



(12) 发明专利申请

(10) 申请公布号 CN 104508114 A

(43) 申请公布日 2015. 04. 08

(21) 申请号 201380032068. 1

代理人 肖善强

(22) 申请日 2013. 06. 17

(51) Int. Cl.

(30) 优先权数据

C12N 1/15(2006. 01)

12172605. 3 2012. 06. 19 EP

C12N 9/18(2006. 01)

C12N 9/26(2006. 01)

(85) PCT国际申请进入国家阶段日

C12N 9/42(2006. 01)

2014. 12. 17

C12N 15/80(2006. 01)

(86) PCT国际申请的申请数据

C12P 1/02(2006. 01)

PCT/EP2013/062490 2013. 06. 17

C12P 21/02(2006. 01)

(87) PCT国际申请的公布数据

W02013/189878 EN 2013. 12. 27

(83) 生物保藏信息

CBS 124902 2009. 07. 01

(71) 申请人 帝斯曼知识产权资产管理有限公司

地址 荷兰海尔伦

(72) 发明人 埃里克·皮特·洛斯

威尔伯特·赫尔曼·马里·海涅

赫尔曼·扬·佩尔

罗波图斯·安东尼厄斯·戴维尔德

布伦达·沃恩克

(74) 专利代理机构 北京东方亿思知识产权代理
有限责任公司 11258

权利要求书2页 说明书31页

序列表39页 附图8页

(54) 发明名称

用于在细胞中表达基因的启动子

(57) 摘要

本发明涉及经分离的 *Rasamsonia* 启动子 DNA 序列, 涉及包含与编码序列可操作相连的这些启动子的 DNA 构建体、载体和宿主细胞。本发明也涉及使用经分离的新启动子来表达基因和 / 或生产生物化合物的方法。本发明也涉及使用本发明的新启动子来改变转录水平和 / 或调节内源基因的方法。

1. *Rasamsonia* 启动子 DNA 序列, 优选 *Rasamsonia emersonii* 启动子 DNA 序列。
2. 根据权利要求 1 所述的 *Rasamsonia* 启动子 DNA 序列, 其连接至能够被过表达的编码序列。
3. 根据权利要求 1 所述的 *Rasamsonia* 启动子 DNA 序列, 其对应于强启动子和 / 或诱导型启动子。
4. 启动子 DNA 序列, 例如 :
 - (a) 以下列表中所示的 DNA 序列 : SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 ;
 - (b) 能够与 (a) 中 DNA 序列的互补体杂交的 DNA 序列 ; 或者
 - (c) 与 (a) 中 DNA 序列至少 50% 同源的 DNA 序列。
5. DNA 构建体, 其包含根据权利要求 1 或 2 所述的启动子 DNA 序列和与所述启动子 DNA 序列可操作相连的编码序列, 从而所述编码序列能够在所述启动子 DNA 序列的控制下被表达。
6. 宿主细胞, 优选真菌宿主细胞, 其包含根据权利要求 3 所述的 DNA 构建体。
7. 根据权利要求 4 所述的宿主细胞, 其中所述宿主细胞是来自以下属的细胞 : *Acremonium*、*Agaricus*、*Aspergillus*、*Aureobasidium*、*Chrysosporium*、*Coprinus*、*Cryptococcus*、*Filobasidium*、*Fusarium*、*Geosmithia*、*Humicola*、*Magnaporthe*、*Mucor*、*Myceliophthora*、*Neocallimastix*、*Neurospora*、*Paecilomyces*、*Penicillium*、*Piromyces*、*Panerochaete*、*Pleurotus*、*Rasamsonia*、*Schizophyllum*、*Talaromyces*、*Thermoascus*、*Thermomyces*、*Thielavia*、*Tolypocladium* 或者 *Trichoderma*, 优选来自 *Rasamsonia*、*Aspergillus*、*Penicillium*、*Chrysosporium* 或者 *Trichoderma* 属, 优选为 *Rasamsonia emersonii*。
8. 一种在合适的宿主细胞中表达编码序列的方法, 其包括 :
 - (a) 提供根据权利要求 3 所述的 DNA 构建体 ;
 - (b) 使用所述 DNA 构建体转化合适的宿主细胞 ; 以及
 - (c) 在有助于所述编码序列表达的培养条件下培养所述合适的宿主细胞。
9. 一种在合适的宿主细胞中生产生物化合物的方法, 其包括 :
 - (a) 提供权利要求 3 所定义的 DNA 构建体 ;
 - (b) 使用所述 DNA 构建体转化合适的宿主细胞 ; 以及
 - (c) 在有助于所述编码序列表达的培养条件下培养所述合适的宿主细胞 ; 以及任选地,
 - (d) 从培养液回收所述生物化合物。
10. 根据权利要求 7 所述的方法, 其中生产的所述生物化合物为多肽或者代谢产物。
11. 根据权利要求 8 所述的方法, 其中生产的所述多肽由存在于权利要求 2 所定义的 DNA 构建体中的编码序列来编码。
12. 根据权利要求 10 所述的方法, 其中存在于权利要求 2 所定义的 DNA 构建体中的所述编码序列编码酶, 所述酶任选地参与代谢产物的生产。
13. 编码葡糖淀粉酶的 DNA 序列, 其包含 :

