可以是 β -木糖苷酶基因。另一方面, β -木糖苷酶基因编码选自以下的 β -木糖苷酶: (i) 包含SEQ ID NO:18的成熟多肽的 β -木糖苷酶; (ii) 包含与SEQ ID NO:18的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的 β -木糖苷酶; (iii) 由包含与SEQ ID NO:17的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多

核苷酸所编码的 β -木糖苷酶;和(iv)由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:17的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的 β -木糖苷酶。

[0140] 另一方面,内源性基因可以是膨胀因子基因。另一方面,膨胀因子基因编码选自以下的膨胀因子:(i)包含SEQ ID NO:20的成熟多肽的膨胀因子;(ii)包含与SEQ ID NO:20的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的膨胀因子;(iii)由包含与SEQ ID NO:19的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的膨胀因子;和(iv)由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:19的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的膨胀因子。

[0141] 另一方面,内源性基因可以是枯草杆菌蛋白酶样丝氨酸蛋白酶基因。另一方面,枯草杆菌蛋白酶样丝氨酸蛋白酶基因编码选自以下的枯草杆菌蛋白酶样丝氨酸蛋白酶:(i)包含SEQ ID NO:22的成熟多肽的枯草杆菌蛋白酶样丝氨酸蛋白酶;(ii)包含与SEQ ID NO:22的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的枯草杆菌蛋白酶样丝氨酸蛋白酶;(iii)由包含与SEQ ID NO:21的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的枯草杆菌蛋白酶样丝氨酸蛋白酶;和(iv)由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:21的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的枯草杆菌蛋白酶样丝氨酸蛋白酶。

[0142] 另一方面,内源性基因可以是天冬氨酸蛋白酶基因。另一方面,天冬氨酸蛋白酶基因编码选自以下的天冬氨酸蛋白酶:(i)包含SEQ ID NO:24的成熟多肽的天冬氨酸蛋白酶;(ii)包含与SEQ ID NO:24的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的天冬氨酸蛋白酶;(iii)由包含与SEQ ID NO:23的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的天冬氨酸蛋白酶;和(iv)由在至少高等严格条件,例如非常高等严格条件下与SEQ ID NO:23的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的天冬氨酸蛋白酶。

[0143] 另一方面,内源性基因可以是胰蛋白酶样丝氨酸蛋白酶基因。另一方面,胰蛋白酶

样丝氨酸蛋白酶基因编码选自以下的胰蛋白酶样丝氨酸蛋白酶：(i) 包含SEQ ID NO:26的成熟多肽的胰蛋白酶样丝氨酸蛋白酶；(ii) 包含与SEQ ID NO:26的成熟多肽具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的胰蛋白酶样丝氨酸蛋白酶；(iii) 由包含与SEQ ID NO:25的成熟多肽编码序列具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的胰蛋白酶样丝氨酸蛋白酶；和(iv) 由在至少高等严格条件，例如非常高等严格条件下与SEQ ID NO:25的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的胰蛋白酶样丝氨酸蛋白酶。

[0144] 另一方面，内源性基因可以是枯草杆菌蛋白酶样丝氨酸蛋白酶基因。另一方面，枯草杆菌蛋白酶样丝氨酸蛋白酶基因编码选自以下的枯草杆菌蛋白酶样丝氨酸蛋白酶：(i) 包含SEQ ID NO:28的成熟多肽的枯草杆菌蛋白酶样丝氨酸蛋白酶；(ii) 包含与SEQ ID NO:28的成熟多肽具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的枯草杆菌蛋白酶样丝氨酸蛋白酶；(iii) 由包含与SEQ ID NO:27的成熟多肽编码序列具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的枯草杆菌蛋白酶样丝氨酸蛋白酶；和(iv) 由在至少高等严格条件，例如非常高等严格条件下与SEQ ID NO:27的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的枯草杆菌蛋白酶样丝氨酸蛋白酶。

[0145] 另一方面，内源性基因可以是天冬氨酸蛋白酶基因。另一方面，天冬氨酸蛋白酶基因编码选自以下的天冬氨酸蛋白酶：(i) 包含SEQ ID NO:30的成熟多肽的天冬氨酸蛋白酶；(ii) 包含与SEQ ID NO:30的成熟多肽具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的天冬氨酸蛋白酶；(iii) 由包含与SEQ ID NO:29的成熟多肽编码序列具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的天冬氨酸蛋白酶；和(iv) 由在至少高等严格条件，例如非常高等严格条件下与SEQ ID NO:29的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的天冬氨酸蛋白酶。

[0146] 串联构建体

[0147] 本发明还涉及串联构建体，其包含：(i) 基因的同源5'区、其同源侧翼区或其组合；

(ii) 一个或多个(例如,几个)可选择性标记物; (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸; (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸; 和 (v) 基因的同源 3' 区、其同源侧翼区或其组合。

[0148] 本发明还涉及串联构建体,其包含: (i) 基因座的同源5' 区、其同源侧翼区或其组合; (ii) 一个或多个第一可选择性标记物; (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸; (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸; 和 (v) 基因座的同源3' 区、其同源侧翼区或其组合。

[0149] 串联构建体可以通过使一个或多个(例如,几个)控制基因与构建体的各个多核苷酸可操作连接而构建,从而在与控制序列相容的条件下引导编码序列在丝状真菌宿主细胞中的表达。取决于表达载体,在插入到载体之前对各个多核苷酸进行操纵可能是合意的或必需的。使用重组DNA方法修饰多核苷酸的技术是本领域公知的。

[0150] 控制序列可以是启动子,即被宿主细胞识别用于编码多肽的多核苷酸的表达的多核苷酸。启动子含有介导多肽表达的转录控制序列。启动子可以是任何在宿主细胞中表现出转录活性的多核苷酸,包括突变的、截短的和杂合的启动子,并且可以得自与宿主细胞同源或异源的编码胞外或胞内多肽的基因。

[0151] 一方面,启动子是不同的启动子。另一方面,两个或更多个(例如,几个)启动子是相同的启动子。

[0152] 用于在丝状真菌宿主细胞中引导构建体的转录的适当启动子的实例是从构巢曲霉(*Aspergillus nidulans*) 乙酰胺酶、黑曲霉(*Aspergillus niger*) 中性 α -淀粉酶、黑曲霉酸性稳定性 α -淀粉酶、黑曲霉或泡盛曲霉(*Aspergillus awamori*) 葡糖淀粉酶(glaA)、米曲霉(*Aspergillus oryzae*) TAKA淀粉酶、米曲霉碱性蛋白酶、米曲霉磷酸丙糖异构酶、尖孢镰孢(*Fusarium oxysporum*) 胰蛋白酶样蛋白酶(WO 96/00787)、镰片镰孢(*Fusarium venenatum*) 淀粉葡萄糖苷酶(WO 00/56900)、镰片镰孢(*Fusarium venenatum*) Daria(WO 00/56900)、镰片镰孢Quinn(WO 00/56900)、米黑根毛霉(*Rhizomucor miehei*) 脂肪酶、米黑根毛霉天冬氨酸蛋白酶、里氏木霉 β -葡萄糖苷酶、里氏木霉纤维二糖水解酶I、里氏木霉纤维二糖水解酶II、里氏木霉内切葡聚糖酶I、里氏木霉内切葡聚糖酶II、里氏木霉内切葡聚糖酶III、里氏木霉内切葡聚糖酶V、里氏木霉木聚糖酶I、里氏木霉木聚糖酶II、里氏木霉木聚糖酶III、里氏木霉 β -木糖苷酶、和里氏木霉翻译延伸因子的基因中获得的启动子,以及NA2-tpi启动子(得自曲霉中性 α -淀粉酶基因的修饰启动子,其中非翻译前导已被曲霉磷酸丙糖异构酶基因的非翻译前导替换;非限制性实例包括黑曲霉中性 α -淀粉酶基因的修饰启动子,其中非翻译前导已被构巢曲霉或米曲霉磷酸丙糖异构酶基因的非翻译前导替换);以及其突变的、截短的和杂合的启动子。其他启动子在美国专利第 6,011,147中描述,其整体并入本文。

[0153] 控制序列也可以是转录终止子,其由宿主细胞识别来终止转录。终止子与编码多肽的多核苷酸的3' 端可操作地连接。可以在宿主细胞中起作用的任何终止子可以用在本发明中。

[0154] 在一个方面,终止子是不同的终止子。在另一个方面,两个或更多个(例如,几个)

终止子是相同的终止子。

[0155] 用于丝状真菌宿主细胞的优选终止子得自构巢曲霉邻氨基苯甲酸合酶、构巢曲霉乙酰胺酶 (amdS)、黑曲霉葡萄糖淀粉酶、黑曲霉 α -葡萄糖苷酶、米曲霉TAKA淀粉酶、尖孢镰孢胰蛋白酶样蛋白酶、里氏木霉 β -葡萄糖苷酶、里氏木霉纤维二糖水解酶I、里氏木霉纤维二糖水解酶II、里氏木霉内切葡聚糖酶I、里氏木霉内切葡聚糖酶II、里氏木霉内切葡聚糖酶III、里氏木霉内切葡聚糖酶V、里氏木霉木聚糖酶I、里氏木霉木聚糖酶II、里氏木霉木聚糖酶III、里氏木霉 β -木糖苷酶、和里氏木霉翻译延伸因子的基因。

[0156] 控制序列还可以是前导序列,即,对由宿主细胞的翻译较为重要的mRNA非翻译区。前导序列可操作地连接到编码多肽的多核苷酸的5'端。任何在宿主细胞中起作用的前导序列均可以使用。

[0157] 用于丝状真菌宿主细胞的优选前导序列得自米曲霉TAKA淀粉酶和构巢曲霉磷酸丙糖异构酶的基因。

[0158] 控制序列也可以是多腺苷酸化序列,即可操作地连接至多核苷酸的3'端的序列,在转录时被宿主细胞识别为向所转录的mRNA添加多腺苷残基的信号。在宿主细胞中起作用的任何多腺苷酸化序列均可以使用。

[0159] 用于丝状真菌宿主细胞的优选多腺苷酸化序列得自构巢曲霉邻氨基苯甲酸合酶、黑曲霉葡萄糖淀粉酶、黑曲霉 α -葡萄糖苷酶、米曲霉TAKA淀粉酶、尖孢镰孢胰蛋白酶样蛋白酶、里氏木霉纤维二糖水解酶I、里氏木霉纤维二糖水解酶II、和里氏木霉内切葡聚糖酶V的基因。

[0160] 控制序列也可以是编码与多肽N端连接的信号肽并引导多肽进入细胞分泌通路的信号肽编码区。多核苷酸的编码序列的5'端可以固有地包含在翻译阅读框中与编码多肽的编码序列片段自然连接的信号肽编码序列。或者,编码序列的5'端可以包含对于编码序列为外来的信号肽编码序列。在编码序列不天然包含信号肽编码序列时,外来信号肽编码序列可能是需要的。或者,为增强多肽的分泌,外来信号肽编码序列可以简单地取代天然信号肽编码序列。然而,任何引导所表达的多肽进入宿主细胞的分泌通路的信号肽编码序列均可以使用。

[0161] 用于丝状真菌宿主细胞的有效信号肽编码序列是得自黑曲霉中性淀粉酶、黑曲霉葡萄糖淀粉酶、米曲霉TAKA淀粉酶、特异腐质霉 (*Humicola insolens*) 纤维素酶、特异腐质霉内切葡聚糖酶V、柔毛腐质霉 (*Humicola lanuginosa*) 脂肪酶、米黑根毛霉 (*Rhizomucor miehei*) 天冬氨酸蛋白酶、里氏木霉纤维二糖水解酶I、里氏木霉纤维二糖水解酶II、里氏木霉内切葡聚糖酶I、里氏木霉内切葡聚糖酶II、里氏木霉内切葡聚糖酶III、和里氏木霉内切葡聚糖酶V的基因的信号肽编码序列。

[0162] 控制序列也可以是编码位于多肽N端的前肽的前肽编码序列。得到的多肽称为酶原或前多肽(或在一些情况下为酵素原)。前多肽一般是失活的且可以通过从前多肽到前肽的催化或自催化切割而转化为活性多肽。前肽编码序列可以得自嗜热毁丝菌 (*Myceliophthora thermophila*) 漆酶 (WO 95/33836) 和米黑根毛霉天冬氨酸蛋白酶的基因。

[0163] 在信号肽和前肽序列均存在时,前肽序列紧邻多肽的N端且信号肽序列紧邻前肽序列的N端。

[0164] 也可能需要添加调控多肽相对于宿主细胞生长的表达的调控序列。调控序列的实例是响应于化学或物理刺激(包括调控化合物的存在)引起基因表达的开启和关闭的那些。丝状真菌中的调控序列包括黑曲霉葡萄糖淀粉酶启动子、米曲霉TAKA α -淀粉酶启动子、米曲霉葡萄糖淀粉酶启动子、里氏木霉纤维二糖水解酶I启动子和里氏木霉纤维二糖水解酶II启动子。调控序列的其他实例是使得基因扩增的那些。在这些情况下,编码多肽的多核苷酸将与调控序列可操作地连接。

[0165] 本发明的串联构建体优选包含一个或多个(例如,几个)允许方便地选择所转化细胞的可选择性标记物。可选择性标记物是基因,其产物提供对生物灭杀剂的抗性或病毒抗性、重金属抗性、相对于营养缺陷型的原养型等。用于丝状真菌宿主细胞的可选择性标记物的实例包括,但不限于,a de A(磷酸核糖基氨基咪唑-琥珀羧胺合酶(phosphoribosylaminoimidazole-succinocarboxamide synthase))、adeB(磷酸核糖基氨基咪唑合酶(phosphoribosylaminoimidazole synthase))、amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰基转移酶)、bar(蒴丝菌素乙酰转移酶)、hph(潮霉素磷酸转移酶)、niaD(硝酸还原酶)、pyrG(乳清酸核苷-5'-磷酸脱羧酶)、sC(硫酸腺苷酰转移酶)、和trpC(邻氨基苯甲酸合成酶)、以及其等价物。优选用在曲霉属细胞中的是构巢曲霉或米曲霉amdS和pyrG基因以及吸水链霉菌(*Streptomyces hygroscopicus*) bar基因。优选用在木霉属细胞中的是adeA、adeB、amdS、hph和pyrG基因。细菌可选择性标记物的实例是赋予抗生素抗性的标记物,例如氨苄青霉素、氯霉素、卡那霉素、新霉素、壮观霉素、或四环素抗性。

[0166] 一个或多个(例如,几个)可选择性标记物可以是如WO 2010/039889 A2(通过引用的方式全部合并入本文)中所述的双选择标记物体系。在一个方面,一个或多个可选择性标记物是hph-tk双选择标记物体系。

[0167] 在本发明的各串联构建体中,一个或多个(例如,几个)可选择性标记物是不同的标记物,除非如本文所述重复使用可选择性标记物。

[0168] 一个或多个(例如,几个)可选择性标记物可以再用于将针对各个基因或基因座的对应串联构建体各自通过靶向整合来取代一个或多个(例如,几个)其他内源性基因或插入到一个或多个其他内源性基因座中。一个或多个串联构建体还可以包含一个或多个可选择性标记物的5'侧翼的第一同源重复以及一个或多个可选择性标记物的3'侧翼的第二同源重复,其中,第一同源重复和第二同源重复进行同源重组以切除一个或多个可选择性标记物。在切除一个或多个可选择性标记物时,一个或多个可选择性标记物可以再用于将针对各个基因或基因座的对应串联构建体各自通过靶向整合来取代一个或多个其他内源性基因或插入到一个或多个其他内源性基因座中。

[0169] 在一个方面,第一和第二同源重复是相同的。在另一个方面,第一和第二同源重复彼此具有至少70%,例如,至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、或至少99%的序列同一性。在另一个方面,第一和第二同源重复各自为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800 bp、至少1000bp、至少1500bp、或至少2000bp。包含一个重复的片段可以比含有另一重复的片段长。

[0170] 本发明的串联构建体还可以包含一个或多个(例如,几个)编码其他具有生物活性

的多肽的其他多核苷酸。例如，串联构建体可以含有一个其他多核苷酸、两个其他多核苷酸、三个其他多核苷酸等。

[0171] 具有生物活性的多肽

[0172] 多肽可以是具有目标生物活性的任何多肽。在本文中，术语“多肽”不是指特定长度的所编码产物，因此涵盖肽、寡肽和蛋白。术语“多肽”还涵盖组合形成所编码产物的两个或更多个（例如，几个）多肽。多肽也包括融合多肽，其包含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（例如，几个）多肽可以对丝状真菌宿主细胞为异源。多肽还可以包括下述多肽的天然存在的等位变化和遗传工程变化以及杂合多肽。

[0173] 在一个方面，多肽选自抗体、抗原、抗微生物肽、酶、生长因子、激素、免疫增强剂、神经递质、受体、报告蛋白、结构蛋白、或转录因子。

[0174] 另一方面，酶选自氧化还原酶、转移酶、水解酶、裂解酶、异构酶和连接酶。在另一个方面，酶选自乙酰甘露聚糖酯酶、乙酰木聚糖酯酶、氨基肽酶、 α -淀粉酶、 α -半乳糖苷酶、 α -葡萄糖苷酶、 α -1,6-葡萄糖苷转移酶、阿拉伯聚糖酶、阿拉伯呋喃糖苷酶、 β -半乳糖苷酶、 β -葡萄糖苷酶、 β -木糖苷酶、糖酶、羧肽酶、过氧化氢酶、纤维二糖水解酶、纤维素酶、几丁质酶、香豆酸酯酶、环糊精糖基转移酶、角质酶、脱氧核糖核酸酶、内切葡聚糖酶、酯酶、阿魏酸酯酶、具有纤维素分解增强活性的GH61多肽、葡萄糖脑苷脂酶、葡萄糖氧化酶、葡萄糖醛酸酶、葡萄糖醛酸酯酶、卤过氧化物酶、半纤维素酶、转化酶、异构酶、漆酶、连接酶、脂肪酶、甘露聚糖酶、甘露糖苷酶、变聚糖酶、氧化酶、果胶分解酶、过氧化物酶、磷脂酶、植酸酶、酚氧化酶、多酚氧化酶、蛋白水解酶、核糖核酸酶、转谷氨酰胺酶、尿激酶和木聚糖酶

[0175] 另一方面，多肽选自白蛋白、胶原蛋白、弹性蛋白原、弹性蛋白、和明胶。

[0176] 另一方面，具有生物活性的多肽可以是不同的多肽。另一方面，具有生物活性的两个或更多个（例如，几个）多肽是相同的多肽。

[0177] 另一方面，多肽包含选自纤维素酶、cip 1蛋白、具有纤维素分解增强活性的GH61、半纤维素酶、酯酶、膨胀素、漆酶、木质素降解酶、果胶酶、过氧化物酶、蛋白酶、和膨胀因子的一种或多种酶。在另一方面，纤维素酶是选自内切葡聚糖酶、纤维二糖水解酶、和 β -葡萄糖苷酶的一种或多种酶。在另一方面，半纤维素酶是选自木聚糖酶、乙酰木聚糖酯酶、阿魏酸酯酶、阿拉伯呋喃糖苷酶、木糖苷酶和葡萄糖醛酸酶的一种或多种酶。

[0178] 在另一方面，多肽之一是纤维素酶。在另一方面，多肽之一是内切葡聚糖酶。在另一方面，多肽之一是纤维二糖水解酶。在另一方面，多肽之一是 β -葡萄糖苷酶。在另一方面，多肽之一是具有纤维素分解增强活性的GH61多肽。在另一方面，多肽之一是膨胀因子蛋白。在另一方面，多肽之一是cip 1蛋白。在另一方面，多肽之一是酯酶。在另一方面，多肽之一是膨胀素。在另一方面，多肽之一是漆酶。在另一方面，多肽之一是木质素降解酶。在另一方面，多肽之一是果胶酶。在另一方面，多肽之一是过氧化物酶。在另一方面，多肽之一是蛋白酶。在另一方面，多肽之一是膨胀因子。

[0179] 在另一方面，多肽之一是半纤维素酶。在另一方面，多肽之一是木聚糖酶。在另一方面，多肽之一是 β -木糖苷酶。在另一方面，多肽之一是乙酰木聚糖酯酶。在另一方面，多肽之一是阿魏酸酯酶。在另一方面，多肽之一是阿拉伯呋喃糖苷酶。在另一方面，多肽之一是葡萄糖醛酸酶。在另一方面，多肽之一是乙酰甘露聚糖酯酶。在另一方面，多肽之一是阿拉伯聚糖酶。在另一方面，多肽之一是香豆酸酯酶。在另一方面，多肽之一是半乳糖苷酶。在另一

方面,多肽之一是葡糖醛酸酯酶。在另一方面,多肽之一是甘露聚糖酶。在另一方面,多肽之一是甘露糖苷酶。

[0180] 作为具有生物活性的多肽的内切葡聚糖酶的实例包括,但不限于,里氏木霉内切葡聚糖酶I (Penttila等人,1986,Gene 45:253-263)、里氏木霉Cel7B内切葡聚糖酶I (GENBANK™登录号M15665)、里氏木霉内切葡聚糖酶II (Saloheimo等人,1988,Gene 63:11-22)、里氏木霉Cel5A 内切葡聚糖酶II (GENBANK™登录号M19373)、里氏木霉内切葡聚糖酶III (Okada等人,1988,Appl. Environ. Microbiol. 64:555-563; GENBANK™登录号AB003694)、里氏木霉内切葡聚糖酶V (Saloheimo 等人,1994,Molecular Microbiology 13:219-228,GENBANK™登录号 Z33381)、棘孢曲霉 (*Aspergillus aculeatus*) 内切葡聚糖酶 (Ooi等人,1990,Nucleic Acids Research 18:5884)、白曲霉 (*Aspergillus kawachii*) 内切葡聚糖酶 (Sakamoto等人,1995,Current Genetics 27:435-439)、软腐欧氏杆菌 (*Erwinia carotovora*) 内切葡聚糖酶 (Saarilahti等人, 1990,Gene 90:9-14)、尖孢镰孢内切葡聚糖酶 (GENBANK™登录号 L29381)、灰腐质霉高温变种 (*Humicola grisea* var. *thermoidea*) 内切葡聚糖酶 (GENBANK™登录号AB003107)、热白丝菌 (*Melanocarpus albomyces*) 内切葡聚糖酶 (GENBANK™登录号MAL515703)、粗糙脉孢菌 (*Neurospora crassa*) 内切葡聚糖酶 (GENBANK™登录号 XM_324477)、特异腐质霉内切葡聚糖酶V、嗜热毁丝菌CBS 117.65 内切葡聚糖酶、担子菌 (basidiomycete) CBS 495.95内切葡聚糖酶、担子菌CBS 494.95内切葡聚糖酶、太瑞斯梭孢壳 (*Thielavia terrestris*) NRRL 8126CEL6B内切葡聚糖酶、太瑞斯梭孢壳NRRL 8126CEL6C 内切葡聚糖酶、太瑞斯梭孢壳NRRL 8126CEL7C内切葡聚糖酶、太瑞斯梭孢壳NRRL 8126CEL7E内切葡聚糖酶、太瑞斯梭孢壳NRRL 8126 CEL7F内切葡聚糖酶、多生枝鼻菌 (*Cladorrhinum foecundissimum*) ATCC 62373CEL7A内切葡聚糖酶、和里氏木霉菌株No. VTT-D-80133 内切葡聚糖酶 (GENBANK™登录号M15665)。

[0181] 作为具有生物活性的多肽的纤维二糖水解酶的实例包括,但不限于,棘孢曲霉纤维二糖水解酶II (WO 2011/059740)、嗜热毛壳菌 (*Chaetomium thermophilum*) 纤维二糖水解酶I、嗜热毛壳菌纤维二糖水解酶II、特异腐质霉纤维二糖水解酶I、嗜热毁丝菌纤维二糖水解酶 II (WO 2009/042871)、*Thielavia hyrcanie*纤维二糖水解酶II (WO 2010/141325)、太瑞斯梭孢壳纤维二糖水解酶II (CEL6A,WO 2006/074435)、里氏木霉纤维二糖水解酶I、里氏木霉纤维二糖水解酶 II和囊状长毛盘菌 (*Trichophaea saccata*) 纤维二糖水解酶II (WO 2010/057086)。

[0182] 作为具有生物活性的多肽的 β -葡糖苷酶的实例包括,但不限于,来自棘孢曲霉 (Kawaguchi等人,1996,Gene 173:287-288)、烟曲霉 (WO 2005/047499)、黑曲霉 (Dan等人, 2000,J. Biol. Chem. 275:4973-4980)、米曲霉 (WO 2002/095014)、巴西青霉 (*Penicillium brasilianum*) IBT 20888 (WO 2007/019442和WO 2010/088387)、太瑞斯梭孢壳 (WO 2011/035029) 和囊状长毛盘菌 (WO 2007/019442) 的 β -葡糖苷酶。

[0183] β -葡糖苷酶还可以是融合蛋白。在一个方面, β -葡糖苷酶是米曲霉 β -葡糖苷酶变体BG融合蛋白 (WO 2008/057637) 或米曲霉 β -葡糖苷酶融合蛋白 (WO 2008/057637)。

[0184] 使用根据Henrissat B.,1991,A classification of glycosyl hydrolases based on amino-acid sequence similarities,Biochem.J.280:309-316、及 Henrissat B. and Bairoch A.,1996,Updating the sequence-based classification of glycosyl

hydrolases, Biochem. J. 316:695-696所述的分类,其他内切葡聚糖酶、纤维二糖水解酶、和 β -葡萄糖苷酶的实例在众多的糖基水解酶家族中公开。

[0185] 可以用于本发明的其他纤维素分解酶在WO 98/13465、WO 98/015619、WO 98/015633、WO 99/06574、WO 99/10481、WO 99/025847、WO 99/031255、WO 2002/101078、WO 2003/027306、WO 2003/052054、WO 2003/052055、WO 2003/052056、WO 2003/052057、WO 2003/052118、WO 2004/016760、WO 2004/043980、WO 2004/048592、WO 2005/001065、WO 2005/028636、WO 2005/093050、WO 2005/093073、WO 2006/074005、WO 2006/117432、WO 2007/071818、WO 2007/071820、WO 2008/008070、WO 2008/008793、美国专利第5,457,046号、美国专利第5,648,263号和美国专利第 5,686,593号中有过记载。

[0186] 作为具有生物活性的多肽的具有纤维素分解增强活性的GH61多肽的实例包括,但不限于,来自太瑞斯梭孢壳(WO 2005/074647、WO 2008/148131、和WO 2011/035027)、嗜热子囊菌(*Thermoascus aurantiacus*) (WO 2005/074656和WO 2010/065830)、里氏木霉(WO 2007/089290)、嗜热毁丝菌(WO 2009/085935、WO 2009/085859、WO 2009/085864、WO 2009/085868)、烟曲霉(WO 2010/138754)的GH61多肽,来自嗜松青霉(*Penicillium pinopuilum*) (WO 2011/005867)、嗜热子囊菌(WO 2011/039319)、青霉(WO 2011/041397)、和甲壳嗜热子囊菌(*Thermoascus crustaceus*) (WO 2011/041504)的GH61多肽。

[0187] 作为具有生物活性的多肽的木聚糖酶的实例包括,但不限于,来自棘孢曲霉(GeneSeqP:AAR63790;WO 94/21785)、烟曲霉(WO 2006/078256)、嗜松青霉(WO 2011/041405)、青霉(WO 2010/126772)、太瑞斯梭孢壳NRRL 8126(WO 2009/079210)、以及囊状长毛盘菌GH10 (WO 2011/057083)的木聚糖酶。

[0188] 作为具有生物活性的多肽的 β -木糖苷酶的实例包括,但不限于,来自粗糙脉孢菌(SwissProt登录号Q7SOW4)、里氏木霉(UniProtKB/TrEMBL登录号Q92458)、和埃默森篮状菌(*Talaromyces emersonii*) (SwissProt登录号Q8X212)的 β -木糖苷酶。

[0189] 作为具有生物活性的多肽的乙酰木聚糖酯酶的实例包括,但不限于,来自棘孢曲霉(WO 2010/108918)、球毛壳菌(*Chaetomium globosum*) (UniProt登录号Q2GWX4)、细丽毛壳菌(*Chaetomium gracile*) (GeneSeqP登录号AAB82124)、特异腐质霉DSM 1800(WO 2009/073709)、红褐肉座菌(*Hypocrea jecorina*) (WO 2005/001036)、嗜热毁丝菌(WO 2010/014880)、粗糙脉孢菌(UniProt登录号q7s259)、颖枯壳针孢(*Phaeosphaeria nodorum*) (UniProt登录号Q0UJH1)、和太瑞斯梭孢壳NRRL 8126(WO 2009/042846)的乙酰木聚糖酯酶。

[0190] 作为具有生物活性的多肽的阿魏酸酯酶的实例包括,但不限于,来自特异腐质霉DSM 1800(WO 2009/076122)、费希新萨托菌(*Neosartorya fischeri*) (UniProt登录号A1D9T4)、粗糙脉孢菌(UniProt 登录号Q9HGR3)、橘灰青霉(*Penicillium aurantiogriseum*) (WO 2009/127729)、和太瑞斯梭孢壳(WO 2010/053838和WO 2010/065448)的阿魏酸酯酶。

[0191] 作为具有生物活性的多肽的阿拉伯呋喃糖苷酶的实例包括,但不限于,来自黑曲霉(GeneSeqP登录号AAR94170)、特异腐质霉DSM 1800(WO 2006/114094和WO 2009/073383)、和大型亚灰树花菌(*M. giganteus*) (WO 2006/114094)的阿拉伯呋喃糖苷酶。

[0192] 作为具有生物活性的多肽的 α -葡萄糖醛酸酶的实例包括,但不限于,来自棒曲霉

(*Aspergillus clavatus*) (UniProt登录号alcc12)、烟曲霉 (SwissProt登录号Q4WW45)、黑曲霉 (UniProt登录号Q96WX9)、土曲霉 (*Aspergillus terreus*) (SwissProt登录号Q0CJP9)、特异腐质霉 (WO 2010/014706)、橘灰青霉 (WO 2009/068565)、埃默森篮状菌 (UniProt登录号Q8X211)、和里氏木霉 (UniProt登录号Q99024) 的 α -葡糖醛酸酶。

[0193] 登录号通过引用的方式整体并入本文。

[0194] 表达载体

[0195] 本发明还涉及包含本发明的串联构建体的表达载体。串联构建体可以插入到载体中或者串联构建体的多个组分可以结合在一起以产生重组表达载体。载体可以包含一个或多个(例如,几个)方便的限制性酶切位点以使得在这些位点插入多核苷酸。在表达载体的创建中,将编码序列置于载体中,从而将编码序列与适当的用于表达的控制序列可操作地连接。

[0196] 重组表达载体可以是任何载体(例如,质粒或病毒),其能够方便地进行重组DNA步骤,并且能够产生多核苷酸的表达。载体的选择将通常依赖于载体与将导入载体的宿主细胞的相容性。载体可以是线状或闭合环状质粒。

[0197] 载体优选地含有一个或多个(例如,几个)可选择性标记物,其使得可以方便选择所转化的细胞。用于丝状真菌宿主细胞的可选择性标记物的实例包括,但不限于,adeA(磷酸核糖基氨基咪唑-琥珀胺合酶)、adeB(磷酸核糖基氨基咪唑合酶)、amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰基转移酶)、bar(蒴丝菌素乙酰转移酶)、hph(潮霉素磷酸转移酶)、niaD(硝酸还原酶)、pyrG(乳清酸核苷-5'-磷酸脱羧酶)、sC(硫酸腺苷酰转移酶)、和trpC(邻氨基苯甲酸合成酶)、以及其等价物。优选用在曲霉属细胞中的是构巢曲霉或米曲霉amdS和pyrG基因以及吸水链霉菌bar基因。优选用在木霉属细胞中的是adeA、adeB、amdS、hph和pyrG基因。细菌可选择性标记物的实例是赋予抗生素抗性的标记物,例如氨苄青霉素、氯霉素、卡那霉素、新霉素、壮观霉素、或四环素抗性。

[0198] 用于连接上述元件以构建重组表达载体的方法为本领域技术人员公知(参见,例如,Sambrook等人,1989,同上)。

[0199] 丝状真菌宿主细胞

[0200] 本发明还涉及丝状真菌菌株,其包含:(a)通过靶向整合用第一串联构建体取代的内源性第一基因,该第一串联构建体包含(i)第一基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个(例如,几个)第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)第一基因的同源3'区、其同源侧翼区或其组合;(b)通过靶向整合用第二串联构建体取代的内源性第二基因,该第二串联构建体包含(i)第二基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个(例如,几个)第二可选择性标记物,(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)第二基因的同源3'区、其同源侧翼区或其组合;或(c) (a)和(b)的组合。

[0201] 本发明还涉及丝状真菌菌株,其包含:(a)通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座,该第一串联构建体包含(i)第一基因座的同源5'区、其同源侧

翼区或其组合, (ii) 一个或多个第一可选择性标记物, (iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸, (iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸, 和 (v) 第一基因座的同源 3' 区、其同源侧翼区或其组合; (b) 通过靶向整合通过经插入第二串联构建体而修饰的内源性第二基因座, 该第二串联构建体包含 (i) 第二基因座的同源 5' 区、其同源侧翼区或其组合, (ii) 一个或多个第二可选择性标记物, (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 第二基因座的同源 3' 区、其同源侧翼区或其组合; 或 (c) (a) 和 (b) 的组合。

[0202] 术语“宿主细胞”包括由于在复制过程中发生突变而与亲本细胞不同的亲本细胞的任何子代。宿主细胞的选择将在很大程度上取决于编码多肽的基因及其来源。

[0203] 宿主细胞可以是在多肽的重组生产中有用的任何丝状真菌细胞。“丝状真菌”包括真菌门和卵菌门的亚门(如由Hawksworth等人, 1995, 同上, 所定义)的所有丝状形式。丝状真菌通常的特征在于由壳质、纤维素、葡聚糖、壳聚糖、甘露聚糖和其他复合多糖组成的菌丝体壁。营养生长是通过菌丝伸长, 碳分解代谢必须是需氧的。相反, 酵母例如酿酒酵母(*Saccharomyces cerevisiae*)的营养生长是通过单细胞菌体的出芽, 且碳分解可以是发酵性的。

[0204] 丝状真菌宿主细胞可以是枝顶孢属(*Acremonium*)、曲霉属(*Aspergillus*)、短梗霉属(*Aureobasidium*)、黑管菌属(*Bjerkandera*)、拟蜡霉属(*Ceriporiopsis*)、金孢属(*Chrysosporium*)、鬼伞属(*Coprinus*)、革盖菌属(*Coriolus*)、隐球菌属(*Cryptococcus*)、线黑粉酵母属(*Filibasidium*)、镰孢菌属(*Fusarium*)、腐质霉属(*Humicola*)、巨座壳属(*Magnaporthe*)、毛霉属(*Mucor*)、毁丝菌属(*Myceliophthora*)、新美鞭菌属(*Neocallimastix*)、脉孢菌属(*Neurospora*)、拟青霉属(*Paecilomyces*)、青霉属(*Penicillium*)、平革菌属(*Phanerochaete*)、白腐菌属(*Phlebia*)、梨囊鞭菌属(*Piromyces*)、侧耳属(*Pleurotus*)、裂褶菌属(*Schizophyllum*)、篮状菌属(*Talaromyces*)、嗜热子囊菌属(*Thermoascus*)、梭孢壳属(*Thielavia*)、弯颈霉属(*Tolypocladium*)、栓菌属(*Trametes*)或木霉属(*Trichoderma*)细胞。

[0205] 例如, 丝状真菌宿主细胞可以是泡盛曲霉(*Aspergillus awamori*)、臭曲霉(*Aspergillus foetidus*)、烟曲霉、日本曲霉(*Aspergillus japonicus*)、构巢曲霉、黑曲霉、米曲霉、烟管菌(*Bjerkandera adusta*)、干拟蜡菌(*Ceriporiopsis aneirina*)、*Ceriporiopsis caregiea*、浅黄拟蜡菌(*Ceriporiopsis gilvescens*)、*Ceriporiopsis pannocinta*、*Ceriporiopsis rivulosa*、*Ceriporiopsis subrufa*、虫拟蜡菌(*Ceriporiopsis subvermispora*)、*Chrysosporium inops*、嗜毛金色孢(*Chrysosporium keratinophilum*)、*Chrysosporium lucknowense*、粪状金孢(*Chrysosporium merdarium*)、*Chrysosporium pannicola*、*Chrysosporium queenslandicum*、热带金孢菌(*Chrysosporium tropicum*)、*Chrysosporium zonatum*、灰盖鬼伞(*Coprinus cinereus*)、粗毛云芝(*Coriolus hirsutus*)、杆孢状镰孢(*Fusarium bactridioides*)、谷类镰孢(*Fusarium cerealis*)、克鲁克威尔镰孢菌(*Fusarium crookwellense*)、黄色镰孢菌(*Fusarium culmorum*)、禾谷镰孢菌(*Fusarium graminearum*)、禾赤镰孢(*Fusarium*

graminum)、异孢镰孢(*Fusarium heterosporum*)、合欢木镰孢(*Fusarium negundi*)、尖孢镰孢、多枝镰孢(*Fusarium reticulatum*)、粉红镰孢(*Fusarium roseum*)、接骨木镰孢(*Fusarium sambucinum*)、肤色镰孢(*Fusarium sarcochroum*)、拟枝孢镰孢菌(*Fusarium sporotrichioides*)、硫色镰孢(*Fusarium sulphureum*)、圆镰孢(*Fusarium torulosum*)、拟丝孢镰孢(*Fusarium trichothecioides*)、镶片镰孢、特异腐质霉、柔毛腐质霉(*Humicola lanuginosa*)、米赫毛霉(*Mucor miehei*)、嗜热毁丝霉、粗糙脉胞菌、产紫青霉(*Penicillium purpurogenum*)、黄孢原毛平革菌(*Phanerochaete chrysosporium*)、脉射菌(*Phlebia radiata*)、刺芹侧耳(*Pleurotus eryngii*)、太瑞斯梭孢壳、长绒毛栓菌(*Trametes villosa*)、变色栓菌(*Trametes versicolor*)、哈茨木霉(*Trichoderma harzianum*)、康氏木霉(*Trichoderma koningii*)、长梗木霉(*Trichoderma longibrachiatum*)、里氏木霉或绿色木霉(*Trichoderma viride*)细胞。

[0206] 在一个方面,丝状真菌宿主细胞是米曲霉。在另一方面,丝状真菌宿主细胞是黑曲霉。在另一方面,丝状真菌宿主细胞是镶片镰孢。在另一方面,丝状真菌宿主细胞是里氏木霉。在另一方面,丝状真菌宿主细胞是长梗木霉。

[0207] 在另一方面,丝状真菌宿主细胞是里氏木霉RutC30。在另一方面,丝状真菌宿主细胞是里氏木霉TV10。在另一方面,丝状真菌宿主细胞是里氏木霉RutC30的突变体。在另一方面,丝状真菌宿主细胞是里氏木霉TV10的突变体。在另一方面,丝状真菌宿主细胞是里氏木霉的形态突变体。参见,例如,WO 97/26330,其通过引用的方式整体并入本文。

[0208] 在另一方面,丝状真菌宿主细胞是包含一个或多个(例如,几个)选自第一枯草杆菌蛋白酶样丝氨酸蛋白酶基因、第一天冬氨酸蛋白酶基因、胰蛋白酶样丝氨酸蛋白酶基因、第二枯草杆菌蛋白酶样丝氨酸蛋白酶基因、和第二天冬氨酸蛋白酶基因的基因的木霉属菌株,其中一个或多个(例如,几个)基因修饰成使突变体菌株当在相同条件下培养时,相对于亲本木霉属菌株,分别在一种或多种(例如,几种)选自第一枯草杆菌蛋白酶样丝氨酸蛋白酶、第一天冬氨酸蛋白酶、胰蛋白酶样丝氨酸蛋白酶、第二枯草杆菌蛋白酶样丝氨酸蛋白酶、和第二天冬氨酸蛋白酶的酶的产生方面有缺陷,如WO 2011/075677中所述,其通过引用的方式整体并入本文。

[0209] 可以将丝状真菌细胞通过涉及原生质体形成、原生质体转化、和细胞壁再生的方法以本身已知的方式进行转化。用于转化曲霉属和木霉属宿主细胞的适当方法在EP 238023、Yelton等人,1984,Proc.Natl. Acad.Sci.USA 81:1470-1474以及Christensen等人,1988,Bio/Technology 6:1419-1422中有过记载。用于转化镰孢属的种的适当方法由Malardier 等人,1989,Gene 78:147-156和WO 1996/00787记载。

[0210] 产生的方法

[0211] 本发明还涉及在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法,其包括:

[0212] (A)在有益于产生多肽的条件下培养丝状真菌宿主细胞,其中该丝状真菌宿主细胞包含:(a)通过靶向整合用第一串联构建体取代的内源性第一基因,该第一串联构建体包含(i)第一基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个(例如,几个)第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的

第二多肽的第二多核苷酸,和(v) 第一基因的同源3' 区、其同源侧翼区或其组合;(b) 通过靶向整合用第二串联构建体取代的内源性第二基因,该第二串联构建体包含(i) 第二基因的同源5' 区、其同源侧翼区或其组合,(ii) 一个或多个(例如,几个)第二可选择性标记物,(iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v) 第二基因的同源3' 区、其同源侧翼区或其组合;或(c) (a) 和(b) 的组合;以及任选地

[0213] (B) 回收多种重组多肽。

[0214] 本发明还涉及在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法,其包括:

[0215] (A) 在有益于产生多肽的条件下培养丝状真菌宿主细胞,其中该丝状真菌宿主细胞包含:(a) 通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座,该第一串联构建体包含(i) 第一基因座的同源5' 区、其同源侧翼区或其组合,(ii) 一个或多个第一可选择性标记物,(iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v) 第一基因座的同源3' 区、其同源侧翼区或其组合;(b) 通过靶向整合经插入第二串联构建体而修饰的内源性第二基因座,该第二串联构建体包含(i) 第二基因座的同源5' 区、其同源侧翼区或其组合,(ii) 一个或多个第二可选择性标记物,(iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v) 第二基因座的同源3' 区、其同源侧翼区或其组合;或(c) (a) 和(b) 的组合;以及任选地

[0216] (B) 回收多种重组多肽。

[0217] 使用本领域已知的方法,在适于产生多肽的营养介质中培养丝状真菌宿主细胞。例如,细胞可以通过在合适的介质中并在使得多肽进行表达和/或分离的条件下进行摇瓶培养或在实验室或工业发酵罐中进行小规模或大规模发酵(包括连续发酵、分批发酵、分批补料发酵或固态发酵)来培养。使用本领域已知的步骤,培养在包含碳源和氮源以及无机盐的合适营养介质中进行。合适的介质可以从商业供应商获得或可以根据公开的组成(例如,在美国典型培养物保藏中心的目录中)来制备。如果多肽被分泌到营养介质中,则可以直接从介质中回收多肽。如果多肽未被分泌,则可以从细胞裂解物中回收。

[0218] 多肽可以使用领域内已知的特定用于多肽的方法来检测。这些检测方法包括但不限于特异性抗体的使用、酶产物的形成、或酶底物的消失。例如,酶检定可以用来判定多肽的活性。

[0219] 多肽可以使用本领域已知的方法来回收。例如,多肽可以通过常规步骤从营养介质中回收,这些常规步骤包括但不限于收集、离心、过滤、萃取、喷雾干燥、蒸发或沉淀。在一个方面,回收整个发酵液。

[0220] 多肽可以通过本领域已知的多种方法进行纯化,以获得基本纯的多肽,方法包括但不限于层析法(例如,离子交换层析、亲和层析、疏水性层析、聚焦层析、和尺寸排阻层析)、电泳法(例如,制备式等电聚焦)、差别溶解度(例如,硫酸铵沉淀)、SDS-PAGE、或萃取

(参见例如,Protein Purification,Janson和Ryden编,VCH Publishers,New York,1989)。

[0221] 本发明通过以下实施例进一步加以说明,这些实施例不应理解成对本发明的范围加以限定。

[0222] 实施例

[0223] 菌株

[0224] 里氏木霉菌株981-0-8 (D4) 是里氏木霉RutC30 (ATCC 56765;Montenecourt and Eveleigh,1979,Adv.Chem.Ser.181:289-301) 的诱变菌株。

[0225] 里氏木霉菌株AgJg115-104-7B1 (PCT/US2010/061105、WO 2011/075677) 是里氏木霉菌株981-0-8 (D4) 的ku70-衍生体。

[0226] 培养基和缓冲溶液

[0227] LB平板由10g胰蛋白胨、5g酵母提取物、5g NaCl、15g Bacto 琼脂和加至1升的去离子水组成。

[0228] 2XYT+氨苄青霉素平板由16g胰蛋白胨、10g酵母提取物、5g 氯化钠、15g Bacto琼脂和加至1升去离子水组成。在高压灭菌的培养基冷却至55℃之后,加入1ml 100mg/ml的氨苄青霉素溶液。

[0229] SOC培养基由20g Bacto-胰蛋白胨、5g Bacto酵母提取物、0.5g NaCl、2.5ml 1M KCl和加至1升的去离子水组成。在高压灭菌之前,用10N NaOH将pH调节至7.0。然后在临用前加入20ml的无菌1M 葡萄糖。

[0230] COVE盐溶液由26g KCl、26g $MgSO_4 \cdot 7H_2O$ 、76g KH_2PO_4 、50ml COVE痕量金属溶液和加至1升的去离子水组成。

[0231] COVE痕量金属溶液由0.04g $NaB_4O_7 \cdot 10H_2O$ 、0.4g $CuSO_4 \cdot 5H_2O$ 、1.2g $FeSO_4 \cdot 7H_2O$ 、0.7g $MnSO_4 \cdot H_2O$ 、0.8g $Na_2MoO_4 \cdot 2H_2O$ 、10g $ZnSO_4 \cdot 7H_2O$ 和加至1升的去离子水组成。

[0232] COVE平板由342.3g蔗糖、20ml COVE盐溶液、10ml 1M乙酰胺、10ml 1.5M CsCl、25g Noble琼脂(Difco) 和加至1升的去离子水组成。

[0233] COVE2平板由30g蔗糖、20ml COVE盐溶液、10ml 1M乙酰胺、25g Noble琼脂(Difco) 和加至1升的去离子水组成。

[0234] 木霉属痕量金属溶液由216g $FeCl_3 \cdot 6H_2O$ 、58g $ZnSO_4 \cdot 7H_2O$ 、27g $MnSO_4 \cdot H_2O$ 、10g $CuSO_4 \cdot 5H_2O$ 、2.4g H_3BO_3 、336g柠檬酸和加至1 升的去离子水组成。

[0235] CIM培养基由20g纤维素、10g玉米浆冻干粉、1.45g $(NH_4)_2SO_4$ 、2.08g KH_2PO_4 、0.28g $CaCl_2$ 、0.42g $MgSO_4 \cdot 7H_2O$ 、0.42ml木霉属痕量金属溶液、1~2滴消泡剂和加至1升的去离子水组成;pH调节至 6.0。

[0236] YP培养基由10g酵母提取物、20g Bacto蛋白胨和加至1升的去离子水组成。

[0237] YPG培养基由4g酵母提取物、1g K_2HPO_4 、0.5g $MgSO_4$ 、15.0g 葡萄糖和加至1升的去离子水组成(pH 6.0)。

[0238] PEG缓冲液由500g聚乙二醇4000 (PEG 4000)、10mM $CaCl_2$ 、10mM Tris-HCl pH 7.5和加至1升的去离子水组成;过滤除菌。

[0239] PDA平板由39g土豆右旋糖琼脂(Difco) 和加至1升的去离子水组成。

[0240] PDA覆层培养基由39g土豆右旋糖琼脂(Difco)、2.44g尿苷和加至1升的去离子水组成。使用之前,将高压灭菌的培养基在微波中融化,然后冷却至55℃。

- [0241] STC由1M山梨糖醇、10mM mM CaCl₂和10mM Tris-HCl,pH 7.5 组成;过滤除菌。
- [0242] TE缓冲液由1M Tris pH 8.0和0.5M EDTApH 8.0组成。
- [0243] 变性溶液由0.5M NaOH和1.5M NaCl组成。
- [0244] 中和溶液由1M Tris pH 8.0和1.5M NaCl组成。
- [0245] 20X SSC由175.3g NaCl、88.2g柠檬酸钠和加至1升的去离子水组成。
- [0246] TrMM-G培养基由20ml COVE盐溶液、6g (NH₄)₂SO₄、0.6g CaCl₂、25g Nobel琼脂(Difco)、20g葡萄糖和加至1升的去离子水组成。
- [0247] NZY+培养基由5g NaCl、3g MgSO₄·7H₂O、5g酵母提取物、10g NZ胺、1.2g MgCl₂、4g葡萄糖和加至1升的去离子水组成。
- [0248] 实施例1:烟曲霉GH61B多肽基因的克隆
- [0249] 烟曲霉部分基因组序列(The Institute for Genomic Research, Rockville, MD, USA)的tblastn搜索(Altschul等人,1997,Nucleic Acids Res.25:3389-3402)使用几种已知的GH61多肽(包括嗜热子囊菌 GH61A多肽(GeneSeqP登录号AEC05922))作为查询序列来进行。基于在氨基酸水平上与查询序列的高度类似性,数个基因被识别为推定的家族GH61同源物。选择一个约850bp的与嗜热子囊菌GH61A多肽序列在氨基酸水平具有高于70%序列同一性的基因组区域进行进一步研究。
- [0250] 烟曲霉NN051616如美国专利第7,244,605号中所述生长并收获。将冷冻的菌丝体用研钵和研杵研磨成细粉末,并使用DNEASY®植物大量提取试剂盒(DNEASY® Plant Maxi Kit)(QIAGEN Inc.,Valencia, CA,USA)根据生产商的说明来分离基因组DNA。
- [0251] 设计以下示出的两个合成的寡核苷酸引物,以从基因组DNA中 PCR扩增出烟曲霉GH61B多肽编码序列。使用IN-FUSION®克隆试剂盒(Clontech Laboratories,Inc., Mountain View,CA,USA)将片段直接克隆入表达载体pAlLo2(WO 2004/099228),而无需进行限制性酶切消化和连接。
- [0252] 正向引物:
- [0253] 5'-ACTGGATTTACCATGACTTTGTCCAAGATCACTTCCA-3' (SEQ ID NO:31)
- [0254] 反向引物:
- [0255] 5'-TCACCTCTAGTTAATTAAGCGTTGAACAGTGCAGGACCA G-3' (SEQ ID NO:32)
- [0256] 粗体字母代表编码序列。剩余序列与pAlLo2的插入位点同源。
- [0257] 将五十皮摩尔的各个上述引物用于PCR反应中,反应由204ng的烟曲霉基因组DNA,1×Pfx扩增缓冲液(Invitrogen,Carlsbad,CA, USA),1.5μl的dATP、dTTP、dGTP和dCTP的10mM混合物,2.5 个单位的PLATINUM®Pfx DNA聚合酶(Invitrogen Corp.,Carlsbad, CA, USA),和1μl的50mM MgSO₄构成,终体积为50μl。扩增使用EPPENDORF® MASTERCYCLER®5333epgradient S(Eppendorf Scientific, Inc.,Westbury,NY,USA)进行,程序为:94℃3分钟,1个循环;和每循环94℃30秒、56℃30秒、和72℃1分钟,30个循环。然后将加热块维持在72℃15分钟,接着进行4℃浸泡循环。反应产物通过使用40mM Tris碱、20mM乙酸钠、1mM EDTA二钠盐(TAE)缓冲液的1.0%琼脂糖凝胶电泳进行分离,其中将约850bp的产物条带从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒(QIAGEN Inc., Valencia,CA,USA)根据生产商的说明进行纯化。
- [0258] 然后使用IN-FUSION®克隆试剂盒将850bp片段克隆入pAlLo2。将质粒pAlLo2

用Nco I和Pac I消化。将质粒片段通过如上的凝胶电泳和QIAQUICK®凝胶纯化试剂盒(QIAGEN Inc.,Valencia,CA,USA) 进行纯化。将基因片段和消化的载体在后述的反应中合并在一起,得到表达质粒pAG43(图1),其中烟曲霉GH61B多肽编码序列的转录处于NA2-tpi启动子的调控之下。NA2-tpi启动子是来自黑曲霉中性 α -淀粉酶基因的修饰启动子,其中非翻译的前导序列被来自构巢曲霉磷酸丙糖异构酶基因的非翻译前导序列替代。重组反应(20 μ l)由 1 \times IN-FUSION®反应缓冲液(Clontech Laboratories,Inc.,Mountain View, CA,USA)、1 \times BSA(Clontech Laboratories,Inc.,Mountain View,CA, USA)、1 μ l的IN-FUSION®酶(稀释1:10)(Clontech Laboratories,Inc., Mountain View,CA,USA)、166ng的用Nco I和Pac I消化的pAlLo2、和110ng的烟曲霉GH61B多肽的纯化PCR产物构成。将反应在37℃下孵育15分钟,接着在50℃孵育15分钟。将反应物用40 μ l的10mM Tris-0.1M EDTA缓冲液稀释,并根据生产商的说明将2.5 μ l的稀释反应物用来转化大肠杆菌(E.coli)XL10SOLOPACK®Gold感受态细胞(Stratagene,La Jolla,CA,USA)。含有pAG43(GH61B多肽编码序列)的大肠杆菌转化体通过限制性酶切消化进行鉴定,并使用BIOROBOT®9600(QIAGEN Inc.,Valencia,CA,USA)制备质粒DNA。

[0259] 862bp PCR片段的DNA测序用Applied Biosystems Model 377XL 自动DNA测序仪(Applied Biosystems,Carlsbad,CA,USA)使用染料-终止子化学法(Giesecke等人,1992,Journal of Virology Methods 38:47-60)和引物步移策略进行。使用下述的载体特异性引物进行测序:

[0260] pAllo2 5Seq:

[0261] 5'-TGTCCTTGTCGATGCG 3' (SEQ ID NO:33)

[0262] pAllo23Seq:

[0263] 5'-CACATGACTTGGCTTCC 3' (SEQ ID NO:34)

[0264] 仔细检查核苷酸序列数据的质量,并将所有序列经 PHRED/PHRAP软件(University of Washington,Seattle,WA,USA)的协助而进行彼此比较。

[0265] 基于所编码的蛋白与嗜热子囊菌GH61A蛋白(GeneSeqP登录号 AEC05922)的相似性来构建用于烟曲霉序列的基因模型。烟曲霉 GH61B多肽编码序列的核苷酸序列和推导的氨基酸序列分别示于SEQ ID NO:35和SEQ ID NO:36。基因组片段编码250个氨基酸的多肽,被2个53和56bp的内含子打断。编码序列和成熟编码序列的G+C含量百分比分别为53.9%和57%。使用SignalP软件程序(Nielsen等人,1997,Protein Engineering 10:1-6),预测21个残基的信号肽。预测的成熟蛋白含有221个氨基酸,其具有23.39kDa的预测分子量。

[0266] 实施例2:用于表达烟曲霉GH61B多肽的pSMai214的构建

[0267] 使用如下所示的基因特异性的正向和反向引物从质粒pAG43(实施例1)中扩增烟曲霉GH61B多肽编码序列。斜体的区域表示与IN-FUSION®反应的插入位点的载体同源性。

[0268] 正向引物:

[0269] 5'-GGACTGCGCACCATGACTTTGTCCAAGATCACTTCCA-3' (SEQ ID NO:37)

[0270] 反向引物:

[0271] 5'-GCCACGGAGCTTAATTAATTAAGCGTTGAACAGTGCAG-3' (SEQ ID NO:38)