- (a) SEQ ID NO:23 所示的 DNA 序列；
 - (b) 能够与 (a) 中 DNA 序列的互补体杂交的 DNA 序列；
 - (c) 与 (a) 中 DNA 序列至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、还更优选至少 90%和最优选至少 95%同源的 DNA 序列；或者
 - (d) 编码葡糖淀粉酶且与 SEQ ID NO:24 至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、还更优选至少 90%和最优选至少 95%同源的 DNA 序列。
14. 葡糖淀粉酶,其具有与 SEQ ID NO:24 至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、还更优选至少 90%和最优选至少 95%同源的 DNA 序列。

用于在细胞中表达基因的启动子

技术领域

[0001] 本发明涉及 DNA 序列,特别是经分离的启动子,以及涉及包含与编码序列可操作相连 (in operative association) 的这些启动子的 DNA 构建体、载体和宿主细胞。本发明还涉及表达基因和 / 或生产生物化合物的方法。

背景技术

[0002] 在宿主细胞中生产重组生物化合物通常是通过构建表达盒来完成的,其中编码生物化合物的 DNA 可操作地连接适合于宿主细胞的启动子。可通过质粒或者载体介导的转化将表达盒引入宿主细胞内。然后可通过在表达盒中含有的启动子正常发挥功能所必需的诱导条件下培养经转化的宿主细胞来完成生物化合物的生产。

[0003] 对于各宿主细胞而言,已通过转化引入宿主内的编码序列的表达和通过该编码序列编码的重组生物化合物的生产需要获得功能性启动子。已知大量启动子在多种宿主细胞中具备功能性。以下为真菌宿主细胞中跨物种使用的启动子的例子: *Aspergillus nidulans* 的启动子 (已知 *A. nidulans* *gpdA* 基因在 *Aspergillus niger* (*A. niger*) 中有功能 (J Biotechnol. 1991Jan;17(1):19-33. Intracellular and extracellular production of proteins in *Aspergillus* under the control of expression signals of the highly expressed *A. nidulans* *gpdA* gene. Punt PJ, Zegers ND, Busscher M, Pouwels PH, van den Hondel CA)。另一个例子是用于 *A. niger* 和 *A. nidulans* 的 *A. niger* β -木糖苷酶 *xlnD* 启动子 (Transcriptional regulation of the xylanolytic enzyme system of *Aspergillus*, van Peij, NNME, PhD-thesis Landbouwniversiteit Wageningen, the Netherlands, ISBN 90-5808-154-0) 以及 *Escherichia coli* β -葡糖醛酸酶基因在 *A. niger*、*A. nidulans* 和 *Cladosporium fulvum* 中的表达,如 Curr Genet. 1989Mar;15(3):177-80: Roberts IN, Oliver RP, Punt PJ, van den Hondel CA. "Expression of the *Escherichia coli* beta-glucuronidase gene in industrial and phytopathogenic filamentous fungi"中所述。