[0272] 将五十皮摩尔的各上述引物用于PCR反应中,该反应由10ng的 pAG43DNA,1×Pfx 扩增缓冲液,1.5μl的dATP、dTTP、dGTP和dCTP 的10mM混合物,2.5个单位的**PLATINUM®** Pfx DNA聚合酶,和1μl 的50mM MgSO₄构成,最终体积为50μl。扩增使用**EPENDORF®** **MASTERCYCLER®**5333epgradient S进行,其程序为:98℃3分钟,1 个循环;和每循环98℃30秒、56℃30秒、和72℃1分钟,30个循环。然后将加热块维持在72℃15分钟。PCR产物通过使用TAE缓冲液的1%琼脂糖凝胶电泳进行分离,其中将约0.9kb的片段从凝胶切出,并使用**MINELUTE®**凝胶提取试剂盒根据生产商的操作指南进行提取。

[0273] 将质粒pMJ09 (WO 2005/047499)用Nco I和Pac I消化,通过在 1mM EDTA二钠盐-50mM Tris碱-50mM硼酸 (TBE) 缓冲液中的1.0%琼脂糖凝胶电泳来分离,从凝胶中切出,并根据生产商的说明使用**QIAQUICK®**凝胶提取试剂盒 (QIAGEN Inc.,Valencia,CA,USA) 提取。

[0274] 根据生产商的操作指南,使用**IN-FUSION®**PCR克隆试剂盒将0.9 kb的PCR产物插入到凝胶纯化的经Nco I/Pac I消化的pMJ09中。**IN-FUSION®**反应由1X **IN-FUSION®**反应缓冲液、100ng的凝胶纯化的经Nco I/Pac I消化的pMJ09、37ng的0.9kb PCR产物、2μl的500 μg/ml BSA、和1μl的**IN-FUSION®**酶构成,反应体积为20μl。将反应在37℃孵育15分钟并在50℃孵育15分钟。在孵育期后,将30μl 的TE缓冲液加入反应中。根据制造商的操作指南,将2.5μl等份用于转化**SOLOPACK®**Gold超级感受态细胞 (Agilent Technologies,Inc., Cedar Creek,TX,USA)。通过测序来筛选转化体,鉴定出含有无PCR 错误的插入的一个克隆并将其定名为pSMai214 (图2)。可以用Pme I 消化质粒pSMai214,以产生用于里氏木霉转化的约5.4kb的片段。该 5.4kb的片段包含由里氏木霉Cel7A纤维二糖水解酶I基因启动子、烟曲霉GH61B多肽编码序列、里氏木霉Cel7A纤维二糖水解酶I基因终止子、和构巢曲霉乙酰胺酶 (amdS) 基因组成的表达盒。

[0275] 实施例3:用于表达烟曲霉CEL3AB-葡糖苷酶和烟曲霉GH61B多肽两者的串联构建体pDM287的构建

[0276] 使用如下所示的基因特异性的正向和反向引物从质粒pSMai214 中扩增烟曲霉GH61B多肽表达盒。斜体的区域表示与**IN-FUSION®**反应的插入位点的载体同源性。

[0277] 正向引物:

[0278] 5'-CGCGGTAGTGGCGCGGTCGACCGAATGTAGGATTGTT-3' (SEQ ID NO:39)

[0279] 反向引物:

[0280] 5'-TTACCAATTGGCGCGCCACTACCGGTTTCGAGAAGA-3' (SEQ ID NO:40)

[0281] 将五十皮摩尔的各上述引物用于PCR反应中,该反应由25ng的 pSMai214DNA,1X PHUSION™高保真热启动DNA聚合酶缓冲液 (Finnzymes Oy,Espoo,Finland),1μl的dATP、dTTP、dGTP和dCTP 的10mM混合物,以及1个单位的PHUSION™高保真热启动DNA聚合酶 (Finnzymes Oy,Espoo,Finland) 构成,最终体积为50μl。扩增使用**EPENDORF®** **MASTERCYCLER®**5333epgradient S进行,其程序为:98℃30秒,1个循环;每循环98℃10秒、60℃30秒、和72℃1 分30秒,35个循环;和72℃10分钟,1个循环。PCR产物通过使用 TAE缓冲液的0.8%琼脂糖凝胶电泳进行分离,其中将约2.3kb的片段从凝胶中切出,并使用**NUCLEOSPIN®**提取II试剂盒 (Macherey-Nagel, Inc.,

Bethlehem, PA, USA) 根据生产商的操作指南进行提取。

[0282] 将约2.3kb的PCR产物使用**IN-FUSION®** Advantage PCR克隆试剂盒(Clontech Laboratories, Inc., Mountain View, CA, USA) 根据生产商的操作指南而插入至经Asc I消化的pEJG107 (WO 2005/047499)。质粒pEJG107包含烟曲霉CEL3A β -葡糖苷酶编码序列(SEQ ID NO: 53 [DNA序列] 和SEQ ID NO: 54 [推导的氨基酸序列])。该**IN-FUSION®**反应由1X **IN-FUSION®**反应缓冲液、125ng的经Asc I消化的 pEJG107、90ng的2.33kb的PCR产物、和1 μ l **IN-FUSION®**酶组成, 反应体积为10 μ l。将反应在37℃孵育15分钟, 接着在50℃孵育15分钟。在孵育期后, 将40 μ l的TE加入反应中。根据制造商的操作指南, 将2 μ l等份用于转化**ONESHOT®** TOP10感受态细胞(Invitrogen, Carlsbad, CA, USA)。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。通过测序来筛选转化体, 鉴定出含有无PCR错误的插入的一个克隆并将其命名为pDM287(图3)。可以用Pme I消化质粒pDM287, 产生用于里氏木霉转化的约9.9kb的片段。该9.9kb的片段包含由(1) 里氏木霉Cel17A纤维二糖水解酶I基因启动子、烟曲霉CEL3A β -葡糖苷酶编码序列、和里氏木霉Cel17A纤维二糖水解酶I基因终止子; 以及(2) 里氏木霉Cel17A纤维二糖水解酶I基因启动子、烟曲霉GH61B 多肽编码序列、和里氏木霉Cel17A纤维二糖水解酶I基因终止子组成的两个表达盒。该9.9kb的片段也包含构巢曲霉乙酰胺酶(amdS) 基因。

[0283] 实施例4: 里氏木霉原生质体的产生和转化

[0284] 原生质体的制备和转化使用Penttila等人, 1987, Gene 61:155-164 的修改操作指南来进行。简要地, 在27℃下于25ml的补充有2% (w/v) 葡萄糖和10mM尿苷的YP培养基中在90rpm的轻柔振荡下培养里氏木霉菌株981-0-8 (D4) 17小时。菌丝体使用真空驱动的一次性过滤系统(Millipore, Bedford, MA, USA) 通过过滤来收集, 并用去离子水洗涤两次和用1.2M山梨糖醇洗涤两次。通过在34℃下于20ml的1.2 M山梨糖醇中以90rpm的轻柔振荡来悬浮所洗涤的菌丝体15~25分钟, 从而产生原生质体, 该20ml的1.2M山梨糖醇含有每ml 15mg的**GLUCANEX®** 200G (Novozymes A/S, Bagsvaerd, Denmark) 和每 ml 0.36个单位的几丁质酶(Sigma Chemical Co., St. Louis, MO, USA)。原生质体通过在400 \times g下离心7分钟而收集, 用冷的1.2M山梨糖醇洗涤两次。将原生质体用血球计数器来计数, 并在STC中再悬浮至 1×10^8 原生质体/ml的最终浓度。将过量的原生质体储存在-80℃的Cryo 1℃冷冻容器(Nalgene, Rochester, NY, USA) 中。

[0285] 用Pme I消化约100 μ g的转化质粒(pSMai214、pDM287或 pEJG107)。消化反应物通过使用TAE缓冲液的0.8%琼脂糖凝胶电泳进行纯化。将含有pSMai214、pDM287或pEJG107的表达盒以及构巢曲霉乙酰胺酶(amdS) 基因的DNA条带从凝胶中切出, 并根据制造商建议的操作指南使用**NUCLEOSPIN®**提取II试剂盒进行提取。

[0286] 将所得的纯化的DNA[1 μ g的9.9kb的经Pme I消化的pDM287 (串联转化) 或1 μ g的7.6kb的经Pme I消化的pEJG107加1 μ g的 5.4kb的经Pme I消化的pSMai214 (共转化)] 加至100 μ l的原生质体溶液中并轻柔地混合。加入PEG缓冲液(250 μ l), 将反应物混合并在 34℃下孵育30分钟。然后加入STC(3ml), 并将反应物混合, 然后涂布在用于amdS选择的COVE平板上。将平板在28℃下孵育6~11天。

[0287] 实施例5: 对表达烟曲霉CEL3A β -葡糖苷酶和烟曲霉GH61B多肽的里氏木霉转化体

的评价

[0288] 使用接种环将里氏木霉转化体(实施例4)从COVE转化平板转移至补充了10mM尿苷的COVE2平板,并在28℃下孵育5~7天。用接种环收集孢子,并将其转移至125ml塑料摇瓶中的25ml CIM培养基中。将摇瓶培养物在28℃、200rpm下孵育5天。将各个培养物的1 ml等份在微量离心机中以13,400×g离心,并回收培养物上清液。根据制造商的说明通过使用**CRITERION®**8~16%Tris-HCl凝胶(Bio-Rad Laboratories,Hercules,CA,USA)的SDS-PAGE对5μl的各个培养物上清液进行分析。将所得凝胶用BIO-SAFE™考马斯(Bio-Rad Laboratories, Hercules,CA,USA)染色。45个pDM287转化体(串联构建体)和45个pEJG107+pSMai214转化体(共转化)的SDS-PAGE图谱示出,转化体产生对应于烟曲霉CEL3Aβ-葡糖苷酶的约130kDa以及对应于烟曲霉GH61B多肽的约24kDa的主要蛋白条带。由未转化的里氏木霉菌株981-0-8(D4)培养物上清液构成的阴性对照样本,未示出约130 kDa和约24kDa的显著条带。

[0289] 以下所示的结果表明,串联构建体pDM287的转化比pEJG107和 pSMai214的共转化得到更多的生产烟曲霉β-葡糖苷酶和烟曲霉 GH61B多肽的阳性转化体。

[0290]

转化 DNA	SDS-PAGE 证实的对于烟曲霉 β-葡糖苷酶和烟曲霉 GH61B 多肽生产为阳性的转化体的数目
pDM287 (串联构建体)	45 个中的 33 个 (73%)
pEJG107 + pSMai214 (共转化)	45 个中的 13 个 (29%)

[0291] 实施例6:表达烟曲霉CEL3Aβ-葡糖苷酶和烟曲霉GH61B多肽的里氏木霉转化体的β-葡糖苷酶检定

[0292] 使用**BIOMEK®**3000、**BIOMEK®**NX、和**ORCA®**机械臂(Beckman Coulter, Inc,Fullerton,CA,USA)对实施例5的培养物上清液检定β-葡糖苷酶活性。培养物上清液在0.1M琥珀酸盐、0.01% **TRITON®**X-100(4-(1,1,3,3-四甲基丁基)苯基-聚乙二醇)缓冲液pH 5.0(样本缓冲液)中适当稀释,接着进行稀释样本的0倍到1/3倍再到1/9倍的一系列稀释。将共20μl的各个稀释液转移至96孔平底板。将两百微升的对硝基苯基-β-D-吡喃葡萄糖苷底物溶液(每ml的0.1M琥珀酸盐pH 5.0中,1mg的对硝基苯基-β-D-吡喃葡萄糖苷)加至各个孔中,然后在环境温度下孵育45分钟。孵育期结束后,将50μl的淬灭缓冲液(1M Tris缓冲液pH 9)加至各个孔中。在405nm的光密度测量96孔板的端点。

[0293] 在图4中示出的结果证实了实施例5的SDS-PAGE结果,即用串联构建体pDM287转化比用pEJG107和pSMai214共转化得到更多的生产烟曲霉β-葡糖苷酶和烟曲霉GH61B多肽的阳性转化体。

[0294] 实施例7:构建表达青霉GH61A多肽的pDM286

[0295] 使用以下所示的基因特异性正向和反向引物由质粒pGH61D23Y4(WO 2011/041397)扩增青霉(*Penicillium emersonii*)GH61A多肽编码序列(SEQ ID NO:43[DNA序列]和SEQ ID NO:44[推导的氨基酸序列])。斜体区表示与**IN-FUSION®**反应的插入位点的

载体同源性。

[0296] 正向引物:

[0297] 5'-CGGACTGCGCACCATGCTGTCTTCGACGACTCGCAC-3' (SEQ ID NO:45)

[0298] 反向引物:

[0299] 5'-TCGCCACGGAGCTTATCGACTTCTTCTAGAACGTC-3' (SEQ ID NO:46)

[0300] 扩增反应物由30ng pGH61D23Y4DNA, 50皮摩尔的各上述引物, 1μl的dATP、dTTP、dGTP和dCTP的10mM混合物, 1X PHUSION™高保真热启动DNA聚合酶缓冲液和1个单位的PHUSION™高保真热启动DNA聚合酶组成, 最终体积为50μl。在EPPENDORF® MASTERCYCLER®5333epgradient S中孵育扩增反应物, 程序为: 98℃30秒, 1个循环; 每循环98℃10秒、60℃30秒和72℃30秒, 35个循环; 以及72℃10分钟, 1个循环。通过使用TAE缓冲液的1%琼脂糖凝胶电泳分离PCR产物, 其中将约0.9kb片段从凝胶中切出, 并使用QIAQUICK®凝胶提取试剂盒根据制造商的操作指南提取。

[0301] 将质粒pMJ09 (WO 2005/047499) 用Nco I和Pac I消化, 使用TBE缓冲液通过1.0%琼脂糖凝胶电泳分离, 从凝胶中切出, 并使用QIAQUICK®凝胶提取试剂盒根据制造商的说明进行提取。

[0302] 使用IN-FUSION®Advantage PCR克隆试剂盒根据制造商的操作指南将0.9kb PCR产物插入到凝胶纯化的Nco I/Pac I消化的pMJ09中。IN-FUSION™反应由1X IN-FUSION™反应缓冲液、180ng凝胶纯化的Nco I/Pac I消化的pMJ09、108ng 0.9kb PCR产物和1μl IN-FUSION™酶组成, 反应体积为10μl。将反应物在37℃下孵育15分钟, 然后在50℃下孵育15分钟。孵育期之后, 将40μl TE加入到反应中。根据制造商的操作指南使用2μl等份来转化ONE SHOT®TOP10感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。通过测序来筛选转化体, 识别出含有没有PCR错误的插入的一个克隆, 并命名为pDM286 (图5)。质粒pDM286可以用Pme I消化, 产生约5.4kb片段以用于里氏木霉转化。5.4kb片段含有由里氏木霉Cel7A纤维二糖水解酶I基因启动子、P.emersonii GH61A多肽编码序列和里氏木霉 Cel7A纤维二糖水解酶I基因终止子组成的表达盒。5.4kb片段还含有构巢曲霉乙酰胺酶 (amdS) 基因。

[0303] 实施例8: 用于表达Penicillium emersonii GH61A多肽和烟曲霉CEL3Aβ-葡萄糖苷酶两者的串联构建体pDM290的构建

[0304] 使用以下示出的基因特异性正向和反向引物从质粒pEJG107扩增烟曲霉CEL3A β-葡萄糖苷酶表达盒。斜体区表示与IN-FUSION®反应的插入位点的载体同源性。

[0305] 正向引物:

[0306] 5'-CGCGGTAGTGGCGCGGTCGACCGAATGTAGGATTGTT-3' (SEQ ID NO:47)

[0307] 反向引物:

[0308] 5'-TTACCAATTGGCGCGCCACTACCGGTTTCGAGAAGA-3' (SEQ ID NO:48)

[0309] 在最终体积50μl的PCR反应中使用50皮摩尔的以上各引物, 该PCR反应由25ng pEJG107DNA, 1X PHUSION™高保真热启动DNA聚合酶缓冲液, 1μl的dATP、dTTP、dGTP和dCTP的10mM混合物和1个单位的PHUSION™高保真热启动DNA聚合酶组成。在EPPENDORF® MASTERCYCLER®5333epgradient S中进行扩增, 程序设定如下:

98℃30秒,1个循环;每循环98℃10秒、60℃30秒和 72℃2分钟30秒,35个循环;以及72℃10分钟,1个循环。

[0310] 使用TAE缓冲液通过0.8%琼脂糖凝胶电泳来分离PCR产物,其中将约4.5kb片段从凝胶中切出,并使用**NUCLEOSPIN® Extract II**试剂盒根据制造商的操作指南进行提取。

[0311] 使用**IN-FUSION™ Advantage PCR**克隆试剂盒根据制造商的操作指南将4.5kb PCR产物插入到Asc I消化的pDM286中。**IN-FUSION™**反应由1X **IN-FUSION™**反应缓冲液、125ng Asc I消化的pDM286、100 ng 4.5kb PCR产物和1μl的**IN-FUSION™**酶组成,反应体积为10μl。将反应物在37℃下孵育15分钟,然后在50℃下孵育15分钟。孵育期之后,将40μl TE加入到反应中。使用2μl等份根据制造商的操作指南转化**ONE SHOT® TOP10**感受态细胞。将大肠杆菌转化反应物铺在 2XYT+氨苄青霉素平板上。通过测序来筛选转化体,识别出包含没有PCR错误的插入的一个克隆,并命名为pDM290(图6)。质粒pDM290 可以用Pme I消化,产生约9.9kb片段,以用于里氏木霉转化。9.9kb 片段包含两种表达盒:(1)里氏木霉Cel7A纤维二糖水解酶I基因启动子、*P.emersonii* GH61A多肽编码序列和里氏木霉Cel7A纤维二糖水解酶I基因终止子;和(2)里氏木霉Cel7A纤维二糖水解酶I基因启动子、烟曲霉CEL3Aβ-葡萄糖苷酶编码序列和里氏木霉Cel7A纤维二糖水解酶I基因终止子。9.9kb片段还含有构巢曲霉乙酰胺酶(amdS) 基因。

[0312] 实施例9:空的里氏木霉cbhII取代构建体pJfyS142的构建

[0313] 为产生构建体,以将里氏木霉cbhII基因(SEQ ID NO:3[DNA序列]和SEQ ID NO:4[推导的氨基酸序列])用烟曲霉cbhII编码序列(SEQ ID NO:49[DNA序列]和SEQ ID NO:50[推导的氨基酸序列])取代,首先使用以下示出的基因特异性正向和反向引物从里氏木霉RutC30基因组DNA扩增出里氏木霉cbhII启动子。斜体区表示与**IN-FUSION®**反应的插入位点的载体同源性。

[0314] 正向引物:

[0315] 5'-acgaattgttttaaacgtcgacCCAAGTATCCAGAGGTGTATGGAAATA TCAGAT-3'(SEQ ID NO:51)

[0316] 反向引物:

[0317] 5'-cgcgtagatctgcggccatGGTGCAATACACAGAGGGTGATCTT-3'(SEQ ID NO:52)

[0318] 将里氏木霉RutC30在28℃下在250ml带挡板的摇瓶中在200rpm 振荡下在50ml补充有2%葡萄糖(w/v)的YP培养基中生长2天。使用**MIRACLOTH®**(Calbiochem,La Jolla,CA,USA)通过过滤收获菌丝体,并将其用去离子水洗涤两次,并在液氮下冷冻。通过研钵及研杵将冷冻的菌丝体研磨成细粉,并使用**DNEASY® Plant Maxi**试剂盒分离基因组DNA,其中裂解孵育延长至2小时。

[0319] 扩增反应由20ng里氏木霉RutC30基因组DNA、200μM dNTP、0.4μM引物、1X **HERCULASE®**反应缓冲液(Stratagene,La Jolla,CA, USA)和1.875个单位的**HERCULASE®**热启动高保真DNA聚合酶(Stratagene,La Jolla,CA,USA)组成,最终体积为50μl。在**EPPENDORF® MASTERCYCLER® 5333epgradient S**中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、55℃30 秒和72℃1分30秒,25个循

环;以及72℃7分钟,1个循环。使用TAE 缓冲液通过1%琼脂糖凝胶电泳来分离PCR产物,其中将1.6kb片段从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒进行提取。

[0320] 使用IN-FUSION® Advantage PCR克隆试剂盒根据制造商的操作指南将1.6kb PCR产物插入到Nco I/Sal I消化的pSMai155 (W0 05/074647) 中。IN-FUSION®反应物由1XIN-FUSION®反应缓冲液、125ng Nco I/Sal I消化的pSMai155、100ng 1.6kb PCR产物和1μl IN-FUSION®酶组成,反应体积为10μl。将反应物在37℃下孵育15 分钟,然后在50℃下孵育15分钟。孵育期之后,将40μl TE加入到反应中。根据制造商的操作指南,使用2μl等份来转化ONESHOT® TOP10感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。通过使用Pci I的限制性酶切分析来筛选所得的转化体,对阳性克隆进行测序,以确保不存在PCR错误。识别出包含没有PCR错误的插入的一个克隆,并命名为pJfyS142-A。质粒pJfyS142-A用于里氏木霉cbhII终止子的插入。

[0321] 使用以下示出的基因特异性正向和反向引物从里氏木霉RutC30基因组DNA扩增cbhII终止子。斜体区表示与IN-FUSION®反应的插入位点的载体同源性。

[0322] 正向引物:

[0323] 5' -atctacgcgtactagtttaattaaGGCTTTCGTGACCGGGCTTCAAACA- 3' (SEQ ID NO:53)

[0324] 反向引物:

[0325] 5' -gcggccgttactagtgatccACTCGGAGTTGTTATACGCTACTCG-3' (SEQ ID NO:54)

[0326] 扩增反应由150ng里氏木霉RutC30基因组DNA、200μM dNTP、0.4μM引物、1X HERCULASE®反应缓冲液和1.875个单位的HERCULASE®热启动高保真DNA聚合酶组成,最终体积为50μl。在EPPENDORF® MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、54℃30 秒和72℃50秒,25个循环;以及72℃7分钟,1个循环。使用TAE 缓冲液通过1%琼脂糖凝胶电泳来分离PCR产物,其中将0.3kb片段从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒进行提取。

[0327] 使用IN-FUSION® Advantage PCR克隆试剂盒根据制造商的操作指南将0.3kb PCR产物插入到Pac I/Bam HI消化的pJfyS142-A中。IN-FUSION®反应由1X IN-FUSION®反应缓冲液、150ng PacI/Bam HI 消化的pJfyS142-A、50ng 0.3kb PCR产物和1μl IN-FUSION®酶组成,反应体积为10μl。将反应物在37℃下孵育15分钟,然后在50℃下孵育15分钟。孵育期之后,将40μl TE加入到反应中。根据制造商的操作指南,使用2μl等份来转化ONESHOT®TOP10感受态细胞。将细胞在42℃下热击30秒,加入250μl SOC培养基。将试管在37℃、200 rpm下孵育1小时,并将250μl铺在直径150mm的2XYT+氨苄青霉素平板上,在37℃下孵育过夜。通过序列分析来筛选转化体,以识别阳性克隆,并确保不存在PCR错误。识别出包含没有PCR错误的插入的一个克隆,并命名为pJfyS142-B。质粒pJfyS142-B用于单纯疱疹 (Herpes simplex) tk基因的插入。

[0328] 通过将质粒用Bgl II和Bam HI消化,将单纯疱疹tk基因从 pJfyS1579-8-6 (W0 2010/039840) 中释放。使用TAE缓冲液对消化物进行1%琼脂糖凝胶电泳,其中将2.3kb条带从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒进行提取。使用QUICK LIGATION™试剂盒根据制造商的操作指南将tk盒插入到Bam HI消化的牛小肠磷酸酶-去磷酸化的

pJfyS142-B中。连接反应物由50ng Bam HI消化的牛小肠磷酸酶-去磷酸化的pJfyS142-B、50ng 2.3kb tk基因插入物、1X QUICK LIGATION™缓冲液和5个单位的QUICK LIGASE™组成,连接体积为 20μl。将反应物在室温下孵育5分钟,并根据制造商的操作指南使用2 μl反应物来转化ONE SHOT®TOP10感受态细胞。将细胞在42℃下热击30秒,加入250μl SOC培养基。将试管在37℃、200rpm下孵育1 小时,并将250μl铺在直径150mm的2XYT+氨苄青霉素平板上,在 37℃下孵育过夜。通过使用Xma I和Bam HI的限制性酶切分析来筛选筛选所得的转化体,以确定插入物的存在和方向,并识别出含有该插入物的克隆,将其命名为pJfyS142-C。质粒pJfyS142-C用于里氏木霉 3' cbhII基因侧翼序列的插入。

[0329] 使用以下示出的正向和反向引物从里氏木霉RutC30基因组DNA 扩增3' cbhII基因侧翼序列。斜体区表示与IN-FUSION®反应的插入位点的载体同源性。

[0330] 正向引物:

[0331] 5' -atccatcacactggcggccgcGCTTCAAACAATGATGTGCGATGGT-3' (SEQ ID NO:55)

[0332] 反向引物:

[0333] 5' -gatgcatgctcgagcggccgcCTACCTTGGCAGCCCTACGAGAGAG- 3' (SEQ ID NO:56)

[0334] 扩增反应由150ng里氏木霉RutC30基因组DNA、200μM dNTP、0.4μM引物、1X HERCULASE®反应缓冲液和1.875个单位的HERCULASE®热启动高保真DNA聚合酶组成,最终体积为50μl。在EPPENDORF® MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、56℃30 秒和72℃下1分50秒,30个循环;以及72℃7分钟,1个循环。使用 TAE缓冲液,PCR反应物进行1%琼脂糖凝胶电泳,其中将1.5kb条带从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒进行提取。使用IN-FUSION®Advantage PCR克隆试剂盒根据制造商的操作指南将3' cbhII基因侧翼序列插入到Not I线性化的pJfyS142-C中。IN-FUSION®反应由1X IN-FUSION®反应缓冲液、150ng Not I线性化的 pJfyS142-C、80ng 1.5kb PCR产物和1 μl IN-FUSION™酶组成,反应体积为10μl。将反应物在37℃下孵育15分钟,然后在50℃下孵育15 分钟。孵育期之后,将40μl TE加入到反应中。根据制造商的操作指南,使用2μl等份来转化ONE SHOT®TOP10感受态细胞。将细胞在 42℃下热击30秒,并加入250μl SOC培养基。将试管在37℃、200rpm 下孵育1小时,并将250μl铺在直径150mm的2XYT+氨苄青霉素平板上,在37℃下孵育过夜。通过使用Bg1 II的限制性酶切分析来筛选所得的转化体,并对阳性克隆进行测序,以确保不存在PCR错误。识别出包含没有PCR错误的插入的一个克隆,并命名为pJfyS142(图7)。质粒pJfyS142用于烟曲霉cbhII编码序列的插入。

[0335] 实施例10:里氏木霉cbhII-烟曲霉cbhII取代构建体pJfyS144的构建

[0336] 使用以下示出的正向和反向引物从pAlLo33(WO 2011/057140) 扩增烟曲霉cbhII编码序列。斜体区表示与IN-FUSION®反应的插入位点的载体同源性。

[0337] 正向引物:

[0338] 5' -ctctgtgtattgcaccATGAAGCACCTTGCATCTTCCATCG-3' (SEQ ID NO:57)

[0339] 反向引物:

[0340] 5' -ccggtcacgaaagccTTAATTAAAAGGACGGGTTAGCGTT-3' (SEQ ID NO:58)

[0341] 扩增反应由20ng pAlLo33、200μM dNTP、0.4μM引物、1mM HERCULASE®反应

缓冲液和1.875个单位的HERCULASE®热启动高保真DNA聚合酶组成,最终体积为50μl。在EPPENDORF®MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、55℃30秒和72℃2分钟,30个循环;以及72℃7分钟,1个循环。

[0342] 使用TAE缓冲液,PCR反应物进行1%琼脂糖凝胶电泳,其中将1.7kb条带从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒提取。使用IN-FUSION®Advantage PCR克隆试剂盒根据制造商的操作指南将1.7kb PCR产物插入到Nco I/Pac I消化的pJfyS142(实施例9)中。IN-FUSION®反应由1X IN-FUSION®反应缓冲液、120ng Nco I/Pac I 消化的pJfyS142、70ng 1.7kb PCR产物和1μl IN-FUSION®酶组成,反应体积为10μl。将反应物在37℃下孵育15分钟,然后在50℃下孵育15分钟。孵育期之后,将40μl TE加入到反应中。根据制造商的操作指南,使用2μl等份来转化ONE SHOT®TOP10感受态细胞。将细胞在42℃下热击30秒,并加入250μl SOC培养基。将试管在37℃、200rpm下孵育1小时,并将250μl铺在直径150mm的2XYT+氨苄青霉素平板上,在37℃下孵育过夜。将所得的转化体进行测序,以确保不存在PCR错误,并确定插入物的存在。识别出序列没有错误的一个克隆,并命名为pJfyS144(图8)。

[0343] 实施例11:里氏木霉cbhI-烟曲霉cbhI取代构建体pJfyS139的构建

[0344] 使用以下的基因特异性正向和反向引物从pEJG93(WO 2011/057140)扩增烟曲霉纤维二糖水解酶I(cbhI)编码序列(SEQ ID NO:59[DNA序列]和SEQ ID NO:60[推导的氨基酸序列])。斜体区表示与IN-FUSION®反应的插入位点的载体同源性,划线部分是导入的Pac I位点。

[0345] 正向引物:

[0346] 5'-cgcgactgcgaccATGCTGGCCTCCACCTTCTCCTACC-3'(SEQ ID NO:61)

[0347] 反向引物:

[0348] 5'-ctttcgccacggagcttaattaaCTACAGGCACTGAGAGTAATAATCA-3'(SEQ ID NO:62)

[0349] 扩增反应由20ng pEJG93、200μM dNTP、0.4μM引物、1X HERCULASE®反应缓冲液和1.875个单位的HERCULASE®热启动高保真DNA聚合酶组成,最终体积为50μl。在EPPENDORF®MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、55℃30秒和72℃1分钟,30个循环;以及72℃7分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳来分离PCR产物,其中将1.6kb片段从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0350] 使用IN-FUSION®Advantage PCR克隆试剂盒根据制造商的操作指南将1.6kb PCR产物插入到Nco I/Pac I消化的pSMai155(WO 05/074647)中。IN-FUSION®反应由1X IN-FUSION®反应缓冲液、125 ng Nco I/Pac I消化的pSMai155、100ng 1.6kb PCR产物和1μl IN-FUSION™酶组成,反应体积为10μl。将反应物在37℃下孵育15分钟,然后在50℃下孵育15分钟。孵育期之后,将40μl TE缓冲液加入到反应中。根据制造商的操作指南使用2μl等份来转化ONE SHOT® TOP10感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。通过测序筛选所得的转化体,并识别出包含没有PCR错误的插入的一个克隆,将其命名为pJfyS139-A。质粒pJfyS139-A用于单纯疱疹病毒胸苷激酶(tk)基因的插入。

[0351] 通过将质粒用Bgl II和Bam HI消化,将单纯疱疹病毒tk编码序列 (SEQ ID NO:63 [DNA序列]和SEQ ID NO:64[推导的氨基酸序列]) 从pJfyS1579-8-6 (WO 2010/039840) 中释放。使用TAE缓冲液,消化物进行1%琼脂糖凝胶电泳,其中将2.3kb条带从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒进行提取。使用QUICK LIGATION™试剂盒根据制造商的操作指南将tk基因盒插入到Bam HI消化的牛小肠磷酸酶处理的pJfyS139-A中。连接反应物由50ng Bam HI消化的牛小肠磷酸酶处理的pJfyS139-A、50ng 2.3kb tk基因插入物、1X QUICK LIGATION™缓冲液和5个单位的QUICK LIGASE™组成,最终体积为 20μl。将反应物在室温下孵育5分钟,并根据制造商的操作指南使用2 μl反应物来转化ONE SHOT®TOP10感受态细胞。将细胞在42℃下热击30秒,加入250μl的SOC培养基。将试管在37℃、200rpm下孵育1小时,并将250μl铺在直径150mm的2XYT+氨苄青霉素平板上,在37℃下孵育过夜。通过使用Xma I的限制性酶切分析来筛选所得的转化体,以确定插入物的存在和方向,并识别出含有插入物的克隆,将其命名为pJfyS139-B。质粒pJfyS139-B用于里氏木霉3' cbhI基因侧翼序列的插入。

[0352] 使用以下示出的正向和反向引物从里氏木霉RutC30基因组DNA (实施例9) 扩增3' cbhI基因侧翼序列。划线部分表示引入的用于克隆的Not I位点。

[0353] 正向引物:

[0354] 5' -ttagactgcggccgcGTGGCGAAAGCCTGACGCACCGGTAGAT-3' (SEQ ID NO:65)

[0355] 反向引物:

[0356] 5' -agtagttagcggccgcACGGCACGGTTAAGCAGGGTCTTGC-3' (SEQ ID NO:66)

[0357] 扩增反应由150ng里氏木霉RutC30基因组DNA、200μM dNTP、0.4μM引物、1X HERCULASE®反应缓冲液和1.875个单位的HERCULASE®热启动高保真DNA聚合酶组成,最终体积为50μl。在EPPENDORF® MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、60℃30秒和72℃下1分30秒,30个循环;以及72℃7分钟,1个循环。

[0358] 根据制造商的操作指南,对PCR反应物使用MINELUTE®核苷酸去除试剂盒(QIAGEN Inc.,Valencia,CA,USA)。将所得的PCR混合物用Not I消化,使用TAE缓冲液通过1%琼脂糖凝胶电泳分离消化的PCR产物。将含有3' cbhI基因侧翼序列的1.3kb片段从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒进行提取。使用QUICK LIGATION™试剂盒将1.3kb片段插入到Not I线性化的牛小肠磷酸酶处理过的 pJfyS139-B中。QUICK LIGATION™反应由100ng Not I线性化的牛小肠磷酸酶处理过的pJfyS139-B、20ng 1.3kb片段、1X QUICK LIGATION™缓冲液和5个单位的QUICK LIGASE™组成,最终体积为 20μl。将反应物在室温下孵育5分钟,并根据制造商的操作指南使用2 μl反应物来转化ONE SHOT®TOP10感受态细胞。将细胞在42℃下热击30秒,加入250μl SOC培养基。将试管在37℃、200rpm下孵育1 小时,并将250μl铺在直径150mm的2XYT+氨苄青霉素平板上,在 37℃下孵育过夜。通过使用Xma I的限制性酶切分析来筛选所得的转化体,以确定插入物的存在和方向,并对阳性克隆进行测序。包含没有PCR错误的3' cbhI基因侧翼序列的克隆被命名为pJfyS139(图9)。

[0359] 实施例12:用于取代里氏木霉cbhI基因的烟曲霉cbhI-烟曲霉cbhII串联表达载体的构建

[0360] 构建串联取代载体pQM21,用表达两种重组蛋白的串联表达盒来取代里氏木霉中

的内生里氏木霉cbhI基因。质粒pQM21含有里氏木霉cbhI 5'侧翼序列、里氏木霉Cel7A纤维二糖水解酶I基因启动子、烟曲霉Cel7A纤维二糖水解酶I编码序列、里氏木霉Cel7A纤维二糖水解酶I基因终止子、里氏木霉Cel6A纤维二糖水解酶II基因启动子、烟曲霉Cel6A纤维二糖水解酶II编码序列、里氏木霉Cel6A纤维二糖水解酶II基因终止子、里氏木霉Cel7A纤维二糖水解酶I基因终止子重复序列、单纯疱疹病毒胸苷激酶(tk)基因、大肠杆菌潮霉素磷酸转移酶(hpt/hygR)选择性标记物、里氏木霉cbhI 3'侧翼序列和氨苄青霉素抗性标记物基因。

[0361] 使用以下示出的基因特异性正向和反向引物从pJfyS144(实施例10)扩增烟曲霉纤维二糖水解酶II表达盒。斜体区表示与IN-FUSION®反应的插入位点的序列同源性,划线部分分别是引入的Bam HI位点和Nhe I位点。

[0362] 正向引物:

[0363] 5'-tcaagcttggtaccgagctcggatCCAAGTATCCAGAGGTGTATGGAAA T-3' (SEQ ID NO: 67)

[0364] 反向引物:

[0365] 5'-ctggcggccggttactagtgctagcACTCGGAGTTGTTATACGCTAC-3' (SEQ ID NO:68)

[0366] 扩增反应由164ng pJfyS144、1μM引物、1X ACCUPRIME™Pfx 反应缓冲液(Invitrogen,Carlsbad,CA,USA)和2.5个单位的 ACCUPRIME™Pfx DNA聚合酶(Invitrogen,Carlsbad,CA,USA)组成,最终体积为50μl。在EPPENDORF® MASTERCYCLER® 5333 epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃15秒、58℃30秒和68℃5分钟,35个循环。使用TAE缓冲液,通过1%琼脂糖凝胶电泳来分离PCR产物,其中将约3.5kb的片段从凝胶中切出,并使用QIAQUICK®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0367] 使用IN-FUSION™Advantage PCR克隆试剂盒根据制造商建议的操作指南将3.5kb PCR产物插入到Bam HI消化的pJfyS139(实施例11)中。IN-FUSION™反应由1X IN-FUSION™反应缓冲液、103ng Bam HI消化的pJfyS139、62ng 3.5kb PCR产物和1μl IN-FUSION™酶组成,反应体积为10μl。将反应物在37℃下孵育15分钟,然后在50℃下孵育15分钟。孵育期之后,向反应中加入15μl TE。根据制造商的操作指南,使用2μl等份来转化SOLOPACK® Gold超级感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。通过测序来筛选转化体,并识别出包含没有PCR错误的插入的一个克隆,将其命名为 pQM18(图10)。质粒pQM18用于在串联表达盒后插入来自里氏木霉 3' cbhI基因侧翼区的同源重复片段,并修饰里氏木霉3' cbhII基因侧翼区。

[0368] 使用以下的正向和反向引物从pJfyS139扩增来自里氏木霉3' cbhI 基因侧翼区的同源重复片段。斜体区表示与IN-FUSION®反应的插入位点的序列同源性,划线部分是引入的用于克隆的Nhe I位点和Xba I 位点。

[0369] 正向引物:

[0370] 5'-gagtagcgtataacaactccgagtgctagcTTTAAGATAACGGAATAGAA GAA AG-3' (SEQ ID NO:69)

[0371] 反向引物:

[0372] 5'-ctggcggccggttactagtgctagaCGGCCACTACCGGTTTCG-3' (SEQ ID NO:70)

[0373] 使用以下的正向和反向引物从pJfyS139扩增里氏木霉3' cbhI基因侧翼序列。斜体区表示与IN-FUSION®反应的插入位点的载体同源性,划线部分是引入的用于克隆的Not I位点。

[0374] 正向引物:

[0375] 5' -tctgcagatatccatcacactggcgccgcTTTAAGATAACGGAATAGAAG AAAG-3' (SEQ ID NO:71)

[0376] 反向引物:

[0377] 5' -aaactctaggatgcatgctcgagcgccgcACGGCACGGTTAAGCAGGGT -3' (SEQ ID NO:72)

[0378] 扩增反应由350ng pJfyS139、1μM引物、1X ACCUPRIME™Pfx 反应缓冲液和2.5个单位的ACCUPRIME™Pfx DNA聚合酶组成,最终体积为50μl。在EPPENDORF® MASTERCYCLER® 5333epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;以及每循环95℃15秒、58℃30秒和68℃5分钟,35个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳纯化PCR产物,其中将约1.1kb的片段和约260bp的片段从凝胶中切出,并使用QIAQUICK®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0379] 通过将质粒用Nhe I和Not I消化,将含有tk基因和hpt(潮霉素磷酸转移酶)选择性标记物的DNA片段从pQM18中释放。使用TAE 缓冲液通过1%琼脂糖凝胶电泳对消化物进行分析,其中将约4.4kb的条带从凝胶中切出,并使用QIAQUICK®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0380] 使用TAE缓冲液通过1%琼脂糖凝胶电泳来分离来自Nhe I和Not I 消化的pQM18中的约9kb DNA片段。将9kb片段从凝胶中切出,并使用QIAQUICK®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0381] IN-FUSION™反应由1X IN-FUSION™反应缓冲液、158ng的9kb Nhe I和Not I消化的pQM18、13ng的260bp来自里氏木霉cbhI 3' 侧翼区的同源重复片段、39ng的1.1kb 3' cbhI侧翼、56ng的4.4kb tk-hpt 片段和1μl IN-FUSION™酶组成,反应体积为10μl。将反应物在37℃下孵育15分钟,并在50℃下孵育15分钟。孵育期之后,向反应中加入40μl TE。根据制造商的操作指南,使用2.5μl等份来转化SOLOPACK®Gold超级感受态细胞。将大肠杆菌转化反应物铺在 2XYT+氨苄青霉素平板上,并在37℃下孵育过夜。通过测序来筛选转化体,并识别出包含没有PCR错误的插入的一个克隆,将其命名为 pQM21(图11)。质粒pQM21用作取代cbhI基因的载体。

[0382] 实施例13:用烟曲霉纤维二糖水解酶I和烟曲霉纤维二糖水解酶II串联表达盒取代内生里氏木霉cbhI基因

[0383] 如实施例4所述,进行里氏木霉菌株AgJg115-104-7B1的原生质体制备和转化。

[0384] 为将内生cbhI基因用烟曲霉cbhI-cbhII串联表达盒取代,将约137 μg的pQM21(实施例12)用Pme I消化。使用TAE缓冲液通过1%琼脂糖凝胶电泳来纯化消化反应物,其中从凝胶中切出含有用于靶向里氏木霉cbhI基因座的烟曲霉CBHI-CBHII串联表达盒的约12kb DNA 条带,并使用NUCLEOSPIN®Extract II纯化试剂盒进行提取。将约1~3 μg的纯化的所得12kb DNA加入到100μl里氏木霉ku70-菌株 AgJg115-104-7B1原生质体溶液中,并轻柔混合。加入PEG缓冲液(250 μl),混合,并在34℃下孵育30分钟。然后加入STC(3ml),混

合,并铺在各个补充有1M蔗糖的PDA平板上。在28℃下孵育16小时后,向各个板中加入20ml补充有35μg潮霉素B/ml的覆层PDA培养基。将平板在28℃下孵育4~7天。

[0385] 得到7个转化体,挑出每一个,并转移至PDA平板,在28℃下孵育7天。采用使用下述操作指南的真菌孢子PCR法来筛选携带取代物的转化体,使用以下所示的对整合的cbhI 5'侧翼区的上游区进行退火的正向引物以及以下所示的对tk区中的区域进行退火的反向引物。

[0386] 正向引物:

[0387] 5'-CAAGCAAAGCGTTCCGTCGCAGTAGCAGGC-3' (SEQ ID NO: 73)

[0388] 反向引物:

[0389] 5'-CAGTGGCGCTTATTACTCAG-3' (SEQ ID NO:74)

[0390] 只有当在cbhII基因座处发生精确的基因取代时,才会产生约7kb 的PCR产物。如果盒已经整合在基因组中的别处,则不会得到扩增物。

[0391] 将来自各转化体的少量孢子悬浮在25μl的TE缓冲液中,并在微波炉中高温加热1分钟。使用各个微波过的孢子悬浮液作为PCR反应中的模板。反应由2μl微波过的孢子悬浮液、200μM dNTP、1μM引物、1X **LONGAMP®**Taq反应缓冲液(New England Biolabs, Inc, Ipswich, MA, USA)和2个单位的**LONGAMP®**Taq DNA聚合酶(New England Biolabs, Inc, Ipswich, MA, USA)组成,最终体积为20μl。在

EPPENDORF® MASTERCYCLER®5333epgradient S中孵育反应物,程序设定为:95℃4分钟,1个循环;每循环95℃15秒、50℃30秒和 68℃7分钟,35个循环;以及68℃15分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳对PCR反应物进行分析。孢子PCR表明,7个转化体中的5个在cbhI基因座处含有取代盒。

[0392] 根据实施例9中所述的过程,从4个阳性转化体中分离出基因组 DNA,并对其进行DNA印迹分析,以确认取代盒是单个拷贝。

[0393] 对于DNA印迹分析,将2μg基因组DNA在20μl反应体积中用10个单位的Bam HI消化,并使用TAE缓冲液对其进行0.7%琼脂糖电泳。将凝胶中的DNA在0.25N HCl中脱嘌呤15分钟,在变性溶液中变性两次,每次15分钟,在中和溶液中中和10~30分钟,并使用TURBOLOTTER™系统(Whatman, Inc., Florham Park, NJ, USA)根据制造商的操作指南转移至**NYTRAN®**Supercharge膜(Whatman, Inc., Florham Park, NJ, USA)。使用STRATALINKER™UV交联剂(Stratagene, La Jolla, CA, USA)将DNA经UV交联在膜上,并在20ml DIG Easy Hyb(Roche Diagnostics Corporation, Indianapolis, IN, USA) 中在42℃下预杂交1小时。

[0394] 使用PCR Dig探针合成试剂盒(Roche Diagnostics Corporation, Indianapolis, IN, USA)根据制造商说明书,用如下所示的正向和反向引物生成与cbhI基因的3'侧翼区杂交的探针。PCR反应由具有MgCl₂的1X **EXPAND®**高保真PCR缓冲液(Roche Diagnostics Corporation, Indianapolis, IN, USA)、1X PCR DIG探针合成混合物(Roche Diagnostics Corporation, Indianapolis, IN, USA)、1μM各引物、100pg的1.1kb 3' cbhI侧翼区和2.625个单位的**EXPAND®**高保真酶混合物(Roche Diagnostics Corporation, Indianapolis, IN, USA)组成。在**EPPENDORF® MASTERCYCLER®**5333epgradient S中进行PCR,程序设定为:95℃ 2分钟,1个循环;每循环95℃30秒、60℃30秒和72℃40秒,10

个循环;每个循环为95℃30秒、60℃30秒和72℃40秒且对于各个连续循环额外有20秒,20个循环;以及72℃7分钟,1个循环。

[0395] 正向引物:

[0396] 5'-GAGAACACAGTGAGACCATAGC-3' (SEQ ID NO:75)

[0397] 反向引物:

[0398] 5'-TCTCAACCCAATCAGCAACATG-3' (SEQ ID NO:76)

[0399] 使用TAE缓冲液通过1%琼脂糖凝胶电泳来纯化探针,其中将与探针对应的720bp条带从凝胶中切出,并使用**NUCLEOSPIN®**Extract II纯化试剂盒进行提取。将探针煮沸5分钟,在冰上冷却2分钟,然后加入到10ml DIG Easy Hyb中,产生杂交溶液。杂交在42℃下进行15~17小时。然后在室温下将膜在低等严格条件下在2X SSC+0.1% SDS中洗涤5分钟,然后,在65℃下在0.5X SSC+0.1% SDS中进行两次高等严格洗涤,每次15分钟。使用化学发光检定(Roche Diagnostics, Indianapolis, IN, USA)根据制造商说明书来检测探针-靶标杂交。DNA印迹分析表明,命名为里氏木霉QMJi029-A5的一个转化体在cbhI基因座处含有取代盒,并选择其用于消除tk和hpt标记物。

[0400] 使用无菌涂布器,在5ml 0.01% **TWEEN®**20中收集在28℃下生长的七日龄PDA平板上的里氏木霉QMJi029-A5的孢子。使用血球计数器确定孢子的浓度,将 10^4 和/或 10^5 孢子铺在150mm平板上,该平板含有补充有 $1\mu\text{M}$ 5-氟-2'-脱氧尿苷(FdU)的TrMM-G培养基。

[0401] 挑选10个FdU抗性的孢子分离株,如上所述从3个孢子分离株中提取基因组DNA。如上所述对分离株进行DNA印迹分析,结果表明所有的孢子分离株已经切除在取代盒的同源重复序列之间的hpt/tk区。选择命名为里氏木霉QMJi030-A5.6的一个菌株用于取代cbhII基因。

[0402] 用 $10\mu\text{l}$ 接种环,收集在28℃下生长的七日龄PDA平板上的里氏木霉QMJi029-A5和里氏木霉QMJi030-A5.6的孢子,并将其转移至125 ml塑料摇瓶中的25ml CIM培养基中。将摇瓶培养物在28℃、200rpm下孵育5天。将各培养物的1ml等份在微型离心机中于13,400xg离心,回收培养上清液。使用**CRITERION®**8~16%Tris-HCl凝胶根据制造商说明书经由SDS-PAGE对 $5\mu\text{l}$ 的各培养上清液进行分析。将所得的凝胶用BIO-SAFE™Coomassie(考马斯)染色。培养物的SDS-PAGE图谱表明,转化体产生50与70kDa之间的两个主要蛋白条带,其分别与烟曲霉纤维二糖水解酶I和烟曲霉纤维二糖水解酶II对应。烟曲霉纤维二糖水解酶I和烟曲霉纤维二糖水解酶II的表达通过质谱分析加以确认(实施例14)。

[0403] 实施例14:用于肽测序的多肽的凝胶内消化

[0404] 使用**MULTIPROBE®**II液体处理机械臂(PerkinElmer Life and Analytical Sciences, Boston, MA, USA)进行凝胶内消化。将实施例13中所述的SDS-PAGE凝胶的一部分切出,其在50和70kDa MW标记物之间含有目标蛋白。将凝胶片在100mM碳酸氢铵pH 8.0中用 $50\mu\text{l}$ 10mM二硫苏糖醇(DTT)还原30分钟。还原之后,将凝胶片在100mM碳酸氢铵pH 8.0中用 $50\mu\text{l}$ 55mM碘乙酰胺烷基化20分钟。使干燥的凝胶条在室温下在含有6ng测序级胰蛋白酶(Promega, Madison, WI, USA) / μl 50mM碳酸氢铵pH 8的25 μl 胰蛋白酶消化液中膨胀30分钟,然后在40℃下消化8小时。上述各个反应步骤之后为,均遵照制造商的标准操作指南用适当的溶液进行多次洗涤和预洗涤。在反应之间使用 $50\mu\text{l}$ 乙腈来使凝胶片脱水,并在步骤之间将凝胶片空气干燥。将肽用HPLC级水中的1%甲酸/2%乙腈提取30分钟,进行两次。将

肽提取液转移至已经冷却至10~15℃的96孔有边缘PCR型板(ABGene, Rochester, NY, USA),并用96孔板盖(PerkinElmer Life and Analytical Sciences, Boston, MA, USA)盖上,以防止蒸发。将板在4℃下进一步储存,直至可以进行质谱分析。

[0405] 蛋白鉴定.对于通过串联质谱法进行肽的从头测序,使用混合正交四极杆飞行时间质谱仪SYNAPTTMMS(Waters Corp., Milford, MA, USA)进行LC/MS/MS分析。SYNAPTTMMS完全是微处理器控制的,其使用MASSLYNX[®]软件4.1版(Waters Corp., Milford, MA, USA)。SYNAPTTMMS装有用于使样本浓缩并脱盐的NANOACQUITY UPLC[®](Waters Corp, Milford, MA, USA)。将样本装载在上样环中安装的捕获柱(180μm ID X 20mm, 5μm SYMMETRY[®]C18)(Waters Corp, Milford, MA, USA)上,并使用二元溶剂控制泵以15μl/分钟用水中的0.1%甲酸洗涤1分钟。在100μm ID x 100mm, C18, 1.7μm, BEH130TMC18纳流融合毛细管柱(Waters Corp, Milford, MA, USA)上以400nl/分钟的流量分离肽。在30分钟时间间隔中,施用0.1%甲酸中1%~85%乙腈的阶梯洗脱梯度。柱洗提液在214nm下监测,并通过装有纳米喷雾接口的电喷雾离子源引入到SYNAPTTMMS。

[0406] 以全元素扫描模式从m/z 250~1900的质量范围获取数据,使用MS到MS/MS的切换标准,以包括大于10.0计数/秒的离子强度和+2、+3和+4的电荷状态。可以获得扫描时间为2.0秒的6个共洗脱种的分析谱图。通常使用45伏的锥孔电压,并将碰撞能量设定为根据洗脱的肽的质量和电荷状态而变化,且在10~60伏的范围内。将得到的谱图以自动的方式合并、平滑化并集中,生成峰列表。使用PROTEINLYNX GLOBAL SERVER[®]2.4软件(Waters Corp, Milford, MA, USA)和MASCOT[®]v.2.2(Matrix Sciences Ltd., London, UK),针对选择的数据库对峰列表进行搜索。对PROTEINLYNX GLOBAL SERVER[®]和MASCOT[®]搜索的结果进行评价,肽的鉴定基于与所预期蛋白序列的肽质量指纹图谱匹配。

[0407] 肽质量指纹图谱证实,里氏木霉QMJi029-A5和里氏木霉 QMJI030A5.6的SDS-PAGE样本含有烟曲霉Cel7A纤维二糖水解酶I、烟曲霉Cel6A纤维二糖水解酶II、里氏木霉纤维二糖水解酶II和其他少量里氏木霉宿主背景蛋白。

[0408] 实施例15:编码Penicillium emersonii GH61A多肽和烟曲霉Cel3Aβ-葡萄糖苷酶的串联基因表达质粒pRRAB01的生成

[0409] 使用以下所示的基因特异性正向和反向引物从质粒pAG121(图 12;SEQ ID NO: 77;来自核苷酸6位至核苷酸620位,参见pAG121 的限制性酶切图谱和序列)扩增里氏木霉纤维二糖水解酶II(cbhII) 基因启动子。正向引物中的斜体区表示与pDM286载体骨架的序列同源性,反向引物中的斜体区表示与用于IN-FUSION[®]反应的下一个插入物即烟曲霉β-葡萄糖苷酶编码序列的序列同源性。

[0410] 正向引物:

[0411] 5'-cgaacgcggtagtgGAATTCTAGGCTAGGTATGC-3'(SEQ ID NO: 78)

[0412] 反向引物:

[0413] 5'-ccaaccgaatctcatGGTGCAATACACAGAGGGTG-3'(SEQ ID NO: 79)

[0414] 扩增反应由1ng的pAG121DNA、100皮摩尔的各上述引物、1μl 的dATP、dTTP、dGTP和dCTP的10mM混合物、1X PHUSIONTM高保真热启动DNA聚合酶缓冲液和1个单位的PHUSIONTM高保真热启动DNA聚合酶组成,最终体积为50μl。在EPPENDORF[®] MASTERCYCLER[®] 5333epgradient S中孵育扩增反应物,程序设定为:98℃下30秒,1个循环;每循环98℃10秒

和72℃20秒,35个循环;以及72℃10分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳分离PCR产物,其中将约640bp片段从凝胶中切出,并使用QIAGEN®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0415] 使用以下所示的基因特异性正向和反向引物从质粒pDM290扩增烟曲霉β-葡萄糖苷酶编码序列。正向引物中的斜体区表示与之前插入物即里氏木霉纤维二糖水解酶II编码基因的序列同源性,反向引物中的斜体区表示与IN-FUSION®反应的下一个插入物即里氏木霉纤维二糖水解酶II基因终止子的序列同源性。

[0416] 正向引物:

[0417] 5'-tctgtgtattgcaccATGAGATTCGGTTGGCTCGA-3' (SEQ ID NO: 80)

[0418] 反向引物:

[0419] 5'-ccggtcacgaaagccCTAGTAGACACGGGGCAGAG-3' (SEQ ID NO:81)

[0420] 扩增反应由1ng的pDM290DNA、100皮摩尔的各上述引物、1μl 的dATP、dTTP、dGTP和dCTP的10mM混合物、1X PHUSION™高保真热启动DNA聚合酶缓冲液和1个单位的PHUSION™高保真热启动DNA聚合酶组成,最终体积为50μl。在EPPENDORF®MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:98℃30秒,1个循环;每循环98℃10秒、65℃30秒和72℃1:35 秒,35个循环;以及72℃10分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳分离PCR产物,其中将约3.1kb片段从凝胶中切出,并使用QIAGEN®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0421] 使用以下所示的基因特异性正向和反向引物从质粒pJfyS144扩增里氏木霉纤维二糖水解酶II基因终止子。正向引物中的斜体区表示与烟曲霉β-葡萄糖苷酶编码序列的序列同源性,反向引物中的斜体区表示与用于IN-FUSION®反应的pDM286骨架的序列同源性。

[0422] 正向引物:

[0423] 5'-ccccgtgtctactagGGCTTTCGTGACCGGGCTTC-3' (SEQ ID NO: 82)

[0424] 反向引物:

[0425] 5'-gtcattaccaattggACTCGGAGTTGTTATACGCT-3' (SEQ ID NO: 83)

[0426] 扩增反应由1ng的pJfyS144DNA、100皮摩尔的各上述引物、1μl 的dATP、dTTP、dGTP和dCTP的10mM混合物、1X PHUSION™高保真热启动DNA聚合酶缓冲液和1个单位PHUSION™高保真热启动DNA聚合酶组成,最终体积为50μl。在EPPENDORF®MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:98℃30秒,1个循环;每循环98℃10秒和72℃10秒,35个循环;以及72℃10分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳分离PCR产物,其中将330bp片段从凝胶中切出,并使用QIAGEN®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0427] 使用以下所示的基因特异性正向和反向引物,将里氏木霉纤维二糖水解酶II基因启动子、烟曲霉β-葡萄糖苷酶编码序列和里氏木霉纤维二糖水解酶II基因终止子在重叠延伸(SOE)PCR反应中结合。斜体区表示与用于IN-FUSION®反应的pDM286中的插入位点的序列同源性。

[0428] 正向引物:

[0429] 5'-cgaacgcggtagtgGAATTCTAGGCTAGGTATGC-3' (SEQ ID NO: 84)

[0430] 反向引物:

[0431] 5'-gtcattaccaattggACTCGGAGTTGTTATACGCT-3' (SEQ ID NO: 85)

[0432] SOE PCR反应由48ng扩增自pAG121的640bp里氏木霉cbhII启动子,228ng扩增自pDM290的3.1kb烟曲霉 β -葡萄糖苷酶基因片段,24ng扩增自pJfyS144的330bp里氏木霉cbhII终止子,100皮摩尔的各上述引物,1 μ l的dATP、dTTP、dGTP和dCTP的10mM混合物,1X PHUSION™高保真热启动DNA聚合酶缓冲液和1个单位的 PHUSION™高保真热启动DNA聚合酶组成,最终体积为50 μ l。在EPPENDORF® MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:98℃30秒,1个循环;每循环98℃10秒和72℃2分5秒,35个循环;和72℃10分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳分离PCR产物,其中将约4.1kb片段从凝胶中切出,并使用QIAGEN®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0433] 将质粒pDM286用Asc I消化,并使用TAE缓冲液通过1%琼脂糖凝胶电泳分离经Asc I消化的pDM286,其中将约8kb的片段从凝胶中切出,并使用QIAGEN®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0434] 使用IN-FUSION™Advantage PCR克隆试剂盒根据制造商的操作指南将4.1kb的PCR产物插入到Asc I消化的pDM286中。IN-FUSION™反应由1X IN-FUSION™反应缓冲液、200ng约8kb凝胶纯化的Asc I消化的pDM286、203.1ng的4.1kb PCR产物和1 μ l IN-FUSION™酶组成,反应体积为10 μ l。将反应物在37℃下孵育15分钟,并在50℃下孵育15分钟。孵育期之后,向反应物中加入40 μ l TE。根据制造商的操作指南,使用2 μ l等份来转化ONE SHOT®TOP10感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。通过测序来筛选转化体,并识别出包含没有PCR错误的插入的一个克隆,将其命名为 pRRAB01 (图13)。

[0435] 实施例16:编码烟曲霉 β -葡萄糖苷酶 (Cel3A) 突变基因的里氏木霉表达载体的生成

[0436] 使用QUIKCHANGE®多点定点诱变试剂盒(Stratagene,La Jolla, CA,USA)通过在pEJG97 (WO 2005/074647) 上进行定点诱变,构建含有替换G142S、Q183R、H266Q和D703G的烟曲霉家族3A β -葡萄糖苷酶的变体。用于定点诱变的寡核苷酸的概要示于表1中。

[0437] 使用BIOROBOT®9600准备所得的变异质粒pDFng128-6。使用 Applied Biosystems 3130xl基因分析仪(Applied Biosystems,Foster City, CA,USA)对变异质粒构建体进行测序,以确认变化。

[0438] 表1

[0439]

氨基酸改变	引物名称	序列	克隆质粒名称
F100D	AfBGmutF100DF	ccctttgggtatccgtGACtgtgagctatacccgcg (SEQ ID NO: 86)	pDFng128-6
S283G	AfBGmutS283GF	cgtcattgagtactggGGCgctcaccacagcggtg (SEQ ID NO: 87)	
N456E	AfBGmutN456EF	gggtagtggtagctgccGAGttcccttacctgtcac (SEQ ID NO: 88)	
F512Y	AfBGmutF512YF	gccgactctggagagggtTACatcagtgctgacggcaac (SEQ ID NO: 89)	

[0440] 设计以下所示的两种合成寡核苷酸引物,从质粒pDFng128-6中 PCR扩增烟曲霉 β -葡萄糖苷酶变体编码序列。使用IN-FUSION™克隆试剂盒将片段直接克隆到表达载体pMJ09中。粗体字母表示编码序列。其余序列与pMJ09的插入位点同源。

[0441] 正向引物:

[0442] 5'-CGGACTGCGCACCATGAGATTTCGGTTGGCTCGA-3' (SEQ ID NO:90)

[0443] 反向引物:

[0444] 5'-TCGCCACGGAGCTTACTAGTAGACACGGGGCAGAG-3' (SEQ ID NO:91)

[0445] 将50皮摩尔的各上述引物用于PCR反应中,该PCR反应由50ng 的pDFng128-6,具有MgCl₂的1X **EXPAND®**高保真PCR缓冲液,各 0.25mM的dATP、dTTP、dGTP和dCTP,和2.6个单位的**EXPAND®**高保真酶混合物组成,最终体积为50 μ l。在**EPPENDORF®****MASTERCYCLER®**5333epgradient S中进行扩增,程序设定为:94℃ 2分钟,1个循环;每循环94℃15秒、65℃30秒和68℃1分钟,30个循环;以及68℃下最终延长7分钟。然后将加热块转至4℃浸泡循环。使用TBE缓冲液通过0.7%琼脂糖凝胶电泳分离反应产物,其中将约 3.1kb的产物条带从凝胶中切出,并使用**QIAQUICK®**凝胶提取试剂盒根据制造商说明书进行提取。

[0446] 将质粒pMJ09用Nco I和Pac I消化,使用TBE缓冲液通过1.0%琼脂糖凝胶电泳分离,从凝胶中切出,并使用**QIAQUICK®**凝胶提取试剂盒根据制造商说明书进行提取。

[0447] 将3.1kb的基因片段和消化的载体使用IN-FUSION™克隆试剂盒连接在一起,得到pDFng113-3(图14),其中 β -葡萄糖苷酶变体编码序列的转录处于来自里氏木霉cbhI基因的启动子的控制下。连接反应(20 μ l)由1X IN-FUSION™反应缓冲液、1X BSA、1 μ l IN-FUSION™酶(1:10稀释)、200ng凝胶纯化的Nco I/Pac I消化的pMJ09以及172.2 ng纯化的3.1kb PCR产物组成。将反应物在37℃下孵育15分钟,然后在50℃下孵育15分钟。使用2 μ l反应物来转化大肠杆菌XL10 **SOLOPACK®**Gold超级感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。使用**BIOROBOT®**9600来准备含有 pDFng133-3的大肠杆菌转

化体。通过DNA测序来确认pDFng133-3中的烟曲霉 β -葡糖苷酶变体插入物。

[0448] 实施例17:编码Penicillium emersonii GH61A多肽和烟曲霉Cel3A β -葡糖苷酶变体的串联表达载体pAmFs074的构建

[0449] 通过将来自pRRAB01和pDFNG133-3的限制性酶切片段组合来生成串联表达载体pAmFs074,以产生用于表达Penicillium emersonii GH61A多肽和烟曲霉Cel3A β -葡糖苷酶变体的单一载体。

[0450] 将1微克经质粒中量试剂盒(QIAGEN Inc., Valencia, CA, USA) 纯化的质粒pRRAB01(实施例15)与最终体积20 μ l中的20个单位 Xho I(New England Biolabs Inc, Ipswich, MA, USA)、1X NEB缓冲液4(New England Biolabs Inc, Ipswich, MA, USA)和1X BSA混合。反应物在37 $^{\circ}$ C下孵育3小时,并随后与4 μ l的5X DNA上样染料(QIAGEN Inc., Valencia, CA, USA)结合。使用TBE缓冲液通过1%琼脂糖凝胶电泳分离限制性酶切消化反应产物,其中将约9.4kb的片段从凝胶中切出,并使用NUCLEOSPIN®Extract II试剂盒根据制造商的操作指南进行提取。9.4kb的片段含有里氏木霉Cel17A纤维二糖水解酶I基因启动子、P.emersonii GH61A多肽编码序列、里氏木霉Cel17A纤维二糖水解酶I基因终止子、里氏木霉Cel16A纤维二糖水解酶II基因启动子、烟曲霉 β -葡糖苷酶Cel3A β -葡糖苷酶编码序列的3'端487bp部分和5'端16bp部分、里氏木霉Cel16A纤维二糖水解酶II基因终止子、构巢曲霉乙酰胺酶(amdS)基因和氨苄青霉素抗性标记物基因。

[0451] 将1微克经质粒中量试剂盒纯化的质粒pDFNG133-3(实施例16)与最终体积20 μ l中的20个单位限制性内切酶Xho I、1X NEB缓冲液4和1X BSA混合。将反应物在37 $^{\circ}$ C下孵育3小时,然后与4 μ l 5X DNA上样染料混结合。使用TBE缓冲液通过1%琼脂糖凝胶电泳分离限制性酶切消化反应产物,其中将约2.6kb的片段从凝胶中切出,并使用NUCLEOSPIN®Extract II试剂盒根据制造商的操作指南进行提取。2.6kb的片段含有烟曲霉 β -葡糖苷酶Cel3A β -葡糖苷酶突变基因的1940bp部分。

[0452] 将纯化的9.4kb pRRAB01Xho I限制性酶切片段和纯化的2.6kb pDFNG133-3Xho I限制性酶切片段使用QUICK LIGATION™试剂盒根据制造商的操作指南连接在一起。将50ng等份的9.4kb pRRAB01片段和50ng等份的2.6kb pDFNG133-3片段混合,用无菌水将体积调节至10 μ l。然后加入10 μ l 2X QUICK LIGATION™缓冲液和1 μ l QUICK LIGASE™,并彻底混合。将反应在25 $^{\circ}$ C下进行5分钟,然后置于冰上。根据制造商的操作指南,使用2 μ l等份的连接反应物来转化ONE SHOT®TOP10感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉素平板上。通过限制性酶切图谱和测序来筛选转化体。识别出包含没有序列错误并方向正确的插入的一个克隆,将其命名为pAmFs074(图15)。

[0453] 质粒pAmFs074可以用Pme I消化,以产生用于里氏木霉转化的约9.35kb片段。9.35kb的片段含有由以下组成的表达盒:(1)里氏木霉Cel17A纤维二糖水解酶I基因启动子、P.emersonii GH61A多肽编码序列和里氏木霉Cel17A纤维二糖水解酶I基因终止子,和(2)里氏木霉Cel16A纤维二糖水解酶II基因启动子、烟曲霉 β -葡糖苷酶Cel3A β -葡糖苷酶突变体编码序列和里氏木霉Cel16A纤维二糖水解酶II基因终止子。9.35kb的片段还含有构巢曲霉乙酰胺酶(amdS)基因。

[0454] 实施例18:用于取代里氏木霉cbhII的Penicillium emersonii GH61A多肽和烟曲霉Cel3A变异 β -葡糖苷酶串联表达载体的产生

[0455] 构建串联取代载体pQM22,用于将里氏木霉中的里氏木霉cbhII 基因用串联表达盒取代,以表达除真菌选择性标记物之外的两种重组蛋白。载体pQM22含有里氏木霉cbhII 5' 侧翼区、里氏木霉Cel7A纤维二糖水解酶I基因启动子、P.emersonii GH61A多肽编码序列、里氏木霉Cel7A纤维二糖水解酶I基因终止子、里氏木霉Cel6A纤维二糖水解酶II基因启动子、烟曲霉 β -葡糖苷酶Cel3A β -葡糖苷酶突变体编码序列、里氏木霉Cel6A纤维二糖水解酶II基因终止子、单纯疱疹病毒胸苷激酶(tk) 基因、大肠杆菌潮霉素磷酸转移酶(hpt/hygR) 选择性标记物、里氏木霉cbhII 3' 侧翼区和氨苄青霉素抗性标记物基因。

[0456] 通过将释放自pAmFs074的约6.4kb串联表达盒和扩增自里氏木霉菌株AgJg115-104-7B1基因组DNA(根据实施例9分离)的约1.5kb 里氏木霉cbhII 5' 侧翼区插入到经Sap I和Pac I消化的pJfyS142(实施例9)的约8.7kb载体片段中,制得载体pQM22。

[0457] 通过将质粒用Pme I和Eco RV消化,从pAmFs074中释放用于P. emersonii GH61A多肽和烟曲霉 β -葡糖苷酶变体的6.4kb串联表达盒。对消化物进行使用TAE缓冲液的1%琼脂糖凝胶电泳,其中将约6.4kb 条带从凝胶中切出,并使用NUCLEOSPIN®Extract II纯化试剂盒根据制造商的操作指南进行提取。

[0458] 使用以下的正向和反向引物,从里氏木霉菌株QMJi030-A5.6基因组DNA扩增5' 里氏木霉cbhII侧翼序列。斜体区表示与用于IN-FUSION®反应的插入位点的序列同源性,划线部分表示引入的用于克隆的Pac I和Psi I位点。

[0459] 正向引物:

[0460] 5' -gcgagtcagtgagcgcgaggaagcggaagagctttaattaaTCTTGAGTGGATGTC TGATCTAG-3' (SEQ ID NO:92)

[0461] 反向引物:

[0462] 5' -gttcggataacaatcctacattcgggtcgacttataaGGATGTATCAATGGGTAT ACG-3' (SEQ ID NO:93)

[0463] 扩增反应由约150ng的里氏木霉QMJi030-A5.6(实施例13)基因组DNA、1 μ M引物、200 μ M GeneAmp®dNTP(Appiled Biosystems, Foster City,USA)、1X PHUSION™高保真热启动DNA聚合酶缓冲液和 1个单位的PHUSION™高保真热启动DNA聚合酶组成,最终体积为50 μ l。在EPPENDORF® MASTERCYCLER®5333epgradient S中孵育扩增反应物,程序设定为:98℃2分钟,1个循环;每循环98℃15秒、60℃30秒和72℃1分钟,35个循环;和在72℃10分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳分离PCR产物,其中将约1.5kb 条带从凝胶中切出,并使用NUCLEOSPIN®Extract II纯化试剂盒根据制造商的操作指南进行提取。

[0464] 将约36 μ g的pJfyS142用Sap I和Pac I消化。使用TAE缓冲液通过1%琼脂糖凝胶电泳分离消化反应物,其中将约8.7kb的条带从凝胶中切出,并使用NUCLEOSPIN®Extract II纯化试剂盒根据制造商的操作指南进行提取。

[0465] IN-FUSION™反应由1X IN-FUSION™反应缓冲液、138ng用Sap I 和Pac I消化的8.7kb pJfyS142片段、205ng来自pAmFs074的6.4kb 片段、49ng里氏木霉cbhII 5' 侧翼区的1.5kb片段和1 μ l IN-FUSION™酶组成,反应体积为10 μ l。将反应物在37℃下孵育15分钟,在50℃下孵育15分钟。孵育期之后,向反应中加入40 μ l TE。根据制造商的操作指南,使用2.5 μ l 1等份来转化ONE SHOT®TOP10感受态细胞。将大肠杆菌转化反应物铺在2XYT+氨苄青霉

素平板上。通过测序来筛选转化体,并识别出包含没有PCR错误的插入的一个克隆,将其命名为 pQM22(图16)。质粒pQM22用作取代里氏木霉cbhII基因的载体。

[0466] 实施例19:用Penicillium emersonii GH61A多肽和烟曲霉β-葡萄糖苷酶突变串联表达盒取代内生里氏木霉cbhII基因

[0467] 为将内生里氏木霉cbhII基因用P.emersonii GH61多肽-烟曲霉β-葡萄糖苷酶突变串联表达盒取代,将里氏木霉QMJi030-A5.6(实施例13) 用6.3μg Pme I线性化的pQM22(实施例18)转化。得到31个转化体,挑取每一个并转移至PDA平板,在28℃下孵育7天。使用实施例13 中所述的真菌孢子PCR法用以下所示的引物组A筛选出含有完整cbhII 基因基因座的转化体之后,根据实施例9的过程从8个转化体中分离出基因组DNA,并使用以下所示的引物组B、与整合的cbhII 5' 侧翼区的上游区进行退火的正向引物、烟曲霉β-葡萄糖苷酶突变区之后的tk区中的反向引物-1和里氏木霉cbhII 3' 侧翼区下游的另一反向引物-2经 PCR加以分析。

[0468] 引物组A正向引物:

[0469] 5'-TCAACCAGCTTCTTTATTGG-3' (SEQ ID NO:94)

[0470] 引物组A反向引物:

[0471] 5'-GATCGCCATAGGCTCATGCTCCGCA-3' (SEQ ID NO:95)

[0472] 引物组B正向引物:

[0473] 5'-GCGGCATCAAACACGAACCTG-3' (SEQ ID NO:96)

[0474] 引物组B反向-1引物:

[0475] 5'-CAGTGGCGCTTATTACTCAG-3' (SEQ ID NO:97)

[0476] 引物组B反向-2引物:

[0477] 5'-GATCGCCATAGGCTCATGCTCCGCA-3' (SEQ ID NO:98)

[0478] 反应由2μl孢子悬浮液、200μM dNTP、1μM引物、1X **LONGAMP®**Taq反应缓冲液和2个单位的**LONGAMP®**Taq DNA聚合酶组成,最终体积为20μl。在**EPPENDORF®** **MASTERCYCLER®** 5333epgradient S中孵育反应物,程序设定为:95℃4分钟,1个循环;每循环95℃15秒、55℃30秒、68℃11分钟,35个循环;以及68℃15分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳对PCR反应物进行分析。PCR结果表明,一个转化体在靶向的cbhII基因座处含有取代盒。如下文所述进行DNA印迹分析,以确认取代盒是P. emersonii GH61A多肽序列和烟曲霉β-葡萄糖苷酶突变编码序列的单一拷贝。

[0479] 根据实施例9中所述的过程从转化体中分离出基因组DNA,并对各转化体进行DNA印迹分析。对于DNA印迹分析,在70μl反应体积中用10个单位的Cln I和/或含有5个单位Stu I和5个单位Sex AI 的限制性内切酶混合物消化6μg基因组DNA。将消化的DNA反应物与14μl的5X DNA上样染料混合,并使用TAE缓冲液使25μl的各混合物进行1%琼脂糖电泳。将凝胶中的DNA在0.25N HCl中脱嘌呤15 分钟,在变性溶液中变性两次,每次15分钟,在中和溶液中中和10~30 分钟,并转移至**NYTRAN®**Supercharge膜。使用**STRATALINKER™** UV交联剂将DNA UV交联到膜上,并在20ml DIG Easy Hyb中在42℃下预杂交1小时。

[0480] 将约1μg的pQM22用Stu I和Xba I消化。使用TAE缓冲液通过 1%琼脂糖凝胶电泳分离消化反应物,其中将来自P.emersonii GH61A 多肽编码序列的约800bp条带从凝胶中切出,并使用**MINELUTE®**凝胶提取试剂盒根据制造商的操作指南进行提取。

[0481] 将约1μg的pQM22用Xho I和Nru I消化。使用TAE缓冲液通过 1%琼脂糖凝胶电泳分离消化反应物,其中将来自烟曲霉Cel3A β-葡萄糖苷酶突变体编码序列的约1kb条带从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0482] 将约1μg的pQM22用Not I和Pvu I消化。使用TAE缓冲液通过 1%琼脂糖凝胶电泳分离消化反应物,其中将来自hpt基因的约1.1kb 条带从凝胶中切出,并使用MINELUTE®凝胶提取试剂盒根据制造商的操作指南进行提取。

[0483] 将266ng的800bp片段、168ng的1kb片段和240ng的1.1kb 片段量各自混合在16μl的最终体积中,以使用DIG-High Prime DNA 标记试剂盒(Roche Diagnostics Corporation, Indianapolis, IN, USA)根据制造商说明书生成探针。将DNA混合物煮沸10分钟,然后迅速在冰上冷却,之后加入4μl的DIG-High引物混合物(Roche Diagnostics Corporation, Indianapolis, IN, USA)。将反应物在37℃下孵育约20小时,之后加入2μl的0.2M EDTA,然后在65℃下加热10分钟,以使反应停止。

[0484] 使用MINELUTE®凝胶提取试剂盒根据制造商的操作指南纯化探针。将探针煮沸5分钟,在冰上冷却2分钟,然后加入到10ml的 DIG Easy Hyb中,产生杂交溶液。杂交在42℃下进行约17小时。然后在室温下将膜在低等严格条件下在2X SSC+0.1%SDS中洗涤5分钟,然后在65℃下在0.5X SSC+0.1%SDS中进行两次高等严格洗涤,每次15分钟。探针-靶标杂交物通过化学发光检定(Roche Diagnostics, Indianapolis, IN, USA)根据制造商说明书进行检测。DNA印迹分析表明,转化体在靶向的cbhII基因座处含有P.emersonii GH61A多肽编码序列和烟曲霉β-葡萄糖苷酶(Cel3A)突变体编码序列的单一拷贝。该转化体命名为里氏木霉QMJi033。

[0485] 实施例20:里氏木霉原生质体的生成和转化

[0486] 原生质体的制备和转化使用Penttila等人,1987, Gene 61:155-164 的改良操作指南进行。简言之,将里氏木霉菌株AgJg115-104-7B1 (PCT/US2010/061105;WO 2011/075677)在25ml补充有2% (w/v) 葡萄糖和10mM尿苷的YP培养基中在27℃下在90rpm轻缓震荡下培养17小时。使用真空驱动的一次性过滤系统通过过滤来收集菌丝体,并用去离子水洗涤两次,再用1.2M山梨糖醇洗涤两次。通过将洗涤的菌丝体在含有15mg GLUCANEX®200G/ml和0.36个单位的几丁质酶/ml的20ml 1.2M山梨糖醇中,在34℃下在90rpm轻缓震荡下悬浮15~25分钟,生成原生质体。通过在400x g下离心7分钟收集原生质体,并将其用冷的1.2M山梨糖醇洗涤两次。将原生质体用血球计数器计数,并在STC中再悬浮至1x10⁸原生质体/ml的最终浓度。将多余的原生质体在-80℃下储存在Cryo 1℃冷冻容器中。

[0487] 将约100μg在以下实施例中描述的转化质粒用Pme I消化。使用 TAE缓冲液通过1%琼脂糖凝胶电泳纯化消化反应物。将DNA条带从凝胶中切出,并使用QIAQUICK®凝胶提取试剂盒进行提取。将所得的纯化DNA加入到100μl原生质体溶液中,并轻缓混合。加入PEG 缓冲液(250μl),混合,并在34℃下孵育30分钟。然后加入STC(3 ml),混合,并铺在补充有1M蔗糖的PDA平板上。在28℃下孵育16 小时之后,向各个板加入20ml补充有35μg潮霉素B/ml的覆层PDA 培养基。将平板在28℃下孵育4~7天。

[0488] 实施例21:用烟曲霉cbhI基因取代内生里氏木霉cbhI基因

[0489] 为将里氏木霉内生cbhI基因(SEQ ID NO:1[DNA序列]和SEQ ID NO:2[推导的氨基

酸序列))用烟曲霉cbhI编码序列(SEQ ID NO:57 [DNA序列]和SEQ ID NO:58[推导的氨基酸序列])取代,根据实施例 20中所述的过程将里氏木霉ku70-菌株AgJg115-104-7B1 (PCT/US2010/061105;WO 2011/075677)用4x 2μg Pme I线性化的 pJfyS139(实施例11)转化。得到7个转化体,挑取每一种并转移至 PDA平板,在28℃下孵育7天。根据实施例9中所述的过程,从转化体中分离基因组DNA,并使各转化体进行DNA印迹分析。

[0490] 对于DNA印迹分析,将2μg基因组DNA在50μl反应体积中用 33个单位的Bgl II消化,并使用TAE缓冲液进行1%琼脂糖电泳。将凝胶中的DNA在0.25N HCl中经一次10分钟洗涤而脱嘌呤,在0.5N NaOH-1.5M NaCl中经两次15分钟洗涤而变性,在1M Tris pH 8-1.5M NaCl经一次30分钟洗涤而中和,并在20X SSC中孵育5分钟。使用 TURBOBLOTTER™系统根据制造商的操作指南将DNA转移至 **NYTRAN®** Supercharge膜。使用STRATALINKER™UV交联剂将DNA UV交联到膜上,并在20ml DIG Easy Hyb中在42℃下预杂交1小时。

[0491] 使用PCR Dig探针合成试剂盒根据制造商说明书用以下所示的正向和反向引物生成与3' cbhI基因侧翼序列杂交的探针。PCR反应由1X **HERCULASE®**反应缓冲液、400nM 各引物、200μM DIG标记的含有 dUTP的dNTP、20ng pJfyS139和1.5个单位的 **HERCULASE®**热启动高保真DNA聚合酶组成。在

EPPENDORF® MASTERCYCLER® 5333 epgradient S中孵育扩增反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、55℃30秒和72℃40秒,25个循环;以及在72℃7分钟,1个循环。

[0492] 正向引物:

[0493] 5'-AAAAACAAACATCCCGTTCATAAC-3' (SEQ ID NO:99)

[0494] 反向引物:

[0495] 5'-ACAAGGTTTACCGGTTTCGAAAAG-3' (SEQ ID NO:100)

[0496] 使用TAE缓冲液通过1%琼脂糖凝胶电泳纯化探针,其中将与探针对应的0.5kb条带从凝胶中切出,并使用**MINELUTE®**凝胶提取试剂盒进行提取。将探针煮沸5分钟,在冰上冷却2分钟,然后加入到 10ml DIG Easy Hyb中,产生杂交溶液。杂交在42℃下进行15~17小时。然后在室温下将膜在低等严格条件下在2X SSC+0.1%SDS中洗涤5分钟,然后在65℃下在0.5X SSC+0.1%SDS中进行两次高等严格洗涤,每次15分钟。探针-靶标杂交物通过化学发光检定(Roche Diagnostics,Indianapolis,IN,USA)根据制造商说明书进行检测。DNA 印迹分析表明,7个转化体中的3个在cbhI基因座处含有取代盒,选择一个转化体里氏木霉JfyS139-8用于消除hpt和tk标记物。

[0497] 通过将28℃下生长的七日龄PDA平板的孢子转移至PDA平板,并在28℃下孵育7天,生成新鲜的孢子平板。使用无菌涂布器将孢子收集在10ml 0.01% **TWEEN®** 20中。使用血球计数器测定孢子浓度,将10⁵孢子铺在含有补充有1μM 5-氟-2'-脱氧尿苷(FdU)的 TrMM-G 培养基的150mm板上。

[0498] 得到300个FDU抗性的孢子分离株,如上所述从2个孢子分离株中提取DNA。如上所述将分离株进行DNA印迹分析,结果表明两个孢子分离株均已切除在取代盒的同源重复序列之间的hpt/tk区。选择命名为里氏木霉JfyS139-8A的一个菌株用于取代cbhII基因。

[0499] 实施例22:用烟曲霉cbhII基因取代内生里氏木霉cbhII基因

[0500] 为将里氏木霉天然cbhII基因用烟曲霉cbhII编码序列取代,根据实施例20中所述

的过程将里氏木霉JfyS139-8A(实施例21)用2 μ g Pme I线性化并经凝胶纯化的pJfyS144(实施例10)转化。得到7个转化体,挑取每一个并转移至PDA平板,在28℃下孵育7天。使用下文所述的真菌孢子PCR法来筛选携带基因取代的转化体,使用以下所示的与整合区外的5' cbhII基因侧翼序列的上游区进行退火的正向引物,以及以下所示的用于烟曲霉cbhII编码序列的反向引物。

[0501] 正向引物:

[0502] 5'-AGCCACATGCCGCATATTGACAAAG-3' (SEQ ID NO:101)

[0503] 反向引物:

[0504] 5'-AGGGATTCA GTGTGCTACAGGCTGC-3' (SEQ ID NO:102)

[0505] 只有当在cbhII基因座处发生精确基因取代时,才会产生大约1.8 kb的PCR产物。如果盒已经整合在基因组中的别处,则不会发生扩增。

[0506] 将来自各转化体的少量孢子悬浮在25 μ l TE缓冲液中,并在微波炉中高温加热1分钟。使用各个微波过的孢子悬浮液作为PCR反应中的模板。反应由1 μ l微波过的孢子悬浮液、1 μ l的10mM dNTP、12.5 μ l的2X **ADVANTAGE®GC-Melt LA**悬浮液(Clontech,Mountain View, CA,USA)、25pmol正向引物、25pmol反向引物、1.25个单位的**ADVANTAGE®GC Genomic LA**聚合酶混合物(Clontech,Mountain View,CA,USA)和9.25 μ l水组成。在**EPPENDORF®MASTERCYCLER®5333epgradient S**中孵育反应物,程序设定为:95℃10分钟,1个循环;每循环95℃30秒、56℃30秒和72℃1分40秒,35个循环;以及72℃7分钟,1个循环;以及保持在4℃。使用TAE缓冲液使PCR反应物进行1%琼脂糖凝胶电泳。孢子PCR表明,7个转化体中的4个在靶向的基因座处含有取代盒,对其中的3个进行DNA印迹分析,以确认取代盒是单个拷贝。

[0507] 根据实施例9中所述的过程,从3个转化体中分离基因组DNA,并对各个转化体进行DNA印迹分析。对于DNA印迹分析,将2 μ g基因组DNA在50 μ l反应体积中用50个单位的Dra I消化,并使用TAE缓冲液使其进行1%琼脂糖电泳。将凝胶中的DNA在0.25N HCl中以一次10分钟洗涤进行脱嘌呤,在0.5N NaOH-1.5M NaCl中以两次15分钟洗涤进行变性,在1M Tris pH 8-1.5M NaCl中以1次30分钟洗涤进行中和,并在20X SSC中孵育5分钟。将DNA转移至**NYTRAN® Supercharge**膜。使用**STRATALINKER™**UV交联剂将DNA UV交联到膜上,并在20ml DIG Easy Hyb中在42℃下预杂交1小时。

[0508] 使用PCR Dig探针合成试剂盒根据制造商说明书,用以下示出的正向和反向引物生成与3' cbhII基因侧翼序列杂交的探针。PCR反应由1X **HERCULASE®**反应缓冲液、400nM各引物、200 μ M DIG标记的含有dUTP的dNTP、150ng里氏木霉RutC30基因组DNA和1.5个单位的**HERCULASE®**热启动高保真DNA聚合酶组成。在**EPPENDORF®MASTERCYCLER®5333epgradient S**中孵育反应物,程序设定为:95℃2分钟,1个循环;每循环95℃30秒、51℃30秒和72℃40秒,30个循环;以及72℃7分钟,1个循环。

[0509] 正向引物:

[0510] 5'-AAAAAACAACATCCCGTTCATAAC-3' (SEQ ID NO:103)

[0511] 反向引物:

[0512] 5'-AACAAGGTTTACCGGTTTCGAAAAG-3' (SEQ ID NO:104)

[0513] 使用TAE缓冲液通过1%琼脂糖凝胶电泳纯化探针,其中将与探针对应的0.5kb条

带从凝胶中切出,并使用QIAQUICK®凝胶提取试剂盒进行提取。将探针煮沸5分钟,在冰上冷却2分钟,并加入到10ml DIG Easy Hyb中,产生杂交溶液。杂交在42℃下进行约17小时。然后在室温下将膜在低等严格条件下在2X SSC+0.1%SDS中洗涤5分钟,然后在65℃下在0.5X SSC+0.1%SDS中进行两次高等严格洗涤,每次15分钟。探针-靶标杂交物通过化学发光检定(Roche Diagnostics, Indianapolis, IN, USA)根据制造商说明书进行检测。DNA印迹分析表明,3个转化体在cbhII基因座处含有取代盒,选择所有三个(命名为 JfyS139/144-5、-6和-10)用于消除hpt和tk标记物。

[0514] 通过将28℃下在PDA平板上生长的七日龄培养物的孢子转移至新的PDA平板,并在28℃下孵育7天,生成新鲜的孢子平板。使用无菌涂布器将孢子收集在10ml 0.01% TWEEN®20中。使用血球计数器测定孢子浓度,将 10^5 和 10^4 孢子铺在含有补充有1μM FdU的 TrMM-G培养基的150mm平板上。

[0515] 从含有 10^5 孢子的平板中得到各个转化体的约500个FdU抗性的孢子分离株,从含有 10^4 孢子的平板中针对各转化体分离约100个FdU抗性的孢子分离株。对于菌株JfyS139/144-5和-6,挑取8个孢子分离株,对于菌株JfyS139/144-10,挑取4个。将来自初级转化体5的各分离株1~8命名为JfyS139/144-5A至-5H。将来自初级转化体6的分离株1~8命名为JfyS139/144-6A至-6H。将来自初级转化体10的分离株1~4命名为JfyS139/144-10A至-10D。使用以下所示的正向和反向引物如上所述进行孢子PCR,以确认hpt和tk标记物已经被正确切除。

[0516] 正向引物:

[0517] 5'-GTTAAGCATACAATTGAACGAGAATGG-3' (SEQ ID NO: 105)

[0518] 反向引物:

[0519] 5'-GATGATATAATGGAGCAAATAAGGG-3' (SEQ ID NO:106)

[0520] 如上所述用以下循环参数进行PCR反应:95℃2分钟,1个循环;每个循环为95℃30秒、55℃30秒和72℃6分钟,30个循环;以及72℃7分钟,1个循环。

[0521] 引物与用于cbhII基因取代的5' (正向) 和3' (反向) 侧翼序列进行退火。已经将hpt/tk盒正确切除的菌株将会显示3.5kb片段,而标记物完整的那些则显示8kb片段。PCR筛选表明,所有的孢子分离株均已正确切除hpt/tk盒。

[0522] 从各初级转化体的A和B孢子分离株提取DNA,并如上所述对其进行DNA印迹分析。DNA印迹分析证实,各个孢子分离株均已正确切除了hpt/tk盒。选择孢子分离株里氏木霉JfyS139/144-10B来代表含有里氏木霉cbhI和cbhII基因二者均分别被烟曲霉同源物取代的菌株。

[0523] 实施例23:里氏木霉ku70基因修复质粒pTH239的生成

[0524] 使用IN-FUSION®Advantage PCR克隆试剂盒将4个DNA片段结合,以生成用内生里氏木霉ku70编码序列(SEQ ID NO:107[DNA序列]和SEQ ID NO:108[推导的氨基酸序列])取代被破坏的里氏木霉ku70编码序列的构建体。使用以下所示的序列特异性正向和反向引物(SEQ ID NO:109和110)从pJfyS139-B(实施例11)扩增包含原核复制原点的氨苄青霉素抗性标记物区。使用以下所示的序列特异性正向和反向引物(SEQ ID NO:111和112)从里氏木霉981-0-8基因组DNA扩增里氏木霉ku70基因上游序列(由来自ku70编码序列上游的989bp和ku70编码序列的前1010bp组成)。使用以下所示的序列特异性正向和反向引

物 (SEQ ID NO:113和114) 从里氏木霉981-0-8基因组DNA扩增里氏木霉ku70基因下游序列 (由从在上游PCR产物中扩增的ku70编码序列的1010 bp片段3'端重复出的500 bp片段、含有ku70 编码序列其余部分的1067 bp片段和来自ku70编码序列下游的461 bp 组成)。根据实施例9中所述的过程来制备里氏木霉981-0-8基因组 DNA。

[0525] 正向引物:

[0526] 5'-GTGTGCGGCCGCTCGAGCATGCATGTTTAAACAGCTTGGC ACTGGCCGTCGTTTT-3' (SEQ ID NO:109)

[0527] 反向引物:

[0528] 5'-ATCAGCCCCGAGACGGCGCCGCGTTTAAACAATTCGTAAT CATGGTCATAGCTGT-3' (SEQ ID NO:110)

[0529] 正向引物:

[0530] 5'-CATGATTACGAATTGTTTAAACGCGGCCGCTCTCGGGGC TGATCTTGTCGAGGA-3' (SEQ ID NO:111)

[0531] 反向引物:

[0532] 5'-GGCGGCCGTTACTAGTGGATCCAGCCCTTGACAGTGATCT TGAGTCCAGGTGCAA-3' (SEQ ID NO:112)

[0533] 正向引物:

[0534] 5'-TGCAGATATCCATCACACTGGCGGCCGAGTTTCCATGTCC AACGTGTTGTTTTGCGC-3' (SEQ ID NO:113)

[0535] 反向引物:

[0536] 5'-GCCAGTGCCAAGCTGTTTAAACATGCATGCTCGAGCGGCC GCACACGCCCTCTCCTCG-3' (SEQ ID NO:114)

[0537] 对于氨苄青霉素抗性标记物和原核复制原点区的扩增,反应由100 ng里氏木霉981-0-8基因组DNA、200μM dNTP、1μM各引物 (SEQ ID NO:109和110)、1X **PHUSION®**高保真热启动DNA聚合酶缓冲液和1.0个单位的**PHUSION®**高保真热启动DNA聚合酶组成,最终体积为50μl。在**EPENDORF® MASTERCYCLER®**5333 epgradient S中孵育扩增反应物,程序设定为:98℃30秒,1个循环;每循环98℃10 秒,55℃30秒和72℃1分30秒,30个循环;以及72℃7分钟,1个循环。使用TAE缓冲液通过1%琼脂糖凝胶电泳分离PCR产物,其中将 2.692 kb片段从凝胶中切出,并使用**MINELUTE®**凝胶提取试剂盒进行提取。

[0538] 对于ku70基因上游序列和下游序列的扩增,反应由100ng pJfyS139-B、200μM dNTP、1μM各引物 (SEQ ID NO:111和112或者113和114)、1X **PHUSION®**高保真热启动DNA聚合酶缓冲液和1.0 个单位的**PHUSION®**高保真热启动DNA聚合酶组成,最终体积为50 μl。在**EPENDORF® MASTERCYCLER®**5333epgradient S中孵育扩增反应物,程序设定为:98℃30秒,1个循环;每循环98℃10秒,55℃ 30秒和72℃1分30秒,30个循环;以及72℃7分钟,1个循环。使用 TAE缓冲液通过1%琼脂糖凝胶电泳分离PCR产物,其中分别将 1.999 kb和2.028kb的片段从凝胶中切出,并使用**MINELUTE®**凝胶提取试剂盒进行提取。

[0539] 从使用Not I和Bam HI的pJfyS139-B的限制性内切酶消化中产生第4个DNA片段。反应由5μg pJfyS139-B、10个单位的Not I、20个单位的Bam HI和20μl限制性内切酶缓冲液

2 (New England Biolabs, Inc., Ipswich, MA, USA) 组成, 总体积为50 μ l。将反应物在37 $^{\circ}$ C下孵育1小时, 然后使用TAE缓冲液通过1%琼脂糖凝胶电泳分离, 其中将4.400kb片段从凝胶中切出, 并使用MINELUTE[®]凝胶提取试剂盒进行提取。

[0540] 使用IN-FUSION[®] Advantage PCR克隆试剂盒根据制造商的操作指南将2,028bp、1,999bp和2,692bp的三种PCR产物插入到Not I和 Bam HI消化的pJfyS139-B中。IN-FUSION[®]反应由1X IN-FUSION[®]反应缓冲液、50ng Not I/Bam HI消化的pJfyS139-B、50ng 1.999kb ku70 基因上游PCR产物、50ng 2.028kb ku70基因下游PCR产物、50ng 2.692 kb氨苄青霉素抗性标记物和原核复制原点PCR产物和1 μ l IN-FUSION[®]酶组成, 反应体积为10 μ l。反应物在37 $^{\circ}$ C下孵育15分钟, 然后在50 $^{\circ}$ C下孵育15分钟。孵育期之后, 向反应物中加入40 μ l TE。使用3 μ l等份根据制造商的操作指南来转化大肠杆菌XL10 GOLD[®]感受态细胞 (Stratagene, La Jolla, CA, USA)。将细胞在42 $^{\circ}$ C下热击30秒, 然后加入500 μ l预热至42 $^{\circ}$ C的NZY+培养基。将管在37 $^{\circ}$ C下以200rpm 震荡孵育40分钟, 然后铺在直径150mm的2XYT+氨苄青霉素平板上, 在37 $^{\circ}$ C下孵育过夜。通过用Hind III和Xba I进行限制性酶切消化分析, 筛选所得的转化体, 并对阳性克隆进行测序, 以确保不存在PCR错误。识别出包含没有PCR错误的插入的一个克隆, 并将其命名为pTH239。

[0541] 实施例24: 烟曲霉cbh1和cbh2取代菌株JfyS139/144-10B中ku70基因的修复

[0542] 在菌株里氏木霉JfyS139/144-10B(实施例22)中对内生里氏木霉 ku70基因进行修复, 以促进需要ku70基因在非源性末端结合中的功能的菌株操纵步骤。根据实施例4中所述的过程, 将里氏木霉 JfyS129/144-10B用23x 2 μ g Pme I线性化的pTH239(实施例23)进行转化。得到19个转化体, 将各个分别转移至PDA平板, 并在28 $^{\circ}$ C下孵育7天。

[0543] 通过PCR来筛选所有19个转化体, 以确认pTH239Pme I片段在破坏的ku70基因基因座处的同源整合。对于各个转化体, 使用无菌接种环从7日龄PDA平板中收集孢子。将孢子转移至含有25 μ l 1mM EDTA-10mM Tris缓冲液的试管, 并用高功率微波处理1分钟。将1 μ l 等份的微波处理过的孢子混合物直接加入到PCR反应中作为模板 DNA。设计以下所示的一组PCR引物, 以跨越ku70编码序列的破坏区进行扩增, 从而区分具有ku70编码序列中的破坏的宿主基因组 (848 bp) 与pTH239靶向目标菌株 (606bp)。PCR反应由1X ADVANTAGE[®] Genomic LA聚合酶反应缓冲液 (Clontech, Mountain View, CA, USA)、400nM各引物、200 μ M dNTP、1 μ l微波处理过的TE-孢子混合物 (如上所述) 和1.0个单位的ADVANTAGE[®] Genomic LA聚合酶 (Clontech, Mountain View, CA, USA) 组成。在EPPENDORF[®] MASTERCYCLER[®] 5333epgradient S中孵育扩增反应物, 程序设定为: 95 $^{\circ}$ C 10分钟, 1个循环; 每循环95 $^{\circ}$ C 30秒, 55 $^{\circ}$ C 30秒和72 $^{\circ}$ C 60秒, 30个循环; 以及72 $^{\circ}$ C 7分钟, 1个循环。

[0544] 正向引物:

[0545] 5' -CAATGACGATCCGCACGCGT-3' (SEQ ID NO:115)

[0546] 反向引物:

[0547] 5' -CAATGACGATCCGCACGCGT-3' (SEQ ID NO:116)

[0548] 仅19个转化体中的1个 (#19) 对于606bp PCR产物为阳性且对 848bp PCR产物为阴性, 表明为包含同源整合在ku70基因座处的pTH239PmeI片段的菌株。

[0549] 使用无菌涂布器, 将来自转化体#19的七日龄PDA平板的孢子收集在10ml 0.01%

TWEEN®20中。使用血球计数器确定孢子浓度,将 10^6 孢子铺在含有补充有 $1\mu\text{M}$ 5-氟-2'-脱氧尿苷 (FdU) 的 TrMM-G 培养基的 150mm 平板上,并在 28°C 下培养 5 天。得到 22 个 FdU 抗性的孢子分离株,将其转化至 PDA 平板,并在 28°C 下培养 5 天。

[0550] 针对存在于修复盒内的在 ku70 编码序列的同源重复之间的 hpt/tk 标记物区的切除,通过 PCR 筛选所有 22 个孢子分离株 (#19A-V)。对于各个孢子分离株,使用无菌接种环从 7 日龄 PDA 平板中收集孢子。将孢子转移至含有 $25\mu\text{l}$ 1mM EDTA- 10mM Tris 缓冲液的试管,并用高功率微波处理 1 分钟。将 $1\mu\text{l}$ 等份的孢子混合物直接加入到 PCR 反应中作为模板基因组 DNA。设计以下所示的一组 PCR 引物,以跨越 hpt/tk 区进行扩增,以区分 hpt/tk 区的存在 (6kb) 或不存在 (1.1kb)。PCR 反应由 1X **ADVANTAGE®** Genomic LA 聚合酶反应缓冲液、 400 nM 各引物 (以下所示)、 $200\mu\text{M}$ dNTP、 $1\mu\text{l}$ 微波处理过的 TE-孢子混合物 (如上所述) 和 1.0 个单位的 **ADVANTAGE®** Genomic LA 聚合酶组成。在

EPPENDORF® MASTERCYCLER®5333epgradient S 中孵育扩增反应物,程序设定为: 95°C 10 分钟, 1 个循环; 每循环 95°C 30 秒, 50°C 30 秒和 72°C 6 分钟, 30 个循环; 以及 72°C 7 分钟, 1 个循环。

[0551] 正向引物:

[0552] 5' -GACACTCTTTTCTCCCATCT-3' (SEQ ID NO:117)

[0553] 反向引物:

[0554] 5' -GAGGAGCAGAAGAAGCTCCG-3' (SEQ ID NO:118)

[0555] 所有 22 个孢子分离株对于与 hpt/tk 标记物区对应的 6kb PCR 产物均为阴性。

[0556] 使用无菌涂布器,将来自分离株 #19A 和 #19L 的七日龄 PDA 平板的孢子收集在 10ml 0.01% **TWEEN®20** 中。使用血球计数器确定孢子浓度,将 10^3 、 10^2 和 10^1 孢子铺在含有 1M 蔗糖的 150mm PDA 平板上,在 28°C 下培养 3 天。从菌株 #19A 和 #19L 的 PDA 平板选择 10 个孢子分离株,并转移至新鲜 PDA 平板,置于 28°C 下。

[0557] 根据实施例 9 中所述的过程从 #19L 和 #19A 的 6 个孢子分离株中提取基因组 DNA,并对其进行 DNA 印迹分析。

[0558] 对于 DNA 印迹分析,将 $2\mu\text{g}$ 基因组 DNA 在 $50\mu\text{l}$ 反应体积中用 (1) 分别 5 个单位和 10 个单位的 Asc I 和 Xho I 或 (2) 分别 5 个单位和 25 个单位的 Asc I 和 Apa I 消化,并使用 TAE 缓冲液进行 1% 琼脂糖电泳。将凝胶中的 DNA 在 0.25N HCl 以一次 10 分钟洗涤进行脱嘌呤,在 0.5N NaOH- 1.5M NaCl 中以两次 15 分钟洗涤进行变性,在 1M Tris pH 8- 1.5M NaCl 中以一次 30 分钟洗涤进行中和,并在 20X SSC 中孵育 5 分钟。使用 TURBOBLOTTER™ 系统根据制造商的操作指南将 DNA 转移至 **NYTRAN®** Supercharge 膜。使用 STRATALINKER™UV 交联剂将 DNA 经 UV 交联到膜上,并在 20ml DIG Easy Hyb 中在 42°C 下预杂交 1 小时。

[0559] 使用 PCR Dig 探针合成试剂盒根据制造商说明书用以下所示的正向和反向引物生成与 ku70 编码序列的 3' 端杂交的探针。为生成纯的用于探针 PCR 反应的模板,从里氏木霉 981-0-8 基因组 DNA 扩增 ku70 编码序列的 3' 端。PCR 反应由 1X **PHUSION®** 高保真热启动 DNA 聚合酶缓冲液、 $1\mu\text{M}$ 各引物、 $200\mu\text{M}$ dNTP、 165ng 里氏木霉 981-0-8 基因组 DNA 和 1.0 个单位的 **PHUSION®** 高保真热启动 DNA 聚合酶组成。在

EPPENDORF® MASTERCYCLER®5333epgradient S 中孵育扩增反应物,程序设定为: 98°C 30 秒, 1 个循环; 每循环 98°C 10 秒、 60°C 30 秒和 72°C 15 秒, 35 个循环; 以及 72°C 10 分

钟,1个循环。

[0560] 正向引物:

[0561] 5'-gcatatataaccactcaagta-3' (SEQ ID NO:119)

[0562] 反向引物:

[0563] 5'-attatcttgaccggccgcagg-3' (SEQ ID NO:120)

[0564] 使用TAE缓冲液通过1%琼脂糖凝胶电泳纯化0.5kb探针模板,并将其从凝胶中切出,使用MINELUTE®凝胶提取试剂盒进行提取。使用上述的引物和扩增条件,如制造商说明书所述,使用纯化的PCR产物生成DIG标记的探针。使用TAE缓冲液通过1%琼脂糖凝胶电泳纯化0.5kb DIG标记的探针,并将其从凝胶中切出,使用MINELUTE®凝胶提取试剂盒进行提取。将探针煮沸5分钟,在冰上冷却2分钟,并加入到10ml DIG Easy Hyb中,产生杂交溶液。杂交在42℃下进行15~17小时。然后在室温下将膜在低等严格条件下在2X SSC+0.1% SDS中洗涤5分钟,然后在65℃下在0.5X SSC+0.1% SDS中进行两次高等严格洗涤,每次15分钟。探针-靶标杂交物通过化学发光检定 (Roche Diagnostics, Indianapolis, IN, USA) 根据制造商说明书进行检测。DNA印迹分析表明,所有孢子分离株均在ku70基因座处含有修复/取代盒,并消除了hpt和tk标记物。选择命名为里氏木霉 981-0-8.5#10B+Ku70#19L3的一个菌株,用于进一步的转化。

[0565] 实施例25:里氏木霉981-0-8.5#10B+Ku70#19L3原生质体的生成和转化

[0566] 原生质体的制备和转化使用Penttila等人,1987, Gene 61:155-164 的改良方案进行。简言之,将里氏木霉菌株981-0-8.5#10B+Ku70#19L3 在25ml补充有2% (w/v) 葡萄糖和10mM尿苷的YP培养基中在27℃下在90rpm轻缓震荡下培养17小时。使用真空驱动的一次性过滤系统通过过滤收集菌丝体,并用去离子水洗涤两次,用1.2M山梨糖醇洗涤两次。通过将洗涤过的菌丝体在含有15mg GLUCANEX®200G/ml 和0.36个单位的几丁质酶/ml的20ml 1.2M山梨糖醇中在34℃下随90 rpm轻缓震荡悬浮15~25分钟。通过在400x g离心7分钟收集原生质体,并将其用冷的1.2M山梨糖醇洗涤两次。将原生质体用血球计数器计数,并再悬浮在STC中至最终浓度为 1×10^8 原生质体/ml。将过量的原生质体在-80℃下储存在Cryo 1℃冷冻容器中。

[0567] 将约100μg质粒pAmFs074用Pme I消化。使用TAE缓冲液通过 0.8%琼脂糖凝胶电泳纯化消化反应物。将含有表达盒的约9.4kb DNA 条带从凝胶中切出,上述表达盒包含Penicillium emersonii GH61A多肽编码序列、烟曲霉β-葡糖苷酶突变体编码序列和amdS标记物,并使用QIAQUICK®凝胶提取试剂盒根据制造商建议的操作指南进行提取。

[0568] 将所得的9.4kb片段(5x 2μg或2x 3.5μg的9.4kb Pme I消化的 pAmFs074)加入到100μl原生质体溶液中,并轻缓混合。加入PEG 缓冲液(250μl),将反应物混合,并在34℃下孵育30分钟。然后加入 STC(3ml),将反应物混合,然后铺在用于amdS选择的COVE平板上。将平板在28℃下孵育6~11天。

[0569] 实施例26:对摇瓶中里氏木霉转化体进行Penicillium emersonii GH61A多肽和烟曲霉β-葡糖苷酶变体的表达的评价

[0570] 使用接种环将112个里氏木霉981-0-8.5#10B+Ku70#19L3转化体从COVE转化平板转移至补充有10mM尿苷的COVE2平板,并在28℃下孵育5~7天。用接种环收集孢子,并将其转移至125ml塑料摇瓶中的25ml CIM培养基中。将摇瓶培养物在28℃、200rpm下孵育5天。将

约14~15ml的各培养物倒入15ml试管中。将试管在863x g下离心 20分钟。将上清液倾倒入新的试管中。使用 **BIOMEK® 3000**、**BIOMEK® NX** 和 **ORCA®** 机械臂 (Beckman Coulter, Inc, Fullerton, CA, USA) 对上清液进行 β -葡萄糖苷酶活性的检定。将上清液适当稀释于0.1M琥珀酸盐、0.01% **TRITON® X-100** (4-(1,1,3,3-四甲基丁基) 苯基-聚乙二醇) 缓冲液pH 5.0 (样本缓冲液) 中,然后将稀释的样本连续稀释0倍至 1/3倍至1/9倍。将共20 μ l的各稀释液共计转移至96孔平底板。向各孔中加入200微升于0.1M琥珀酸盐pH 5.0缓冲液中的1mg/ml对硝基苯基- β -D-吡喃葡萄糖苷底物,然后在环境温度下孵育45分钟。孵育期一结束,向各孔中加入50 μ l的淬灭缓冲液 (1M Tris pH 9)。对于96 孔板,在405nm的光密度处测量终点。通过SDS-PAGE对33个示出大于16,000 μ M/min/ml的活性的样本进行分析,以确定GH61多肽的表达。

[0571] 使用 **CRITERION® 8~16%** Tris-HCl凝胶根据制造商说明书经 SDS-PAGE对10 μ l的各培养上清液进行分析。将所得的凝胶用 BIO-SAFE™考马斯染色。培养物的SDS-PAGE图谱示出,转化体产生与烟曲霉 β -葡萄糖苷酶变体对应的约130kDa、以及与 *Penicillium emersonii* GH61A GH61A多肽对应的约27kDa的主要蛋白条带。基于 SDS-PAGE图谱上GH61A多肽的表达,选择21个转化体,用于进一步的孢子纯化和分析。

[0572] 孢子纯化的进行是通过用1 μ l接种环接触孢子平板,之后将环浸在1ml 0.01% **TWEEN® 20**中,然后进行涡旋。将1微升和10微升的该孢子混合物与100 μ l 0.01% **TWEEN® 20**混合,并铺在大的COVE 平板上。将平板在28℃下孵育,直至在板上出现单个的孢子分离株。

[0573] 将所有21个候选物的单个孢子分离株均转移至COVE2+10mM 尿苷平板,并在28℃下孵育约1周,之后接种至含有25ml CIM培养基的125ml塑料摇瓶。如上所述通过SDS-PAGE对所得的摇瓶培养物进行分析,并与初级转化体的摇瓶培养物进行比较。选择6个孢子纯化的转化体,即597A、676D、679C、680A、683A和686C,用于2 升发酵,以测定收率。

[0574] 实施例27:发酵

[0575] 将前6个孢子纯化的里氏木霉981-0-8.5#10B+Ku70#19L3转化体 (597A、676D、679C、680A、683A和686C) 和里氏木霉QMj1033 在28℃下在PDA平板上培养约1周。

[0576] 摇瓶培养基由20g右旋糖、10g玉米浆冻干粉、1.45g (NH₄)₂SO₄、2.08g KH₂PO₄、0.36g CaCl₂、0.42g MgSO₄·7H₂O和0.42ml痕量金属溶液组成。痕量金属溶液由216g FeCl₃·6H₂O、58g ZnSO₄·7H₂O、27g MnSO₄·H₂O、10g CuSO₄·5H₂O、2.4g H₃BO₃、336g柠檬酸和加至1 升的去离子水组成。

[0577] 将100ml摇瓶培养基加入到500ml摇瓶中。用两个塞子从固体平板培养物接种摇瓶,并在28℃下在定轨摇床上在200rpm下孵育48小时。使用50ml摇瓶肉汤来接种2升的发酵容器。

[0578] 分批发酵培养基每升由30g纤维素、4g右旋糖、10g玉米浆冻干粉、3.8g (NH₄)₂SO₄、2.8g KH₂PO₄、2.64g CaCl₂、1.63g MgSO₄·7H₂O、1.8ml消泡剂和0.66ml痕量金属溶液组成。痕量金属溶液每升由216g FeCl₃·6H₂O、58g ZnSO₄·7H₂O、27g MnSO₄·H₂O、10g CuSO₄·5H₂O、2.4g H₃BO₃和336g柠檬酸组成。发酵进料培养基由右旋糖组成。

[0579] 将共1.8升的分批发酵培养基加入到 **APPLIKON® 3** 升玻璃夹套发酵罐 (Applikon Biotechnology Inc., Foster City CA USA) 中。将发酵进料培养基以0~4g/l/

hr的速率给予185小时的时间段。将发酵容器保持在28℃温度下,使用**APPLIKON®**1030控制系统(Applikon Biotechnology Inc., Foster City CA USA)将pH控制在4.5+/-0.1的设定点。以1vvm的速率将空气加入至容器中,通过以1100~1300rpm转动的Rushton桨搅动肉汤。发酵结束时,从容器中收获全部肉汤,并在 3000x g下离心,以除去生物质。将上清液无菌过滤,并在5~10℃下储存。

[0580] 实施例28:对2升发酵肉汤中的里氏木霉转化体进行Penicillium emersonii GH61A多肽和烟曲霉β-葡糖苷酶变体在2升发酵肉汤中表达的评价

[0581] 通过SDS-PAGE对前6个孢子纯化的里氏木霉 981-0-8.5#10B+Ku70#19L3转化体(597A、676D、679C、680A、683A 和686C)和里氏木霉QMJi033的发酵肉汤进行分析,以评价Penicillium emersonii GH61A多肽和烟曲霉β-葡糖苷酶变体的表达水平。将0.5μl 体积的各菌株的发酵肉汤上清液与25μl去离子水以及25μl含有5%2- 巯基乙醇的Laemmli染料(Bio-Rad Laboratories, Hercules, CA, USA) 混合。将混合物在95℃加热5分钟,将10μl(相当于0.1μl发酵肉汤) 根据制造商说明书装载在不含染料的**CRITERION®**8~16%Tris-HCl凝胶(Bio-Rad Laboratories, Hercules, CA, USA) 上。将凝胶在200伏特下跑大约55分钟,然后通过Gel Doc™EZ成像器(Bio-Rad Laboratories, Hercules, CA, USA) 分析。发酵肉汤上清液的SDS-PAGE图谱示出,里氏木霉QMJi033中Penicillium emersonii GH61A多肽和烟曲霉β-葡糖苷酶变体的表达与前6个里氏木霉981-0-8.5#10B+Ku70#19L3转化体(597A、676D、679C、680A、683A和686C)是相当的。

[0582] 通过以下编号的段落对本发明进行进一步的说明:

[0583] [1]一种构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,其包括:(a)通过向丝状真菌菌株中导入第一串联构建体,通过靶向整合取代内源性第一基因,该第一串联构建体包含(i) 第一基因的同源5' 区、其同源侧翼区或其组合,(ii) 一个或多个第一可选择性标记物,(iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v) 第一基因的同源3' 区、其同源侧翼区或其组合;(b)通过向丝状真菌菌株中导入第二串联构建体,通过靶向整合取代内源性第二基因,该第二串联构建体包含(i) 第二基因的同源5' 区、其同源侧翼区或其组合,(ii) 一个或多个第二可选择性标记物,(iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v) 第二基因的同源3' 区、其同源侧翼区或其组合;或(c) (a) 和(b) 的组合。

[0584] [2]段落1的方法,其中第一基因是纤维二糖水解酶I基因。

[0585] [3]段落2的方法,其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I:(i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶 I;(ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I;(iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至

少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I;和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0586] [4]段落1的方法,其中第二基因是纤维二糖水解酶II基因。

[0587] [5]段落4的方法,其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II;(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II;(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II;和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0588] [6]段落1~5中任一段的方法,其中各个串联构建体通过同源重组整合到丝状真菌菌株的染色体中。

[0589] [7]段落1~6中任一段的方法,其中第一基因的同源5'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0590] [8]段落1~7中任一段的方法,其中第一基因的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0591] [9]段落1~8中任一段的方法,其中第二基因的同源5'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0592] [10]段落1~9中任一段的方法,其中第二基因的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0593] [11]段落1~10中任一段的方法,其还包括通过向丝状真菌菌株中导入针对各个基因的对应串联构建体各自经靶向整合取代一个或多个其他内源性基因,该串联构建体包含:(i)基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个可选择性标记物;(iii)与启动子和终止子可操作连接的编码具有生物活性的多肽的多核苷酸;(iv)与另一启动子和另一终止子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸;和(v)基因的同源3'区、其同源侧翼区或其组合。

[0594] [12]段落1~11中任一段的方法,其中一个或多个串联构建体还包含一个或多个可选择性标记物的5'侧翼的第一同源重复和一个或多个可选择性标记物的3'侧翼的第二

同源重复,其中第一同源重复和第二同源重复进行同源重组,以切除一个或多个可选择性标记物。

[0595] [13]段落12的方法,其中第一和第二同源重复是相同的,或者彼此具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0596] [14]段落12或13的方法,其中第一和第二同源重复各自为至少 50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0597] [15]段落12~14中任一段的方法,其中在切除一个或多个可选择性标记物时,一个或多个可选择性标记物可以再用于将针对各个基因的对应串联构建体各自通过靶向整合来取代一个或多个其他内源性基因。

[0598] [16]段落1~15中任一段的方法,其还包括将丝状真菌宿主细胞用串联构建体转化,该串联构建体包含:(i) 一个或多个可选择性标记物;(ii) 与第五启动子和第五终止子可操作连接的编码具有生物活性的第五多肽的第五多核苷酸;和(iii) 与第六启动子和第六终止子可操作连接的编码具有生物活性的第六多肽的第六多核苷酸,其中该串联构建体通过异位整合而整合。

[0599] [17]段落1~16中任一段的方法,其中具有生物活性的多肽是不同的多肽。

[0600] [18]段落1~16中任一段的方法,其中两个或更多个具有生物活性的多肽是相同的多肽。

[0601] [19]段落1~18中任一段的方法,其中启动子是不同的启动子。

[0602] [20]段落1~18中任一段的方法,其中两个或更多个启动子是相同的启动子。

[0603] [21]段落1~20中任一段的方法,其中终止子是不同的终止子。

[0604] [22]段落1~20中任一段的方法,其中两个或更多个终止子是相同的终止子。

[0605] [23]段落1~22中任一段的方法,其中一个或多个串联构建体包含在表达载体中。

[0606] [24]段落1~23中任一段的方法,其中丝状真菌菌株是枝顶孢属、曲霉属、短梗霉属、黑管菌属、拟蜡霉属、金孢属、鬼伞属、革盖菌属、隐球菌属、线黑粉酵母属、镰孢菌属、腐质霉属、巨座壳属、毛霉属、毁丝菌属、新美鞭菌属、脉孢菌属、拟青霉属、青霉属、平革菌属、白腐菌属、梨囊鞭菌属、侧耳属、裂褶菌属、篮状菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属或木霉属菌株。

[0607] [25]段落24的方法,其中木霉属菌株选自哈茨木霉、康氏木霉、长梗木霉、里氏木霉或绿色木霉。

[0608] [26]段落24的方法,其中木霉属菌株是里氏木霉。

[0609] [27]一种丝状真菌菌株,根据段落1~26中任一段的方法得到。

[0610] [28]一种丝状真菌菌株,其包含:(a) 通过靶向整合用第一串联构建体取代的内源性第一基因,该第一串联构建体包含(i) 第一基因的同源5' 区、其同源侧翼区或其组合,(ii) 一个或多个可选择性标记物,(iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v) 第一基因的同源3' 区、其同源侧翼区或其组合;(b) 通过靶向整合用第二串联构建体取代的内源性第二基因,该第二串联构建体包

含 (i) 第二基因的同源5' 区、其同源侧翼区或其组合, (ii) 一个或多个可选择性标记物, (iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 第二基因的同源3' 区、其同源侧翼区或其组合; 或 (c) (a) 和 (b) 的组合。

[0611] [29]段落28的丝状真菌菌株, 其中第一基因是纤维二糖水解酶I 基因。

[0612] [30]段落29的丝状真菌菌株, 其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I: (i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I; (ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I; (iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I; 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0613] [31]段落28的丝状真菌菌株, 其中第二基因是纤维二糖水解酶II 基因。

[0614] [32]段落31的丝状真菌菌株, 其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II: (i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II; (ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II; (iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II; 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0615] [33]段落28~32中任一段的丝状真菌菌株, 其中各个串联构建体通过同源重组整合到丝状真菌菌株的染色体中。

[0616] [34]段落28~33中任一段的丝状真菌菌株, 其中第一串联构建体的第一基因的同源5' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0617] [35]段落28~34中任一段的丝状真菌菌株, 其中第一串联构建体的第一基因的同源3' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0618] [36]段落28~35中任一段的丝状真菌菌株, 其中第二串联构建体的第二基因的同源5' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少

800bp、至少1000bp、至少1500bp或至少2000bp。

[0619] [37]段落28~36中任一段的丝状真菌菌株,其中第二串联构建体的第二基因的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0620] [38]段落28~37中任一段的丝状真菌菌株,其还包含一个或多个各自通过靶向整合用针对各基因的对应串联构建体取代的其他内源性基因,该串联构建体包含:(i)基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个可选择性标记物;(iii)与启动子和终止子可操作连接的编码具有生物活性的多肽的多核苷酸;(iv)与另一启动子和另一终止子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸;和(v)基因的同源3'区、其同源侧翼区或其组合。

[0621] [39]段落28~38中任一段的丝状真菌菌株,其中一个或多个串联构建体还包含一个或多个可选择性标记物的5'侧翼的第一同源重复和一个或多个可选择性标记物的3'侧翼的第二同源重复,其中第一同源重复和第二同源重复进行同源重组,以切除一个或多个可选择性标记物。

[0622] [40]段落39的丝状真菌菌株,其中第一和第二同源重复是相同的,或者彼此具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0623] [41]段落39或40的丝状真菌菌株,其中第一和第二同源重复各自为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0624] [42]段落39~41中任一段的丝状真菌菌株,其中在切除一个或多个可选择性标记物时,一个或多个可选择性标记物可以再用于将针对各个基因的对应串联构建体各自通过靶向整合来取代一个或多个其他内源性基因。

[0625] [43]段落28~42中任一段的丝状真菌菌株,其还包含串联构建体,该串联构建体包含:(i)一个或多个可选择性标记物;(ii)与第五启动子和第五终止子可操作连接的编码具有生物活性的第五多肽的第五多核苷酸;和(iii)与第六启动子和第六终止子可操作连接的编码具有生物活性的第六多肽的第六多核苷酸,其中该串联构建体通过异位整合而整合。

[0626] [44]段落28~43中任一段的丝状真菌菌株,其中具有生物活性的多肽是不同的多肽。

[0627] [45]段落28~43中任一段的丝状真菌菌株,其中两个或更多个具有生物活性的多肽是相同的多肽。

[0628] [46]段落28~45中任一段的丝状真菌菌株,其中启动子是不同的启动子。

[0629] [47]段落28~45中任一段的丝状真菌菌株,其中两个或更多个启动子是相同的启动子。

[0630] [48]段落28~47中任一段的丝状真菌菌株,其中终止子是不同的终止子。

[0631] [49]段落28~47中任一段的丝状真菌菌株,其中两个或更多个终止子是相同的终止子。

[0632] [50]段落28~49中任一段的丝状真菌菌株,其中一个或多个串联构建体包含在表达载体中。

[0633] [51]段落28~50中任一段的丝状真菌菌株,其中丝状真菌菌株是枝顶孢属、曲霉属、短梗霉属、黑管菌属、拟蜡霉属、金孢属、鬼伞属、革盖菌属、隐球菌属、线黑粉酵母属、镰孢菌属、腐质霉属、巨座壳属、毛霉属、毁丝菌属、新美鞭菌属、脉孢菌属、拟青霉属、青霉属、平革菌属、白腐菌属、梨囊鞭菌属、侧耳属、裂褶菌属、篮状菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属或木霉属菌株。

[0634] [52]段落51的丝状真菌菌株,其中木霉属菌株选自哈茨木霉、康氏木霉、长梗木霉、里氏木霉或绿色木霉。

[0635] [53]段落51的丝状真菌菌株,其中木霉属菌株是里氏木霉。

[0636] [54]一种在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法,其包括在有益于产生多肽的条件下培养丝状真菌宿主细胞,其中丝状真菌宿主细胞包含:(a)通过靶向整合用第一串联构建体取代的内源性第一基因,该第一串联构建体包含(i)第一基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)第一基因的同源3'区、其同源侧翼区或其组合;(b)通过靶向整合用第二串联构建体取代的内源性第二基因,该第二串联构建体包含(i)第二基因的同源5'区、其同源侧翼区或其组合,(ii)一个或多个可选择性标记物,(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)第二基因的同源3'区、其同源侧翼区或其组合;或(c)(a)和(b)的组合。

[0637] [55]权利要求54的方法,其还包括回收多种重组多肽。

[0638] [56]段落54或55的方法,其中第一基因是纤维二糖水解酶I基因。

[0639] [57]段落56的方法,其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I:(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I;(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I;(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I;和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0640] [58]段落54或55的方法,其中第二基因是纤维二糖水解酶II基因。

[0641] [59]段落58的方法,其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II;(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至

少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II；(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II；和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0642] [60]段落54~59中任一段的方法，其中第一串联构建体的第一基因的同源5'区、其同源侧翼区或其组合为至少50bp，例如，至少100 bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500 bp或至少2000bp。

[0643] [61]段落54~60中任一段的方法，其中第一串联构建体的第一基因的同源3'区、其同源侧翼区或其组合为至少50bp，例如，至少100 bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500 bp或至少2000bp。

[0644] [62]段落54~61中任一段的方法，其中第二串联构建体的第二基因的同源5'区、其同源侧翼区或其组合为至少50bp，例如，至少100 bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500 bp或至少2000bp。

[0645] [63]段落54~62中任一段的方法，其中第二串联构建体的第二基因的同源3'区、其同源侧翼区或其组合为至少50bp，例如，至少100 bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500 bp或至少2000bp。

[0646] [64]段落54~63中任一段的方法，其中丝状真菌宿主细胞还包含一个或多个各自通过靶向整合用针对各个基因的对应串联构建体取代的其他内源性基因，该串联构建体包含：(i)基因的同源5'区、其同源侧翼区或其组合；(ii)一个或多个可选择性标记物；(iii)与启动子和终止子可操作连接的编码具有生物活性的多肽的多核苷酸；(iv)与另一启动子和另一终止子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸；和(v)基因的同源3'区、其同源侧翼区或其组合。

[0647] [65]段落54~64中任一段的方法，其中一个或多个串联构建体还包含一个或多个可选择性标记物的5'侧翼的第一同源重复和一个或多个可选择性标记物的3'侧翼的第二同源重复，其中第一同源重复和第二同源重复进行同源重组，以切除一个或多个可选择性标记物。

[0648] [66]段落65的方法，其中第一和第二同源重复是相同的，或者彼此具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的的序列同一性。

[0649] [67]段落65或66的方法，其中第一和第二同源重复各自为至少 50bp，例如，至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0650] [68]段落65~67中任一段的方法，其中在切除一个或多个可选择性标记物时，一个或多个可选择性标记物可以再用于将针对各个基因的对应串联构建体各自通过靶向整

合来取代一个或多个其他内源性基因。

[0651] [69]段落54~68中任一段的方法,其中丝状真菌宿主细胞还包含串联构建体,该串联构建体包含:(i)一个或多个可选择性标记物;(ii)与第五启动子和第五终止子可操作连接的编码具有生物活性的第五多肽的第五多核苷酸;和(iii)与第六启动子和第六终止子可操作连接的编码具有生物活性的第六多肽的第六多核苷酸,其中该串联构建体通过异位整合而整合。

[0652] [70]段落54~69中任一段的方法,其中具有生物活性的多肽是不同的多肽。

[0653] [71]段落54~69中任一段的方法,其中两个或更多个具有生物活性的多肽是相同的多肽。

[0654] [72]段落54~71中任一段的方法,其中启动子是不同的启动子。

[0655] [73]段落54~71中任一段的方法,其中两个或更多个启动子是相同的启动子。

[0656] [74]段落54~73中任一段的方法,其中终止子是不同的终止子。

[0657] [75]段落54~73中任一段的方法,其中两个或更多个终止子是相同的终止子。

[0658] [76]段落54~75中任一段的方法,其中一个或多个串联构建体包含在表达载体中。

[0659] [77]段落54~76中任一段的方法,其中丝状真菌菌株是枝顶孢属、曲霉属、短梗霉属、黑管菌属、拟蜡霉属、金孢属、鬼伞属、革盖菌属、隐球菌属、线黑粉酵母属、镰孢菌属、腐质霉属、巨座壳属、毛霉属、毁丝菌属、新美鞭菌属、脉孢菌属、拟青霉属、青霉属、平革菌属、白腐菌属、梨囊鞭菌属、侧耳属、裂褶菌属、篮状菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属或木霉属菌株。

[0660] [78]段落77的方法,其中木霉属菌株选自哈茨木霉、康氏木霉、长梗木霉、里氏木霉和绿色木霉。

[0661] [79]段落77的方法,其中木霉属菌株是里氏木霉。

[0662] [80]一种串联构建体,其包含:(i)基因的同源5'区、其同源侧翼区或其组合;(ii)一个或多个第一可选择性标记物;(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸;(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸;和(v)基因的同源3'区、其同源侧翼区或其组合。

[0663] [81]段落80的串联构建体,其中基因是纤维二糖水解酶I基因。

[0664] [82]段落81的串联构建体,其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I:(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I;(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I;(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I;和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ

ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0665] [83]段落80的串联构建体,其中基因是纤维二糖水解酶II基因。

[0666] [84]段落83的串联构建体,其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II;(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II;(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II;和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0667] [85]段落80~84中任一段的串联构建体,其中基因的同源5'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0668] [86]段落80~85中任一段的串联构建体,其中基因的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0669] [87]段落80~86中任一段的串联构建体,其中具有生物活性的多肽是不同的多肽。

[0670] [88]段落80~86中任一段的串联构建体,其中两个或更多个具有生物活性的多肽是相同的多肽。

[0671] [89]段落80~88中任一段的串联构建体,其中启动子是不同的启动子。

[0672] [90]段落80~88中任一段的串联构建体,其中两个或更多个启动子是相同的启动子。

[0673] [91]段落80~90中任一段的串联构建体,其中终止子是不同的终止子。

[0674] [92]段落80~90中任一段的串联构建体,其中两个或更多个终止子是相同的终止子。

[0675] [93]一种表达载体,其包含段落80~92中任一段的串联构建体。

[0676] [94]一种构建用于产生具有生物活性的多种重组多肽的丝状真菌菌株的方法,其包括:(a)通过靶向整合向内源性第一基因座中插入第一串联构建体,该第一串联构建体包含(i)第一基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)第一基因座的同源3'区、其同源侧翼区或其组合;(b)通过靶向整合向内源性第二基因座中插入第二串联构建体,该第二串联构建体包含(i)第二基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个第二可选择性标记物,(iii)与第三启动子

和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸, (iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸, 和 (v) 第二基因座的同源3' 区、其同源侧翼区或其组合; 或 (c) (a) 和 (b) 的组合。

[0677] [95]段落94的方法, 其中第一基因座是纤维二糖水解酶I基因。

[0678] [96]段落95的方法, 其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I: (i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I; (ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%, 例如至少 75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少 85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少 91%、至少92%、至少 93%、至少94%、至少95%、至少96%、至少 97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I; (iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少 86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I; 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0679] [97]段落94的方法, 其中第二基因座是纤维二糖水解酶II基因。

[0680] [98]段落97的方法, 其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II: (i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II; (ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少 85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少 93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II; (iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少 86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II; 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0681] [99]段落94~98中任一段的方法, 其中各个串联构建体通过同源重组整合到丝状真菌菌株的染色体中。

[0682] [100]段落94~99中任一段的方法, 其中第一基因座的同源5' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少 1500bp或至少2000bp。

[0683] [101]段落94~100中任一段的方法, 其中第一基因座的同源3' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200 bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000 bp。

[0684] [102]段落94~101中任一段的方法, 其中第二基因座的同源5' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200 bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000 bp。

[0685] [103]段落94~102中任一段的方法,其中第二基因座的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200 bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000 bp。

[0686] [104]段落94~103中任一段的方法,其还包括各自通过靶向整合向一个或多个其他内源性基因座中插入针对各个基因座的对应串联构建体,该串联构建体包含:(i) 基因座的同源5'区、其同源侧翼区或其组合;(ii) 一个或多个可选择性标记物;(iii) 与启动子和终止子可操作连接的编码具有生物活性的多肽的多核苷酸;(iv) 与另一启动子和另一终止子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸;和(v) 基因座的同源3'区、其同源侧翼区或其组合。

[0687] [105]段落94~104中任一段的方法,其中一个或多个串联构建体还包含一个或多个可选择性标记物的5'侧翼的第一同源重复和一个或多个可选择性标记物的3'侧翼的第二同源重复,其中第一同源重复和第二同源重复进行同源重组,以切除一个或多个可选择性标记物。

[0688] [106]段落105的方法,其中第一和第二同源重复是相同的,或者彼此具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0689] [107]段落105或106的方法,其中第一和第二同源重复各自为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0690] [108]段落105~107中任一段的方法,其中在切除一个或多个可选择性标记物时,一个或多个可选择性标记物可以再用于将针对各个基因的对应串联构建体各自通过靶向整合经插入来修饰一个或多个其他内源性基因座。

[0691] [109]段落94~108中任一段的方法,其还包括用串联构建体转化丝状真菌宿主细胞,该串联构建体包含:(i) 一个或多个可选择性标记物;(ii) 与第五启动子和第五终止子可操作连接的编码具有生物活性的第五多肽的第五多核苷酸;和(iii) 与第六启动子和第六终止子可操作连接的编码具有生物活性的第六多肽的第六多核苷酸,其中该串联构建体通过异位整合而整合。

[0692] [110]段落94~109中任一段的方法,其中具有生物活性的多肽是不同的多肽。

[0693] [111]段落94~109中任一段的方法,其中两个或更多个具有生物活性的多肽是相同的多肽。

[0694] [112]段落94~111中任一段的方法,其中启动子是不同的启动子。

[0695] [113]段落94~111中任一段的方法,其中两个或更多个启动子是相同的启动子。

[0696] [114]段落94~113中任一段的方法,其中终止子是不同的终止子。

[0697] [115]段落94~113中任一段的方法,其中两个或更多个终止子是相同的终止子。

[0698] [116]段落94~115中任一段的方法,其中一个或多个串联构建体包含在表达载体中。

[0699] [117]段落94~116中任一段的方法,其中丝状真菌菌株是枝顶孢属、曲霉属、短梗霉属、黑管菌属、拟蜡霉属、金孢属、鬼伞属、革盖菌属、隐球菌属、线黑粉酵母属、镰孢菌属、腐质霉属、巨座壳属、毛霉属、毁丝菌属、新美鞭菌属、脉孢菌属、拟青霉属、青霉属、平革菌

属、白腐菌属、梨囊鞭菌属、侧耳属、裂褶菌属、篮状菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属或木霉属菌株。

[0700] [118]段落117的方法,其中木霉属菌株选自哈茨木霉、康氏木霉、长梗木霉、里氏木霉和绿色木霉。

[0701] [119]段落117的方法,其中木霉属菌株是里氏木霉。

[0702] [120]一种丝状真菌菌株,由段落94~119中任一段的方法得到。

[0703] [121]一种丝状真菌菌株,其包含:(a)通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座,该第一串联构建体包含(i)第一基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个第一可选择性标记物,(iii)与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸,(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸,和(v)第一基因座的同源3'区、其同源侧翼区或其组合;(b)通过靶向整合经插入第二串联构建体而修饰的内源性第二基因座,该第二串联构建体包含(i)第二基因座的同源5'区、其同源侧翼区或其组合,(ii)一个或多个可选择性标记物,(iii)与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸,(iv)与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸,和(v)第二基因座的同源3'区、其同源侧翼区或其组合;或(c)(a)和(b)的组合。

[0704] [122]段落121的丝状真菌菌株,其中第一基因座是纤维二糖水解酶I基因。

[0705] [123]段落122的丝状真菌菌株,其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I:(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I;(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I;(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I;和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0706] [124]段落121的丝状真菌菌株,其中第二基因座是纤维二糖水解酶II基因。

[0707] [125]段落124的丝状真菌菌株,其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II:(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II;(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II;(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、

至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II；和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0708] [126]段落121~125中任一段的丝状真菌菌株,其中各个串联构建体通过同源重组整合到丝状真菌菌株的染色体中。

[0709] [127]段落121~126中任一段的丝状真菌菌株,其中第一串联构建体的第一基因座的同源5'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000 bp、至少1500bp或至少2000bp。

[0710] [128]段落121~127中任一段的丝状真菌菌株,其中第一串联构建体的第一基因座的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000 bp、至少1500bp或至少2000bp。

[0711] [129]段落121~128中任一段的丝状真菌菌株,其中第二串联构建体的第二基因座的同源5'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000 bp、至少1500bp或至少2000bp。

[0712] [130]段落121~129中任一段的丝状真菌菌株,其中第二串联构建体的第二基因座的同源3'区、其同源侧翼区或其组合为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000 bp、至少1500bp或至少2000bp。

[0713] [131]段落121~130中任一段的丝状真菌菌株,其还包含一个或多个各自通过靶向整合用针对各个基因座的对应串联构建体经插入而修饰的其他内源基因座,该串联构建体包含:(i)基因座的同源5'区、其同源侧翼区或其组合;(ii)一个或多个可选择性标记物;(iii)与启动子和终止子可操作连接的编码具有生物活性的多肽的多核苷酸;(iv)与另一启动子和另一终止子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸;和(v)基因座的同源3'区、其同源侧翼区或其组合。

[0714] [132]段落121~131中任一段的丝状真菌菌株,其中一个或多个串联构建体还包含一个或多个可选择性标记物的5'侧翼的第一同源重复和一个或多个可选择性标记物的3'侧翼的第二同源重复,其中第一同源重复和第二同源重复进行同源重组,以切除一个或多个可选择性标记物。

[0715] [133]段落132的丝状真菌菌株,其中第一和第二同源重复是相同的,或者彼此具有至少70%,例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0716] [134]段落132或133的丝状真菌菌株,其中第一和第二同源重复各自为至少50bp,例如,至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0717] [135]段落132~134中任一段的丝状真菌菌株,其中在切除一个或多个可选择性标记物时,一个或多个可选择性标记物可以再用于各自通过靶向整合向一个或多个其他内源性基因座插入针对各个基因座的对应串联构建体。

[0718] [136]段落121~135中任一段的丝状真菌菌株,其还包含串联构建体,该串联构建

体包含：(i) 一个或多个可选择性标记物；(ii) 与第五启动子和第五终止子可操作连接的编码具有生物活性的第五多肽的第五多核苷酸；和 (iii) 与第六启动子和第六终止子可操作连接的编码具有生物活性的第六多肽的第六多核苷酸，其中该串联构建体通过异位整合而整合。

[0719] [137]段落121~136中任一段的丝状真菌菌株，其中具有生物活性的多肽是不同的多肽。

[0720] [138]段落121~136中任一段的丝状真菌菌株，其中两个或更多个具有生物活性的多肽是相同的多肽。

[0721] [139]段落121~138中任一段的丝状真菌菌株，其中启动子是不同的启动子。

[0722] [140]段落121~138中任一段的丝状真菌菌株，其中两个或更多个启动子是相同的启动子。

[0723] [141]段落121~140中任一段的丝状真菌菌株，其中终止子是不同的终止子。

[0724] [142]段落121~140中任一段的丝状真菌菌株，其中两个或更多个终止子是相同的终止子。

[0725] [143]段落121~142中任一段的丝状真菌菌株，其中一个或多个串联构建体包含在表达载体中。

[0726] [144]段落121~143中任一段的丝状真菌菌株，其中丝状真菌菌株是枝顶孢属、曲霉属、短梗霉属、黑管菌属、拟蜡霉属、金孢属、鬼伞属、革盖菌属、隐球菌属、线黑粉酵母属、镰孢菌属、腐质霉属、巨座壳属、毛霉属、毁丝菌属、新美鞭菌属、脉孢菌属、拟青霉属、青霉属、平革菌属、白腐菌属、梨囊鞭菌属、侧耳属、裂褶菌属、篮状菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属或木霉属菌株。

[0727] [145]段落144的丝状真菌菌株，其中木霉属菌株选自哈茨木霉、康氏木霉、长梗木霉、里氏木霉和绿色木霉。

[0728] [146]段落144的丝状真菌菌株，其中木霉属菌株是里氏木霉。

[0729] [147]一种在丝状真菌菌株中产生具有生物活性的多种重组多肽的方法，其包括：在有益于产生多肽的条件下培养丝状真菌宿主细胞，其中丝状真菌宿主细胞包含：(a) 通过靶向整合经插入第一串联构建体而修饰的内源性第一基因座，该第一串联构建体包含 (i) 第一基因座的同源5' 区、其同源侧翼区或其组合，(ii) 一个或多个第一可选择性标记物，(iii) 与第一启动子和第一终止子可操作连接的编码具有生物活性的第一多肽的第一多核苷酸，(iv) 与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸，和 (v) 第一基因座的同源3' 区、其同源侧翼区或其组合；(b) 通过靶向整合经插入第二串联构建体而修饰的内源性第二基因座，该第二串联构建体包含 (i) 第二基因座的同源5' 区、其同源侧翼区或其组合，(ii) 一个或多个可选择性标记物，(iii) 与第三启动子和第三终止子可操作连接的编码具有生物活性的第三多肽的第三多核苷酸，(iv) 与第四启动子和第四终止子可操作连接的编码具有生物活性的第四多肽的第四多核苷酸，和 (v) 第二基因座的同源3' 区、其同源侧翼区或其组合；或 (c) (a) 和 (b) 的组合。

[0730] [148]段落147的方法，其还包括回收多种重组多肽。

[0731] [149]段落147或148的方法，其中第一基因座是纤维二糖水解酶 I 基因。

[0732] [150]段落149的方法，其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解

酶I: (i) 包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I; (ii) 包含与SEQ ID NO:2的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I; (iii) 由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I; 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0733] [151]段落147或148的方法, 其中第二基因座是纤维二糖水解酶 II 基因。

[0734] [152]段落151的方法, 其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II: (i) 包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II; (ii) 包含与SEQ ID NO:4的成熟多肽具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II; (iii) 由包含与SEQ ID NO:3的成熟多肽编码序列具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II; 和 (iv) 由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0735] [153]段落147~152中任一段的方法, 其中第一串联构建体的第一基因座的同源5' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0736] [154]段落147~153中任一段的方法, 其中第一串联构建体的第一基因座的同源3' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0737] [155]段落147~154中任一段的方法, 其中第二串联构建体的第二基因座的同源5' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0738] [156]段落147~155中任一段的方法, 其中第二串联构建体的第二基因座的同源3' 区、其同源侧翼区或其组合为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0739] [157]段落147~156中任一段的方法, 其中丝状真菌宿主细胞还包含一个或多个各自通过靶向整合用针对各个基因座的对应串联构建体经插入而修饰的其他内源性基因座, 该串联构建体包含: (i) 基因座的同源5' 区、其同源侧翼区或其组合; (ii) 一个或多个可选择性标记物; (iii) 与启动子和终止子可操作连接的编码具有生物活性的多肽的多核苷酸。

酸; (iv) 与另一启动子和另一终止子可操作连接的编码具有生物活性的另一多肽的另一多核苷酸; 和 (v) 基因座的同源3' 区、其同源侧翼区或其组合。

[0740] [158]段落147~157中任一段的方法, 其中一个或多个串联构建体还包含一个或多个可选择性标记物的5' 侧翼的第一同源重复和一个或多个可选择性标记物的3' 侧翼的第二同源重复, 其中第一同源重复和第二同源重复进行同源重组, 以切除一个或多个可选择性标记物。

[0741] [159]段落158的方法, 其中第一和第二同源重复是相同的, 或者彼此具有至少70%, 例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%的序列同一性。

[0742] [160]段落158或159的方法, 其中第一和第二同源重复各自为至少50bp, 例如, 至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0743] [161]段落158~160中任一段的方法, 其中在切除一个或多个可选择性标记物时, 一个或多个可选择性标记物可以再用于将针对各个基因的对应串联构建体各自通过靶向整合经插入来修饰一个或多个其他内源性基因座。

[0744] [162]段落147~161中任一段的方法, 其中丝状真菌宿主细胞还包含串联构建体, 该串联构建体包含: (i) 一个或多个可选择性标记物; (ii) 与第五启动子和第五终止子可操作连接的编码具有生物活性的第五多肽的第五多核苷酸; 和 (iii) 与第六启动子和第六终止子可操作连接的编码具有生物活性的第六多肽的第六多核苷酸, 其中该串联构建体通过异位整合而整合。

[0745] [163]段落147~162中任一段的方法, 其中具有生物活性的多肽是不同的多肽。

[0746] [164]段落147~162中任一段的方法, 其中两个或更多个具有生物活性的多肽是相同的多肽。

[0747] [165]段落147~164中任一段的方法, 其中启动子是不同的启动子。

[0748] [166]段落147~164中任一段的方法, 其中两个或更多个启动子是相同的启动子。

[0749] [167]段落147~166中任一段的方法, 其中终止子是不同的终止子。

[0750] [168]段落147~166中任一段的方法, 其中两个或更多个终止子是相同的终止子。

[0751] [169]段落147~168中任一段的方法, 其中一个或多个串联构建体包含在表达载体中。

[0752] [170]段落147~169中任一段的方法, 其中丝状真菌菌株是枝顶孢属、曲霉属、短梗霉属、黑管菌属、拟蜡霉属、金孢属、鬼伞属、革盖菌属、隐球菌属、线黑粉酵母属、镰孢菌属、腐质霉属、巨座壳属、毛霉属、毁丝菌属、新美鞭菌属、脉孢菌属、拟青霉属、青霉属、平革菌属、白腐菌属、梨囊鞭菌属、侧耳属、裂褶菌属、篮状菌属、嗜热子囊菌属、梭孢壳属、弯颈霉属、栓菌属或木霉属菌株。

[0753] [171]段落170的方法, 其中木霉属菌株选自哈茨木霉、康氏木霉、长梗木霉、里氏木霉和绿色木霉。

[0754] [172]段落170的方法, 其中木霉属菌株是里氏木霉。

[0755] [173]一种串联构建体, 其包含: (i) 基因座的同源5' 区、其同源侧翼区或其组合; (ii) 一个或多个第一可选择性标记物; (iii) 与第一启动子和第一终止子可操作连接的编

码具有生物活性的第一多肽的第一多核苷酸；(iv)与第二启动子和第二终止子可操作连接的编码具有生物活性的第二多肽的第二多核苷酸；和(v)基因座的同源3'区、其同源侧翼区或其组合。

[0756] [174]段落173的串联构建体，其中基因座是纤维二糖水解酶I基因。

[0757] [175]段落174的串联构建体，其中纤维二糖水解酶I基因编码选自以下的纤维二糖水解酶I：(i)包含SEQ ID NO:2的成熟多肽的纤维二糖水解酶I；(ii)包含与SEQ ID NO:2的成熟多肽具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶I；(iii)由包含与SEQ ID NO:1的成熟多肽编码序列具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶I；和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:1的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶I。

[0758] [176]段落173的串联构建体，其中基因座是纤维二糖水解酶II 基因。

[0759] [177]段落176的串联构建体，其中纤维二糖水解酶II基因编码选自以下的纤维二糖水解酶II：(i)包含SEQ ID NO:4的成熟多肽的纤维二糖水解酶II；(ii)包含与SEQ ID NO:4的成熟多肽具有至少70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的氨基酸序列的纤维二糖水解酶II；(iii)由包含与SEQ ID NO:3的成熟多肽编码序列具有至少 70%，例如至少75%、至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列的多核苷酸所编码的纤维二糖水解酶II；和(iv)由在至少高等严格条件例如非常高等严格条件下与SEQ ID NO:3的成熟多肽编码序列或其全长互补体杂交的多核苷酸所编码的纤维二糖水解酶II。

[0760] [178]段落173~177中任一段的串联构建体，其中基因座的同源 5' 区、其同源侧翼区或其组合为至少50bp，例如，至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0761] [179]段落173~178中任一段的串联构建体，其中基因座的同源 3' 区、其同源侧翼区或其组合为至少50bp，例如，至少100bp、至少200bp、至少400bp、至少800bp、至少1000bp、至少1500bp或至少2000bp。

[0762] [180]段落173~179中任一段的串联构建体，其中具有生物活性的多肽是不同的多肽。

[0763] [181]段落173~179中任一段的串联构建体，其中两个或更多个具有生物活性的多肽是相同的多肽。

[0764] [182]段落173~181中任一段的串联构建体，其中启动子是不同的启动子。

[0765] [183]段落173~181中任一段的串联构建体,其中两个或更多个启动子是相同的启动子。

[0766] [184]段落173~183中任一段的串联构建体,其中终止子是不同的终止子。

[0767] [185]段落173~183中任一段的串联构建体,其中两个或更多个终止子是相同的终止子。

[0768] [186]一种表达载体,其包含段落173~185中任一段的串联构建体。

[0769] 本文所说明并要求保护的发明在范围上并不受本文所述的具体方面的限定,因为这些方面意在作为本发明多个方面的示例。任何等同方面均意在包含于本发明的范围内。确实,除本文示出并说明的那些以外,基于以上的描述,本发明的多种修改对本领域技术人员来说将会是显而易见的。这些修改也意在落入所附权利要求的范围之内。在冲突的情况下,包括定义在内的当前公开内容将做出控制。

序列表

<110> 诺维信股份有限公司.

D·亚韦

金绮明

<120> 在丝状真菌宿主细胞中产生多种重组多肽的方法

<130> 12280-W0-PCT

<150> US 61/526,809

<151> 2011-08-24

<160> 120

<170> PatentIn版本3.5

<210> 1

<211> 1545

<212> DNA

<213> 里氏木霉(*Trichoderma reesei*)

<400> 1

```

atgtatcgga agttggccgt catctcggcc ttcttggcca cagctcgtgc tcagtcggcc 60
tgcactctcc aatcggagac tcacccgcct ctgacatggc agaaatgctc gtctggtggc 120
acgtgcactc aacagacagg ctccgtggtc atcgacgcca actggcgctg gactcacgct 180
acgaacagca gcacgaactg ctacgatggc aacacttggg gctcgaccct atgtcctgac 240
aacgagacct gcgcaagaa ctgctgtctg gacggtgccg cctacgcgtc cacgtacgga 300
gttaccacga gcggtaacag cctctccatt ggctttgtca cccagtctgc gcagaagaac 360
gttggcgctc gcctttacct tatggcgagc gacacgacct accaggaatt caccctgctt 420
ggcaacgagt tctctttcga tgttgatggt tcgcagctgc cgtgcggctt gaacggagct 480
ctctacttcg tgtccatgga cgcggtatgg ggctgagca agtatccac caacaccgct 540
ggcgccaagt acggcacggg gtactgtgac agccagtgtc cccgcgatct gaagttcatc 600
aatggccagg ccaacgttga gggctgggag ccgtcatcca acaacgcgaa cacgggcatt 660
ggaggacacg gaagctgctg ctctgagatg gatctctggg aggccaactc catctccgag 720
gctcttacct cccacccttg cacgactgtc ggccaggaga tctgcgaggg tgatgggtgc 780
ggcggaactt actccgataa cagatatggc ggcaacttgc atcccgatgg ctgcgactgg 840
aaccataacc gcctgggcaa caccagcttc tacggccctg gctcaagctt taccctcgat 900
accaccaaga aattgaccgt tgtcaccag ttcgagacgt cgggtgccat caaccgatac 960
tatgtccaga atggcgctac tttccagcag cccaacgccg agcttggtag ttactctggc 1020
aacgagctca acgatgatta ctgcacagct gaggaggcag aattcggcgg atcctctttc 1080
tcagacaagg gcggcctgac tcagttcaag aaggtacct ctggcgcat ggttctggtc 1140
atgagtctgt gggatgatta ctacgccaac atgctgtggc tggactccac ctacccgaca 1200
aacgagacct cctccacacc cgggtgccgt cgcggaagct gctccaccag ctccggtgtc 1260
cctgctcagg tcgaatctca gtctcccaac gccaaagtca cttcttccaa catcaagttc 1320
ggaccattg gcagcaccgg caaccctagc ggcggaacc ctcccggcgg aaaccgcct 1380

```

ggcaccacca ccacccgccg cccagccact accactggaa gctctcccgg acctacccag 1440
 tctcactacg gccagtgcgg cggatttggc tacagcggcc ccacggtctg cgccagcggc 1500
 acaacttgcc aggtcctgaa cccttactac tctcagtgcc tgtaa 1545

<210> 2

<211> 514

<212> PRT

<213> 里氏木霉

<400> 2

Met	Tyr	Arg	Lys	Leu	Ala	Val	Ile	Ser	Ala	Phe	Leu	Ala	Thr	Ala	Arg
1				5					10					15	
Ala	Gln	Ser	Ala	Cys	Thr	Leu	Gln	Ser	Glu	Thr	His	Pro	Pro	Leu	Thr
				20				25						30	
Trp	Gln	Lys	Cys	Ser	Ser	Gly	Gly	Thr	Cys	Thr	Gln	Gln	Thr	Gly	Ser
				35				40						45	
Val	Val	Ile	Asp	Ala	Asn	Trp	Arg	Trp	Thr	His	Ala	Thr	Asn	Ser	Ser
				50				55						60	
Thr	Asn	Cys	Tyr	Asp	Gly	Asn	Thr	Trp	Ser	Ser	Thr	Leu	Cys	Pro	Asp
65					70					75					80
Asn	Glu	Thr	Cys	Ala	Lys	Asn	Cys	Cys	Leu	Asp	Gly	Ala	Ala	Tyr	Ala
				85					90						95
Ser	Thr	Tyr	Gly	Val	Thr	Thr	Ser	Gly	Asn	Ser	Leu	Ser	Ile	Gly	Phe
				100					105						110
Val	Thr	Gln	Ser	Ala	Gln	Lys	Asn	Val	Gly	Ala	Arg	Leu	Tyr	Leu	Met
				115					120						125
Ala	Ser	Asp	Thr	Thr	Tyr	Gln	Glu	Phe	Thr	Leu	Leu	Gly	Asn	Glu	Phe
				130				135							140
Ser	Phe	Asp	Val	Asp	Val	Ser	Gln	Leu	Pro	Cys	Gly	Leu	Asn	Gly	Ala
145					150					155					160
Leu	Tyr	Phe	Val	Ser	Met	Asp	Ala	Asp	Gly	Gly	Val	Ser	Lys	Tyr	Pro
				165					170						175
Thr	Asn	Thr	Ala	Gly	Ala	Lys	Tyr	Gly	Thr	Gly	Tyr	Cys	Asp	Ser	Gln
				180					185						190
Cys	Pro	Arg	Asp	Leu	Lys	Phe	Ile	Asn	Gly	Gln	Ala	Asn	Val	Glu	Gly
				195					200						205
Trp	Glu	Pro	Ser	Ser	Asn	Asn	Ala	Asn	Thr	Gly	Ile	Gly	Gly	His	Gly
				210					215						220
Ser	Cys	Cys	Ser	Glu	Met	Asp	Ile	Trp	Glu	Ala	Asn	Ser	Ile	Ser	Glu
225					230					235					240
Ala	Leu	Thr	Pro	His	Pro	Cys	Thr	Thr	Val	Gly	Gln	Glu	Ile	Cys	Glu

				245					250					255			
Gly	Asp	Gly	Cys	Gly	Gly	Thr	Tyr	Ser	Asp	Asn	Arg	Tyr	Gly	Gly	Thr		
				260					265					270			
Cys	Asp	Pro	Asp	Gly	Cys	Asp	Trp	Asn	Pro	Tyr	Arg	Leu	Gly	Asn	Thr		
				275					280					285			
Ser	Phe	Tyr	Gly	Pro	Gly	Ser	Ser	Phe	Thr	Leu	Asp	Thr	Thr	Lys	Lys		
				290					295					300			
Leu	Thr	Val	Val	Thr	Gln	Phe	Glu	Thr	Ser	Gly	Ala	Ile	Asn	Arg	Tyr		
305						310				315					320		
Tyr	Val	Gln	Asn	Gly	Val	Thr	Phe	Gln	Gln	Pro	Asn	Ala	Glu	Leu	Gly		
				325					330					335			
Ser	Tyr	Ser	Gly	Asn	Glu	Leu	Asn	Asp	Asp	Tyr	Cys	Thr	Ala	Glu	Glu		
				340					345					350			
Ala	Glu	Phe	Gly	Gly	Ser	Ser	Phe	Ser	Asp	Lys	Gly	Gly	Leu	Thr	Gln		
				355					360					365			
Phe	Lys	Lys	Ala	Thr	Ser	Gly	Gly	Met	Val	Leu	Val	Met	Ser	Leu	Trp		
				370					375					380			
Asp	Asp	Tyr	Tyr	Ala	Asn	Met	Leu	Trp	Leu	Asp	Ser	Thr	Tyr	Pro	Thr		
385						390				395					400		
Asn	Glu	Thr	Ser	Ser	Thr	Pro	Gly	Ala	Val	Arg	Gly	Ser	Cys	Ser	Thr		
				405					410					415			
Ser	Ser	Gly	Val	Pro	Ala	Gln	Val	Glu	Ser	Gln	Ser	Pro	Asn	Ala	Lys		
				420					425					430			
Val	Thr	Phe	Ser	Asn	Ile	Lys	Phe	Gly	Pro	Ile	Gly	Ser	Thr	Gly	Asn		
				435					440					445			
Pro	Ser	Gly	Gly	Asn	Pro	Pro	Gly	Gly	Asn	Pro	Pro	Gly	Thr	Thr	Thr		
				450					455					460			
Thr	Arg	Arg	Pro	Ala	Thr	Thr	Thr	Gly	Ser	Ser	Pro	Gly	Pro	Thr	Gln		
465						470				475					480		
Ser	His	Tyr	Gly	Gln	Cys	Gly	Gly	Ile	Gly	Tyr	Ser	Gly	Pro	Thr	Val		
				485					490					495			
Cys	Ala	Ser	Gly	Thr	Thr	Cys	Gln	Val	Leu	Asn	Pro	Tyr	Tyr	Ser	Gln		
				500					505					510			

Cys Leu

<210> 3

<211> 1611

<212> DNA

<213> 里氏木霉

<400> 3

```

atgattgtcg gcattctcac cacgctggct acgctggcca cactcgcagc tagtgtgcct 60
ctagaggagc ggcaagcttg ctcaagcgtc tggtaattat gtgaaccctc tcaagagacc 120
caaatactga gatatgtcaa ggggccaatg tggtagccag aattggtcgg gtccgacttg 180
ctgtgcttcc ggaagcacat gcgtctactc caacgactat tactcccagt gtcttcccgg 240
cgctgcaagc tcaagctcgt ccacgcgcgc cgctcgacg acttctcgag tatccccac 300
aacatcccgg tcgagctccg cgacgcctcc acctggttct actactacca gactacctcc 360
agtcggatcg ggaaccgcta cgtattcagg caaccctttt gttggggtca ctcttgggc 420
caatgcatat tacgcctctg aagtttagcag cctcgtctatt cctagcttga ctggagccat 480
ggccactgct gcagcagctg tcgcaaaggt tccctctttt atgtggctgt aggtcctccc 540
ggaaccaagg caatctgtta ctgaaggctc atcattcact gcagagatac tcttgacaag 600
acccctctca tggagcaaac cttggccgac atccgcaccg ccaacaagaa tggcggtaac 660
tatgccggac agtttgtggt gtatgacttg ccgtagcgcg attgcgctgc cttgcctcg 720
aatggcgaat actctattgc cgatggtggc gtcgccaat ataagaacta tategacacc 780
attcgtcaaa ttgtcgtgga atattccgat atccggacc ccttggttat tggtagagt 840
ttaaacacct gcctcccccc ccccttcctt tcttttcccg ccggcatctt gtcgttgtgc 900
taactattgt tccctcttcc agagcctgac tctcttgcca acctggtgac caacctcgtt 960
actccaaagt gtgccaatgc tcagtcagcc taccttgagt gcatcaacta cgccgtcaca 1020
cagctgaacc ttccaaatgt tgcgatgtat ttggacgctg gccatgcagg atggcttggc 1080
tggccggcaa accaagaccc ggccgctcag ctatttgcaa atgtttacaa gaatgcatcg 1140
tctccgagag ctcttcgcgg attggcaacc aatgtcgcca actacaacgg gtggaacatt 1200
accagcccc catcgtacac gcaaggcaac gctgtctaca acgagaagct gtacatccac 1260
gctattggac gtcttcttgc caatcacggc tggccaacg ctttcttcat cactgatcaa 1320
ggtcgatcgg gaaagcagcc tacgggacag caacagtggg gagactggtg caatgtgatc 1380
ggcaccggat ttggtattcg cccatccgca aacactgggg actcgttgct ggattcgttt 1440
gtctgggtca agccaggcgg cgagtgtgac ggcaccagcg acagcagtgc gccacgattt 1500
gactcccact gtgcgtcccc agatgccttg caaccggcgc ctcaagctgg tgcttggttc 1560
caagcctact ttgtgcagct tctcacaac gcaaaccat cgttcctgta a 1611

```

<210> 4

<211> 471

<212> PRT

<213> 里氏木霉

<400> 4

Met Ile Val Gly Ile Leu Thr Thr Leu Ala Thr Leu Ala Thr Leu Ala

1 5 10 15

Ala Ser Val Pro Leu Glu Glu Arg Gln Ala Cys Ser Ser Val Trp Gly

20 25 30

Gln Cys Gly Gly Gln Asn Trp Ser Gly Pro Thr Cys Cys Ala Ser Gly

35 40 45

Ser Thr Cys Val Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu Pro Gly

50	55	60
Ala Ala Ser Ser Ser Ser Ser Thr Arg Ala Ala Ser Thr Thr Ser Arg		
65	70	75
Val Ser Pro Thr Thr Ser Arg Ser Ser Ser Ala Thr Pro Pro Pro Gly		80
	85	90
Ser Thr Thr Thr Arg Val Pro Pro Val Gly Ser Gly Thr Ala Thr Tyr		95
	100	105
Ser Gly Asn Pro Phe Val Gly Val Thr Pro Trp Ala Asn Ala Tyr Tyr		110
	115	120
Ala Ser Glu Val Ser Ser Leu Ala Ile Pro Ser Leu Thr Gly Ala Met		125
	130	135
Ala Thr Ala Ala Ala Ala Val Ala Lys Val Pro Ser Phe Met Trp Leu		140
145	150	155
Asp Thr Leu Asp Lys Thr Pro Leu Met Glu Gln Thr Leu Ala Asp Ile		160
	165	170
Arg Thr Ala Asn Lys Asn Gly Gly Asn Tyr Ala Gly Gln Phe Val Val		175
	180	185
Tyr Asp Leu Pro Asp Arg Asp Cys Ala Ala Leu Ala Ser Asn Gly Glu		190
	195	200
Tyr Ser Ile Ala Asp Gly Gly Val Ala Lys Tyr Lys Asn Tyr Ile Asp		205
	210	215
Thr Ile Arg Gln Ile Val Val Glu Tyr Ser Asp Ile Arg Thr Leu Leu		220
225	230	235
Val Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu Gly Thr		240
	245	250
Pro Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Ile Asn Tyr		255
	260	265
Ala Val Thr Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala		270
	275	280
Gly His Ala Gly Trp Leu Gly Trp Pro Ala Asn Gln Asp Pro Ala Ala		285
	290	295
Gln Leu Phe Ala Asn Val Tyr Lys Asn Ala Ser Ser Pro Arg Ala Leu		300
305	310	315
Arg Gly Leu Ala Thr Asn Val Ala Asn Tyr Asn Gly Trp Asn Ile Thr		320
	325	330
Ser Pro Pro Ser Tyr Thr Gln Gly Asn Ala Val Tyr Asn Glu Lys Leu		335
	340	345
Tyr Ile His Ala Ile Gly Arg Leu Leu Ala Asn His Gly Trp Ser Asn		350
	355	360
		365

Ala Phe Phe Ile Thr Asp Gln Gly Arg Ser Gly Lys Gln Pro Thr Gly
 370 375 380
 Gln Gln Gln Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly
 385 390 395 400
 Ile Arg Pro Ser Ala Asn Thr Gly Asp Ser Leu Leu Asp Ser Phe Val
 405 410 415
 Trp Val Lys Pro Gly Gly Glu Cys Asp Gly Thr Ser Asp Ser Ser Ala
 420 425 430
 Pro Arg Phe Asp Ser His Cys Ala Leu Pro Asp Ala Leu Gln Pro Ala
 435 440 445
 Pro Gln Ala Gly Ala Trp Phe Gln Ala Tyr Phe Val Gln Leu Leu Thr
 450 455 460
 Asn Ala Asn Pro Ser Phe Leu
 465 470

<210> 5

<211> 1377

<212> DNA

<213> 里氏木霉

<400> 5

atggcgccct cagttacact gccgttgacc acggccatcc tggccattgc ccggtcgtc 60
 gccgcccagc aaccgggtac cagcaccccc gaggtccatc ccaagttgac aacctacaag 120
 tgtacaaagt ccgggggggtg cgtggcccag gacacctcgg tggtccttga ctggaactac 180
 cgctggatgc acgacgcaaa ctacaactcg tgcaccgtca acggcggcgt caacaccacg 240
 ctctgccctg acgaggcgac ctgtggcaag aactgcttca tcgagggcgt cgactacgcc 300
 gcctcgggcg tcacgacctc gggcagcagc ctcaccatga accagtacat gccagcagc 360
 tctggcggct acagcagcgt ctctcctcgg ctgtatctcc tggactctga cggtgagtac 420
 gtgatgctga agctcaacgg ccaggagctg agcttcgacg tcgacctctc tgctctgccg 480
 tgtggagaga acggctcgtc ctacctgtct cagatggacg agaacggggg cgccaaccag 540
 tataacacgg ccggtgccaa ctacgggagc ggctactgcg atgctcagtg ccccgctccag 600
 acatggagga acggcaccct caacactagc caccagggtc tctgctgcaa cgagatggat 660
 atcctggagg gcaactcgag ggcgaatgcc ttgaccctc actcttgac ggccacggcc 720
 tgcgactctg ccggttgccg cttcaacccc tatggcagcg gctacaaaag ctactacggc 780
 cccggagata ccgttgacac ctccaagacc ttcaccatca tcaccagtt caacacggac 840
 aacggctcgc cctcgggcaa ccttgtgagc atcaccgcga agtaccagca aaacggcgtc 900
 gacatcccca gcgcccagcc cggcgggcagc accatctcgt cctgcccgtc cgctcagcc 960
 tacggcggcc tcgccaccat gggcaaggcc ctgagcagcg gcatggtgct cgtgttcagc 1020
 atttggaaacg acaacagcca gtacatgaac tggctcgaca gcggcaacgc cgccccctgc 1080
 agcagcaccg agggcaaccc atccaacatc ctggccaaca accccaacac gcacgtcgtc 1140
 ttctccaaca tccgctgggg agacattggg tctactacga actcgactgc gccccgccc 1200

ccgcctgcgt ccagcacgac gttttcgact acacggagga gctcgacgac ttcgagcagc 1260
 ccgagctgca cgcagactca ctggggggcag tgcggtggca ttgggtacag cgggtgcaag 1320
 acgtgcacgt cgggcactac gtgccagtat agcaacgact actactcgca atgcctt 1377

<210> 6

<211> 459

<212> PRT

<213> 里氏木霉

<400> 6

Met	Ala	Pro	Ser	Val	Thr	Leu	Pro	Leu	Thr	Thr	Ala	Ile	Leu	Ala	Ile
1				5					10					15	
Ala	Arg	Leu	Val	Ala	Ala	Gln	Gln	Pro	Gly	Thr	Ser	Thr	Pro	Glu	Val
			20					25					30		
His	Pro	Lys	Leu	Thr	Thr	Tyr	Lys	Cys	Thr	Lys	Ser	Gly	Gly	Cys	Val
		35					40					45			
Ala	Gln	Asp	Thr	Ser	Val	Val	Leu	Asp	Trp	Asn	Tyr	Arg	Trp	Met	His
	50					55				60					
Asp	Ala	Asn	Tyr	Asn	Ser	Cys	Thr	Val	Asn	Gly	Gly	Val	Asn	Thr	Thr
65				70					75					80	
Leu	Cys	Pro	Asp	Glu	Ala	Thr	Cys	Gly	Lys	Asn	Cys	Phe	Ile	Glu	Gly
			85					90					95		
Val	Asp	Tyr	Ala	Ala	Ser	Gly	Val	Thr	Thr	Ser	Gly	Ser	Ser	Leu	Thr
			100					105					110		
Met	Asn	Gln	Tyr	Met	Pro	Ser	Ser	Ser	Gly	Gly	Tyr	Ser	Ser	Val	Ser
		115				120						125			
Pro	Arg	Leu	Tyr	Leu	Leu	Asp	Ser	Asp	Gly	Glu	Tyr	Val	Met	Leu	Lys
	130					135					140				
Leu	Asn	Gly	Gln	Glu	Leu	Ser	Phe	Asp	Val	Asp	Leu	Ser	Ala	Leu	Pro
145				150					155					160	
Cys	Gly	Glu	Asn	Gly	Ser	Leu	Tyr	Leu	Ser	Gln	Met	Asp	Glu	Asn	Gly
			165					170					175		
Gly	Ala	Asn	Gln	Tyr	Asn	Thr	Ala	Gly	Ala	Asn	Tyr	Gly	Ser	Gly	Tyr
		180						185				190			
Cys	Asp	Ala	Gln	Cys	Pro	Val	Gln	Thr	Trp	Arg	Asn	Gly	Thr	Leu	Asn
	195					200						205			
Thr	Ser	His	Gln	Gly	Phe	Cys	Cys	Asn	Glu	Met	Asp	Ile	Leu	Glu	Gly
	210				215					220					
Asn	Ser	Arg	Ala	Asn	Ala	Leu	Thr	Pro	His	Ser	Cys	Thr	Ala	Thr	Ala
225				230					235				240		
Cys	Asp	Ser	Ala	Gly	Cys	Gly	Phe	Asn	Pro	Tyr	Gly	Ser	Gly	Tyr	Lys

	245		250		255
Ser Tyr Tyr Gly Pro Gly Asp Thr Val Asp Thr Ser Lys Thr Phe Thr					
	260		265		270
Ile Ile Thr Gln Phe Asn Thr Asp Asn Gly Ser Pro Ser Gly Asn Leu					
	275		280		285
Val Ser Ile Thr Arg Lys Tyr Gln Gln Asn Gly Val Asp Ile Pro Ser					
	290		295		300
Ala Gln Pro Gly Gly Asp Thr Ile Ser Ser Cys Pro Ser Ala Ser Ala					
305		310		315	320
Tyr Gly Gly Leu Ala Thr Met Gly Lys Ala Leu Ser Ser Gly Met Val					
	325		330		335
Leu Val Phe Ser Ile Trp Asn Asp Asn Ser Gln Tyr Met Asn Trp Leu					
	340		345		350
Asp Ser Gly Asn Ala Gly Pro Cys Ser Ser Thr Glu Gly Asn Pro Ser					
	355		360		365
Asn Ile Leu Ala Asn Asn Pro Asn Thr His Val Val Phe Ser Asn Ile					
	370		375		380
Arg Trp Gly Asp Ile Gly Ser Thr Thr Asn Ser Thr Ala Pro Pro Pro					
385		390		395	400
Pro Pro Ala Ser Ser Thr Thr Phe Ser Thr Thr Arg Arg Ser Ser Thr					
	405		410		415
Thr Ser Ser Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly					
	420		425		430
Gly Ile Gly Tyr Ser Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys					
	435		440		445
Gln Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu					
	450		455		

<210> 7

<211> 1254

<212> DNA

<213> 里氏木霉

<400> 7

```

atgaacaagt ccggtggetcc attgetgett gcagcgtcca tactatatgg cggcgccgtc 60
gcacagcaga ctgtctgggg ccagtgtgga ggtattggtt ggagcggacc tacgaattgt 120
gtctctggct cagcttggtc gacctcaat cttattatg cgcaatgtat tccgggagcc 180
actactatca ccacttcgac ccggccacca tccggtccaa ccaccaccac cagggtacc 240
tcaacaagct catcaactcc acccagcagc tctgggggtcc gatttgccgg cgttaacatc 300
gcgggttttg actttggctg taccacagat ggcacttgcg ttacctcgaa ggttttatcct 360
ccgttgaaga acttcaccgg ctcaaacaac taccctgatg gcatcggcca gatgcagcac 420

```

ttcgtcaacg aggacgggat gactattttc cgcttacctg tcggatggca gtacctcgctc 480
 aacaacaatt tgggcggcaa tcttgattcc acgagcattt ccaagtatga tcagcttggtt 540
 caggggtgcc tgtctctggg cgcatactgc atcgtcgaca tccacaatta tgctcgatgg 600
 aacggtggga tcattgggtca gggcggccct actaatgctc aattcacgag cctttggctc 660
 cagttggcat caaagtacgc atctcagtcg aggggtgtgg tccggcatcat gaatgagccc 720
 cacgacgtga acatcaacac ctgggctgcc acggccaag aggttgtaac cgcaatccgc 780
 aacgctgggtg ctacgtcgca attcatctct ttgcctggaa atgattggca atctgctggg 840
 gctttcatat ccgatggcag tgcagccgcc ctgtctcaag tcacgaaccc ggatgggtca 900
 acaacgaatc tgatttttga cgtgcacaaa tacttggact cagacaactc cggtactcac 960
 gccgaatgta ctacaaataa cattgacggc gccttttctc cgcttgccac ttggctccga 1020
 cagaacaatc gccaggtat cctgacagaa accggtgggtg gcaacgttca gtctgcata 1080
 caagacatgt gccagcaaat ccaatatctc aaccagaact cagatgtcta tcttggctat 1140
 gttggttggg gtgccggatc atttgatagc acgtatgtcc tgacggaaac accgactagc 1200
 agtggttaact catggacgga cacatccttg gtcagctcgt gtctcgcaag aaag 1254

<210> 8

<211> 418

<212> PRT

<213> 里氏木霉

<400> 8

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

Asn Asn Asn Leu Gly Gly Asn Leu Asp Ser Thr Ser Ile Ser Lys Tyr		
165	170	175
Asp Gln Leu Val Gln Gly Cys Leu Ser Leu Gly Ala Tyr Cys Ile Val		
180	185	190
Asp Ile His Asn Tyr Ala Arg Trp Asn Gly Gly Ile Ile Gly Gln Gly		
195	200	205
Gly Pro Thr Asn Ala Gln Phe Thr Ser Leu Trp Ser Gln Leu Ala Ser		
210	215	220
Lys Tyr Ala Ser Gln Ser Arg Val Trp Phe Gly Ile Met Asn Glu Pro		
225	230	235
His Asp Val Asn Ile Asn Thr Trp Ala Ala Thr Val Gln Glu Val Val		
245	250	255
Thr Ala Ile Arg Asn Ala Gly Ala Thr Ser Gln Phe Ile Ser Leu Pro		
260	265	270
Gly Asn Asp Trp Gln Ser Ala Gly Ala Phe Ile Ser Asp Gly Ser Ala		
275	280	285
Ala Ala Leu Ser Gln Val Thr Asn Pro Asp Gly Ser Thr Thr Asn Leu		
290	295	300
Ile Phe Asp Val His Lys Tyr Leu Asp Ser Asp Asn Ser Gly Thr His		
305	310	315
Ala Glu Cys Thr Thr Asn Asn Ile Asp Gly Ala Phe Ser Pro Leu Ala		
325	330	335
Thr Trp Leu Arg Gln Asn Asn Arg Gln Ala Ile Leu Thr Glu Thr Gly		
340	345	350
Gly Gly Asn Val Gln Ser Cys Ile Gln Asp Met Cys Gln Gln Ile Gln		
355	360	365
Tyr Leu Asn Gln Asn Ser Asp Val Tyr Leu Gly Tyr Val Gly Trp Gly		
370	375	380
Ala Gly Ser Phe Asp Ser Thr Tyr Val Leu Thr Glu Thr Pro Thr Ser		
385	390	395
Ser Gly Asn Ser Trp Thr Asp Thr Ser Leu Val Ser Ser Cys Leu Ala		
405	410	415

Arg Lys

<210> 9

<211> 2615

<212> DNA

<213> 里氏木霉

<400> 9

atgggatttg gccgcaatgc tgccgagccc gagtgtttct gcaacgttat ccaggagatt 60

```

tgcgcttgcc caagagggag ttgacgggga gagtcccaac tggttccttc agtaacgcca 120
ccctggcaga ctatataact tgtggacaag actctgcttt gttgagttct tcctaccagt 180
cttgaccaag accattctgt tgagcccaat cagaaatgcg ttaccgaaca gcagctgcgc 240
tggcacttgc cactgggccc ttgctaggg cagacagtca gtatagctgg tccatactgg 300
gatgtatatg tatcctggag acaccatgct gactcttgaa tcaaggtagc tcaacatcgg 360
gggcctcggc tgaggcagtt gtacctcctg cagggactcc atggggaacc gcgtacgaca 420
aggcgaaggc cgcattggca aagctcaatc tccaagataa ggtcggcatc gtgagcgggtg 480
tcggctggaa cggcggtcct tgcgttggaa acacatctcc ggcctccaag atcagctatc 540
catcgctatg ccttcaagac ggacccctcg gtgttcgata ctcgacaggc agcacagcct 600
ttacgccggg cgttcaagcg gcctcgacgt gggatgtcaa tttgatccgc gaacgtggac 660
agttcatcgg tgaggaggtg aaggcctcgg ggattcatgt catacttggg cctgtggctg 720
ggccgctggg aaagactccg cagggcggtc gcaactggga gggcttcggg gtcgatccat 780
atctcacggg cattgccatg ggtcaaacca tcaacggcat ccagtcggta ggcgtgcagg 840
cgacagcgaa gcactatatc ctcaacgagc aggagctcaa tcgagaaacc atttcgagca 900
accagatga ccgaactctc catgagctgt atacttggcc atttgccgac gcggttcagg 960
ccaatgtcgc ttctgtcatg tgctcgtaca acaaggtcaa taccacctgg gcctgcgagg 1020
atcagtacac gctgcagact gtgctgaaag accagctggg gttcccaggc tatgtcatga 1080
cggactggaa cgcacagcac acgactgtcc aaagcgcgaa ttctgggctt gacatgtcaa 1140
tgcttgccac agacttcaac ggtaacaatc ggctctgggg tccagctctc accaatgcgg 1200
taaatagcaa tcagggtccc acgagcagag tcgacgatat ggtgactcgt atcctcgccg 1260
catggtactt gacaggccag gaccaggcag gctatccgtc gttcaacatc agcagaaatg 1320
ttcaaggaaa ccacaagacc aatgtcaggg caattgccag ggacggcatc gttctgtctc 1380
agaatgacgc caacatcctg ccgctcaaga agcccgctag cattgccgtc gttggatctg 1440
ccgcaatcat tggtaaccac gccagaaact cgccctcgtg caacgacaaa ggctgcgacg 1500
acggggcctt gggcatgggt tggggttccg gcgccgtcaa ctatccgtac ttcgtcgcgc 1560
cctacgatgc catcaatacc agagcgtctt cgcagggcac ccaggttacc ttgagcaaca 1620
ccgacaacac gtcctcaggc gcactcgcag caagaggaaa ggacgtcgcc atcgtcttca 1680
tcaccgccga ctcgggtgaa ggctacatca ccgtggaggg caacgcgggc gatcgcaaca 1740
acctggatcc gtggcacaac ggcaatgccc tgggtccaggc ggtggccggg gccaacagca 1800
acgtcattgt tgttgtccac tccgttggcg ccatcattct ggagcagatt cttgctcttc 1860
cgcagggtcaa ggccgttgte tgggcgggtc ttctttctca ggagagcggc aatgcgctcg 1920
tcgacgtgct gtggggagat gtcagccctt ctggcaagct ggtgtacacc attgcgaaga 1980
gccccaatga ctataaact cgcactgctt ccggcggcag tgacagcttc agcgagggac 2040
tgttcatcga ctataagcac ttcgacgacg ccaatatcac gccgcggtac gaggtcggct 2100
atggactgtg taagtttgct aacctgaaca atctattaga caggttgact gacgatgac 2160
tgtggaatga tagcttacac caagttcaac tactcacgcc tctccgtctt gtcgaccgcc 2220
aagtctggtc ctgcgactgg ggccgttggt ccgggaggcc cgagtgatct gttccagaat 2280
gtcgcgacag tcaccgttga catcgcaaac tctggccaag tgactggtgc cgaggtagcc 2340
cagctgtaca tcacctaccc atcttcagca cccaggaccc ctccgaagca gctgcgaggc 2400

```

tttgccaagc tgaacctcac gcctggtcag agcggaaacag caacgttcaa catccgacga 2460
cgagatctca gctactggga cacggcttcg cagaaatggg tgggtgccgtc ggggtcgttt 2520
ggcatcagcg tgggagcgag cagccgggat atcaggctga cgagcactct gtcggtagcg 2580
tagcgcgagg aggggtgaagg cggttgacct gtgac 2615

<210> 10

<211> 744

<212> PRT

<213> 里氏木霉

<400> 10

Met	Arg	Tyr	Arg	Thr	Ala	Ala	Ala	Leu	Ala	Leu	Ala	Thr	Gly	Pro	Phe
1				5				10					15		
Ala	Arg	Ala	Asp	Ser	His	Ser	Thr	Ser	Gly	Ala	Ser	Ala	Glu	Ala	Val
			20					25					30		
Val	Pro	Pro	Ala	Gly	Thr	Pro	Trp	Gly	Thr	Ala	Tyr	Asp	Lys	Ala	Lys
			35					40					45		
Ala	Ala	Leu	Ala	Lys	Leu	Asn	Leu	Gln	Asp	Lys	Val	Gly	Ile	Val	Ser
			50					55					60		
Gly	Val	Gly	Trp	Asn	Gly	Gly	Pro	Cys	Val	Gly	Asn	Thr	Ser	Pro	Ala
65				70				75					80		
Ser	Lys	Ile	Ser	Tyr	Pro	Ser	Leu	Cys	Leu	Gln	Asp	Gly	Pro	Leu	Gly
			85					90					95		
Val	Arg	Tyr	Ser	Thr	Gly	Ser	Thr	Ala	Phe	Thr	Pro	Gly	Val	Gln	Ala
			100					105					110		
Ala	Ser	Thr	Trp	Asp	Val	Asn	Leu	Ile	Arg	Glu	Arg	Gly	Gln	Phe	Ile
			115					120					125		
Gly	Glu	Glu	Val	Lys	Ala	Ser	Gly	Ile	His	Val	Ile	Leu	Gly	Pro	Val
			130					135					140		
Ala	Gly	Pro	Leu	Gly	Lys	Thr	Pro	Gln	Gly	Gly	Arg	Asn	Trp	Glu	Gly
145				150				155					160		
Phe	Gly	Val	Asp	Pro	Tyr	Leu	Thr	Gly	Ile	Ala	Met	Gly	Gln	Thr	Ile
			165					170					175		
Asn	Gly	Ile	Gln	Ser	Val	Gly	Val	Gln	Ala	Thr	Ala	Lys	His	Tyr	Ile
			180					185					190		
Leu	Asn	Glu	Gln	Glu	Leu	Asn	Arg	Glu	Thr	Ile	Ser	Ser	Asn	Pro	Asp
			195					200					205		
Asp	Arg	Thr	Leu	His	Glu	Leu	Tyr	Thr	Trp	Pro	Phe	Ala	Asp	Ala	Val
			210					215					220		
Gln	Ala	Asn	Val	Ala	Ser	Val	Met	Cys	Ser	Tyr	Asn	Lys	Val	Asn	Thr
225				230				235					240		

Thr Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys Asp		
245	250	255
Gln Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln His		
260	265	270
Thr Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro Gly		
275	280	285
Thr Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr Asn		
290	295	300
Ala Val Asn Ser Asn Gln Val Pro Thr Ser Arg Val Asp Asp Met Val		
305	310	315
Thr Arg Ile Leu Ala Ala Trp Tyr Leu Thr Gly Gln Asp Gln Ala Gly		
325	330	335
Tyr Pro Ser Phe Asn Ile Ser Arg Asn Val Gln Gly Asn His Lys Thr		
340	345	350
Asn Val Arg Ala Ile Ala Arg Asp Gly Ile Val Leu Leu Lys Asn Asp		
355	360	365
Ala Asn Ile Leu Pro Leu Lys Lys Pro Ala Ser Ile Ala Val Val Gly		
370	375	380
Ser Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys Asn		
385	390	395
Asp Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser Gly		
405	410	415
Ala Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn Thr		
420	425	430
Arg Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp Asn		
435	440	445
Thr Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile Val		
450	455	460
Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly Asn		
465	470	475
Ala Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala Leu		
485	490	495
Val Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val His		
500	505	510
Ser Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln Val		
515	520	525
Lys Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn Ala		
530	535	540
Leu Val Asp Val Leu Trp Gly Asp Val Ser Pro Ser Gly Lys Leu Val		

545	550	555	560
Tyr Thr Ile Ala Lys Ser Pro Asn Asp Tyr Asn Thr Arg Ile Val Ser			
	565	570	575
Gly Gly Ser Asp Ser Phe Ser Glu Gly Leu Phe Ile Asp Tyr Lys His			
	580	585	590
Phe Asp Asp Ala Asn Ile Thr Pro Arg Tyr Glu Phe Gly Tyr Gly Leu			
	595	600	605
Ser Tyr Thr Lys Phe Asn Tyr Ser Arg Leu Ser Val Leu Ser Thr Ala			
	610	615	620
Lys Ser Gly Pro Ala Thr Gly Ala Val Val Pro Gly Gly Pro Ser Asp			
625	630	635	640
Leu Phe Gln Asn Val Ala Thr Val Thr Val Asp Ile Ala Asn Ser Gly			
	645	650	655
Gln Val Thr Gly Ala Glu Val Ala Gln Leu Tyr Ile Thr Tyr Pro Ser			
	660	665	670
Ser Ala Pro Arg Thr Pro Pro Lys Gln Leu Arg Gly Phe Ala Lys Leu			
	675	680	685
Asn Leu Thr Pro Gly Gln Ser Gly Thr Ala Thr Phe Asn Ile Arg Arg			
	690	695	700
Arg Asp Leu Ser Tyr Trp Asp Thr Ala Ser Gln Lys Trp Val Val Pro			
705	710	715	720
Ser Gly Ser Phe Gly Ile Ser Val Gly Ala Ser Ser Arg Asp Ile Arg			
	725	730	735
Leu Thr Ser Thr Leu Ser Val Ala			
	740		

<210> 11

<211> 752

<212> DNA

<213> 里氏木霉

<400> 11

```

atggttgccct tttccagcct catctgcgct ctcaccagca tcgccagtac tctggcgatg 60
cccacaggcc tcgagcctga gagcagtgtc aacgtcacag agcgtggcat gtacgacttt 120
gttcttggag ctcaaatga tcatcgccgt cgtgctagca tcaactacga ccaaaactac 180
caaactggcg gacaagtcag ctattcgctt tccaacactg gcttctcagt gaactggaac 240
actcaagatg actttgttgt gggcgtttgt tggacgactg gatcttctgc gtaggaggac 300
tcctcatcat tctgcacttt gaaagcatct tctgaccaa agcttctctt agtcccatca 360
actttggcgg ctcttttagt gtcaacagcg gaactggcct gctttccgtc tatggctgga 420
gcaccaaccc actggttgag tactacatca tggaggacaa ccacaactac ccagcacagg 480
gtaccgtcaa gggaaccgtc accagcgacg gagccactta caccatctgg gagaataccc 540

```


gtgtcaacga gccttccatc cagggcacag cgaccttcaa ccagtacatt tccgtgcgga 600
 actcgcccgag gaccagcggg actgttactg tgcagaacca cttcaatgct tgggcctcgc 660
 ttggcctgca ccttgggcag atgaactacc aggttgtcgc tgtcgaaggc tggggtggta 720
 gtggttctgc ctcacagagt gtcagcaact ag 752

<210> 12

<211> 229

<212> PRT

<213> 里氏木霉

<400> 12

Met	Val	Ala	Phe	Ser	Ser	Leu	Ile	Cys	Ala	Leu	Thr	Ser	Ile	Ala	Ser
1				5					10					15	
Thr	Leu	Ala	Met	Pro	Thr	Gly	Leu	Glu	Pro	Glu	Ser	Ser	Val	Asn	Val
			20					25					30		
Thr	Glu	Arg	Gly	Met	Tyr	Asp	Phe	Val	Leu	Gly	Ala	His	Asn	Asp	His
		35					40					45			
Arg	Arg	Arg	Ala	Ser	Ile	Asn	Tyr	Asp	Gln	Asn	Tyr	Gln	Thr	Gly	Gly
		50				55					60				
Gln	Val	Ser	Tyr	Ser	Pro	Ser	Asn	Thr	Gly	Phe	Ser	Val	Asn	Trp	Asn
65				70					75					80	
Thr	Gln	Asp	Asp	Phe	Val	Val	Gly	Val	Gly	Trp	Thr	Thr	Gly	Ser	Ser
			85						90				95		
Ala	Pro	Ile	Asn	Phe	Gly	Gly	Ser	Phe	Ser	Val	Asn	Ser	Gly	Thr	Gly
		100						105					110		
Leu	Leu	Ser	Val	Tyr	Gly	Trp	Ser	Thr	Asn	Pro	Leu	Val	Glu	Tyr	Tyr
		115						120					125		
Ile	Met	Glu	Asp	Asn	His	Asn	Tyr	Pro	Ala	Gln	Gly	Thr	Val	Lys	Gly
		130				135					140				
Thr	Val	Thr	Ser	Asp	Gly	Ala	Thr	Tyr	Thr	Ile	Trp	Glu	Asn	Thr	Arg
145				150						155				160	
Val	Asn	Glu	Pro	Ser	Ile	Gln	Gly	Thr	Ala	Thr	Phe	Asn	Gln	Tyr	Ile
			165						170				175		
Ser	Val	Arg	Asn	Ser	Pro	Arg	Thr	Ser	Gly	Thr	Val	Thr	Val	Gln	Asn
		180						185					190		
His	Phe	Asn	Ala	Trp	Ala	Ser	Leu	Gly	Leu	His	Leu	Gly	Gln	Met	Asn
		195				200					205				
Tyr	Gln	Val	Val	Ala	Val	Glu	Gly	Trp	Gly	Gly	Ser	Gly	Ser	Ala	Ser
		210				215					220				
Gln	Ser	Val	Ser	Asn											
225															

<210> 13

<211> 796

<212> DNA

<213> 里氏木霉

<400> 13

caagaagaca tcaacatggt ctccttcacc tccctcctcg ccggcgctgc cgccatctcg 60
 ggcgtcttgg ccgctcccg cgcggaggtc gaatccgtgg ctgtggagaa gcgccagacg 120
 attcagcccc gcacgggcta caacaacggc tactttctact cgtactggaa cgatggccac 180
 ggcggcggtga cgtacaccaa tgggtcccg gcggcagttct ccgtcaactg gtccaactcg 240
 ggcaactttg tcggcgga gggatggcag cccggcacca agaacaagta agactaccta 300
 ctcttaccctt ctttgaccaa cacagcaca cacaatacaa cacatgtgac taccaatcat 360
 ggaatcggat ctaacagctg tgttttcaaa aaaaagggtc atcaacttct cgggcagcta 420
 caaccccaac ggcaacagct acctctccgt gtacggctgg tcccgaacc ccctgatcga 480
 gtactacatc gtcgagaact ttggcaccta caaccgctc acgggcgcca ccaagctggg 540
 cgaggtcacc tccgacggca gcgtctacga catttaccgc acgcagcgcg tcaaccagcc 600
 gtccatcatc ggcaccgcca ccttttacca gtactggtcc gtccgcccga accaccgctc 660
 gagcggtcc gtcaacacgg cgaaccactt caacgcgtgg gctcagcaag gcctgacgct 720
 cgggacgatg gattaccaga ttgttgccgt ggagggttac tttagctctg gctctgcttc 780
 catcaccgctc agctaa 796

<210> 14

<211> 223

<212> PRT

<213> 里氏木霉

<400> 14

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

115	120	125
Tyr Asn Pro Ser Thr Gly Ala Thr Lys Leu Gly Glu Val Thr Ser Asp		
130	135	140
Gly Ser Val Tyr Asp Ile Tyr Arg Thr Gln Arg Val Asn Gln Pro Ser		
145	150	155
Ile Ile Gly Thr Ala Thr Phe Tyr Gln Tyr Trp Ser Val Arg Arg Asn		
165	170	175
His Arg Ser Ser Gly Ser Val Asn Thr Ala Asn His Phe Asn Ala Trp		
180	185	190
Ala Gln Gln Gly Leu Thr Leu Gly Thr Met Asp Tyr Gln Ile Val Ala		
195	200	205
Val Glu Gly Tyr Phe Ser Ser Gly Ser Ala Ser Ile Thr Val Ser		
210	215	220

<210> 15

<211> 1352

<212> DNA

<213> 里氏木霉

<400> 15

atgaaagcaa acgtcatctt gtgcctcctg gccccctgg tcgccgtctt cccacccgaa 60
accatccacc tcgaccccga gctcgccgct ctccgcgcca acctcaccga gcgaacagcc 120
gacctctggg accgccaagc ctctcaaagc atcgaccagc tcatcaagag aaaaggcaag 180
ctctactttg gcaccgccac cgaccgcggc ctctccaac gggaaaagaa cgcggccatc 240
atccaggcag acctcggcc ggtgacgccg gagaacagca tgaagtggca gtcgctcgag 300
aacaaccaag gccagctgaa ctggggagac gccgactatc tcgtcaactt tgcccagcaa 360
aacggcaagt cgatacgcg ccacactctg atctggcact cgcagctgcc tgcgtgggtg 420
aacaatatca acaacgcgga tactctgcgg caagtcattc gcacccatgt ctctactgtg 480
gttgggcggt acaagggcaa gattcgtgct tgggtgagtt ttgaacacca catgcccctt 540
ttcttagtcc gctcctcctc ctcttgaac ttctcacagt tatagccgta tacaacattc 600
gacaggaaat ttaggatgac aactactgac tgacttgtgt gtgtgatggc gataggacgt 660
ggtcaatgaa atcttcaacg aggatggaac gtcgcgtctt tcagtctttt ccaggctcct 720
cggcgaggag tttgtctcga ttgcctttcg tgctgctcga gatgctgacc cttctgccc 780
tctttacatc aacgactaca atctcgaccg cgccaactat ggcaaggtca acgggttgaa 840
gacttacgtc tccaagtgga tctetcaagg agttcccatt gacggtattg gtgagccacg 900
accctaaat gtccccatt agagtctctt tctagagcca aggttgaag ccattcaggg 960
actgacacga gagccttctc tacaggaagc cagtccatc tcagcggcgg cggaggtctt 1020
ggtacgtgg gtgcgtcca gcagctggca acggtaccg tcaccgagct ggccattacc 1080
gagctggaca ttcagggggc accgacgacg gattacacc aagttgttca agcatgcctg 1140
agcgtctcca agtgcgtcgg catcaccgtg tggggcatca gtgacaaggt aagttgcttc 1200
ccctgtctgt gcttatcaac tgtaagcagc aacaactgat gctgtctgtc ttacactagg 1260

actcgtggcg tgccagcacc aaccctcttc tgtttgacgc aaacttcaac cccaagccgg 1320
catataacag cattgttggc atcttacaat ag 1352
<210> 16
<211> 347
<212> PRT
<213> 里氏木霉
<400> 16
Met Lys Ala Asn Val Ile Leu Cys Leu Leu Ala Pro Leu Val Ala Ala
1 5 10 15
Leu Pro Thr Glu Thr Ile His Leu Asp Pro Glu Leu Ala Ala Leu Arg
20 25 30
Ala Asn Leu Thr Glu Arg Thr Ala Asp Leu Trp Asp Arg Gln Ala Ser
35 40 45
Gln Ser Ile Asp Gln Leu Ile Lys Arg Lys Gly Lys Leu Tyr Phe Gly
50 55 60
Thr Ala Thr Asp Arg Gly Leu Leu Gln Arg Glu Lys Asn Ala Ala Ile
65 70 75 80
Ile Gln Ala Asp Leu Gly Gln Val Thr Pro Glu Asn Ser Met Lys Trp
85 90 95
Gln Ser Leu Glu Asn Asn Gln Gly Gln Leu Asn Trp Gly Asp Ala Asp
100 105 110
Tyr Leu Val Asn Phe Ala Gln Gln Asn Gly Lys Ser Ile Arg Gly His
115 120 125
Thr Leu Ile Trp His Ser Gln Leu Pro Ala Trp Val Asn Asn Ile Asn
130 135 140
Asn Ala Asp Thr Leu Arg Gln Val Ile Arg Thr His Val Ser Thr Val
145 150 155 160
Val Gly Arg Tyr Lys Gly Lys Ile Arg Ala Trp Asp Val Val Asn Glu
165 170 175
Ile Phe Asn Glu Asp Gly Thr Leu Arg Ser Ser Val Phe Ser Arg Leu
180 185 190
Leu Gly Glu Glu Phe Val Ser Ile Ala Phe Arg Ala Ala Arg Asp Ala
195 200 205
Asp Pro Ser Ala Arg Leu Tyr Ile Asn Asp Tyr Asn Leu Asp Arg Ala
210 215 220
Asn Tyr Gly Lys Val Asn Gly Leu Lys Thr Tyr Val Ser Lys Trp Ile
225 230 235 240
Ser Gln Gly Val Pro Ile Asp Gly Ile Gly Ser Gln Ser His Leu Ser
245 250 255

Gly Gly Gly Gly Ser Gly Thr Leu Gly Ala Leu Gln Gln Leu Ala Thr		
260	265	270
Val Pro Val Thr Glu Leu Ala Ile Thr Glu Leu Asp Ile Gln Gly Ala		
275	280	285
Pro Thr Thr Asp Tyr Thr Gln Val Val Gln Ala Cys Leu Ser Val Ser		
290	295	300
Lys Cys Val Gly Ile Thr Val Trp Gly Ile Ser Asp Lys Asp Ser Trp		
305	310	315
Arg Ala Ser Thr Asn Pro Leu Leu Phe Asp Ala Asn Phe Asn Pro Lys		
325	330	335
Pro Ala Tyr Asn Ser Ile Val Gly Ile Leu Gln		
340	345	

<210> 17

<211> 2564

<212> DNA

<213> 里氏木霉

<400> 17

```

ggacagccgg acgcaatggt gaataacgca gctcttctcg ccgccctgtc ggctctcctg 60
cccacggccc tggcgcagaa caatcaaaca tacgccaact actctgctca gggccagcct 120
gatctctacc ccgagacact tgccacgctc acactctcgt tccccgactg cgaacatggc 180
ccccctcaaga acaatctcgt ctgtgactca tcggccgggt atgtagagcg agcccaggcc 240
ctcatctcgc tcttcaccct cgaggagctc attctcaaca cgcaaaactc gggccccggc 300
gtgcctcgcc tgggtcttcc gaactaccaa gtctggaatg aggctctgca cggcttggaac 360
cgcgccaact tcgccaccaa gggcggccag ttcgaatggg cgacctcggt ccccatgccc 420
atcctcacta cggcggccct caaccgcaca ttgatccacc agattgccga catcatctcg 480
acccaagctc gagcattcag caacagcggc cgttacggtc tcgacgtcta tgcgcaaacc 540
gtcaatggct tccgaagccc cctctggggc cgtggccagg agacgcccgg cgaagacgcc 600
tttttctca gctccgccta tacttacgag tacatcacgg gcatccaggg tggcgtcgac 660
cctgagcacc tcaaggttgc cgccacgggt aagcactttg ccggatacga cctcgagaac 720
tggaacaacc agtcccgtct cggtttcgac gccatcataa ctcagcagga cctctccgaa 780
tactacactc ccagttcct cgctgcggcc cgttatgcaa agtcacgcag cttgatgtgc 840
gcatacaact ccgtcaacgg cgtgcccagc tgtgccaaca gcttcttctt gcagacgctt 900
ttgcgcgaga gctggggctt ccccgaatgg ggatacgtct cgtccgattg cgatgccgtc 960
tacaacgttt tcaaccctca tgactacgcc agcaaccagt cgtcagccgc cgccagctca 1020
ctgcgagccg gcaccgatat cgactgcggt cagacttacc cgtggcacct caacgagtec 1080
tttgtggccg gcgaagtctc ccgcggcgag atcgagcggc ccgtcaccgc tctgtacgcc 1140
aacctcgtec gtctcgata cttcgacaag aagaaccagt accgctcgct cggttgaag 1200
gatgtcgtea agactgatgc ctggaacatc tcgtacgagg ctgctgttga gggcatcgtc 1260
ctgctcaaga acgatggcac tctccctctg tccaagaagg tgcgcagcat tgctctgatc 1320

```

```

ggaccatggg ccaatgccac aacccaaatg caaggcaact actatggccc tgccccatac 1380
ctcatcagcc ctctggaagc tgctaagaag gccggctatc acgtcaactt tgaactcggc 1440
acagagatcg ccggcaacag caccactggc tttgccaagg ccattgctgc cgccaagaag 1500
tcggatgcca tcatctacct cggtggaatt gacaacacca ttgaacagga gggcgctgac 1560
cgcacggaca ttgcttggcc cggtaatcag ctggatctca tcaagcagct cagcgaggtc 1620
ggcaaacccc ttgtcgtcct gcaaattggc ggtggtcagg tagactcatc ctcgctcaag 1680
agcaacaaga aggtcaactc cctcgtctgg ggcgatatac ccggccagtc gggaggcggt 1740
gccctcttcg acattctctc tggcaagcgt gtcctgccg gccgactggg caccactcag 1800
taccggctg agtatgttca ccaattcccc cagaatgaca tgaacctccg accgatgga 1860
aagtcaaacc ctggacagac ttacatctgg tacaccggca aaccgtcta cgagtttggc 1920
agtgtctct tctacaccac cttcaaggag actctcgcca gccaccccaa gagcctcaag 1980
ttcaaacct catcgatcct ctctgtcct caccctggat acattacag cgagcagatt 2040
cccgcttca ccttcgagc caacatcaag aactcgggca agacggagtc cccatatacg 2100
gccatgctgt ttgttcgcac aagcaacgct ggcccagccc cgtaccgaa caagtggctc 2160
gtcggattcg accgacttgc cgacatcaag cctggtcact cttccaagct cagcatcccc 2220
atccctgtca gtgctctcgc ccgtgttgat tctcacgaa accgattgt ataccctggc 2280
aagtatgagc tagccttgaa caccgacgag tctgtgaagc ttgagtttga gttggtggga 2340
gaagaggtaa cgattgagaa ctggccgttg gaggagcaac agatcaagga tgctacacct 2400
gacgcataag ggttttaatg atgttggtat gacaaacggg tagagtagtt aatgatggaa 2460
taggaagagg ccatagtttt ctgtttgcaa accatttttg ccattgcgaa aaaaaaaaaa 2520
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 2564

```

<210> 18

<211> 780

<212> PRT

<213> 里氏木霉

<400> 18

```

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

```

100	105	110
Gly Leu Asp Arg Ala Asn Phe	Ala Thr Lys Gly Gly Gln Phe Glu Trp	
115	120	125
Ala Thr Ser Phe Pro Met Pro	Ile Leu Thr Thr Ala Ala Leu Asn Arg	
130	135	140
Thr Leu Ile His Gln Ile Ala Asp	Ile Ile Ser Thr Gln Ala Arg Ala	
145	150	155
Phe Ser Asn Ser Gly Arg Tyr Gly	Leu Asp Val Tyr Ala Pro Asn Val	
165	170	175
Asn Gly Phe Arg Ser Pro Leu Trp	Gly Arg Gly Gln Glu Thr Pro Gly	
180	185	190
Glu Asp Ala Phe Phe Leu Ser Ser	Ala Tyr Thr Tyr Glu Tyr Ile Thr	
195	200	205
Gly Ile Gln Gly Gly Val Asp Pro	Glu His Leu Lys Val Ala Ala Thr	
210	215	220
Val Lys His Phe Ala Gly Tyr Asp	Leu Glu Asn Trp Asn Asn Gln Ser	
225	230	235
Arg Leu Gly Phe Asp Ala Ile Ile	Thr Gln Gln Asp Leu Ser Glu Tyr	
245	250	255
Tyr Thr Pro Gln Phe Leu Ala Ala	Ala Arg Tyr Ala Lys Ser Arg Ser	
260	265	270
Leu Met Cys Ala Tyr Asn Ser Val	Asn Gly Val Pro Ser Cys Ala Asn	
275	280	285
Ser Phe Phe Leu Gln Thr Leu Leu	Arg Glu Ser Trp Gly Phe Pro Glu	
290	295	300
Trp Gly Tyr Val Ser Ser Asp Cys	Asp Ala Val Tyr Asn Val Phe Asn	
305	310	315
Pro His Asp Tyr Ala Ser Asn Gln	Ser Ser Ala Ala Ser Ser Leu	
325	330	335
Arg Ala Gly Thr Asp Ile Asp Cys	Gly Gln Thr Tyr Pro Trp His Leu	
340	345	350
Asn Glu Ser Phe Val Ala Gly Glu	Val Ser Arg Gly Glu Ile Glu Arg	
355	360	365
Ser Val Thr Arg Leu Tyr Ala Asn	Leu Val Arg Leu Gly Tyr Phe Asp	
370	375	380
Lys Lys Asn Gln Tyr Arg Ser Leu	Gly Trp Lys Asp Val Val Lys Thr	
385	390	395
Asp Ala Trp Asn Ile Ser Tyr Glu	Ala Ala Val Glu Gly Ile Val Leu	
405	410	415

Leu Lys Asn Asp Gly Thr Leu Pro Leu Ser Lys Lys Val Arg Ser Ile			
420	425	430	
Ala Leu Ile Gly Pro Trp Ala Asn Ala Thr Thr Gln Met Gln Gly Asn			
435	440	445	
Tyr Tyr Gly Pro Ala Pro Tyr Leu Ile Ser Pro Leu Glu Ala Ala Lys			
450	455	460	
Lys Ala Gly Tyr His Val Asn Phe Glu Leu Gly Thr Glu Ile Ala Gly			
465	470	475	480
Asn Ser Thr Thr Gly Phe Ala Lys Ala Ile Ala Ala Ala Lys Lys Ser			
485	490	495	
Asp Ala Ile Ile Tyr Leu Gly Gly Ile Asp Asn Thr Ile Glu Gln Glu			
500	505	510	
Gly Ala Asp Arg Thr Asp Ile Ala Trp Pro Gly Asn Gln Leu Asp Leu			
515	520	525	
Ile Lys Gln Leu Ser Glu Val Gly Lys Pro Leu Val Val Leu Gln Met			
530	535	540	
Gly Gly Gly Gln Val Asp Ser Ser Ser Leu Lys Ser Asn Lys Lys Val			
545	550	555	560
Asn Ser Leu Val Trp Gly Gly Tyr Pro Gly Gln Ser Gly Gly Val Ala			
565	570	575	
Leu Phe Asp Ile Leu Ser Gly Lys Arg Ala Pro Ala Gly Arg Leu Val			
580	585	590	
Thr Thr Gln Tyr Pro Ala Glu Tyr Val His Gln Phe Pro Gln Asn Asp			
595	600	605	
Met Asn Leu Arg Pro Asp Gly Lys Ser Asn Pro Gly Gln Thr Tyr Ile			
610	615	620	
Trp Tyr Thr Gly Lys Pro Val Tyr Glu Phe Gly Ser Gly Leu Phe Tyr			
625	630	635	640
Thr Thr Phe Lys Glu Thr Leu Ala Ser His Pro Lys Ser Leu Lys Phe			
645	650	655	
Asn Thr Ser Ser Ile Leu Ser Ala Pro His Pro Gly Tyr Thr Tyr Ser			
660	665	670	
Glu Gln Ile Pro Val Phe Thr Phe Glu Ala Asn Ile Lys Asn Ser Gly			
675	680	685	
Lys Thr Glu Ser Pro Tyr Thr Ala Met Leu Phe Val Arg Thr Ser Asn			
690	695	700	
Ala Gly Pro Ala Pro Tyr Pro Asn Lys Trp Leu Val Gly Phe Asp Arg			
705	710	715	720
Leu Ala Asp Ile Lys Pro Gly His Ser Ser Lys Leu Ser Ile Pro Ile			

				725				730				735			
Pro	Val	Ser	Ala	Leu	Ala	Arg	Val	Asp	Ser	His	Gly	Asn	Arg	Ile	Val
				740				745				750			
Tyr	Pro	Gly	Lys	Tyr	Glu	Leu	Ala	Leu	Asn	Thr	Asp	Glu	Ser	Val	Lys
				755				760				765			
Leu	Glu	Phe	Glu	Leu	Val	Gly	Glu	Glu	Val	Thr	Ile				
				770				775				780			
<210> 19															
<211> 2779															
<212> DNA															
<213> 里氏木霉															
<400> 19															
atgtccgcgt ctagtgattt atatgtagaa tgatcacaat tcatgtaact gcgttttcgc 60															
acatgcaaaa agccctaacg tgagactgag ccacttccta gttttcgtat catgtcagtt 120															
gcaaggttac accacaatgc agctcaacga gacaacgctg ccagcccata ataatggata 180															
gctggttgta gagagattaa gaagagaatg ctgtttcaga aggaagacta tatcatagca 240															
gctgctacat ttccctcttt ccctctttcc atcccttaat agatacgtac ccttgcaatt 300															
ggccgtttcg gaagagcttt tctgcttata ctaaccacct acgccagaat accggtggaa 360															
taatcagtg tcaataggga aacccccaac tgcagcatat aagcctatta acaagacgtc 420															
ccaacatgca ttttccttca gtccgcagca ggctatagag agtaggcaat ttacacacca 480															
cttttagcct ctgcacatat ctcaccacat ttgcattacg gcattccacta ttacaaccac 540															
ttggcacctg atggctttgc tctaccata tcggttttta cgttccgctg tgttcagtcg 600															
ttaaattccgt ggtggagcag aacgaccagc ttctcgtatc gggaactccg cttatccgat 660															
accctcagtc gaacccttcg tgatactcag cctaaataac atcgcatcgt agcagacaac 720															
ttcagtaatt ttgtggggt aatggtcaga tcgctcctct tatatataaa gcagaggtta 780															
gtgggctaag gaaattcgtg gttcgcttat agtagagctg tcagttgccc ttcccgaact 840															
gttagacggg atggctggta agcttatcct cgtggctcta gcaagccttg tatcactctc 900															
tattcagcag aattgcgcag cattattgta agagtgttga gcgtgttgag taccatctgt 960															
atcgttgcta acgtaggctt ttagtggcca atgtggaggc atagggtggg ccggcaccac 1020															
atgttgcggt gctggcgccc agtgcagttt tgtcaatgac tggtactccc agtgccctgc 1080															
gtcaaccgta tgagctccga tccgggccgt caatatcttc taactccaga ctgtacaggg 1140															
cggaaccccc ccaaacggaa caacttcctc tagcttggtt tcacggacgt cgtcagcatc 1200															
ctcatccgtc ggctcgtctt caccggcgcg caactcacca actggcagtg cttccaccta 1260															
cacaaccaca gatacagcta ccgtggctcc tcattcgcag tctccttacc ccagcattgc 1320															
cgcattccagt tgcggatcgt ggaccctcgt ggataatgtt tgctgcccac catattgtgc 1380															
taatgatgac acatccgagt catgctcagg ctgcggtacc tgcactacgc cgccctcggc 1440															
ggactgcaaa tccggaacca tgtatccaga ggtccatcac gtatccagca acgagagctg 1500															
gcactacagt gtaagatgac caacgctggg gtatctaate ctttgtcttc ctcggcgtgc 1560															
tgaccttgga gcatttagag atcaaccac tttggcctaa cgagcggcgg ggctgtggc 1620															

```

tttggcctgt acggtctctg cacaaagggc agtggttacag ccagctggac ggatcccatg 1680
cttggcgcga cgtgtgacgc tttttgtaca gcgtatcccc tgctttgcaa agaccctacc 1740
ggcactaccc ttcgtggcaa cttcgcagct ccaaacggcg attactacac ccaagttggg 1800
gaccccgaga ggcaatcatt ttctggtgta gtattcactg acagtgcgat agttctggtc 1860
ctcgttgcca ggagccctcg ataactacct gtctgcggc gagtgcattg agctgataca 1920
aacaaagccc gatgggaccg attatgctgt cggagaagcc ggctacacgg atccaattac 1980
tctcgagatt gtggacagct gcccgtgcag cgcgaaactcc aagtggtgct gtcagagagc 2040
cccgtccatc cgtgccattg tactacatgc gccaacgaa tggccctggc taacatctcg 2100
caggtgggcc gggcgccgat cattgcggag agatcgactt caaatacggc tgtcctcttc 2160
ctgctgacag cattcatctc gacctgtcag acattgccat gggccgtttg cagggcaatg 2220
gatcactaac caatggcgtc atcccgaact gatatagaag agtccaatgc cccaaagttg 2280
ggaacgccta catttggett cgaaatggcg gagggcetta ctattttgct ctcacggcag 2340
tcaacaccaa cggaccgggc tcagtcacca aaatcgagat caagggcgca gacaccgaca 2400
actgggttgc cttggtccat gacccaaact atacgagtag ccgcccacaa gaacgctatg 2460
gcagttgggt aatcccacag ggatcagggc cttttaactt gcctgttga attcgtctga 2520
ctagcccaac gggggaacag attgtgaatg aacaggccat caagacattc actcctccgg 2580
ccacaggtga ccccaatttt tactacattg acattggtgt gcagtttagc cagaattgat 2640
ggcaagcatt gggcaatggg cttcttgctg tgggacaatg atgtaggcta gattctcaat 2700
gcttcaagta tgtggtgtac gtcttcgtgt gtatagatag gtatgctgtt cacttaaata 2760
cacatccttt ggtacgttg 2779

```

<210> 20

<211> 493

<212> PRT

<213> 里氏木霉

<400> 20

```

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

```

Ser	Gln	Ser	Pro	Tyr	Pro	Ser	Ile	Ala	Ala	Ser	Ser	Cys	Gly	Ser	Trp
115				120				125							
Thr	Leu	Val	Asp	Asn	Val	Cys	Cys	Pro	Ser	Tyr	Cys	Ala	Asn	Asp	Asp
130				135				140							
Thr	Ser	Glu	Ser	Cys	Ser	Gly	Cys	Gly	Thr	Cys	Thr	Thr	Pro	Pro	Ser
145				150				155				160			
Ala	Asp	Cys	Lys	Ser	Gly	Thr	Met	Tyr	Pro	Glu	Val	His	His	Val	Ser
				165				170				175			
Ser	Asn	Glu	Ser	Trp	His	Tyr	Ser	Arg	Ser	Thr	His	Phe	Gly	Leu	Thr
180				185				190							
Ser	Gly	Gly	Ala	Cys	Gly	Phe	Gly	Leu	Tyr	Gly	Leu	Cys	Thr	Lys	Gly
195				200				205							
Ser	Val	Thr	Ala	Ser	Trp	Thr	Asp	Pro	Met	Leu	Gly	Ala	Thr	Cys	Asp
210				215				220							
Ala	Phe	Cys	Thr	Ala	Tyr	Pro	Leu	Leu	Cys	Lys	Asp	Pro	Thr	Gly	Thr
225				230				235				240			
Thr	Leu	Arg	Gly	Asn	Phe	Ala	Ala	Pro	Asn	Gly	Asp	Tyr	Tyr	Thr	Gln
				245				250				255			
Phe	Trp	Ser	Ser	Leu	Pro	Gly	Ala	Leu	Asp	Asn	Tyr	Leu	Ser	Cys	Gly
260				265				270							
Glu	Cys	Ile	Glu	Leu	Ile	Gln	Thr	Lys	Pro	Asp	Gly	Thr	Asp	Tyr	Ala
275				280				285							
Val	Gly	Glu	Ala	Gly	Tyr	Thr	Asp	Pro	Ile	Thr	Leu	Glu	Ile	Val	Asp
290				295				300							
Ser	Cys	Pro	Cys	Ser	Ala	Asn	Ser	Lys	Trp	Cys	Cys	Gly	Pro	Gly	Ala
305				310				315				320			
Asp	His	Cys	Gly	Glu	Ile	Asp	Phe	Lys	Tyr	Gly	Cys	Pro	Leu	Pro	Ala
				325				330				335			
Asp	Ser	Ile	His	Leu	Asp	Leu	Ser	Asp	Ile	Ala	Met	Gly	Arg	Leu	Gln
340				345				350							
Gly	Asn	Gly	Ser	Leu	Thr	Asn	Gly	Val	Ile	Pro	Thr	Arg	Tyr	Arg	Arg
355				360				365							
Val	Gln	Cys	Pro	Lys	Val	Gly	Asn	Ala	Tyr	Ile	Trp	Leu	Arg	Asn	Gly
370				375				380							
Gly	Gly	Pro	Tyr	Tyr	Phe	Ala	Leu	Thr	Ala	Val	Asn	Thr	Asn	Gly	Pro
385				390				395				400			
Gly	Ser	Val	Thr	Lys	Ile	Glu	Ile	Lys	Gly	Ala	Asp	Thr	Asp	Asn	Trp
				405				410				415			
Val	Ala	Leu	Val	His	Asp	Pro	Asn	Tyr	Thr	Ser	Ser	Arg	Pro	Gln	Glu

420	425	430
Arg Tyr Gly Ser Trp Val Ile Pro Gln Gly Ser Gly Pro Phe Asn Leu		
435	440	445
Pro Val Gly Ile Arg Leu Thr Ser Pro Thr Gly Glu Gln Ile Val Asn		
450	455	460
Glu Gln Ala Ile Lys Thr Phe Thr Pro Pro Ala Thr Gly Asp Pro Asn		
465	470	475
Phe Tyr Tyr Ile Asp Ile Gly Val Gln Phe Ser Gln Asn		
485	490	

<210> 21

<211> 2777

<212> DNA

<213> 里氏木霉

<400> 21

```

atgggtgcgct cgcacctatt cgtgtcgtcg ctcgcgacct tctccggagt cattgcccgt 60
gtctccgggc atgggtcaaa gatcggtccc ggcgcgtaga tcttcgaatt cgaggattca 120
caggtagagtc tcgctcggcc gttggcaccg gatgacatgc ctccagctgt ggcaatgctg 180
acgggactcc atccaggaca cggccgattt ctacaagaag ctcaacggcg agggctcaac 240
gcgcctgaag ttcgactaca agctgttcaa gggcggtctc gtccagctca aggacctaga 300
caaccatgag gcaaaggccc agcagatggc ccagctgcct gctgtcaaga acgtgtggcc 360
cgtcaccctc atcgacgccc ccaaccccaa ggtcgagtgg gttgccggca gcacggcgcc 420
tactctggag agcagggcga tcaagaagcc accgatcccg aacgactcga gcgacttccc 480
cacgcaccag atgacccaaa tcgacaagct gcgagccaag ggctacacgg gcaagggcgt 540
cagggttgcc gtcatgata caggcgtgag tacaagccca ctgtcccaag caagtcgtgt 600
agacgctcac atacggccag attgactaca cccaccctgc tctcggcggc tgcttttgta 660
ggggctgtct ggtctccttt ggcaccgatt tggtcggtga cgactacacc ggctttaaca 720
cgctgttccc cgatgatgac cccgtcgact gcgcgggcca cggttctcac gttgctggta 780
tcattgtctg gcaggagaat ccgtacggct tcaactggcg cgctcccgat gtcaccctcg 840
gcgcttatcg agtctttggc tgcgacggcc aggcgggtaa cgatgtcctg atttccgctt 900
acaaccaggc ctttgaggac ggtgcccaga tcatactgc ctccattggc ggtccctctg 960
gctgggctga ggagccgtgg gccgttgccg tcaccgcgat cgttgaggca ggtgttccct 1020
gcacggtctc tgccggcaac gagggcgact ctgggtctct ctttgccagc acggcagcca 1080
atggcaagaa agtcattgct gtcgcctccg tcgacaacga gaacatccct tcagtgtgtg 1140
ccgtggcctc ttacaaaatt gacagcggcg ctgcccagga ctttggttac gtctctctct 1200
ccaaggcgtg ggacggcgtg agcaagcccc tgtatgtgtg gtcgttcgac actactattc 1260
ccgacgatgg ctgctgcct ctccttgaca gcactccga cctctctgac tacattgtcc 1320
ttgtccgccg tggcacctgc acctttgtcc agaaagccca aatgtcgtc gcaaaggcg 1380
ccaagtacct gctctattat aacaacattc ccggtgcgtg ggccgtcgat gtcagcgccg 1440
tccccgagat tgaggctgtc ggcatggtcg atgacaagac ggggtgctacc tggattgccg 1500

```

ccctcaagga tggaaagacc gtcaccctga cactgactga cccgatcgag agcgagaagc 1560
 aaattcagtt cagcgacaac ccgacaactg gcggtgctct gagcggctac acaacctggg 1620
 gccctacctg ggagctggac gtcaagcctc agatcagctc tcccggcggc aacattctct 1680
 ccacgtaccc cgtggctctc ggaggatatg ccaccctgtc cggtagctcc atggcctgcc 1740
 ccctgacggc ggctgctgtt gctctgattg gacaagctcg tggcaccttt gaccctgcct 1800
 tgatcgacaa cttgttggca acgactgcca acccccagct gttcaacgac ggcgagaagt 1860
 tctacgactt cctcgcccc gttccccaac agggcggtgg cctcatccag gcctacgatg 1920
 ccgcctttgc gaccactctc ctgtcaccgt ccagcctgtc gttcaacgac actgaccact 1980
 tcatacaaga gaagcagatc accctcaaga acaccagcaa gcagagggtc acctacaagc 2040
 tcaaccacgt ccccaaccaac accttttaca ctctggcacc cggtaacggc tatccagctc 2100
 cctttcctaa cgacgccgtt gccgtcacg ccaatctcaa gtttaatctg cagcaagtga 2160
 ccctgccccg cggcaggtcc atcaactgtc agttcttccc tactcccccc agggacgtcg 2220
 acgccaagcg cctggcgctt tggtcgggct acatcacggt caacggcacg gatggcacca 2280
 gtctgtctgt cccgtaccag ggccctaccg gctccctgca caagcagaag gtgctctatc 2340
 cggaggactc ctggatcgcc gattccaccg atgaaagcct ggcccctgtt gagaacggca 2400
 ccgtcttcac cattcccgcg ccgggcaacg ctggccccga tgacaagctc ccatcgctcg 2460
 tcgtcagccc tgcccttggc tctcgttatg tccgcgttga tctcgtctc ctgtccgcgc 2520
 ctctcatgg caccaagctc aagacggtca agttctcga caccacctcc atcggccagc 2580
 ctgccggatc accgctcctc tggatcagcc gtggcgccaa ccctattgct tggaccggcg 2640
 agctgtctga caacaagttt gctccccctg gaacgtacaa ggccgtgttc catgctctgc 2700
 gtattttcgg caacgagaag aagaaggagg actgggatgt gagcgaatct cctgccttca 2760
 ccatcaagta tgcgtag 2777

<210> 22

<211> 882

<212> PRT

<213> 里氏木霉

<400> 22

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

Val Thr Leu Ile Asp Ala Pro Asn Pro Lys Val Glu Trp Val Ala Gly		
100	105	110
Ser Thr Ala Pro Thr Leu Glu Ser Arg Ala Ile Lys Lys Pro Pro Ile		
115	120	125
Pro Asn Asp Ser Ser Asp Phe Pro Thr His Gln Met Thr Gln Ile Asp		
130	135	140
Lys Leu Arg Ala Lys Gly Tyr Thr Gly Lys Gly Val Arg Val Ala Val		
145	150	155
Ile Asp Thr Gly Ile Asp Tyr Thr His Pro Ala Leu Gly Gly Cys Phe		
165	170	175
Gly Arg Gly Cys Leu Val Ser Phe Gly Thr Asp Leu Val Gly Asp Asp		
180	185	190
Tyr Thr Gly Phe Asn Thr Pro Val Pro Asp Asp Asp Pro Val Asp Cys		
195	200	205
Ala Gly His Gly Ser His Val Ala Gly Ile Ile Ala Ala Gln Glu Asn		
210	215	220
Pro Tyr Gly Phe Thr Gly Gly Ala Pro Asp Val Thr Leu Gly Ala Tyr		
225	230	235
Arg Val Phe Gly Cys Asp Gly Gln Ala Gly Asn Asp Val Leu Ile Ser		
245	250	255
Ala Tyr Asn Gln Ala Phe Glu Asp Gly Ala Gln Ile Ile Thr Ala Ser		
260	265	270
Ile Gly Gly Pro Ser Gly Trp Ala Glu Glu Pro Trp Ala Val Ala Val		
275	280	285
Thr Arg Ile Val Glu Ala Gly Val Pro Cys Thr Val Ser Ala Gly Asn		
290	295	300
Glu Gly Asp Ser Gly Leu Phe Phe Ala Ser Thr Ala Ala Asn Gly Lys		
305	310	315
Lys Val Ile Ala Val Ala Ser Val Asp Asn Glu Asn Ile Pro Ser Val		
325	330	335
Leu Ser Val Ala Ser Tyr Lys Ile Asp Ser Gly Ala Ala Gln Asp Phe		
340	345	350
Gly Tyr Val Ser Ser Ser Lys Ala Trp Asp Gly Val Ser Lys Pro Leu		
355	360	365
Tyr Ala Val Ser Phe Asp Thr Thr Ile Pro Asp Asp Gly Cys Ser Pro		
370	375	380
Leu Pro Asp Ser Thr Pro Asp Leu Ser Asp Tyr Ile Val Leu Val Arg		
385	390	395
Arg Gly Thr Cys Thr Phe Val Gln Lys Ala Gln Asn Val Ala Ala Lys		

				405					410					415			
Gly	Ala	Lys	Tyr	Leu	Leu	Tyr	Tyr	Asn	Asn	Ile	Pro	Gly	Ala	Leu	Ala		
				420					425					430			
Val	Asp	Val	Ser	Ala	Val	Pro	Glu	Ile	Glu	Ala	Val	Gly	Met	Val	Asp		
				435					440					445			
Asp	Lys	Thr	Gly	Ala	Thr	Trp	Ile	Ala	Ala	Leu	Lys	Asp	Gly	Lys	Thr		
				450					455					460			
Val	Thr	Leu	Thr	Leu	Thr	Asp	Pro	Ile	Glu	Ser	Glu	Lys	Gln	Ile	Gln		
465									470					475			480
Phe	Ser	Asp	Asn	Pro	Thr	Thr	Gly	Gly	Ala	Leu	Ser	Gly	Tyr	Thr	Thr		
				485					490					495			
Trp	Gly	Pro	Thr	Trp	Glu	Leu	Asp	Val	Lys	Pro	Gln	Ile	Ser	Ser	Pro		
				500					505					510			
Gly	Gly	Asn	Ile	Leu	Ser	Thr	Tyr	Pro	Val	Ala	Leu	Gly	Gly	Tyr	Ala		
				515					520					525			
Thr	Leu	Ser	Gly	Thr	Ser	Met	Ala	Cys	Pro	Leu	Thr	Ala	Ala	Ala	Val		
				530					535					540			
Ala	Leu	Ile	Gly	Gln	Ala	Arg	Gly	Thr	Phe	Asp	Pro	Ala	Leu	Ile	Asp		
545									550					555			560
Asn	Leu	Leu	Ala	Thr	Thr	Ala	Asn	Pro	Gln	Leu	Phe	Asn	Asp	Gly	Glu		
				565					570					575			
Lys	Phe	Tyr	Asp	Phe	Leu	Ala	Pro	Val	Pro	Gln	Gln	Gly	Gly	Gly	Leu		
				580					585					590			
Ile	Gln	Ala	Tyr	Asp	Ala	Ala	Phe	Ala	Thr	Thr	Leu	Leu	Ser	Pro	Ser		
				595					600					605			
Ser	Leu	Ser	Phe	Asn	Asp	Thr	Asp	His	Phe	Ile	Lys	Lys	Lys	Gln	Ile		
				610					615					620			
Thr	Leu	Lys	Asn	Thr	Ser	Lys	Gln	Arg	Val	Thr	Tyr	Lys	Leu	Asn	His		
625																	
				630					635					640			
Val	Pro	Thr	Asn	Thr	Phe	Tyr	Thr	Leu	Ala	Pro	Gly	Asn	Gly	Tyr	Pro		
				645					650					655			
Ala	Pro	Phe	Pro	Asn	Asp	Ala	Val	Ala	Ala	His	Ala	Asn	Leu	Lys	Phe		
				660					665					670			
Asn	Leu	Gln	Gln	Val	Thr	Leu	Pro	Ala	Gly	Arg	Ser	Ile	Thr	Val	Asp		
				675					680					685			
Val	Phe	Pro	Thr	Pro	Pro	Arg	Asp	Val	Asp	Ala	Lys	Arg	Leu	Ala	Leu		
				690					695					700			
Trp	Ser	Gly	Tyr	Ile	Thr	Val	Asn	Gly	Thr	Asp	Gly	Thr	Ser	Leu	Ser		
705																	
				710					715					720			

Val	Pro	Tyr	Gln	Gly	Leu	Thr	Gly	Ser	Leu	His	Lys	Gln	Lys	Val	Leu		
					725						730						735
Tyr	Pro	Glu	Asp	Ser	Trp	Ile	Ala	Asp	Ser	Thr	Asp	Glu	Ser	Leu	Ala		
					740						745						750
Pro	Val	Glu	Asn	Gly	Thr	Val	Phe	Thr	Ile	Pro	Ala	Pro	Gly	Asn	Ala		
					755						760						765
Gly	Pro	Asp	Asp	Lys	Leu	Pro	Ser	Leu	Val	Val	Ser	Pro	Ala	Leu	Gly		
					770						775						780
Ser	Arg	Tyr	Val	Arg	Val	Asp	Leu	Val	Leu	Leu	Ser	Ala	Pro	Pro	His		
					785						790						800
Gly	Thr	Lys	Leu	Lys	Thr	Val	Lys	Phe	Leu	Asp	Thr	Thr	Ser	Ile	Gly		
					805						810						815
Gln	Pro	Ala	Gly	Ser	Pro	Leu	Leu	Trp	Ile	Ser	Arg	Gly	Ala	Asn	Pro		
					820						825						830
Ile	Ala	Trp	Thr	Gly	Glu	Leu	Ser	Asp	Asn	Lys	Phe	Ala	Pro	Pro	Gly		
					835						840						845
Thr	Tyr	Lys	Ala	Val	Phe	His	Ala	Leu	Arg	Ile	Phe	Gly	Asn	Glu	Lys		
					850						855						860
Lys	Lys	Glu	Asp	Trp	Asp	Val	Ser	Glu	Ser	Pro	Ala	Phe	Thr	Ile	Lys		
					865						870						880

Tyr Ala

<210> 23

<211> 1302

<212> DNA

<213> 里氏木霉

<400> 23

```

atgcagacct ttggagcttt tctcgtttcc ttctcgccg ccagcggcct ggccgcgcc 60
ctccccaccg agggtcagaa gacggcttcc gtcgaggtcc agtacaacaa gaactacgtc 120
ccccacggcc ctactgctct cttcaaggcc aagagaaagt atggcgctcc catcagcgac 180
aacctgaagt ctctcgtggc tgccaggcag gccaaagcagg ctctcgccaa gcgccagacc 240
ggctcggcgc ccaaccaccc cagtgcagc gccgattcgg agtacatcac ctccgtctcc 300
atcggcactc cggtcaggt cctccccctg gactttgaca ccggtcctc cgacctgtgg 360
gtcttttagct ccgagacgcc caagtcttcg gccaccggcc acgcatcta cacgccctcc 420
aagtcgtcca cctccaagaa ggtgtctggc gccagctggt ccatcagcta cggcgacggc 480
agcagctcca gcggcgatgt ctacaccgac aaggtcacca tcggaggctt cagcgtcaac 540
accagggcgc tcgagtctgc caccgcgtg tccaccgagt tcgtccagga cacggtcatc 600
tctggcctcg tcggccttgc ctttgacagc ggcaaccagg tcaggccgca cccgcagaag 660
acgtggttct ccaacgccgc cagcagcctg gctgagcccc ttttactgc cgacctgagg 720
cacggacaga gtaagtagac actcactgga attcgttctt ttcccgatca tcatgaaagc 780

```


aagtagactg actgaaccaa acaactagac ggcagctaca actttggcta catcgacacc 840
agcgctcgcca agggcccccgt tgcctacacc cccgttgaca acagccaggg cttctgggag 900
ttcactgcct cgggctactc tgtcggcggc ggcaagctca accgcaactc catcgacggc 960
attgccgaca ccggcaccac cctgctcctc ctgcagaca acgtcgtcga tgcctactac 1020
gccaacgtcc agtcggccca gtacgacaac cagcaggagg gtgtcgtctt cgactgcgac 1080
gaggacctcc cttcgttcag cttcgggtgtt ggaagctcca ccatcaccat ccctggcgat 1140
ctgctgaacc tgactcccct cgaggagggc agtccacct gcttcgggtgg cctccagagc 1200
agctccggca ttggcatcaa catctttggt gacgttgccc tcaaggctgc cctggttgctc 1260
tttgacctcg gcaacgagcg cctgggctgg gctcagaaat aa 1302

<210> 24

<211> 407

<212> PRT

<213> 里氏木霉

<400> 24

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

195	200	205
Asp Ser Gly Asn Gln Val Arg Pro His Pro Gln Lys Thr Trp Phe Ser		
210	215	220
Asn Ala Ala Ser Ser Leu Ala Glu Pro Leu Phe Thr Ala Asp Leu Arg		
225	230	235
His Gly Gln Asn Gly Ser Tyr Asn Phe Gly Tyr Ile Asp Thr Ser Val		
245	250	255
Ala Lys Gly Pro Val Ala Tyr Thr Pro Val Asp Asn Ser Gln Gly Phe		
260	265	270
Trp Glu Phe Thr Ala Ser Gly Tyr Ser Val Gly Gly Gly Lys Leu Asn		
275	280	285
Arg Asn Ser Ile Asp Gly Ile Ala Asp Thr Gly Thr Thr Leu Leu Leu		
290	295	300
Leu Asp Asp Asn Val Val Asp Ala Tyr Tyr Ala Asn Val Gln Ser Ala		
305	310	315
Gln Tyr Asp Asn Gln Gln Glu Gly Val Val Phe Asp Cys Asp Glu Asp		
325	330	335
Leu Pro Ser Phe Ser Phe Gly Val Gly Ser Ser Thr Ile Thr Ile Pro		
340	345	350
Gly Asp Leu Leu Asn Leu Thr Pro Leu Glu Glu Gly Ser Ser Thr Cys		
355	360	365
Phe Gly Gly Leu Gln Ser Ser Ser Gly Ile Gly Ile Asn Ile Phe Gly		
370	375	380
Asp Val Ala Leu Lys Ala Ala Leu Val Val Phe Asp Leu Gly Asn Glu		
385	390	395
Arg Leu Gly Trp Ala Gln Lys		400
405		

<210> 25

<211> 933

<212> DNA

<213> 里氏木霉

<400> 25

atggctcccg cttcccaagt cgtctcagct ctcatgctgc ccgctctcgc cttgggagcc 60
gccatccagc cccgtggcgc tgacatcgtg ggaggaaccg ccgcctcgtc cggcgagttc 120
ccctacattg tcagtctgca gaacccaac cagggcggcc acttctgcgg tgggtgtcttg 180
gtcaacgcca acaccgtcgt taccgccgct cactgctccg ttgtctaccc tgccctgcag 240
atccgcgtcc gcgccgtac tcttgtaagt ttgcttgttt cgagtcctcg aaaagacatg 300
aacctgcgat ggctaacc aa agcacctcct ctctgataga cctggaactc tggcggtacc 360
ctggtcggcg tctcccagat catcgtgaac ccgtcctaca acgaccgcac caccgacttt 420

gacgttgccg tctggcacct gtccagccct atccgcgaga gctccacat tggctacgcc 480
 actcttccccg cccagggtc cgaccccggtg gccggctcga ccgtcaccac cgctggctgg 540
 taagcatcat catcattgat agccgggaca tgctggcgtc aaatccgagt ttgctaacca 600
 ttctttccaaa aaaacagggg caccaccagc gagaactcca actccatccc ctcccgctg 660
 aacaaggtct ccgtccccgt cgtcgccgc tccacctgcc aggccgacta ccgcagccag 720
 gggctcagtg tcaccaacaa catgttctgc gccggcctca cccagggcgg caaggactct 780
 tgctctggcg actctggcgg ccccatcggt gacgccaacg gtgtcctcca ggggtgcgtt 840
 tcttggggta tcggctgtgc tgaggccggt ttccctgggt tctacaccag aatcggaac 900
 tttgtcaact acatcaacca gaacctcgca taa 933

<210> 26

<211> 259

<212> PRT

<213> 里氏木霉

<400> 26

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

195	200	205
Ser Cys Ser Gly Asp Ser Gly Gly Pro Ile Val Asp Ala Asn Gly Val		
210	215	220
Leu Gln Gly Val Val Ser Trp Gly Ile Gly Cys Ala Glu Ala Gly Phe		
225	230	235
Pro Gly Val Tyr Thr Arg Ile Gly Asn Phe Val Asn Tyr Ile Asn Gln		
245	250	255

Asn Leu Ala

<210> 27

<211> 1684

<212> DNA

<213> 里氏木霉

<400> 27

```

atgcggtccg ttgtcgccct ctccatggcg gccgttgccc aggccagcac attccagatt 60
ggcaccatcc acgagaagtc ggccccctg ctgagcaacg tcgaggccaa cgccatcccc 120
gatgcctaca tcatcaagtt caaggaccac gtgggtgagg atgatgcctc caagcaccac 180
gactggatcc agagcatcca cacaacggtt gagcaggagc gccttgagct ccgcaagcga 240
agcaacgtct ttggcgccga cgacgtcttt gacggtctga agcacacttt caagattggc 300
gacggcttca agggctacgc cggtcacttc cagagctctg tcattgagca ggtccggaac 360
caccctgacg taagttttgc acagccgccc tcccttttgg ctccccaaca aagctaacc 420
ctcccagggt gagtacctcg agcgcgacag cattgtgcac accatgcttc ccctcgagtc 480
caaggacagc atcatcgttg aggactcgtg caacggcgag acggagaagc aggctccctg 540
gggtcttgcc cgtatctctc accgagagac gctcaacttt ggctccttca acaagtacct 600
ctacaccgct gatggtggtg aggggtgtga tgcctatgtc attgacaccg gcaccaacat 660
cgagcacgtc gactttgagg gtcgtgccaa gtggggcaag accatccctg ccggcgatga 720
ggacgaggac ggcaacggcc acggcactca ctgctctggt accgttgctg gtaagaagta 780
cggtgttgcc aagaaggccc acgtctacgc cgtcaagggt ctccgatcca acggatccgg 840
caccatgtct gacgtcgtca agggcgtcga gtacgtgct ctctcccaca ttgagcaggt 900
gaagaaggcc aagaagggca agcgggaagg cttcaagggc tccgtcgcca acatgtccct 960
cggtggtggc aagaccagc ctcttgacgc tgccgtcaac gccgccgtcc gcgccggtgt 1020
ccactttgcc gttgctgccg gcaacgacaa cgctgatgct tgcaactact cccccgctgc 1080
cgccactgag cccctcaccg tcggtgcttc tgctctcgat gacagccgtg cttacttctc 1140
caactacggc aagtgcactg acatcttcgc cctgggtctg agcatccagt ccacctggat 1200
tggtccaag tatgccgtca acaccatctc tggtacctcc atggcctctc ctacatctg 1260
cggctctcct gcctactacc tgtctctcca gcccgctggt gactctgagt tcgctgttgc 1320
ccccatcacc cccaagaagc tcaaggagag cgtcatctct gtcgccacca agaagccct 1380
ctctgacctg cccgactctg acacccccaa cctgctcgcc tggaacggcg gtggtgcag 1440
caacttctcc cagattgtcg aggccggcag ctacactgtc aagccaagc agaacaagca 1500
ggccaagctc ccagcacca ttgaggagct cgaggaggcc atcgagggtg actttgaggt 1560

```

cgtctctggc gagatcgtca agggtgccaa gagctttggc tccaaggcgg agaagtttgc 1620
caagaagatc cacgatctcg tcgaggagga gattgaggag ttcattctctg agctctccga 1680
gtaa 1684

<210> 28

<211> 540

<212> PRT

<213> 里氏木霉

<400> 28

Met	Arg	Ser	Val	Val	Ala	Leu	Ser	Met	Ala	Ala	Val	Ala	Gln	Ala	Ser
1				5					10					15	
Thr	Phe	Gln	Ile	Gly	Thr	Ile	His	Glu	Lys	Ser	Ala	Pro	Val	Leu	Ser
				20				25					30		
Asn	Val	Glu	Ala	Asn	Ala	Ile	Pro	Asp	Ala	Tyr	Ile	Ile	Lys	Phe	Lys
				35				40					45		
Asp	His	Val	Gly	Glu	Asp	Asp	Ala	Ser	Lys	His	His	Asp	Trp	Ile	Gln
				50				55					60		
Ser	Ile	His	Thr	Asn	Val	Glu	Gln	Glu	Arg	Leu	Glu	Leu	Arg	Lys	Arg
65				70					75					80	
Ser	Asn	Val	Phe	Gly	Ala	Asp	Asp	Val	Phe	Asp	Gly	Leu	Lys	His	Thr
				85					90					95	
Phe	Lys	Ile	Gly	Asp	Gly	Phe	Lys	Gly	Tyr	Ala	Gly	His	Phe	His	Glu
				100				105					110		
Ser	Val	Ile	Glu	Gln	Val	Arg	Asn	His	Pro	Val	Glu	Tyr	Ile	Glu	Arg
				115				120					125		
Asp	Ser	Ile	Val	His	Thr	Met	Leu	Pro	Leu	Glu	Ser	Lys	Asp	Ser	Ile
				130				135					140		
Ile	Val	Glu	Asp	Ser	Cys	Asn	Gly	Glu	Thr	Glu	Lys	Gln	Ala	Pro	Trp
145				150					155					160	
Gly	Leu	Ala	Arg	Ile	Ser	His	Arg	Glu	Thr	Leu	Asn	Phe	Gly	Ser	Phe
				165					170					175	
Asn	Lys	Tyr	Leu	Tyr	Thr	Ala	Asp	Gly	Gly	Glu	Gly	Val	Asp	Ala	Tyr
				180					185					190	
Val	Ile	Asp	Thr	Gly	Thr	Asn	Ile	Glu	His	Val	Asp	Phe	Glu	Gly	Arg
				195				200					205		
Ala	Lys	Trp	Gly	Lys	Thr	Ile	Pro	Ala	Gly	Asp	Glu	Asp	Glu	Asp	Gly
				210				215					220		
Asn	Gly	His	Gly	Thr	His	Cys	Ser	Gly	Thr	Val	Ala	Gly	Lys	Lys	Tyr
225					230				235					240	
Gly	Val	Ala	Lys	Lys	Ala	His	Val	Tyr	Ala	Val	Lys	Val	Leu	Arg	Ser

	245		250		255
Asn Gly Ser Gly Thr Met Ser Asp Val Val Lys Gly Val Glu Tyr Ala					
	260		265		270
Ala Leu Ser His Ile Glu Gln Val Lys Lys Ala Lys Lys Gly Lys Arg					
	275		280		285
Lys Gly Phe Lys Gly Ser Val Ala Asn Met Ser Leu Gly Gly Gly Lys					
	290		295		300
Thr Gln Ala Leu Asp Ala Ala Val Asn Ala Ala Val Arg Ala Gly Val					
305		310		315	320
His Phe Ala Val Ala Ala Gly Asn Asp Asn Ala Asp Ala Cys Asn Tyr					
	325		330		335
Ser Pro Ala Ala Ala Thr Glu Pro Leu Thr Val Gly Ala Ser Ala Leu					
	340		345		350
Asp Asp Ser Arg Ala Tyr Phe Ser Asn Tyr Gly Lys Cys Thr Asp Ile					
	355		360		365
Phe Ala Pro Gly Leu Ser Ile Gln Ser Thr Trp Ile Gly Ser Lys Tyr					
	370		375		380
Ala Val Asn Thr Ile Ser Gly Thr Ser Met Ala Ser Pro His Ile Cys					
385		390		395	400
Gly Leu Leu Ala Tyr Tyr Leu Ser Leu Gln Pro Ala Gly Asp Ser Glu					
	405		410		415
Phe Ala Val Ala Pro Ile Thr Pro Lys Lys Leu Lys Glu Ser Val Ile					
	420		425		430
Ser Val Ala Thr Lys Asn Ala Leu Ser Asp Leu Pro Asp Ser Asp Thr					
	435		440		445
Pro Asn Leu Leu Ala Trp Asn Gly Gly Gly Cys Ser Asn Phe Ser Gln					
	450		455		460
Ile Val Glu Ala Gly Ser Tyr Thr Val Lys Pro Lys Gln Asn Lys Gln					
465		470		475	480
Ala Lys Leu Pro Ser Thr Ile Glu Glu Leu Glu Glu Ala Ile Glu Gly					
	485		490		495
Asp Phe Glu Val Val Ser Gly Glu Ile Val Lys Gly Ala Lys Ser Phe					
	500		505		510
Gly Ser Lys Ala Glu Lys Phe Ala Lys Lys Ile His Asp Leu Val Glu					
	515		520		525
Glu Glu Ile Glu Glu Phe Ile Ser Glu Leu Ser Glu					
	530		535		540
<210> 29					
<211> 1341					

<212> DNA

<213> 里氏木霉

<400> 29

```

atgaagagcg cggtacttgc cgccgcggcg cttgtcggct ccgcccgaagc cggcattcac 60
aagatgaagc tgcagaaggt ctccctggag cagcagctgg taagacgaca ccctcatcca 120
cggcctcgta ctctagccaa gcgcaatcac tgacacgccg cctctctcat ctaggagggt 180
tcgagcatcg aggcccacgt ccagcagctc ggccagaagt acatgggcgt ccgccctact 240
agccgtgccg aggtcatgtt caacgacaag ccgcccgaag tccagggcgg gcacccggtt 300
cccgtcacca acttcatgaa tgcccaatgt aagtcgtgat gcgcagcaca gcacgagagt 360
cccgtccca ggtagcgagc acatgcttac taacttgctc ggacagactt ctctgagatt 420
accatcgga cccccctca gtcgttcaag gttgtctcg acacgggaag ctctaacctc 480
tggttccct ctcatcgtg caacagcatc gctgtcttc tgcactccac gtacgattcg 540
tcttcatcgt cgacgtacaa gcccacggc tccgattttg agatccacta cggatcaggt 600
agcttgactg gcttcatctc caacgatgtc gtgacgattg gcgacctcaa gatcaagggg 660
caggactttg ccgaggcaac cagcgagccc ggccttgctt ttgctttcgg ccgcttcgac 720
ggcattcttg gccttggtc cgataccatc tcggtcaatg gcattgtccc ccccttttac 780
cagatgggtc accagaagct gatcgacgag cccgtctttg ctttctacct gggaagcagc 840
gacgagggtt ccgaggctgt ctttggcggc gtcgacgatg ctactacga gggcaagatt 900
gagtacattc ccctgcgccg caaggcctac tgggaggtgg accttgactc cattgccttc 960
ggtgacgagg tcgccgagct cgagaacact ggcccatcc ttgacaccgg cacctctctc 1020
aacgtctcc cctcgggcct cgccgagctc ctgaacgctg agattggcgc caagaagggc 1080
tttggcggtc agtacactgt tgactgtccc aagcgtgatt ccctccccga catcaccttc 1140
agcctggccg gctccaagta cagccttccc gccagcgact acatcattga gatgtctggc 1200
aactgcattt cgtccttcca gggcatggac ttccccgagc ccgtgggccc cctggtcatt 1260
ctgggtgatg ctttcttgcg ccgtactac tccgtctacg accttggcag ggacgccgtt 1320
ggtcttgcca aggccaaata a 1341

```

<210> 30

<211> 395

<212> PRT

<213> 里氏木霉

<400> 30

```

Met Lys Ser Ala Leu Leu Ala Ala Ala Leu Val Gly Ser Ala Gln
1           5           10           15
Ala Gly Ile His Lys Met Lys Leu Gln Lys Val Ser Leu Glu Gln Gln
          20           25           30
Leu Glu Gly Ser Ser Ile Glu Ala His Val Gln Gln Leu Gly Gln Lys
          35           40           45
Tyr Met Gly Val Arg Pro Thr Ser Arg Ala Glu Val Met Phe Asn Asp
          50           55           60

```

Lys	Pro	Pro	Lys	Val	Gln	Gly	Gly	His	Pro	Val	Pro	Val	Thr	Asn	Phe
65					70				75						80
Met	Asn	Ala	Gln	Tyr	Phe	Ser	Glu	Ile	Thr	Ile	Gly	Thr	Pro	Pro	Gln
					85				90						95
Ser	Phe	Lys	Val	Val	Leu	Asp	Thr	Gly	Ser	Ser	Asn	Leu	Trp	Val	Pro
					100				105						110
Ser	Gln	Ser	Cys	Asn	Ser	Ile	Ala	Cys	Phe	Leu	His	Ser	Thr	Tyr	Asp
					115				120						125
Ser	Ser	Ser	Ser	Ser	Thr	Tyr	Lys	Pro	Asn	Gly	Ser	Asp	Phe	Glu	Ile
					130				135						140
His	Tyr	Gly	Ser	Gly	Ser	Leu	Thr	Gly	Phe	Ile	Ser	Asn	Asp	Val	Val
145						150				155					160
Thr	Ile	Gly	Asp	Leu	Lys	Ile	Lys	Gly	Gln	Asp	Phe	Ala	Glu	Ala	Thr
					165				170						175
Ser	Glu	Pro	Gly	Leu	Ala	Phe	Ala	Phe	Gly	Arg	Phe	Asp	Gly	Ile	Leu
					180				185						190
Gly	Leu	Gly	Tyr	Asp	Thr	Ile	Ser	Val	Asn	Gly	Ile	Val	Pro	Pro	Phe
					195				200						205
Tyr	Gln	Met	Val	Asn	Gln	Lys	Leu	Ile	Asp	Glu	Pro	Val	Phe	Ala	Phe
					210				215						220
Tyr	Leu	Gly	Ser	Ser	Asp	Glu	Gly	Ser	Glu	Ala	Val	Phe	Gly	Gly	Val
225						230				235					240
Asp	Asp	Ala	His	Tyr	Glu	Gly	Lys	Ile	Glu	Tyr	Ile	Pro	Leu	Arg	Arg
					245				250						255
Lys	Ala	Tyr	Trp	Glu	Val	Asp	Leu	Asp	Ser	Ile	Ala	Phe	Gly	Asp	Glu
					260				265						270
Val	Ala	Glu	Leu	Glu	Asn	Thr	Gly	Ala	Ile	Leu	Asp	Thr	Gly	Thr	Ser
					275				280						285
Leu	Asn	Val	Leu	Pro	Ser	Gly	Leu	Ala	Glu	Leu	Leu	Asn	Ala	Glu	Ile
					290				295						300
Gly	Ala	Lys	Lys	Gly	Phe	Gly	Gly	Gln	Tyr	Thr	Val	Asp	Cys	Ser	Lys
305						310				315					320
Arg	Asp	Ser	Leu	Pro	Asp	Ile	Thr	Phe	Ser	Leu	Ala	Gly	Ser	Lys	Tyr
					325				330						335
Ser	Leu	Pro	Ala	Ser	Asp	Tyr	Ile	Ile	Glu	Met	Ser	Gly	Asn	Cys	Ile
					340				345						350
Ser	Ser	Phe	Gln	Gly	Met	Asp	Phe	Pro	Glu	Pro	Val	Gly	Pro	Leu	Val
					355				360						365
Ile	Leu	Gly	Asp	Ala	Phe	Leu	Arg	Arg	Tyr	Tyr	Ser	Val	Tyr	Asp	Leu

370	375	380
Gly Arg Asp Ala Val Gly Leu Ala Lys Ala Lys		
385	390	395
<210> 31		
<211> 37		
<212> DNA		
<213> 烟曲霉 (<i>Aspergillus fumigatus</i>)		
<400> 31		
actggatttta ccatgacttt gtccaagatc acttcca 37		
<210> 32		
<211> 40		
<212> DNA		
<213> 烟曲霉		
<400> 32		
tcacctctag ttaattaagc gttgaacagt gcaggaccag 40		
<210> 33		
<211> 17		
<212> DNA		
<213> 烟曲霉		
<400> 33		
tgtcccttgt cgatgcg 17		
<210> 34		
<211> 17		
<212> DNA		
<213> 烟曲霉		
<400> 34		
cacatgactt ggcttcc 17		
<210> 35		
<211> 862		
<212> DNA		
<213> 烟曲霉		
<400> 35		
atgactttgt ccaagatcac ttccattgct ggccttctgg cctcagcgtc tctcgtggct 60		
ggccacggct ttgtttctgg cattgttget gatgggaaat agtatgtgct tgaaccacac 120		
aaatgacagc tgcaacagct aacttctatt ccagttacgg aggttacctt gttaaccaat 180		
acctctacat gagcaaccct cccgacacca ttgcctggtc caccaccgcc accgacctcg 240		
gctttgtgga cggcaccggc taccagtctc cggatattat ctgccacaga gacgcaaaga 300		
atggcaagtt gaccgcaacc gttgcagccg gttcacagat cgaattccag tggacgacgt 360		
ggccagagtc tcacatgga ccggtacgac gccgaagaga agagaacata ttgtgaccag 420		

ataggctaac atagcatagt tgattactta cctcgctcca tgcaacggcg actgtgccac 480
 cgtggacaag accaccctga agtttgtcaa gatcgccgct caaggcttga tcgacggctc 540
 caaccacact ggtgtttggg ctgatgatga aatgatcgcc aacaacaaca cggccacagt 600
 gaccattcct gcctcctatg cccccgaaa ctacgtcctt cgccacgaga tcatcgccct 660
 tcactctgcg ggtaacctga acggcgcgca gaactacccc cagtgtttca acatccaaat 720
 caccgggtggc ggcagtgctc agggatctgg caccgctggc acgtccctgt acaagaatac 780
 tgatcctggc atcaagtttg acatctactc ggatctgagc ggtggatacc ctattcctgg 840
 tcctgcaactg ttcaacgctt aa 862

<210> 36

<211> 250

<212> PRT

<213> 烟曲霉

<400> 36

Met	Thr	Leu	Ser	Lys	Ile	Thr	Ser	Ile	Ala	Gly	Leu	Leu	Ala	Ser	Ala
1				5					10					15	
Ser	Leu	Val	Ala	Gly	His	Gly	Phe	Val	Ser	Gly	Ile	Val	Ala	Asp	Gly
			20					25					30		
Lys	Tyr	Tyr	Gly	Gly	Tyr	Leu	Val	Asn	Gln	Tyr	Pro	Tyr	Met	Ser	Asn
			35					40					45		
Pro	Pro	Asp	Thr	Ile	Ala	Trp	Ser	Thr	Thr	Ala	Thr	Asp	Leu	Gly	Phe
		50				55					60				
Val	Asp	Gly	Thr	Gly	Tyr	Gln	Ser	Pro	Asp	Ile	Ile	Cys	His	Arg	Asp
65				70						75				80	
Ala	Lys	Asn	Gly	Lys	Leu	Thr	Ala	Thr	Val	Ala	Ala	Gly	Ser	Gln	Ile
				85					90					95	
Glu	Phe	Gln	Trp	Thr	Thr	Trp	Pro	Glu	Ser	His	His	Gly	Pro	Leu	Ile
			100					105					110		
Thr	Tyr	Leu	Ala	Pro	Cys	Asn	Gly	Asp	Cys	Ala	Thr	Val	Asp	Lys	Thr
			115				120					125			
Thr	Leu	Lys	Phe	Val	Lys	Ile	Ala	Ala	Gln	Gly	Leu	Ile	Asp	Gly	Ser
			130				135					140			
Asn	Pro	Pro	Gly	Val	Trp	Ala	Asp	Asp	Glu	Met	Ile	Ala	Asn	Asn	Asn
145				150						155				160	
Thr	Ala	Thr	Val	Thr	Ile	Pro	Ala	Ser	Tyr	Ala	Pro	Gly	Asn	Tyr	Val
				165						170				175	
Leu	Arg	His	Glu	Ile	Ile	Ala	Leu	His	Ser	Ala	Gly	Asn	Leu	Asn	Gly
			180					185					190		
Ala	Gln	Asn	Tyr	Pro	Gln	Cys	Phe	Asn	Ile	Gln	Ile	Thr	Gly	Gly	Gly
			195					200					205		

Ser Ala Gln Gly Ser Gly Thr Ala Gly Thr Ser Leu Tyr Lys Asn Thr
 210 215 220
 Asp Pro Gly Ile Lys Phe Asp Ile Tyr Ser Asp Leu Ser Gly Gly Tyr
 225 230 235 240
 Pro Ile Pro Gly Pro Ala Leu Phe Asn Ala
 245 250

<210> 37

<211> 37

<212> DNA

<213> 烟曲霉

<400> 37

ggactgcgca ccatgacttt gtccaagatc acttcca 37

<210> 38

<211> 38

<212> DNA

<213> 烟曲霉

<400> 38

gccacggagc ttaattaatt aagcgttgaa cagtgcag 38

<210> 39

<211> 37

<212> DNA

<213> 烟曲霉

<400> 39

cgcggtagtgc gcgcggctcga ccgaatgtag gattggtt 37

<210> 40

<211> 36

<212> DNA

<213> 烟曲霉

<400> 40

ttaccaattg gcgcgccact accgcgttcg agaaga 36

<210> 41

<211> 3060

<212> DNA

<213> 烟曲霉

<400> 41

atgagattcg gttggctcga ggtggccgct ctgacggccg cttctgtagc caatgccag 60
 gtttgtgatg ctttcccgctc attgtttcgg atatagtga caatagtcag ggaaataatc 120
 aggaattggc tttctctcca ccattctacc cttgccttg ggctgatggc caggagagat 180
 gggcagatgc ccatcgacgc gccgtcgaga tcgtttctca gatgacactg gcggagaagg 240

```

ttaaacccttac aacgggtact ggggtgggttg cgactttttt gttgacagtg agcttttcttc 300
actgaccatc tacacagatg ggaaatggac cgatgcgtcg gtcaaaccgg cagcgttccc 360
aggtaagctt gcaattctgc aacaacgtgc aagtgtagtt gctaaaacgc ggtggtgcag 420
acttgggtatc aactggggtc tttgtggcca ggattcccct ttgggtatcc gtttctgtga 480
gctatacccg cggagtcttt cagtccttgt attatgtgct gatgattgtc tctgtatagc 540
tgacctcaac tccgccttcc ctgctggtac taatgtcgcc gcgacatggg acaagacact 600
cgcctacctt cgtggcaagg ccatgggtga ggaattcaac gacaaggcg tggacatttt 660
gctggggcct gctgctggtc ctctcgga ataccggac ggcggcagaa tctgggaagg 720
cttctctcct gatccggttc tcaactggtg acttttcgcc gaaactatca agggatatcca 780
agacgcgggt gtgattgcta ctgccaagca ttacattctg aatgaacagg agcatttccg 840
acaggttggc gaggcccagg gatatggtta caacatcacg gagacgatca gctccaacgt 900
ggatgacaag accatgcacg agttgtacct ttggtgagta gttgacactg caaatgagga 960
ccttgattga tttgactgac ctggaatgca ggccctttgc agatgctgtg cgcggtaaga 1020
ttttccgtag acttgacctc gcgacgaaga aatcgctgac gaaccatcgt agctggcggt 1080
ggcgctgtca tgtgttccta caatcaaate aacaacagct acggttgtca aaacagtcaa 1140
actctcaaca agctcctcaa ggctgagctg ggcttccaag gcttcgtcat gactgactgg 1200
agcgtcacc acagcgggtg cggcgctgcc ctcgctgggt tggatatgtc gatgcctgga 1260
gacatttcct tcgacgacgg actctccttc tggggcacga acctaactgt cagtgttctt 1320
aacggcaccg ttccagcctg gcgtgtcgat gacatggctg ttcgtatcat gaccgcgtac 1380
tacaaggttg gtcgtgaccg tcttcgtatt cccctaact tcagctcctg gaccgggat 1440
gagtacggct gggagcattc tgctgtctcc gagggagcct ggaccaaggt gaacgacttc 1500
gtcaatgtgc agcgcagtca ctctcagatc atccgtgaga ttggtgccgc tagtacagt 1560
ctcttgaaga acacgggtgc tcttcctttg accggcaagg aggttaaagt ggggtgttctc 1620
ggatgaagacg ctggttccaa cccgtgggggt gctaacggct gccccgaccg cggctgtgat 1680
aacggcactc ttgctatggc ctggggtagt ggtactgcca acttccctta cttgtcacc 1740
cccgagcagg ctatccagcg agaggtcatc agcaacggcg gcaatgtctt tgctgtgact 1800
gataacgggg ctctcagcca gatggcagat gttgcatctc aatccagggt agtgccggct 1860
cttagaaaaa gaacgttctc tgaatgaagt ttttaacca ttgcgaacag cgtgtctttg 1920
gtgtttgtca acgccgactc tggagagggt ttcacagtgc tcgacggcaa cgagggtgac 1980
cgcaaaaaatc tcaactctgtg gaagaacggc gagggcgta ttgacactgt tgtcagccac 2040
tgcaacaaca cgatttgtgt tattcacagt gttgggcccc tcttgatcga ccggtggtat 2100
gataacccca acgtcactgc catcatctgg gccggttgc ccggtcagga gagggtgcaac 2160
tccctggtcg acgtgctcta tggccgcgtc aacccagcg ccaagacccc gttcacctgg 2220
ggcaagactc gggagtctta cggggctccc ttgctcaccg agcctaacaa tggcaatggt 2280
gctccccagg atgatttcaa cgagggcgtc ttcattgact accgtcactt tgacaagcgc 2340
aatgagaccc ccatttatga gtttggccat ggcttgagct acaccacctt tggttactct 2400
caccttcggg ttcaggccct caatagtctg agttcggcat atgtcccgac tagcggagag 2460
accaagcctg cgccaaccta tgggtgagatc ggtagtgccg ccgactacct gtatcccag 2520
ggtctcaaaa gaattaccaa gtttatttac ctttggtcga actcgaccga cctcgaggat 2580

```

tcttctgacg acccgaacta cggctgggag gactcggagt acattcccga aggcgctagg 2640
 gatgggtctc ctcaaccctt cctgaaggct ggcggcgctc ctggtggtaa ccctaccctt 2700
 tatcaggatc ttgttagggt gtcggccacc ataaccaaca ctggtaacgt cgccggttat 2760
 gaagtccttc aattggtgag tgaccgcgat gttccttgcg ttgcaatttg gctaactcgc 2820
 ttctagtatg tttcactggg cggaccgaac gagcctcggg tcgttctgcg caagttcgac 2880
 cgaatcttcc tggctcctgg ggagcaaaag gtttggacca cgactcttaa ccgtcgtgat 2940
 ctcgccaatt gggatgtgga ggctcaggac tgggtcatca caaagtaccc caagaaagtg 3000
 cacgtcggca gtcctcgcg taagctgcct ctgagagcgc ctctgccccg tgtctactag 3060

<210> 42

<211> 863

<212> PRT

<213> 烟曲霉

<400> 42

Met Arg Phe Gly Trp Leu Glu Val Ala Ala Leu Thr Ala Ala Ser Val
 1 5 10 15
 Ala Asn Ala Gln Glu Leu Ala Phe Ser Pro Pro Phe Tyr Pro Ser Pro
 20 25 30
 Trp Ala Asp Gly Gln Gly Glu Trp Ala Asp Ala His Arg Arg Ala Val
 35 40 45
 Glu Ile Val Ser Gln Met Thr Leu Ala Glu Lys Val Asn Leu Thr Thr
 50 55 60
 Gly Thr Gly Trp Glu Met Asp Arg Cys Val Gly Gln Thr Gly Ser Val
 65 70 75 80
 Pro Arg Leu Gly Ile Asn Trp Gly Leu Cys Gly Gln Asp Ser Pro Leu
 85 90 95
 Gly Ile Arg Phe Ser Asp Leu Asn Ser Ala Phe Pro Ala Gly Thr Asn
 100 105 110
 Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Lys Ala
 115 120 125
 Met Gly Glu Glu Phe Asn Asp Lys Gly Val Asp Ile Leu Leu Gly Pro
 130 135 140
 Ala Ala Gly Pro Leu Gly Lys Tyr Pro Asp Gly Gly Arg Ile Trp Glu
 145 150 155 160
 Gly Phe Ser Pro Asp Pro Val Leu Thr Gly Val Leu Phe Ala Glu Thr
 165 170 175
 Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
 180 185 190
 Ile Leu Asn Glu Gln Glu His Phe Arg Gln Val Gly Glu Ala Gln Gly
 195 200 205

Tyr Gly Tyr Asn Ile Thr Glu Thr Ile Ser Ser Asn Val Asp Asp Lys			
210	215	220	
Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala			
225	230	235	240
Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr			
	245	250	255
Gly Cys Gln Asn Ser Gln Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu			
	260	265	270
Gly Phe Gln Gly Phe Val Met Ser Asp Trp Ser Ala His His Ser Gly			
	275	280	285
Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Ile			
	290	295	300
Ser Phe Asp Asp Gly Leu Ser Phe Trp Gly Thr Asn Leu Thr Val Ser			
305	310	315	320
Val Leu Asn Gly Thr Val Pro Ala Trp Arg Val Asp Asp Met Ala Val			
	325	330	335
Arg Ile Met Thr Ala Tyr Tyr Lys Val Gly Arg Asp Arg Leu Arg Ile			
	340	345	350
Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Trp Glu His			
	355	360	365
Ser Ala Val Ser Glu Gly Ala Trp Thr Lys Val Asn Asp Phe Val Asn			
	370	375	380
Val Gln Arg Ser His Ser Gln Ile Ile Arg Glu Ile Gly Ala Ala Ser			
385	390	395	400
Thr Val Leu Leu Lys Asn Thr Gly Ala Leu Pro Leu Thr Gly Lys Glu			
	405	410	415
Val Lys Val Gly Val Leu Gly Glu Asp Ala Gly Ser Asn Pro Trp Gly			
	420	425	430
Ala Asn Gly Cys Pro Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met			
	435	440	445
Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu			
	450	455	460
Gln Ala Ile Gln Arg Glu Val Ile Ser Asn Gly Gly Asn Val Phe Ala			
465	470	475	480
Val Thr Asp Asn Gly Ala Leu Ser Gln Met Ala Asp Val Ala Ser Gln			
	485	490	495
Ser Ser Val Ser Leu Val Phe Val Asn Ala Asp Ser Gly Glu Gly Phe			
	500	505	510
Ile Ser Val Asp Gly Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp			

515	520	525
Lys Asn Gly Glu Ala Val Ile Asp Thr Val Val Ser His Cys Asn Asn		
530	535	540
Thr Ile Val Val Ile His Ser Val Gly Pro Val Leu Ile Asp Arg Trp		
545	550	555
Tyr Asp Asn Pro Asn Val Thr Ala Ile Ile Trp Ala Gly Leu Pro Gly		
565	570	575
Gln Glu Ser Gly Asn Ser Leu Val Asp Val Leu Tyr Gly Arg Val Asn		
580	585	590
Pro Ser Ala Lys Thr Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr		
595	600	605
Gly Ala Pro Leu Leu Thr Glu Pro Asn Asn Gly Asn Gly Ala Pro Gln		
610	615	620
Asp Asp Phe Asn Glu Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys		
625	630	635
Arg Asn Glu Thr Pro Ile Tyr Glu Phe Gly His Gly Leu Ser Tyr Thr		
645	650	655
Thr Phe Gly Tyr Ser His Leu Arg Val Gln Ala Leu Asn Ser Ser Ser		
660	665	670
Ser Ala Tyr Val Pro Thr Ser Gly Glu Thr Lys Pro Ala Pro Thr Tyr		
675	680	685
Gly Glu Ile Gly Ser Ala Ala Asp Tyr Leu Tyr Pro Glu Gly Leu Lys		
690	695	700
Arg Ile Thr Lys Phe Ile Tyr Pro Trp Leu Asn Ser Thr Asp Leu Glu		
705	710	715
Asp Ser Ser Asp Asp Pro Asn Tyr Gly Trp Glu Asp Ser Glu Tyr Ile		
725	730	735
Pro Glu Gly Ala Arg Asp Gly Ser Pro Gln Pro Leu Leu Lys Ala Gly		
740	745	750
Gly Ala Pro Gly Gly Asn Pro Thr Leu Tyr Gln Asp Leu Val Arg Val		
755	760	765
Ser Ala Thr Ile Thr Asn Thr Gly Asn Val Ala Gly Tyr Glu Val Pro		
770	775	780
Gln Leu Tyr Val Ser Leu Gly Gly Pro Asn Glu Pro Arg Val Val Leu		
785	790	795
Arg Lys Phe Asp Arg Ile Phe Leu Ala Pro Gly Glu Gln Lys Val Trp		
805	810	815
Thr Thr Thr Leu Asn Arg Arg Asp Leu Ala Asn Trp Asp Val Glu Ala		
820	825	830

Gln Asp Trp Val Ile Thr Lys Tyr Pro Lys Lys Val His Val Gly Ser
835 840 845

Ser Ser Arg Lys Leu Pro Leu Arg Ala Pro Leu Pro Arg Val Tyr
850 855 860

<210> 43

<211> 835

<212> DNA

<213> 青霉(Penicillium sp.)

<400> 43

atgctgtctt cgacgactcg caccctcgcc ttacaggcc ttgcgggcct tctgtccgct 60
cccctggtca aggcccatgg ctttgtccag ggcattgtca tcggtgacca attgtaagtc 120
cctctcttgc agttctgtcg attaaactgct ggactgcttg cttgactccc tgctgactcc 180
caacagctac agcgggtaca tcgtcaactc gtteccctac gaatccaacc cccccccgt 240
catcggtcgg gccacgaccg ccaccgacct gggcttcgtc gacggcacag gataccaagg 300
cccggacatc atctgccacc ggaatgcgac gcccgcgcgc ctgacagccc ccgtggccgc 360
cggcggcacc gtcgagctgc agtggacgcc gtggccggac agccaccacg gaccgcgtcat 420
cacctacctg gcgccgtgca acggcaactg ctcgaccgtc gacaagacga cgctggagtt 480
cttcaagatc gaccagcagg gcctgatcga cgacacgagc ccgccgggca cctgggcgtc 540
ggacaacctc atcgccaaca acaatagctg gaccgtcacc attcccaaca gcgtcgcccc 600
cggcaactac gtcttgcgcc acgagatcat cgccctgcac tcggccaaca acaaggacgg 660
cgcccagaac taccgccagt gcatcaacat cgaggtcacg ggcggcggct ccgacgcgcc 720
tgagggtact ctgggcgagg atctctacca tgacaccgac ccgggcattc tggtcgacat 780
ttacgagccc attgcgacgt ataccattcc ggggcgcgcct gagccgacgt tctag 835

<210> 44

<211> 253

<212> PRT

<213> 青霉

<400> 44

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

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

45

⟨211⟩ 36

⟨212⟩ DNA

〈213〉 青霉

<400> 45

cggactgcgc accatgctgt cttcgacgac tcgcac 36

⟨210⟩ 46

<211> 35

⟨212⟩ DNA

〈213〉 青霉

<400> 46

tgcgcacgga gcttatcgac ttctttctaga acgtc 35

<210> 47

<211> 37

⟨212⟩ DNA

〈213〉 烟曲霉

<400> 47

cgcggttagtg ggcggttcga ccgaatgtag gattggtt 37

<210> 48

<211> 36

<212> DNA

<213> 烟曲霉

<400> 48

ttaccaattg gcgcgccact accgcgttcg agaaga 36

<210> 49

<211> 1713

<212> DNA

<213> 烟曲霉

<400> 49

atgaagcacc ttgcatcttc catcgcatcg actctactgt tgcctgccgt gcaggcccag 60
 cagaccgtat ggggccaatg tatgttcttg ctgtcactgg aataagactg tatcaactgc 120
 tgatatgctt ctaggtggcg gccaaagctg gtctggcccc acgagctgtg ttgccggcgc 180
 agcctgtagc aactgaatc cctgtatgtt agatatcgct ctgagtggag atttatactg 240
 acttccttag actacgtca gtgtatcccc ggagccaccg cgacgtccac caccctcacg 300
 acgacgacgg cggcgacgac gacatcccag accaccacca aacctaccac gactggtcca 360
 actacatccg caccaccgt gaccgcatcc ggtaaccctt tcagcggcta ccagctgtat 420
 gccaacccct actactcctc cgaggtccat actctggcca tgccttctct gccagctcg 480
 ctgcagccca aggctagtgc tgttgctgaa gtgccctcat ttgtttggct gtaagtggcc 540
 ttatcccaat actgagacca actctctgac agtcgtagcg acgttgccgc caaggtgccc 600
 actatgggaa cctacctggc cgacattcag gccaagaaca aggccggcgc caaccctcct 660
 atcgctggta tcttcgtggt ctacgacttg ccggaccgtg actgcgccgc tctggccagt 720
 aatggcgagt actcaattgc caacaacggt gtggccaact acaaggcgta cattgacgcc 780
 atccgtgctc agctggtgaa gtactctgac gttcacacca tcctcgatcat cggtaggccg 840
 tacacctccg ttgcgcgccg cttttctctg acatcttgca gaaccgaca gcttgccaa 900
 cctggtgacc aacctcaacg tcgccaaatg cgccaatgcg cagagcgctt acctggagt 960
 tgtcgactat gctctgaagc agtcaacct gcccaacgtc gccatgtacc tcgacgcagg 1020
 tatgcctcac ttcccgcat ctgtatccct tccagacact aactcatcag gccatgcggg 1080
 ctggctcgga tggcccgcca acttggggcc cgccgcaaca ctcttcgcca aagtctacac 1140
 cgacgcgggt tccccgcgg ctgttcgtgg cctggccacc aacgtcgcca actacaacgc 1200
 ctggtcgctc agtacctgcc cctcctacac ccaggagac cccaactgcg acgagaagaa 1260
 gtacatcaac gccatggcgc ctcttctcaa ggaagccggc ttcgatgccc acttcatcat 1320
 ggatacctgt aagtgttat tccaatcgcc gatgtgtgcc gactaatcaa tgtttcagcc 1380
 cggaatggcg tccagccac gaagcaaac gcctggggtg actggtgcaa cgtcatcggc 1440
 accgcttcg gtgttcgccc ctcgactaac accggcgatc cgctccagga tgcctttgtg 1500
 tggatcaagc ccggtggaga gagtgatggc acgtccaact cgacttcccc ccggtatgac 1560
 gcgcaactgc gatatagtga tgctctgcag cctgctcctg aggctggtac ttggttccag 1620
 gtatgtcatc cattagccag atgagggata agtgactgac ggacctaggc ctactttgag 1680

cagcttctga ccaacgctaa cccgtccttt taa 1713

<210> 50

<211> 454

<212> PRT

<213> 烟曲霉

<400> 50

Met	Lys	His	Leu	Ala	Ser	Ser	Ile	Ala	Leu	Thr	Leu	Leu	Leu	Pro	Ala
1				5					10					15	
Val	Gln	Ala	Gln	Gln	Thr	Val	Trp	Gly	Gln	Cys	Gly	Gly	Gln	Gly	Trp
			20					25					30		
Ser	Gly	Pro	Thr	Ser	Cys	Val	Ala	Gly	Ala	Ala	Cys	Ser	Thr	Leu	Asn
		35					40					45			
Pro	Tyr	Tyr	Ala	Gln	Cys	Ile	Pro	Gly	Ala	Thr	Ala	Thr	Ser	Thr	Thr
	50					55					60				
Leu	Thr	Thr	Thr	Thr	Ala	Ala	Thr	Thr	Thr	Ser	Gln	Thr	Thr	Thr	Lys
65					70					75					80
Pro	Thr	Thr	Thr	Gly	Pro	Thr	Thr	Ser	Ala	Pro	Thr	Val	Thr	Ala	Ser
				85					90					95	
Gly	Asn	Pro	Phe	Ser	Gly	Tyr	Gln	Leu	Tyr	Ala	Asn	Pro	Tyr	Tyr	Ser
		100						105					110		
Ser	Glu	Val	His	Thr	Leu	Ala	Met	Pro	Ser	Leu	Pro	Ser	Ser	Leu	Gln
		115					120						125		
Pro	Lys	Ala	Ser	Ala	Val	Ala	Glu	Val	Pro	Ser	Phe	Val	Trp	Leu	Asp
	130					135					140				
Val	Ala	Ala	Lys	Val	Pro	Thr	Met	Gly	Thr	Tyr	Leu	Ala	Asp	Ile	Gln
145				150						155				160	
Ala	Lys	Asn	Lys	Ala	Gly	Ala	Asn	Pro	Pro	Ile	Ala	Gly	Ile	Phe	Val
			165						170					175	
Val	Tyr	Asp	Leu	Pro	Asp	Arg	Asp	Cys	Ala	Ala	Leu	Ala	Ser	Asn	Gly
		180						185					190		
Glu	Tyr	Ser	Ile	Ala	Asn	Asn	Gly	Val	Ala	Asn	Tyr	Lys	Ala	Tyr	Ile
	195						200					205			
Asp	Ala	Ile	Arg	Ala	Gln	Leu	Val	Lys	Tyr	Ser	Asp	Val	His	Thr	Ile
	210					215					220				
Leu	Val	Ile	Glu	Pro	Asp	Ser	Leu	Ala	Asn	Leu	Val	Thr	Asn	Leu	Asn
225				230						235				240	
Val	Ala	Lys	Cys	Ala	Asn	Ala	Gln	Ser	Ala	Tyr	Leu	Glu	Cys	Val	Asp
			245						250					255	
Tyr	Ala	Leu	Lys	Gln	Leu	Asn	Leu	Pro	Asn	Val	Ala	Met	Tyr	Leu	Asp

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

450

<210> 51

<211> 54

<212> DNA

<213> 里氏木霉

<400> 51

acgaattgtt taaacgtcga cccaagtatc cagaggtgta tggaaatatc agat 54

<210> 52

<211> 44

<212> DNA

<213> 里氏木霉

<400> 52

cgcgtagatc tgcggccatg gtgcaatata cagagggtga tctt 44

<210> 53

<211> 48

<212> DNA
<213> 里氏木霉
<400> 53
atctacgcgt actagttaat taaggctttc gtgaccgggc ttcaaaca 48
<210> 54
<211> 46
<212> DNA
<213> 里氏木霉
<400> 54
gcggccgtta ctagtggatc cactcggagt tgttatacgc tactcg 46
<210> 55
<211> 46
<212> DNA
<213> 里氏木霉
<400> 55
atccatcaca ctggcggccg cgcttcaaac aatgatgtgc gatggt 46
<210> 56
<211> 46
<212> DNA
<213> 里氏木霉
<400> 56
gatgcatgct cgagcggccg cctaccttgg cagccctacg agagag 46
<210> 57
<211> 41
<212> DNA
<213> 烟曲霉
<400> 57
ctctgtgtat tgcacatga agcaccttgc atcttccatc g 41
<210> 58
<211> 40
<212> DNA
<213> 烟曲霉
<400> 58
ccggtcacga aagccttaat taaaaggacg ggtagcggtt 40
<210> 59
<211> 1599
<212> DNA
<213> 烟曲霉
<400> 59

atgctggcct ccaccttctc ctaccgcatg tacaagaccg cgctcatcct ggccgcccctt 60
 ctgggctctg gccaggctca gcaggtcggt acttcccagg cggaagtgca tccgtccatg 120
 acctggcaga gctgcacggc tggcggcagc tgcaccacca acaacggcaa ggtggtcatc 180
 gacgcgaact ggcgttgggt gcacaaagtc ggcgactaca ccaactgcta caccggcaac 240
 acctgggaca cgactatctg ccctgacgat gcgacctgcg catccaactg cgcccttgag 300
 ggtgccaact acgaatccac ctatggtgtg accgccagcg gcaattccct ccgcctcaac 360
 ttcgtcacca ccagccagca gaagaacatt ggctcgcgtc tgtacatgat gaaggacgac 420
 tcgacctacg agatgtttaa gctgctgaac caggagtcca ccttcgatgt cgatgtctcc 480
 aacctcccct gcggtctcaa cggtgctctg tactttgtcg ccatggacgc cgacgggtggc 540
 atgtccaagt acccaaccaa caaggccggt gccaaagtac gtactggata ctgtgactcg 600
 cagtgccctc gcgacctcaa gttcatcaac ggtcaggcca acgtcgaagg gtggcagccc 660
 tcctccaacg atgccaatgc gggtagcggc aaccacgggt cctgctgcgc ggagatggat 720
 atctgggagg ccaacagcat ctccacggcc ttcaccccc atccgtgcca cacgcccggc 780
 caggatgatgt gcaccggtga tgcctgcggt ggcacctaca gctccgaccg ctacggcggc 840
 acctgcgacc ccgacggatg tgatttcaac tccttcgcgc agggcaacaa gacattctac 900
 ggccctggca tgaccgtcga caccaagagc aagtttaccg tcgtcaccca gttcatcacc 960
 gacgacggca cctccagcgg caccctcaag gagatcaagc gcttctacgt gcagaacggc 1020
 aaggatgatcc ccaactcgga gtcgacctgg accggcgcca gcggcaactc catcaccacc 1080
 gagtactgca ccgcccagaa gagcctgttc caggaccaga acgtcttcga aaagcacggc 1140
 ggccctcgagg gcatgggtgc tgccctcgcc cagggtatgg ttctcgtcat gtccctgtgg 1200
 gatgatcact cggccaacat gctctggctc gacagcaact acccgaccac tgcctcttcc 1260
 accactcccc gcgtcgcccc tggtagctgc gacatctcct ccggcgctccc tgcggatgtc 1320
 gaggcgaacc accccgacgc ctacgtcgtc tactccaaca tcaaggtcgg ccccatcggc 1380
 tcgaccttca acagcgggtg ctcgaacccc ggtggcgga ccaccacgac aactaccacc 1440
 cagcctacta ccaccacgac cacggctgga aaccctggcg gcaccggagt cgcacagcac 1500
 tatggccagt gtggtggaat cggatggacc ggaccacaa cctgtgccag cccttatacc 1560
 tgccagaagc tgaatgatta ttactctcag tgccctgtag 1599

<210> 60

<211> 532

<212> PRT

<213> 烟曲霉

<400> 60

Met Leu Ala Ser Thr Phe Ser Tyr Arg Met Tyr Lys Thr Ala Leu Ile

1 5 10 15

Leu Ala Ala Leu Leu Gly Ser Gly Gln Ala Gln Gln Val Gly Thr Ser

20 25 30

Gln Ala Glu Val His Pro Ser Met Thr Trp Gln Ser Cys Thr Ala Gly

35 40 45

Gly Ser Cys Thr Thr Asn Asn Gly Lys Val Val Ile Asp Ala Asn Trp

50	55	60
Arg Trp Val His Lys	Val Gly Asp Tyr Thr	Asn Cys Tyr Thr Gly Asn
65	70	75
Thr Trp Asp Thr Thr	Ile Cys Pro Asp Asp	Ala Thr Cys Ala Ser Asn
85	90	95
Cys Ala Leu Glu Gly	Ala Asn Tyr Glu Ser Thr	Tyr Gly Val Thr Ala
100	105	110
Ser Gly Asn Ser Leu Arg	Leu Asn Phe Val Thr Thr	Ser Gln Gln Lys
115	120	125
Asn Ile Gly Ser Arg Leu	Tyr Met Met Lys Asp Asp	Ser Thr Tyr Glu
130	135	140
Met Phe Lys Leu Leu Asn	Gln Glu Phe Thr Phe	Asp Val Asp Val Ser
145	150	155
Asn Leu Pro Cys Gly Leu	Asn Gly Ala Leu Tyr Phe	Val Ala Met Asp
165	170	175
Ala Asp Gly Gly Met Ser	Lys Tyr Pro Thr Asn Lys	Ala Gly Ala Lys
180	185	190
Tyr Gly Thr Gly Tyr Cys	Asp Ser Gln Cys Pro Arg	Asp Leu Lys Phe
195	200	205
Ile Asn Gly Gln Ala Asn	Val Glu Gly Trp Gln Pro	Ser Ser Asn Asp
210	215	220
Ala Asn Ala Gly Thr Gly	Asn His Gly Ser Cys Cys	Ala Glu Met Asp
225	230	235
Ile Trp Glu Ala Asn Ser	Ile Ser Thr Ala Phe Thr	Pro His Pro Cys
245	250	255
Asp Thr Pro Gly Gln Val	Met Cys Thr Gly Asp Ala	Cys Gly Gly Thr
260	265	270
Tyr Ser Ser Asp Arg Tyr	Gly Gly Thr Cys Asp Pro	Asp Gly Cys Asp
275	280	285
Phe Asn Ser Phe Arg Gln	Gly Asn Lys Thr Phe Tyr	Gly Pro Gly Met
290	295	300
Thr Val Asp Thr Lys Ser	Lys Phe Thr Val Val Thr	Gln Phe Ile Thr
305	310	315
Asp Asp Gly Thr Ser Ser	Gly Thr Leu Lys Glu Ile	Lys Arg Phe Tyr
325	330	335
Val Gln Asn Gly Lys Val	Ile Pro Asn Ser Glu Ser	Thr Trp Thr Gly
340	345	350
Val Ser Gly Asn Ser Ile	Thr Thr Glu Tyr Cys Thr	Ala Gln Lys Ser
355	360	365

Leu Phe Gln Asp Gln Asn Val Phe Glu Lys His Gly Gly Leu Glu Gly
 370 375 380
 Met Gly Ala Ala Leu Ala Gln Gly Met Val Leu Val Met Ser Leu Trp
 385 390 395 400
 Asp Asp His Ser Ala Asn Met Leu Trp Leu Asp Ser Asn Tyr Pro Thr
 405 410 415
 Thr Ala Ser Ser Thr Thr Pro Gly Val Ala Arg Gly Thr Cys Asp Ile
 420 425 430
 Ser Ser Gly Val Pro Ala Asp Val Glu Ala Asn His Pro Asp Ala Tyr
 435 440 445
 Val Val Tyr Ser Asn Ile Lys Val Gly Pro Ile Gly Ser Thr Phe Asn
 450 455 460
 Ser Gly Gly Ser Asn Pro Gly Gly Gly Thr Thr Thr Thr Thr Thr Thr
 465 470 475 480
 Gln Pro Thr Thr Thr Thr Thr Thr Ala Gly Asn Pro Gly Gly Thr Gly
 485 490 495
 Val Ala Gln His Tyr Gly Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro
 500 505 510
 Thr Thr Cys Ala Ser Pro Tyr Thr Cys Gln Lys Leu Asn Asp Tyr Tyr
 515 520 525
 Ser Gln Cys Leu
 530
 <210> 61
 <211> 40
 <212> DNA
 <213> 烟曲霉
 <400> 61
 cgcggaactgc gcaccatgct ggcctccacc ttctcctacc 40
 <210> 62
 <211> 48
 <212> DNA
 <213> 烟曲霉
 <400> 62
 ctttcgccac ggagcttaat taactacagg cactgagagt aataatca 48
 <210> 63
 <211> 1131
 <212> DNA
 <213> 单纯疱疹病毒
 <400> 63

atggcttcgt accccggcca tcaacacgcg tctgcgttcg accaggctgc gcgttctcgc 60
 ggccatagca accgacgtac ggcgttgccg cctcgccggc agcaagaagc cacggaagtc 120
 cgccccgagc agaaaatgcc cacgctactg cgggtttata tagacgttcc ccacgggatg 180
 gggaaaacca ccaccacgca actgctggtg gccctgggtt cgcgcgacga tatcgtctac 240
 gtacccgagc cgatgactta ctggcgggtg ctgggggctt ccgagacaat cgcgaaacatc 300
 tacaccacac aacaccgcct cgaccagggt gagatatcgg ccgggggacgc ggcggtggta 360
 atgacaagcg cccagataac aatgggcatg ctttatgccg tgaccgacgc cgttctggct 420
 cctcatatcg ggggggaggc tgggagctca catgccccgc cccgggccct caccctcatc 480
 ttcgaccgcc atcccatcgc cgccctcctg tgctaccggc ccgcgcggta ctttatgggc 540
 agcatgaccc cccaggccgt gctggcggtc gtggccctca tcccgcgcac cttgcccggc 600
 accaacatcg tgcttggggc cttccggag gacagacaca tcgaccgcct ggccaaacgc 660
 cagcgccccg gcgagcggtt ggacctggtt atgctggctg cgattcgccg cgtttacggg 720
 ctacttgcca atacggtgcg gtatctgcag tgcggcgggt cgtggcggga ggactgggga 780
 cagctttcgg ggacggccgt gccgccccag ggtgccgagc cccagagcaa cgcgggcccc 840
 cgaccccata tcggggacac gttatttacc ctgtttcggg gccccagatt gctggcccc 900
 aacggcgacc tgtataacgt gtttgcttg gccttgacg tcttgccaa acgcctccgt 960
 tccatgcacg tctttatcct ggattacgac caatcgccc cgggtgccg ggacgccttg 1020
 ctgcaactta cctccgggat ggtccagacc cacgtacca cccccgctc cataccgacg 1080
 atatgcgacc tggcgcgcac gtttgcccgg gagatggggg aggctaactg a 1131

<210> 64

<211> 376

<212> PRT

<213> 单纯疱疹病毒

<400> 64

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

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

<210> 65

<211> 43

<212> DNA

<213> 里氏木霉

<400> 65

ttagactgcg gccgcgtggc gaaagcctga cgcaccgcta gat 43

<210> 66
<211> 41
<212> DNA
<213> 里氏木霉
<400> 66
agtagttagc ggccgcacgg cacggttaag cagggtcttg c 41
<210> 67
<211> 50
<212> DNA
<213> 烟曲霉
<400> 67
tcaagcttgg taccgagctc ggatccaagt atccagaggt gtatggaaat 50
<210> 68
<211> 46
<212> DNA
<213> 烟曲霉
<400> 68
ctggcggccg ttactagtgc tagcactcgg agttgttata cgctac 46
<210> 69
<211> 55
<212> DNA
<213> 里氏木霉
<400> 69
gagtagcgta taacaactcc gagtgctagc ttttaagataa cggaatagaa gaaag 55
<210> 70
<211> 42
<212> DNA
<213> 里氏木霉
<400> 70
ctggcggccg ttactagtct agacgcgcca ctaccgcgtt cg 42
<210> 71
<211> 55
<212> DNA
<213> 里氏木霉
<400> 71
tctgcagata tccatcacac tggcggccgc ttttaagataa cggaatagaa gaaag 55
<210> 72
<211> 50
<212> DNA

<213> 里氏木霉
<400> 72
aaactctagg atgcatgctc gagcggccgc acggcacggt taagcagggt 50
<210> 73
<211> 30
<212> DNA
<213> 里氏木霉
<400> 73
caagcaaagc gttccgtcgc agtagcaggc 30
<210> 74
<211> 20
<212> DNA
<213> 里氏木霉
<400> 74
cagtggcgct tattactcag 20
<210> 75
<211> 22
<212> DNA
<213> 里氏木霉
<400> 75
gagaacacag tgagaccata gc 22
<210> 76
<211> 22
<212> DNA
<213> 里氏木霉
<400> 76
tctcaaccca atcagcaaca tg 22
<210> 77
<211> 15014
<212> DNA
<213> 里氏木霉
<400> 77
tcgacgaatt ctaggctagg tatgcgaggc acgcggatct agggcagact gggcattgca 60
agctgcttaa gatccgatcc atacgtccg tgcgcctaga tcccgtctga cccgtaacgt 120
tagctatggg gtagtagaac tcccgtcaac ggctattctc acctagactt tccccttcga 180
atcgatacca catcatcttg agggcagttg ccgataagag tggatctgaa aggggaagct 240
actgacaagt tgttatattg cctgtgtacc aagcgcta atgtggacagga ttaatgccag 300
tgactgttca acaatataac ggacacatgg ttcgcgatta cacctgtcct aattacggtc 360
agttcattag cctcaagtag agcctatttc ctgcgccgaa agtcatctct cttattgcat 420

tcaagtaatc ggagttcatc tcggataaag gagcggcctt tcagtagaga gaataacgta 480
 ttctgccctt cccactaact caggggtgcag cgcaacacta cacgcaacat atacacttta 540
 aagacgggaa ggggtgattga gtcccacgtc gcgttgatgat gtgcgttgta tatgtgaaat 600
 ttagccgtgc aacaaggcta ttctacgaaa aatgctacac tccacatgtt aaaggcgcgt 660
 aatcggcacg ttgttccgat aagatgcttt ttacgatgtg aggtgtacaa tttccgcgta 720
 tcaaccagct tctttattgg gtaatatata gccaggcggg gatgaagctc attagccgcc 780
 agttggtcga agaaataacc cattatatgt cggtccgccc ctacttcgag taatcggcgg 840
 actcaaggct atacaatgtt gccaactctc cgggctttat cctgtgctcc cgaataccac 900
 tgagttccga tatgttacaa cggttgagag gcccgaaata ggacacgagg gcttatgggtg 960
 atcgtgatga tgcttcagcg cacggaagtc acagacaccg cctgtataaa agggggactg 1020
 tagcactact acgaagtcgc gtgccttcag tgtctgtggc ggacatattt tccccctgac 1080
 tgaccctgta tgaggcgcaa catggtctca cagcagctca cctgaagagg cttgtaagat 1140
 actgggacat actccgcgtt gtaccagagt gtcgtcagat ggacttctcc gaacattcta 1200
 caccctctgt gtattgcacc atggcgttcc tccccctcc tccgctccgc cgttgtggcc 1260
 gtgggagaca cataacgtgg taccgcaagg aggggggagg aggcgaggcg gcaacaccgg 1320
 gccctgccgg tgttgccct tgccgtgat ggcaggtcca cccgctactg ggactgctgc 1380
 cgggacggcc acaaccggga acggcgacta ccgtccaggt gggcgatgac cctgacgacg 1440
 aagccttcgt gcggctgggc caagaaggct cccgtgaacc agcctgtctt ttctgcaac 1500
 ttcggaagca cgccgaccg gttcttccga gggcacttgg tcggacagaa aaggacgttg 1560
 gccaacttcc agcgtatcac ggacttcgac gccaagtccg gctgcgagcc gggcgggtgtc 1620
 cggttgaagg tcgcatagt cctgaagctg cggttcaggc cgacgctcgg cccgccacag 1680
 gcctactcgt gcgccacca gaccccatgg gctgtgaacg acgacttcgc gctcggtttt 1740
 cggatgagca cgcggtggt ctggggtacc cgacacttgc tgctgaagcg cgagccaaaa 1800
 gctgccacct ctattgccgg cagcaatgag gcgggctggt gctgcgcctg ctacgagctc 1860
 cgacgggtgga gataacggcc gtcgttactc cgcccgacca cgacgcggac gatgctcgag 1920
 accttcacat ccggtcctgt tgctggcaag aagatggctg tccagtccac cagcactggc 1980
 tggaagtgtg ggccaggaca acgaccgtt ttctaccagc aggtcaggtg gtcgtgaccg 2040
 ggtgatcttg gcagcaacca cttcgatctc aacatccccg gcggcggcgt cggcatcttc 2100
 ccactagaac cgtcgttggt gaagctagag ttgtaggggc cgccgccgca gccgtagaag 2160
 gacggatgca ctccccagtt cggcggtctg cccggccagc gctacggcgg catctcgtcc 2220
 ctgcctacgt gaggggtcaa gccgccagac gggccggctc gcatgccgc gtagagcagg 2280
 cgcaacgagt gcgatcggt ccccgacgcc ctcaagcccc gctgctactg gcgcttcgac 2340
 gcgttgctca cgctagccaa ggggctgcgg gagttcgggc cgacgatgac cgcgaagctg 2400
 tggttcaaga acgccgacaa tccgagcttc agcttcctgc aggtccagt cccagccgag 2460
 accaagttct tgcggctgtt aggtcgaag tcgaaggcag tccaggtcac gggtcggctc 2520
 ctcgtcgtc gcaccggatg ccgccgcaac gacgacggca acttccctgc cgtccagatc 2580
 gagcagcgag cgtggcctac ggcggcgttg ctgctgccgt tgaagggacg gcaggtctag 2640
 cctccagca gcaccagctc tccggtcaac cagcctacca gcaccagcac cacgtccacc 2700
 gggaggctcgt cgtggtcgag aggccagttg gtcggatggt cgtggtcgtg gtgcaggtgg 2760

```

tccaccacct cgagcccgcc agtccagcct acgactccca gcggtgcac tgctgagagg 2820
aggtgggtgga gctcgggcgg tcaggtcgga tgctgagggt cgccgacgtg acgactctcc 2880
tgggctcagt gcggcggcaa tggctggagc ggctgcacca cctgcgtcgc tggcagcact 2940
acccgagtca cgccgccgtt accgacctcg ccgacgtggt ggacgcagcg accgtcgtga 3000
tgcacgaaga ttaatgactg gtaccatcag tgcctgtaga attcgcggcc gcagatctac 3060
acgtgcttct aattactgac catggtagtc acggacatct taagcgccgg cgtctagatg 3120
gcgtactagt agctccgtgg cgaaagcctg acgcaccggt agattcttgg tgagcccgtg 3180
cgcatgatca tcgaggcacc gctttcggac tgcgtggcca tctaagaacc actcgggcat 3240
tcatgacggc ggccgggagct acatggcccc gggtgattta ttttttttgt atctacttct 3300
agtactgccg ccgccctcga tgtaccgggg ccactaaat aaaaaaaca tagatgaaga 3360
gaccttttcc aaatatacgg tcaactcatc tttcactgga gatgcggcct gcttgggtatt 3420
ctgggaaaag tttatatgcc agttgagtag aaagtacct ctacgccgga cgaaccataa 3480
gcgatgttgt cagcttgcca aattgtggct ttcgaaaaca caaacgatt ccttaattaa 3540
cgctacaaca gtcgaaccgt ttaacaccga aagcttttgt gttttgctaa ggaattaatt 3600
ctagagggtg ctgacacctg gcggtagaca atcaatccat ttcgctatag ttaaaggatg 3660
gatctccact gactgtggac cgccatctgt tagttaggtg aagcgatata aatttcctac 3720
gggatgaggg caattgggta tatgatcatg tatgtagtgg gtgtgcataa tagtagtgaa 3780
ccctactccc gttaaccaat atactagtac atacatcacc cacacgtatt atcatcactt 3840
atggaagcca agtcatgtga ttgtaatcga ccgacggaat tgaggatata cggaataaca 3900
taccttcggt tcagtacact aacattagct ggctgcctta actcctatag gcctttatgt 3960
gacaccgtga aagccatgct ctttccttcg tgtagaagac cagacagaca gtccctgatt 4020
ctgtggcact ttcggtacga gaaaggaagc acatcttctg gtctgtctgt cagggactaa 4080
taccttgca caaagcacta gaaaattagc attccatcct tctctgcttg ctctgctgat 4140
atgggaacgt gtttcgtgat cttttaatcg taaggtagga agagacgaac gagacgacta 4200
atcactgtca ttcaatgcat ctggaaacgc aaccctgaag ggattcttcc tttgagagat 4260
tagtgacagt aagttacgta gacctttgcg ttgggacttc cctaagaagg aaactctcta 4320
ggaagcgtgt catatctctt cggttctacg gcaggttttt ttctgctctt tcgtagcatg 4380
ccttcgcaca gtatagagaa gccaagatgc cgtccaaaaa aagacgagaa agcatcgtac 4440
gcatgggtcac ttcagcgctt atttacagtt gctgggtattg atttcttggtg caaattgcta 4500
cgtaccagtg aagtcgcgaa taaatgtcaa cgaccataac taaagaacac gtttaacgat 4560
tctgacactt attagctatg gagtaccac atttcccagc aacttcccca cttcctctgc 4620
agactgtgaa taatcgatac ctcagtgggtg taaagggtcg ttgaaggggt gaaggagacg 4680
aatcgccaac gtcctctctt cactgagtct ccgtccgata acctgcactg caaccggtgc 4740
ttagcggttg caggagagaa gtgactcaga ggcaggctat tggacgtgac gttggccacg 4800
cccatgatac gcctccggat catactcttc ctgcacgagg gcatcaagct cactaaccgc 4860
gggtactatg cggaggccta gtatgagaag gacgtgctcc cgtagttcga gtgattggcg 4920
cttgaaactc tcattcttct tatcgatgtt cttatccgca aaghtaaccg gaacaaccac 4980
gaactttgag agtaagaaga atagctacaa gaataggcgt ttccattggc cttgttggtg 5040
gctcgtgaaa tccagcaggt tgatcacaga ggcataccca tagtaccgga actggtcatg 5100

```

cgagcacttt aggtcgtcca actagtgtct ccgtatgggt atcatggcct tgaccagtac 5160
 ccgtaccgca gcggtaggcg taatcggcgc gatgatggcg tccagttcct tcccggcctt 5220
 ggcatggcgt cgccatccgc attagccgcg ctactaccgc aggtcaagga agggccggaa 5280
 ttcttcagcc tcccgccatt tctcaaggta ctccatctgg taattccact tctggagatg 5340
 aagaagtcgg agggcggtaa agagttccat gaggtagacc attaaggtga agacctctac 5400
 cgtgtcccag agctcgttca tgtaaacagc tttgatgttc gggttcagta ggtctttgat 5460
 gcacagggtc tcgagcaagt acaattgtcg aaactacaag cccaagtcac ccagaaacta 5520
 atttgagtc gccggctcgc cggatgcact gatatcgcg attacgtcgg cgctgccgtc 5580
 taaacctcag cgcccgagcg gcctacgtga ctatagcgcg taatgcagcc gcgacggcag 5640
 agccgcgtag atatgggaga tgagatcgtg gccgaaatcg tgcttgtagt gcgtccacgg 5700
 tcggcgcac tataccctct actctagcac cggcttttag acgaacatac cgcaggtgcc 5760
 ggtcacggtg tgaccggctt tggcagatgc ggcgacggtg gtttccacgc cgcgcaggat 5820
 ccagtgccac actggccgaa accgtcacg ccgtgccac caaaggtcgc gcgcgtccta 5880
 aggagggtgt ggaaggacat tgccgtcgaa gttgtagtag ccgatattga gcccgcctt 5940
 tcctcccaca ccttcctgta acggcagctt caacatcac ggctataact cgggcggcaa 6000
 cttgatcttg gaggcaataa tgtccgactc ggactggcg cagggcattg ggatgacctt 6060
 gaactagaac ctccgttatt acaggtgag cctgaccgcg gtcccgtacc cctactggaa 6120
 ggagtcgtat ttccaaggct cctgaccgag gacggatttg gtgaagaggc ggaggtctaa 6180
 cctcagcata aaggttccga ggactggctc ctgcctaaac cacttctccg cctccagatt 6240
 catacttcat cagtgactgc cgtctcgta tatagtataa aaagcaagaa aggaggacag 6300
 gtatgaagta gtcactgacg gccagagcat atatcatatt tttcgttctt tcctcctgtc 6360
 tggaggcctg gtatagagca ggaaaagaag gaagaggcga aggactcacc ctcaacagag 6420
 acctccggac catatctcgt cttttcttc cttctccgt tcctgagtgg gagttgtctc 6480
 tgcgtaatcg gcccgacaac gctgtgcacc gtctcctgac cctccatgct gttcgccatc 6540
 acgcattagc cgggctgttg cgacacgtgg cagaggactg ggaggtacga caagcggtag 6600
 tttgcatacg gcagccgccc atgactcggc cttagaccgt acaggaagtt gaacgcggcc 6660
 aaacgtatgc cgtcggcggg tactgagccg gaatctggca tgtccttcaa cttgcgccgg 6720
 ggcactcgaa tcgagccacc gatatccgtt cctacaccga tgacgccacc acgaatccca 6780
 ccgtgagctt agctcggtag ctataggcaa ggatgtggct actgcggtag tgcttagggt 6840
 acgatcgac cctcaccacc agaactgccg ccgcacgacc agttcttggt gcgtgggttg 6900
 tgctagcgtg ggagtggtag tcttgacggc ggctgtctgg tcaagaacaa cgcacccaac 6960
 acggtgcgcc cgatgatgtt gttgactgtc tcgcagacca tcagggtctg cgggacagag 7020
 tgccacgcgg gctactacaa caactgacag agcgtctggt agtcccagac gccctgtctc 7080
 gtcttgacgt agaagacggc accggtttg cggagcatgg ttgtcagaac cgagtccctt 7140
 cagaactgca tcttctgccg tggccgaaac gcctcgtacc aacagtcttg gctcagggga 7200
 tcgtcgtact tgtttagcca tgagatgtag ccattgatg tttcgtagcc ctggtggcat 7260
 agcagcatga acaaatcggc actctacatc gggttaactac aaagcatcgg gaccaccgta 7320
 atgttagctg acaaaaaggg acatctaacg acttaggggc aacggtgtac cttgactcga 7380
 tacaatcgac tgtttttccc ttagattgc tgaatccccg ttgccacatg gaactgagct 7440

```

agctgggtcctt tgagagagat ggggaggcca tgaagtggac caacgggtct cttgtgcttt 7500
tcgaccagaa actctctcta cccctccggt acttcacctg gttgcccaga gaacacgaaa 7560
gcgtagtatt catcgagttc ccttgccctgc gcgagagcgg cgtcaggga gaactcgtgg 7620
cgcatcataa gtagctcaag ggaacggacg cgctctcgcc gcagtccctt cttgagcacc 7680
gcgcagtttg tctgcacaga agccagcgtc agcttgatag tcccataagg tggcgttgtt 7740
cgcgctcaaac agacgtgtct tcggtcgcag tcgaactatc agggtattcc accgcaacaa 7800
acatctccct gagaggtaga ggggacccta ctaactgctg ggcgattgct gcccgtttac 7860
tgtagaggga ctctccatct cccctgggat gattgacgac ccgctaacga cgggcaaagt 7920
agaatgctag cgtaacttcc accgaggtca actctccggc cgccagcttg gacacaagat 7980
tcttacgata gcattgaagg tggctccagt tgagaggccg gcggtcgaac ctgtgttcta 8040
ctgcagcggg ggctctgtg atcttcagtt cggcctctga aaggatcccc gatttctttg 8100
gacgtcgctt ccggagacac tagaagtcaa gccggagact ttcctagggg ctaaagaaac 8160
ggaaatcaat aacgctgtct tccgcaggca gcgtctggac tttccattca tcagggatgg 8220
cctttagtta ttgcgacaga aggctccgt cgacagacct aaagtaagt agtccctacc 8280
tttttgcgag gcgggcgcgc ttatcagcgg ccagttcttc ccaggattga ggcattctgt 8340
aaaaacgctc cgcccgcgcg aatagtcgcc ggtcaagaag ggtcctaact ccgtaagaca 8400
gttagcttat agtcaggatg ttggctcgac gagtgtaaac tgggagttgg catgagggtt 8460
caatcgaata tcagtcctac aaccgagctg ctcacatttg accctcaacc gtactcccaa 8520
atgtaggctt ctttagcccc gcatccccct cattctctc attgatcccg ggggagcggg 8580
tacatccgaa gaaatcgggg cgtaggggga gtaagaggag taactagggc cccctcgcct 8640
tggtgttgat aagagactaa ttatagggtt tagctggtgc ctagctggtg attggctggc 8700
accacaacta ttctctgatt aatatcccaa atcgaccac gatcgaccac taaccgaccg 8760
ttcgccgaat tttaggggcc aaggaaagct gcagaaccgc ggcaactggt aacggtaatt 8820
aagcggctta aaatgcccgg ttcctttcga cgtcttggcg ccgtgaccat ttgccattaa 8880
aagctatcag ccccatgcta acgagtttaa attacgtgta ttgctgataa acaccaacag 8940
ttcgatagtc ggggtacgat tgctcaaatt taatgcacat aacgactatt tgtggttgct 9000
agctttactg aaagatggga gtcacggtgt ggcttcccca ctgcgattat tgcacaagca 9060
tcgaaatgac tttctaccct cagtgccaca ccgaaggggt gacgctaata acgtgttcgt 9120
gcgagggcga acttgactgt cgtcgtgag cagcctgcag tcaaacatac atatatatca 9180
cgctcccgct tgaactgaca gcagcgactc gtcggacgtc agtttgtagt tatatatagt 9240
accgcaaga cgtctggcct tgtagaacac gacgtccct agcaacacct gccgtgtcag 9300
tggcgcttct gcagaccgga acatcttggt ctgcgaggga tcgttggtgga cggcacagtc 9360
cctctacggg tgttacttgc attcaggatg ctctccagcg ggcgagctat tcaaaatatt 9420
ggagatgcca acaatgaacg taagtccctac gagaggtcgc ccgctcgata agttttataa 9480
caaagcagg atctcgtatt gccaggattc agctgaagca acaggtgcca aggaaatctg 9540
gtttcgtcca tagagcataa cggtcctaag tcgacttcgt tgtccacggg tcccttagac 9600
cgtcggttct catctgggct tgctcggtcc tggcgtagaa tgcacctag agtttaaaaca 9660
gcagccaaga gtagaccga acgagccagg accgcatctt acgtaggatc tcaaatgtgt 9720
gcttggcact ggccgtcgtt ttacaacgtc gtgactggga aaaccctggc gttaccaaac 9780

```


cgaaccgtga ccggcagcaa aatgttgacg cactgaccct tttgggaccg caatgggttg 9840
 ttaatcgccct tgcagcacat ccccttttcg ccagctggcg taatagcgaa gagggccgca 9900
 aattagcgga acgtcgtgta gggggaaagc ggtcgaccgc attatcgctt ctccgggcgt 9960
 ccgatcgccc ttcccaacag ttgcgcagcc tgaacggcga atggcgccctg atgcggtatt 10020
 ggctagcggg aagggttgtc aacgcgtcgg acttgccgct taccgcggac tacgccataa 10080
 ttctccttac gcatctgtgc ggtattttcac accgcatatg gtgcactctc agtacaatct 10140
 aagaggaatg cgtagacacg ccataaagtg tggcgtatac cacgtgagag tcatgttaga 10200
 gctctgatgc cgcatagtta agccagcccc gacacccgcc aacacccgct gacgcgccct 10260
 cgagactacg gcgtatcaat tcggtcgggg ctgtggggcg ttgtgggcga ctgcgcggga 10320
 gacgggcttg tctgctcccg gcatccgctt acagacaagc tgtgaccgtc tccgggagct 10380
 ctgcccgaac agacgagggc cgtaggcgaa tgtctgttcg aactggcag aggccctcga 10440
 gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc gagacgaaag ggccctcgtga 10500
 cgtacacagt ctccaaaagt ggtagtagtg gctttgcgcg ctctgctttc ccggagcact 10560
 tacgcctatt tttatagggt aatgtcatga taataatggt ttcttagacg tcaggtggca 10620
 atgcggataa aaatatccaa ttacagtact attattacca aagaatctgc agtccaccgt 10680
 cttttcgggg aaatgtgcgc ggaacccta tttgtttatt tttctaaata cattcaaata 10740
 gaaaagcccc ttacacgcg ccttggggat aaacaaataa aaagatttat gtaagtttat 10800
 tgtatccgct catgagacaa taaccctgat aaatgcttca ataatttga aaaaggaaga 10860
 acataggcga gtactctgtt attgggacta tttacgaagt tattataact ttttccttct 10920
 gtatgagtat tcaacatttc cgtgtcgccc ttattccctt ttttgcggca ttttgccttc 10980
 catactcata agttgtaaag gcacagcggg aataaggga aaacgcctt aaaacggaag 11040
 ctgtttttgc tcaccagaa acgctggtga aagtaaaaga tgctgaagat cagttgggtg 11100
 gacaaaaacg agtgggtcct tgcgaccact ttcattttct acgacttcta gtcaaccac 11160
 cacgagtggg ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc 11220
 gtgctcacc aatgtagctt gacctagagt tgcgccatt ctaggaactc taaaagcgg 11280
 ccgaagaacg ttttccaatg atgagcactt ttaaagtctt gctatgtggc gcggtattat 11340
 ggcttcttgc aaaaggttac tactcgtgaa aatttcaaga cgatacaccg cgccataata 11400
 cccgtattga cgccgggcaa gagcaactcg gtcgccgat acactattct cagaatgact 11460
 gggcataact gcggccggtt ctgcttgagc cagcggcgta tgtgataaga gtcttactga 11520
 tggttgagta ctaccagtc acagaaaagc atcttacgga tggcatgaca gtaagagaat 11580
 accaactcat gagtggtcag tgtcttttcg tagaatgcct accgtactgt cattctctta 11640
 tatgcagtgc tgccataacc atgagtgata aactgcggc caacttactt ctgacaacga 11700
 atacgtcacg acggtattgg tactcactat tgtgacgccg gttgaatgaa gactgttgct 11760
 tcggaggacc gaaggagcta accgtttttt tgcacaacat gggggatcat gtaactcgcc 11820
 agcctcctgg ctctctcgat tggcgaaaaa acgtgttgta cccctagta cattgagcgg 11880
 ttgatcggtt ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt gacaccacga 11940
 aactagcaac ccttggcctc gacttacttc ggtatggttt gctgctcgca ctgtggtgct 12000
 tgctgtagc aatggcaaca acgttgcgca aactattaac tggcgaacta cttactctag 12060
 acggacatcg ttaccgttgt tgcaacgcgt ttgataattg accgcttgat gaatgagatc 12120

cttccccggca acaattaata gactggatgg aggcggataa agttgcagga ccacttctgc 12180
gaagggccgt tgtaattat ctgacctacc tccgcctatt tcaacgtcct ggtgaagacg 12240
gctcggccct tccggctggc tggtttattg ctgataaatc tggagccggt gagcgtgggt 12300
cgagccggga aggccgaccg accaaataac gactatntag acctcggcca ctgcaccca 12360
ctcgcggtat cattgcagca ctggggccag atggtaagcc ctcccgtatc gtagttatct 12420
gagcgccata gtaacgtcgt gaccccggc taccattcgg gagggcatag catcaataga 12480
acacgacggg gagtcaggca actatggatg aacgaaatag acagatcgct gagataggtg 12540
tgtgtgccc ctgagtcctg tgatacctac ttgctttatc tgtctagcga ctctatccac 12600
cctcactgat taagcattgg taactgtcag accaagttta ctcatatata ctttagattg 12660
ggagtgacta attcgtaacc attgacagtc tggttcaaat gagtatatat gaaatctaac 12720
atttaaaact tcatttttaa tttaaaagga tctaggtgaa gatccttttt gataatctca 12780
taaattttga agtaaaaatt aaattttcct agatccactt ctaggaaaaa ctattagagt 12840
tgacaaaaat cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga 12900
actggtttta gggaattgca ctcaaaaagca aggtgactcg cagtctgggg catcttttct 12960
tcaaaggatc ttcttgagat ctttttttct tgccgctaata ctgctgcttg caaacaaaaa 13020
agtttcctag aagaactcta ggaaaaaaag acgcgcatta gacgacgaac gtttgttttt 13080
aaccaccgct accagcgggt gtttgtttgc cggatcaaga gctaccaact ctttttccga 13140
ttgggtggcga tggtcgccac caaacaacg gcctagtctc cgatggttga gaaaaaggct 13200
aggtaactgg cttcagcaga gcgcagatac caaatactgt tcttctagt tagccgtagt 13260
tccattgacc gaagtcgtct cgcgtctatg gtttatgaca agaagatcac atcggcacatca 13320
taggccacca cttcaagaac tctgtagcac gcctacata cctcgctctg ctaatcctgt 13380
atccggtggt gaagttcttg agacatcgtg gcggatgtat ggagcgagac gattaggaca 13440
taccagtggc tgctgccagt ggcgataagt cgtgtcttac cgggttgac tcaagacgat 13500
atggtcaccg acgacggta ccgctattca gcacagaatg gccaacctg agttctgcta 13560
agttaccgga taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct 13620
tcaatggcct attccgcgtc gccagcccga cttgcccccc aagcacgtgt gtcgggtcga 13680
tgagcgcaac gacctacacc gaactgagat acctacagcg tgagctatga gaaagcgcca 13740
acctcgcttg ctggatgtgg cttgactcta tggatgtcgc actcgatact ctttcgcggt 13800
cgcttcccga agggagaaaag gcggacaggt atccggtgaa cggcagggtc ggaacaggag 13860
gcgaagggtc tccctcttct cgctgtcca taggccattc gccgtcccag cttgtcctc 13920
agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc 13980
tcgcgtgctc cctcgaaggc ccccttttgc ggaccataga aatatcagga cagcccaaag 14040
gccacctctg acttgagcgt cgatttttgt gatgctcgtc aggggggagg agcctatgga 14100
cggtggagac tgaactcgca gctaaaaaca ctacgagcag tccccccg ccgataacct 14160
aaaacgccag caacgcggcc tttttacggt tcctggcctt ttgctggcct tttgtcaca 14220
ttttgcggtc gttgcgccg aaaaatgcca aggaccggaa aacgaccgga aaacgagtgt 14280
tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagttag 14340
acaagaaagg acgcaatagg ggactaagac acctattggc ataatggcgg aaactcactc 14400
ctgataaccg tcgccgcagc cgaacgaccg agcgacgca gtcagtgagc gaggaagcgg 14460

gactatggcgc agcggcgctcg gcttgctggc tcgcgctcgt cagtcactcg ctccttcgcc 14520
aagagcgccc aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat taatgcagct 14580
ttctcgcggg ttatgcgttt ggcgagagg ggcgcgcaac cggctaagta attacgtcga 14640
ggcacgacag gtttcccgcac tggaaagcgg gcagtgcgcg caacgcaatt aatgtgagtt 14700
ccgtgctgtc caaagggctg acctttcgcc cgtcactcgc gttgcgttaa ttacactcaa 14760
agctcactca ttaggcaccc caggctttac actttatgct tccggctcgt atgttggtgtg 14820
tcgagtgcgt aatccgtggg gtccgaaatg tgaaatacga aggccgagca tacaacacac 14880
gaattgtgag cggataacaa ttccacacag gaaacagcta tgaccatgat tacgaattgt 14940
cttaacactc gcctattgtt aaagtgtgtc ctttgtcgat actggtacta atgcttaaca 15000
ttaaacgaat ttgc 15014

<210> 78

<211> 35

<212> DNA

<213> 里氏木霉

<400> 78

cgaacgcggt agtgggaatt ctaggctagg tatgc 35

<210> 79

<211> 35

<212> DNA

<213> 里氏木霉

<400> 79

ccaaccgaat ctcatggtgc aatacacaga ggggtg 35

<210> 80

<211> 35

<212> DNA

<213> 烟曲霉

<400> 80

tctgtgtatt gcacatgag attcggttgg ctcga 35

<210> 81

<211> 35

<212> DNA

<213> 烟曲霉

<400> 81

ccggtcacga aagccctagt agacacgggg cagag 35

<210> 82

<211> 35

<212> DNA

<213> 里氏木霉

<400> 82

ccccgtgtct actagggctt tcgtgaccgg gcttc 35
<210> 83
<211> 35
<212> DNA
<213> 里氏木霉
<400> 83
gtcattacca attggactcg gagttgttat acgct 35
<210> 84
<211> 35
<212> DNA
<213> 里氏木霉
<400> 84
cgaacgcggt agtgggaatt ctaggctagg tatgc 35
<210> 85
<211> 35
<212> DNA
<213> 里氏木霉
<400> 85
gtcattacca attggactcg gagttgttat acgct 35
<210> 86
<211> 36
<212> DNA
<213> 烟曲霉
<400> 86
ccctttgggt atccgtgact gtgagctata cccgcg 36
<210> 87
<211> 35
<212> DNA
<213> 烟曲霉
<400> 87
cgatcatgagt gactggggcg ctcaccacag cgggtg 35
<210> 88
<211> 36
<212> DNA
<213> 烟曲霉
<400> 88
gggtagtggt actgccgagt tcccttacct tgtcac 36
<210> 89
<211> 39

<212> DNA
<213> 烟曲霉
<400> 89
gccgactctg gagagggtta catcagtgtc gacggcaac 39
<210> 90
<211> 33
<212> DNA
<213> 烟曲霉
<400> 90
cggactgcgc accatgagat tcggttggct cga 33
<210> 91
<211> 35
<212> DNA
<213> 烟曲霉
<400> 91
tcgccacgga gcttactagt agacacgggg cagag 35
<210> 92
<211> 61
<212> DNA
<213> 里氏木霉
<400> 92
gcgagtcagt gagcgaggaa gcggaagagc ttaattaatc ttgagtggat gtctgatcta 60
g 61
<210> 93
<211> 58
<212> DNA
<213> 里氏木霉
<400> 93
gttcggataa caatcctaca ttcggtcgac ttataaggat gtatcaatgg gttatacg 58
<210> 94
<211> 20
<212> DNA
<213> 烟曲霉
<400> 94
tcaaccagct tctttattgg 20
<210> 95
<211> 25
<212> DNA
<213> 烟曲霉

<400> 95
gatcgccata ggctcatgct ccgca 25
<210> 96
<211> 21
<212> DNA
<213> 烟曲霉
<400> 96
gcggcatcaa acacgaacct g 21
<210> 97
<211> 20
<212> DNA
<213> 烟曲霉
<400> 97
cagtggcgct tattactcag 20
<210> 98
<211> 25
<212> DNA
<213> 烟曲霉
<400> 98
gatcgccata ggctcatgct ccgca 25
<210> 99
<211> 25
<212> DNA
<213> 里氏木霉
<400> 99
aaaaaaca aaa catcccggttc ataac 25
<210> 100
<211> 25
<212> DNA
<213> 里氏木霉
<400> 100
aacaagggttt accgggtttcg aaaag 25
<210> 101
<211> 25
<212> DNA
<213> 烟曲霉
<400> 101
agccacatgc cgcattatga caaag 25
<210> 102

<211> 25
<212> DNA
<213> 里氏木霉
<400> 102
agggattcag tgtgctacag gctgc 25
<210> 103
<211> 25
<212> DNA
<213> 里氏木霉
<400> 103
aaaaaaca aaa catcccggttc ataac 25
<210> 104
<211> 25
<212> DNA
<213> 里氏木霉
<400> 104
aacaaggttt accggtttcg aaaag 25
<210> 105
<211> 27
<212> DNA
<213> 里氏木霉
<400> 105
gttaagcata caattgaacg agaattgg 27
<210> 106
<211> 25
<212> DNA
<213> 里氏木霉
<400> 106
gatgatataa tggagcaaat aaggg 25
<210> 107
<211> 2077
<212> DNA
<213> 里氏木霉
<400> 107
atgacggatg cacaaaagaa ttggaggaga gacgaaaacg acgaggacga tgaagcagag 60
caggagctcg atgaggctgt aagtcgccgc gaggcgcac tggtctgaca agcgtcgtct 120
gacactcttt tctcccatct agagcctcaa ggccgagaaa gatgcaattc ttctagccat 180
tgaagtcagt ccgtcgatgc ttgagcctcc gccagtctcc agctctagga aagctgatcg 240
ggacagcccc gttcaagctg cgctgaaatg cgccccccac ctgatggagc agcgcacatcat 300

ctccaacccc aaagacatga tgggaatcct cctctttggg acagaaaaga ccaagttccg 360
 ggacgacaat ggccgcagtg ggctcgggta tccgaattgc tacctcttta tggacctcga 420
 cattccggca gctgaagacg tcaaagcggt gaaggcgctg accgaggacg aagacgaaga 480
 cgaagtgtctg aagcccgcca ccaccgacac agtttccatg tccaacgtgt tgttttgcgc 540
 caaccagata ttcaccacaa aggcggccaa ctttggcagc cggcgacttt tcattgtgac 600
 ggacaatgac gatccgcacg cgtcggacaa ggcgggcagg tctgctgccg ctgttcgggc 660
 aaaggacttg tacgatctgg gcatcacgat cgacttggtt ccaatcacca caggagactc 720
 caagtttgat ctacgcaaat tttacgatgt aagctatatt tcttcgtttc ttcgctctaa 780
 aatcacccac cctccgtcgt gacatagact gacaaggaac taggatattg tctatcgca 840
 cccgaatgcc gaggccaatc gcaccgaagt gcgagcctca aaatcgggag atggactgtc 900
 tctttcaaac tcgtcattt caaacatcaa ttccaagcag acgccaagc gagcattgtt 960
 ccatctgcca tttgagattg cacctggact caagatcact gtcaagggt acaacattgt 1020
 gcatcggcaa acgccggcga gaacgtgcta catctggctg gaaggggaga aggtcagat 1080
 tgcaacaggc gaaacgacgc gagttgcaga ggattctgcc agaacagtcg aaaagcaaga 1140
 gataaaaaag gcctacaagt ttggtggcga atacgtatac tttacgcccg aggagcagaa 1200
 gaagctccgg gatthttggc cgcccacgat ccgatcatt ggattcaaga agcgcagcat 1260
 gattcccgtc tgggccagcg tcaagaagtc gacctttatc tttcccagcg aagaggatta 1320
 catcgatcg acacgcgtct tttcagccct atggcagaag cttctaaagg atgacaagat 1380
 cggcctcgtc tgggtcgtgc ttcgatctaa cgcgcagccc atgtttgccg ctctgattcc 1440
 atcaagagag cagtccgaag acgacgcggg gacaccatat ctaccagctg gcctgtggct 1500
 gtatcctctc cctacggctg acgacctgcg agatataaat gtcgaacgaa agctcactg 1560
 ctcggaggac ctaaaaacca aaatgagagt cattgtacaa cagctcaatc tccccagg 1620
 catatataac ccactcaagt acccgaaccc ggctctgcaa tggcactaca agatcctcca 1680
 gacctcgcg ttggaggagg agatgccgga agaaccgaa gacttgacgg agccccaaaa 1740
 caaggcgata agcaaacgcg tcggagggtta cttggaggag tgggtccgaga ctctgaaaga 1800
 cgaggcggac agggccactc gatccaggtc cttgaagcga gagattgaag atgatcccc 1860
 ggagcgcccc gcaaagcaga gaaaggtagc tggagagcgg cccagcggat cgaatcttag 1920
 catggcgcag cttagggatg ccattgagag cgggagcatc tcgaagatga cagtggcaca 1980
 gctgaaggat gtcgctggcg ccagaggact cagcacgggt ggtaagaagg ctgatttgct 2040
 ggagcggata gagcagtggg ttgaggagaa cagctga 2077

<210> 108

<211> 649

<212> PRT

<213> 里氏木霉

<400> 108

Met Thr Asp Ala Gln Lys Asn Trp Arg Arg Asp Glu Asn Asp Glu Asp

1 5 10 15

Asp Glu Ala Glu Gln Glu Leu Asp Glu Ala Val Ser Arg Arg Asp Leu

20 25 30

Lys	Ala	Gln	Lys	Asp	Ala	Ile	Leu	Leu	Ala	Ile	Glu	Val	Ser	Pro	Ser
35							40					45			
Met	Leu	Glu	Pro	Pro	Pro	Val	Ser	Ser	Ser	Arg	Lys	Ala	Asp	Arg	Asp
50						55					60				
Ser	Pro	Val	Gln	Ala	Ala	Leu	Lys	Cys	Ala	Arg	His	Leu	Met	Glu	Gln
65				70						75				80	
Arg	Ile	Ile	Ser	Asn	Pro	Lys	Asp	Met	Met	Gly	Ile	Leu	Leu	Phe	Gly
			85					90						95	
Thr	Glu	Lys	Thr	Lys	Phe	Arg	Asp	Asp	Asn	Gly	Arg	Ser	Gly	Leu	Gly
			100					105					110		
Tyr	Pro	Asn	Cys	Tyr	Leu	Phe	Met	Asp	Leu	Asp	Ile	Pro	Ala	Ala	Glu
			115				120					125			
Asp	Val	Lys	Ala	Leu	Lys	Ala	Leu	Thr	Glu	Asp	Glu	Asp	Glu	Asp	Glu
			130				135					140			
Val	Leu	Lys	Pro	Ala	Thr	Thr	Asp	Thr	Val	Ser	Met	Ser	Asn	Val	Leu
145					150					155					160
Phe	Cys	Ala	Asn	Gln	Ile	Phe	Thr	Thr	Lys	Ala	Ala	Asn	Phe	Gly	Ser
			165						170					175	
Arg	Arg	Leu	Phe	Ile	Val	Thr	Asp	Asn	Asp	Asp	Pro	His	Ala	Ser	Asp
			180					185					190		
Lys	Ala	Ala	Arg	Ser	Ala	Ala	Ala	Val	Arg	Ala	Lys	Asp	Leu	Tyr	Asp
			195				200					205			
Leu	Gly	Ile	Thr	Ile	Asp	Leu	Phe	Pro	Ile	Thr	Thr	Gly	Asp	Ser	Lys
			210				215					220			
Phe	Asp	Leu	Ser	Lys	Phe	Tyr	Asp	Asp	Ile	Val	Tyr	Arg	Asp	Pro	Asn
225					230					235				240	
Ala	Glu	Ala	Asn	Arg	Thr	Glu	Val	Arg	Ala	Ser	Lys	Ser	Gly	Asp	Gly
			245						250					255	
Leu	Ser	Leu	Leu	Asn	Ser	Leu	Ile	Ser	Asn	Ile	Asn	Ser	Lys	Gln	Thr
			260					265					270		
Pro	Lys	Arg	Ala	Leu	Phe	His	Leu	Pro	Phe	Glu	Ile	Ala	Pro	Gly	Leu
			275				280					285			
Lys	Ile	Thr	Val	Lys	Gly	Tyr	Asn	Ile	Val	His	Arg	Gln	Thr	Pro	Ala
			290				295				300				
Arg	Thr	Cys	Tyr	Ile	Trp	Leu	Glu	Gly	Glu	Lys	Ala	Gln	Ile	Ala	Thr
305					310					315				320	
Gly	Glu	Thr	Thr	Arg	Val	Ala	Glu	Asp	Ser	Ala	Arg	Thr	Val	Glu	Lys
			325						330					335	
Gln	Glu	Ile	Lys	Lys	Ala	Tyr	Lys	Phe	Gly	Gly	Glu	Tyr	Val	Tyr	Phe

340	345	350
Thr Pro Glu Glu Gln Lys Lys Leu Arg Asp Phe Gly Ala Pro Thr Ile		
355	360	365
Arg Ile Ile Gly Phe Lys Lys Arg Ser Met Ile Pro Val Trp Ala Ser		
370	375	380
Val Lys Lys Ser Thr Phe Ile Phe Pro Ser Glu Glu Asp Tyr Ile Gly		
385	390	395
Ser Thr Arg Val Phe Ser Ala Leu Trp Gln Lys Leu Leu Lys Asp Asp		
405	410	415
Lys Ile Gly Leu Ala Trp Cys Val Leu Arg Ser Asn Ala Gln Pro Met		
420	425	430
Phe Ala Ala Leu Ile Pro Ser Arg Glu Gln Ser Glu Asp Asp Ala Gly		
435	440	445
Thr Pro Tyr Leu Pro Ala Gly Leu Trp Leu Tyr Pro Leu Pro Thr Ala		
450	455	460
Asp Asp Leu Arg Asp Ile Asn Val Glu Arg Lys Leu Asp Cys Ser Glu		
465	470	475
Asp Leu Lys Thr Lys Met Arg Val Ile Val Gln Gln Leu Asn Leu Pro		
485	490	495
Lys Gly Ile Tyr Asn Pro Leu Lys Tyr Pro Asn Pro Ala Leu Gln Trp		
500	505	510
His Tyr Lys Ile Leu Gln Thr Leu Ala Leu Glu Glu Glu Met Pro Glu		
515	520	525
Glu Pro Glu Asp Leu Thr Glu Pro Lys Asn Lys Ala Ile Ser Lys Arg		
530	535	540
Val Gly Gly Tyr Leu Glu Glu Trp Ser Glu Thr Leu Lys Asp Glu Ala		
545	550	555
Asp Arg Ala Thr Arg Ser Arg Ser Leu Lys Arg Glu Ile Glu Asp Asp		
565	570	575
Ala Pro Glu Arg Pro Ala Lys Gln Arg Lys Val Ala Gly Glu Arg Pro		
580	585	590
Ser Gly Ser Asn Leu Ser Met Ala Gln Leu Arg Asp Ala Ile Glu Ser		
595	600	605
Gly Ser Ile Ser Lys Met Thr Val Ala Gln Leu Lys Asp Val Ala Gly		
610	615	620
Ala Arg Gly Leu Ser Thr Gly Gly Lys Lys Ala Asp Leu Leu Glu Arg		
625	630	635
Ile Glu Gln Trp Val Glu Glu Asn Ser		
645		

<210> 109
<211> 55
<212> DNA
<213> 里氏木霉
<400> 109
gtgtgcggcc gctcgagcat gcatgtttaa acagcttggc actggccgtc gtttt 55
<210> 110
<211> 55
<212> DNA
<213> 里氏木霉
<400> 110
atcagccccg agacggcgcc gcgtttaaac aattcgtaat catggtcata gctgt 55
<210> 111
<211> 55
<212> DNA
<213> 里氏木霉
<400> 111
catgattacg aattgtttaa acgcggcgcc gtctcggggc tgatcttgtc gagga 55
<210> 112
<211> 55
<212> DNA
<213> 里氏木霉
<400> 112
ggcggccggtt actagtggat ccagcccttg acagtgatct tgagtccagg tgcaa 55
<210> 113
<211> 58
<212> DNA
<213> 里氏木霉
<400> 113
tgcagatata catcacactg gcggccgcag tttccatgtc caacgtgttg ttttgcgc 58
<210> 114
<211> 58
<212> DNA
<213> 里氏木霉
<400> 114
gccagtgcc a gctgtttaa acatgcatgc tcgagcggcc gcacacgccc tctcctcg 58
<210> 115
<211> 20
<212> DNA

<213> 里氏木霉
<400> 115
caatgacgat ccgcacgcgt 20
<210> 116
<211> 20
<212> DNA
<213> 里氏木霉
<400> 116
caatgacgat ccgcacgcgt 20
<210> 117
<211> 20
<212> DNA
<213> 里氏木霉
<400> 117
gacactcttt tctcccatct 20
<210> 118
<211> 20
<212> DNA
<213> 里氏木霉
<400> 118
gaggagcaga agaagctccg 20
<210> 119
<211> 22
<212> DNA
<213> 里氏木霉
<400> 119
gcatatataa ccactcaag ta 22
<210> 120
<211> 22
<212> DNA
<213> 里氏木霉
<400> 120
attatcttgg accggccgca gg 22

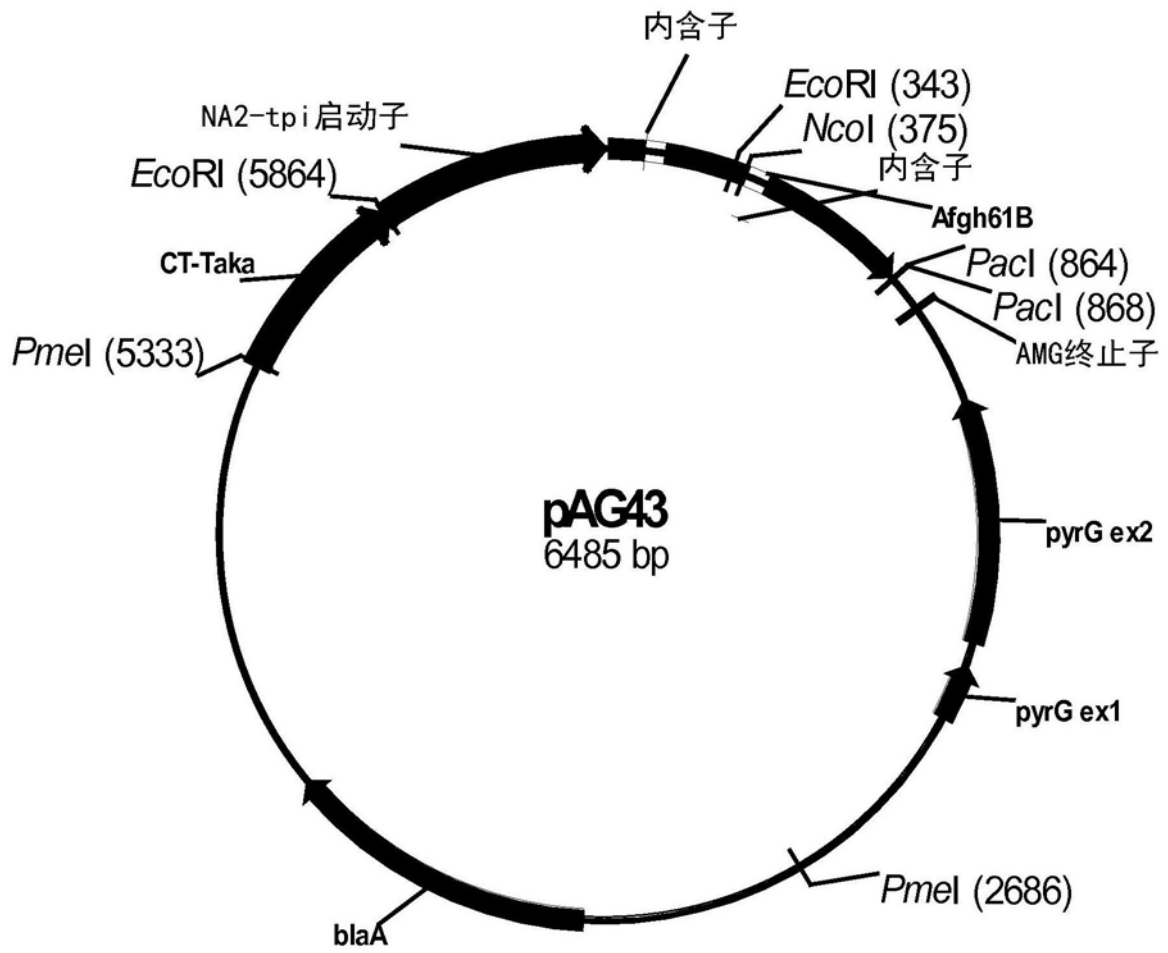


图1

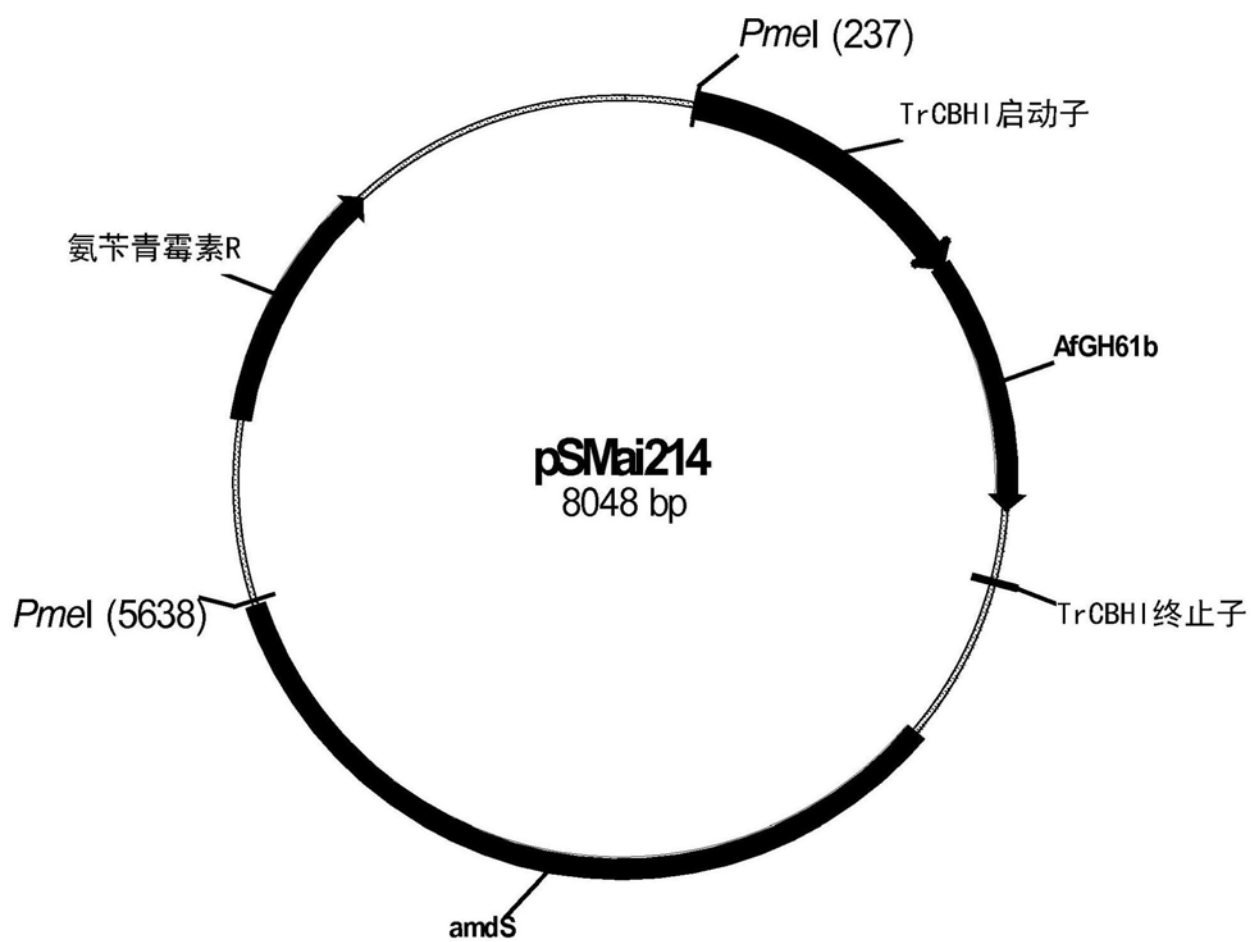


图2

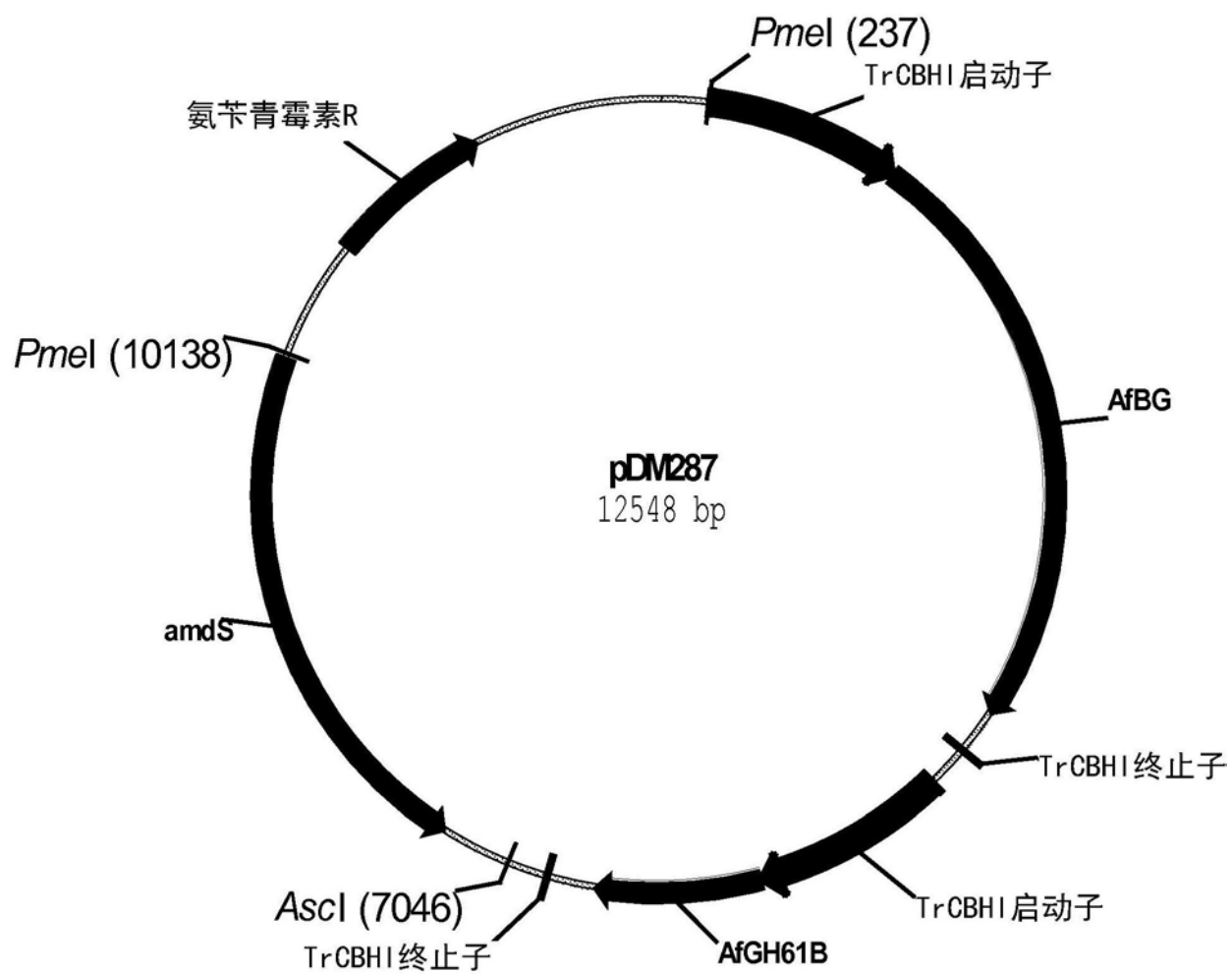


图3

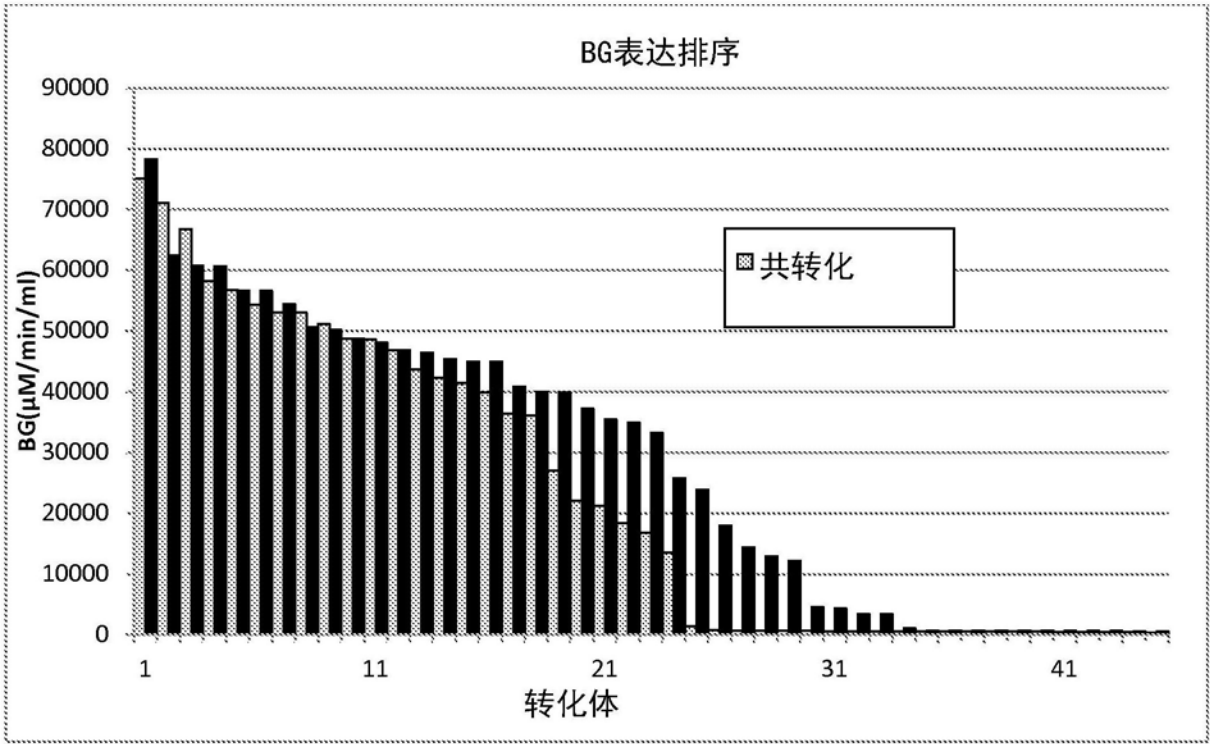


图4

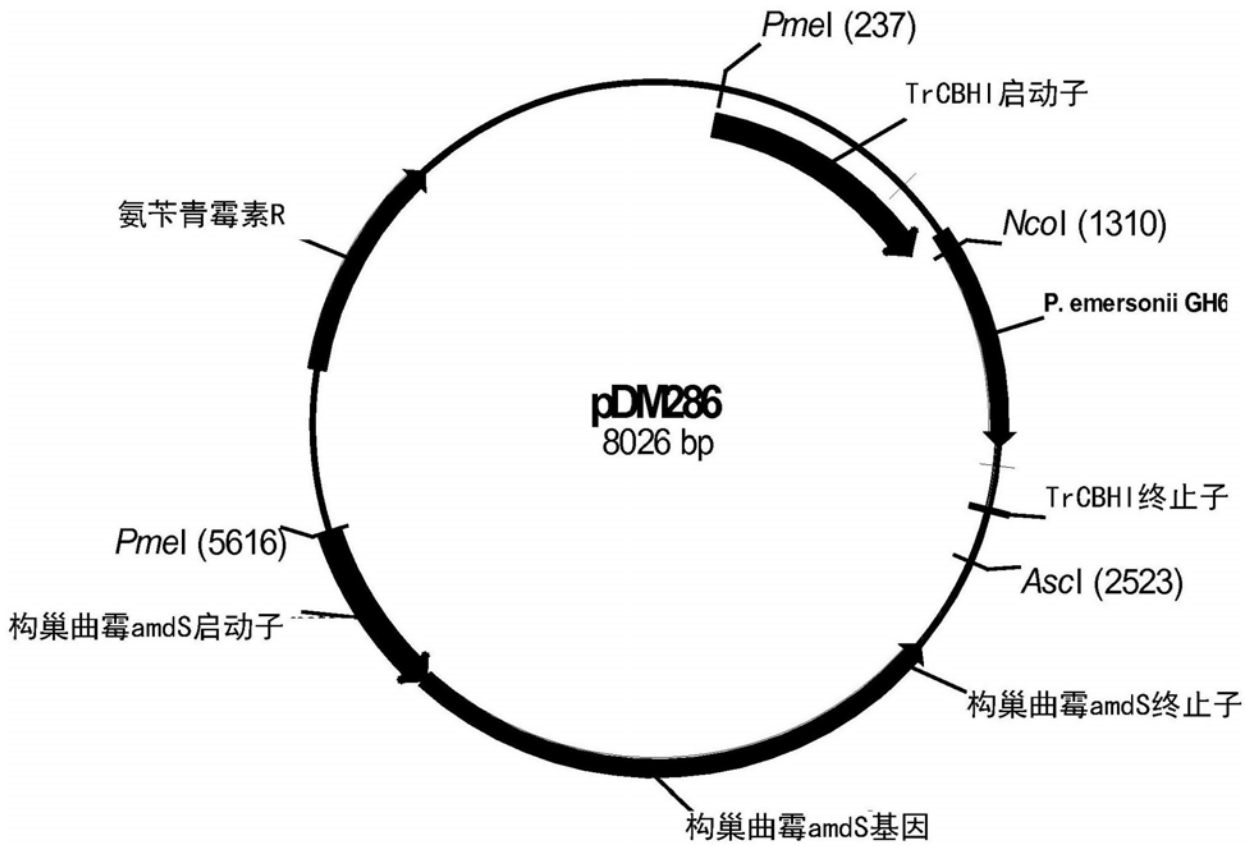


图5

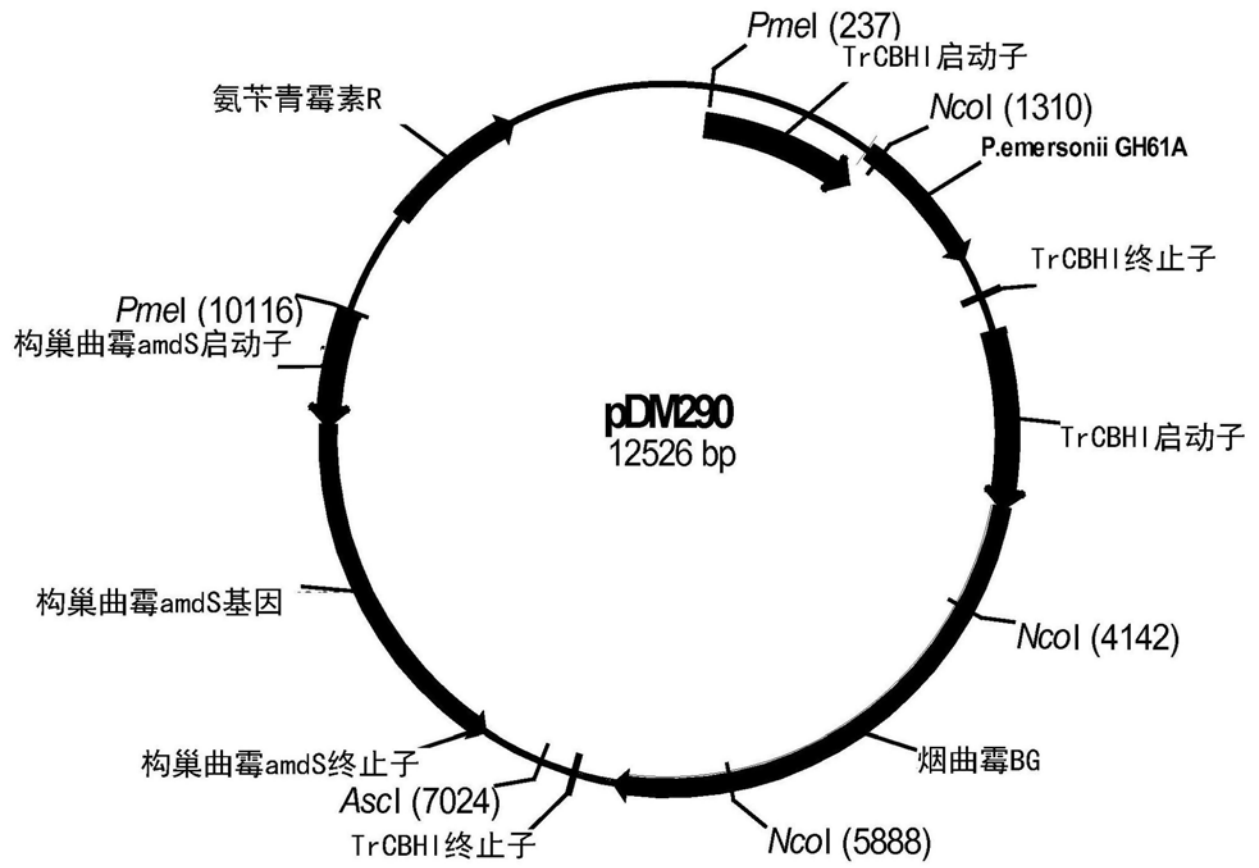


图6

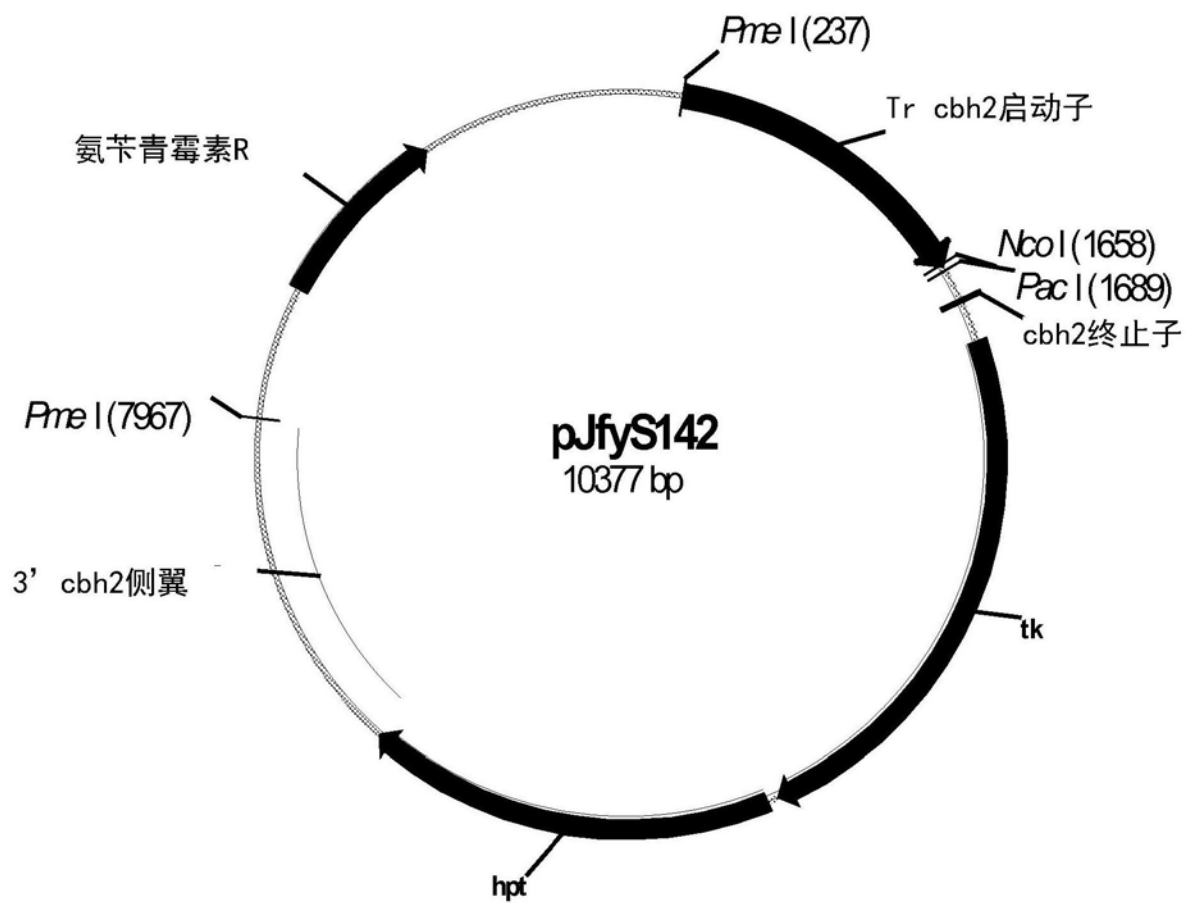


图7

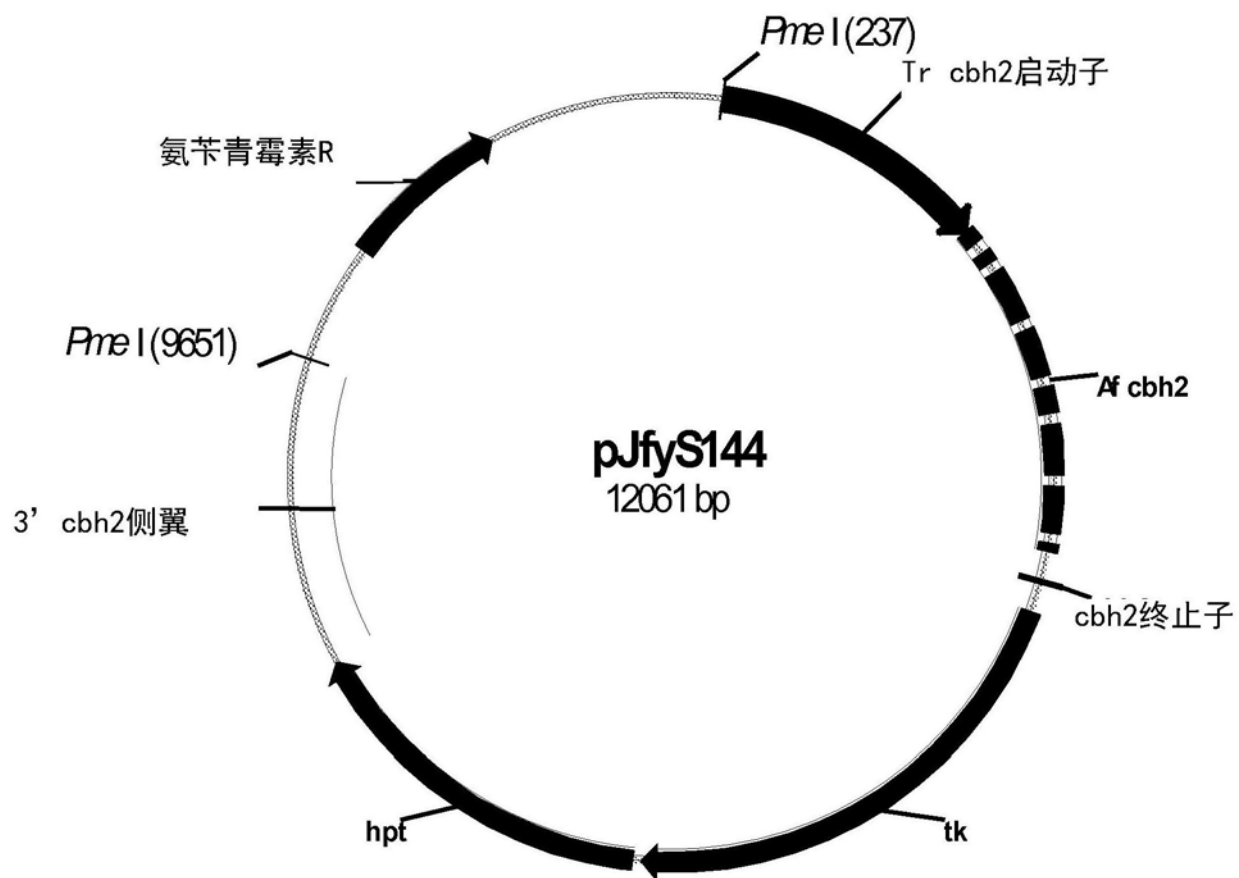


图8

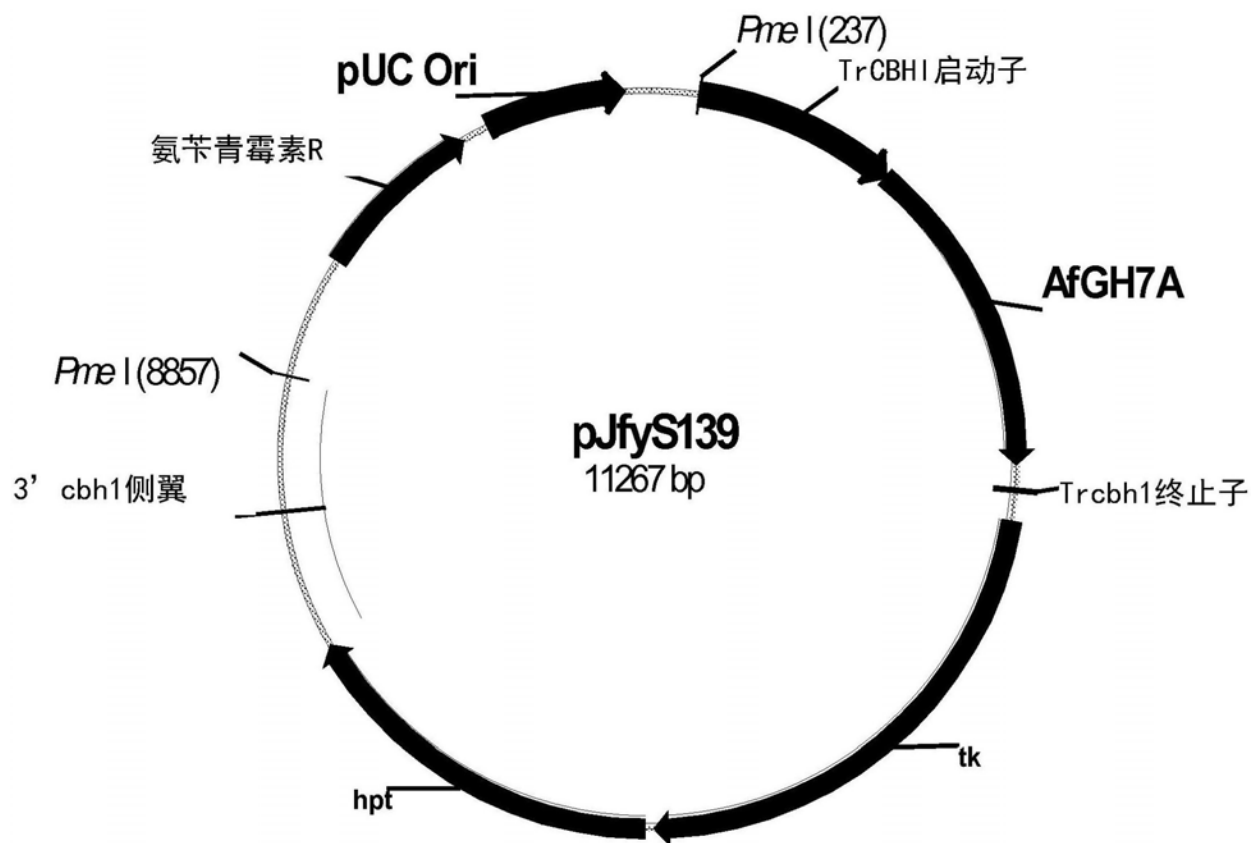


图9

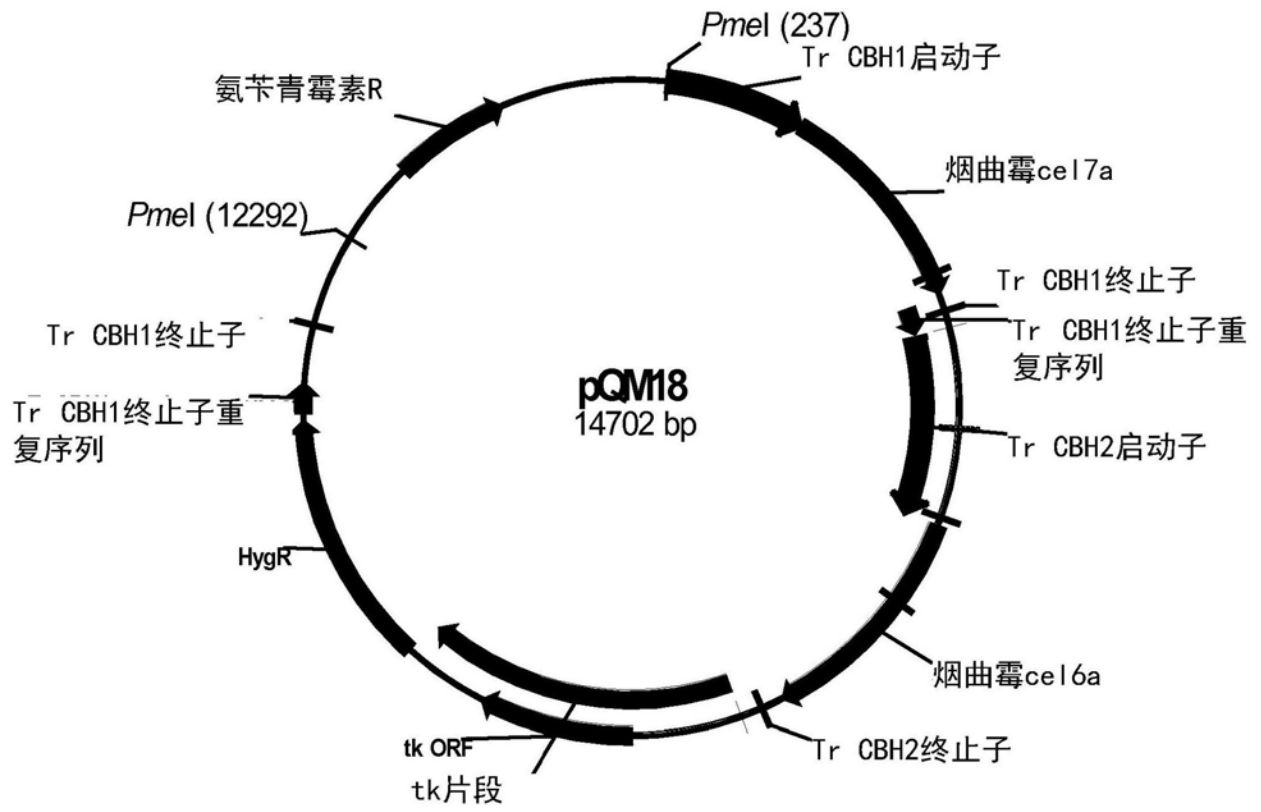


图10

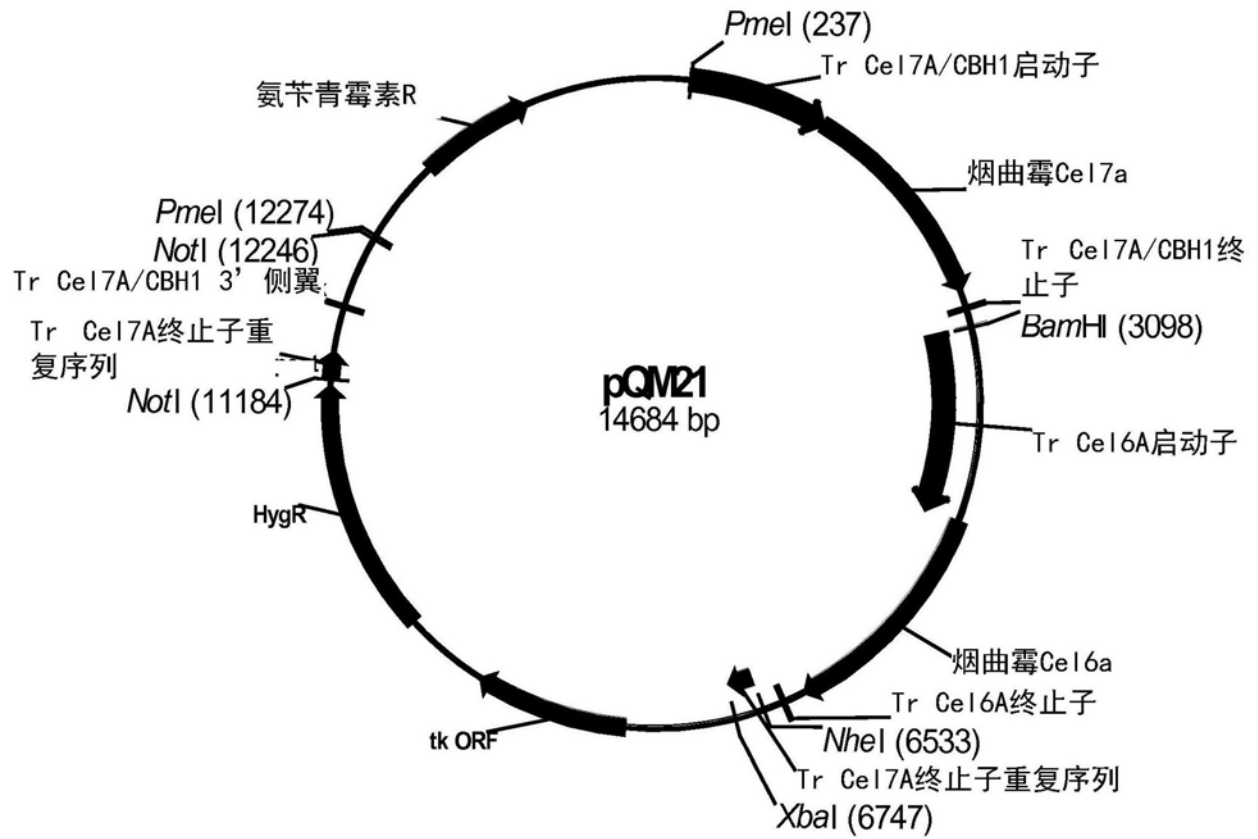


图11

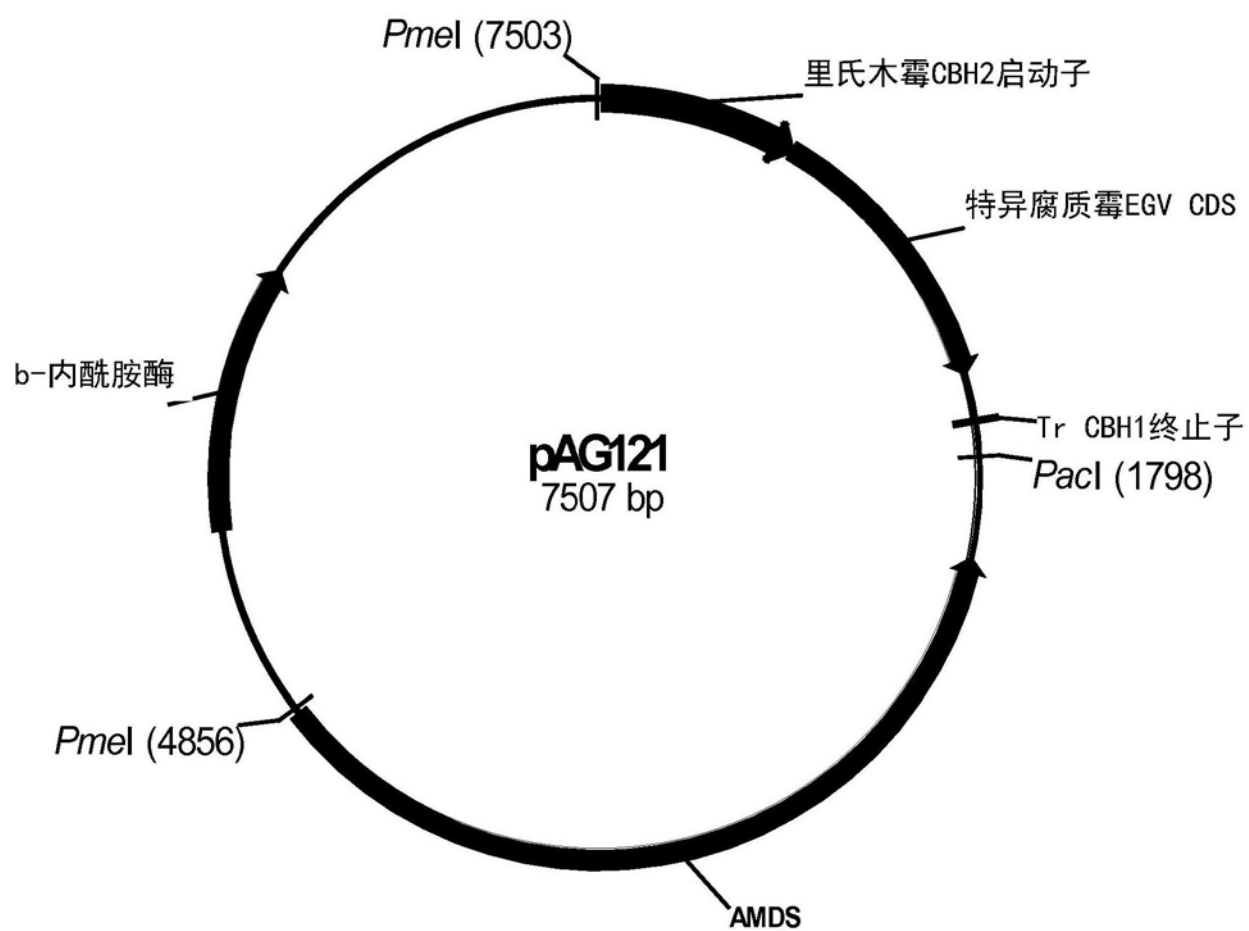


图12

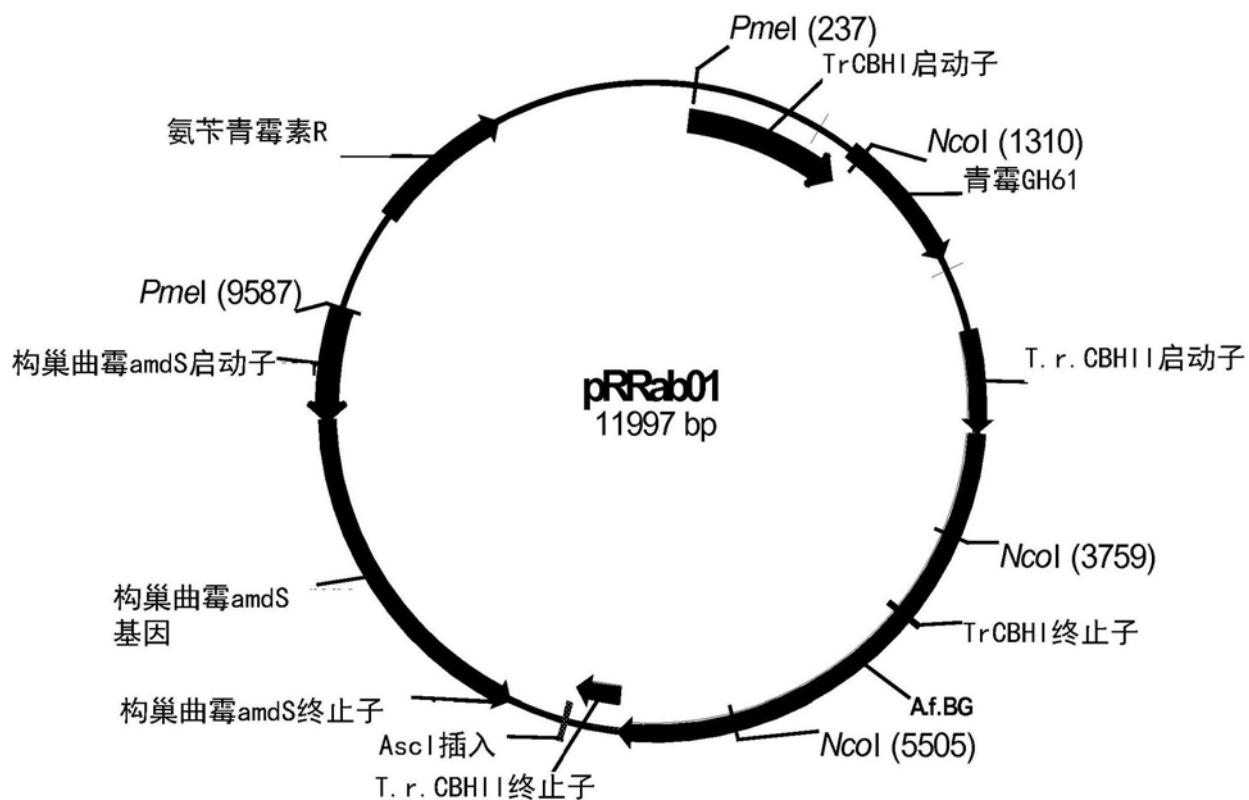


图13

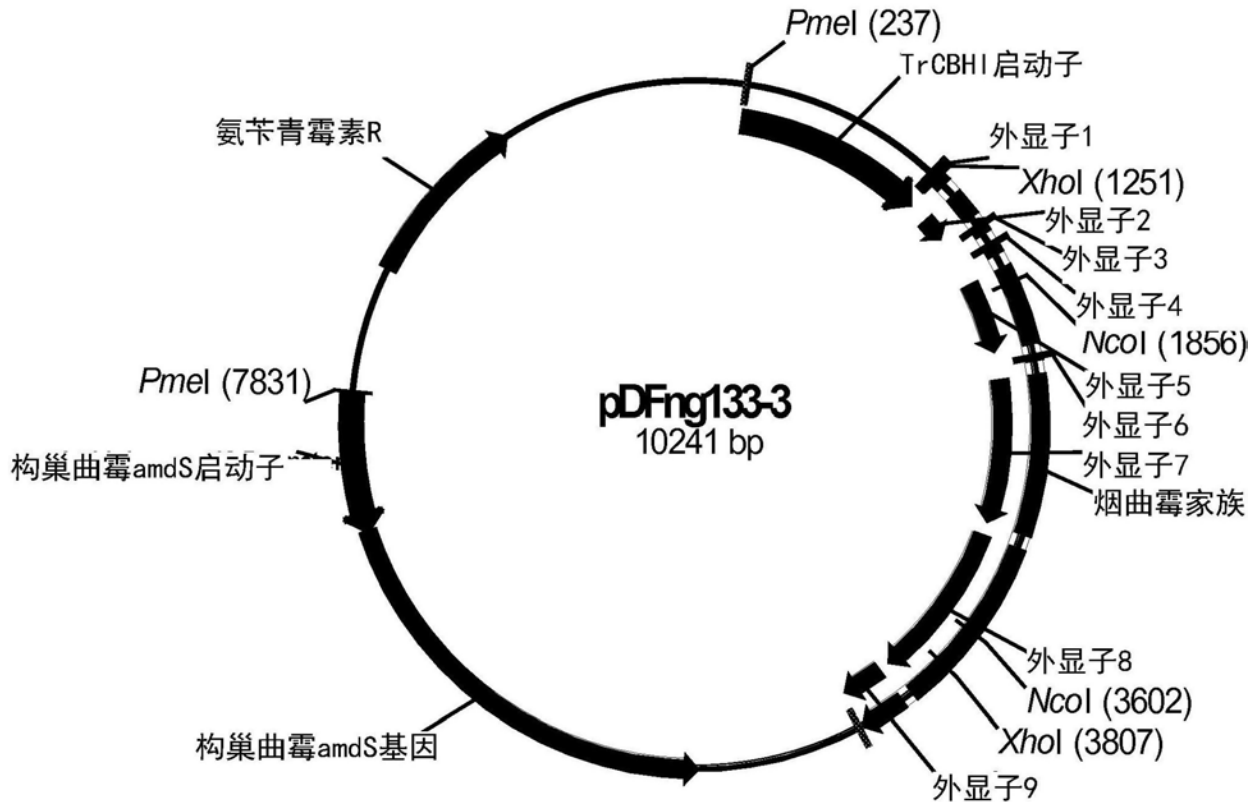


图14

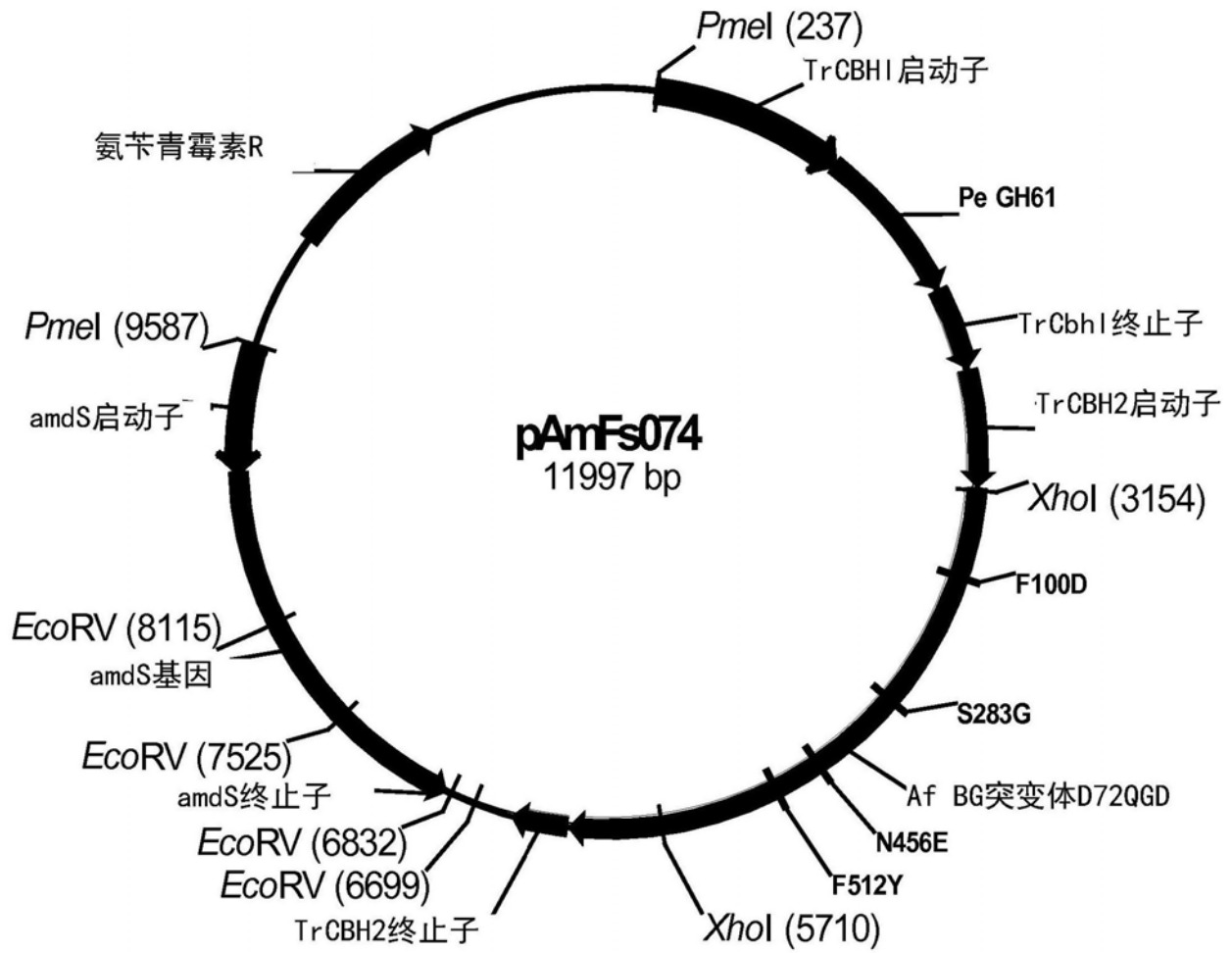


图15

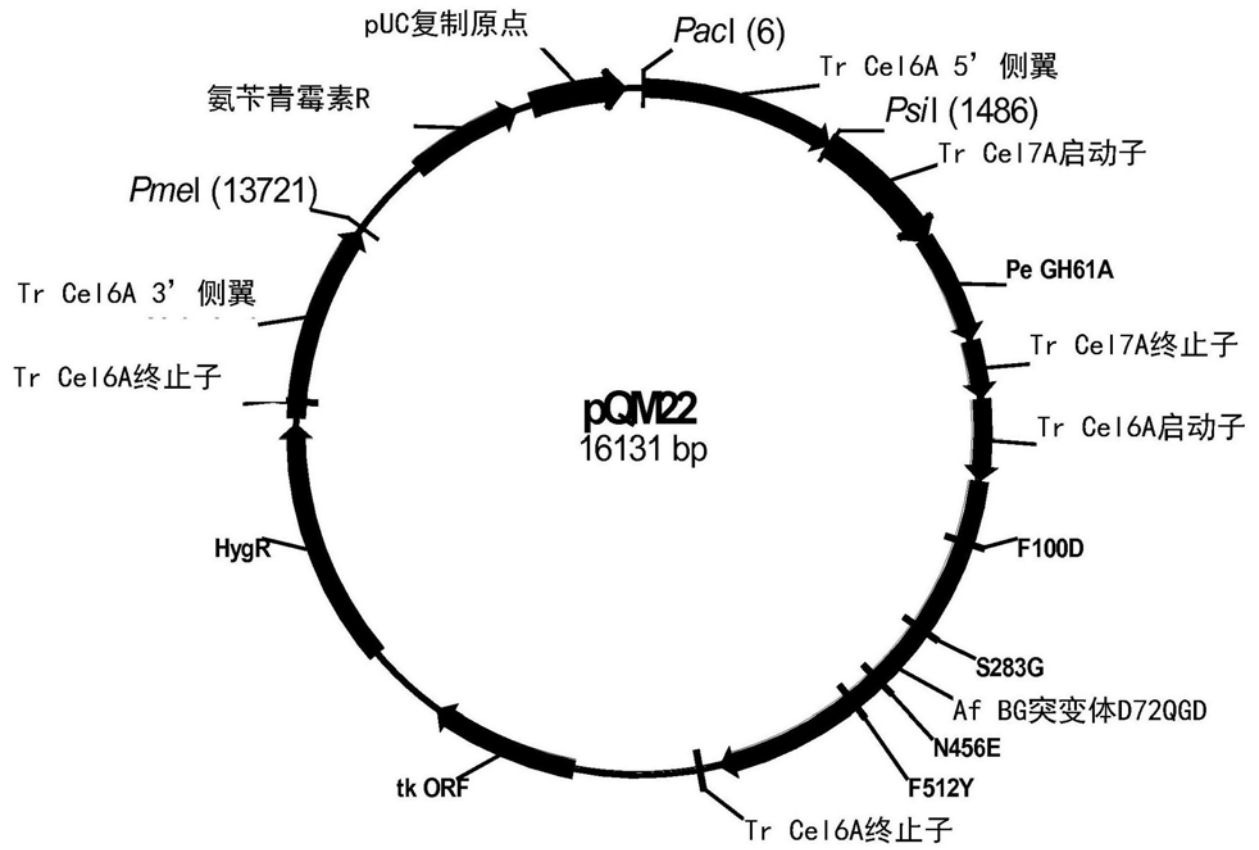


图16