[0004] 迄今为止,没有 *Rasamsonia emersonii* 启动子用于形成重组产品,而仅仅使用跨物种使用的启动子。

[0005] 仍然需要以下的启动子:其用于控制经引入基因的表达,用于控制内源基因的表达水平,用于控制内源基因的表达调节或者用于介导内源基因的失活,或者用于生产多肽,或者用于前述应用的组合。这些启动子,优选为经改进的启动子,可以例如比之前已知的启动子更强。它们也可被特定便利的底物或者化合物诱导。当期望在单种宿主中同时过表达多种基因时,知道若干种功能性启动子也是有利的。为了防止压制 (squenching) (特定转录因子的滴定 (titration)), 优选使用多种不同的启动子,例如每个待表达的基因使用一个特定的启动子。

[0006] 发明简述

[0007] 根据第一方面,本发明提供 *Rasamsonia* 启动子 DNA 序列,优选 *Rasamsonia*

emersonii 启动子 DNA 序列,更优选地,其连接可被过表达的编码序列。本发明的 Rasamsonia 启动子 DNA 优选连接可被过表达的编码序列。有利地,本发明的 Rasamsonia 启动子对应于强启动子和 / 或诱导型启动子。

[0008] 根据另一方面,本发明提供启动子 DNA 序列,例如:

[0009] (a) 以下列表中所示的 DNA 序列:SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17;

[0010] (b) 能够与 (a) 中 DNA 序列的互补体 (complement) 杂交的 DNA 序列;或者

[0011] (c) 与 (a) 中 DNA 序列至少 50% 同源的 DNA 序列。

[0012] 另一方面,本发明提供 DNA 构建体,其包含本发明的启动子 DNA 序列以及与该启动子 DNA 序列可操作相连的编码序列,从而编码序列可在启动子 DNA 序列的控制下被表达。

[0013] 又一方面,本发明提供宿主细胞,优选为真菌宿主细胞,其包含本发明的 DNA 构建体。该宿主细胞优选为经转化的宿主细胞,例如经转化的真菌宿主细胞,并且有利地使用重组技术生产。宿主细胞优选为来自以下属的细胞: Acremonium、Agaricus、Aspergillus、Aureobasidium、Chrysosporium、Coprinus、Cryptococcus、Filobasidium、Fusarium、Geosmithia、Humicola、Magnaporthe、Mucor、Myceliophthora、Neocallimastix、Neurospora、Paecilomyces、Penicillium、Piromyces、Panerochaete、Pleurotus、Rasamsonia、Schizophyllum、Talaromyces、Thermoascus、Thermomyces、Thielavia、Tolypocladium 或者 Trichoderma, 优选来自 Rasamsonia、Aspergillus、Penicillium、Chrysosporium 或者 Trichoderma 属,优选为 Rasamsonia emersonii。

[0014] 再一方面,本发明提供在合适的宿主细胞中表达编码序列的方法,其包括:

[0015] (a) 提供本发明的 DNA 构建体;

[0016] (b) 使用所述 DNA 构建体来转化合适的宿主细胞;以及

[0017] (c) 在有助于编码序列表达的培养条件下培养所述合适的宿主细胞。

[0018] 而且,本发明提供在合适的宿主细胞中生产生物化合物的方法,其包括:

[0019] (a) 提供本发明的 DNA 构建体;

[0020] (b) 使用所述 DNA 构建体来转化合适的宿主细胞;以及

[0021] (c) 在有助于编码序列表达的培养条件下培养合适的宿主细胞;以及任选地,

[0022] (d) 从培养液中回收生物化合物。

[0023] 有利地,生产的生物化合物为多肽或者代谢产物。

[0024] 在本发明的方法中,生产的多肽优选通过本发明的 DNA 构建体中存在的编码序列来编码。

[0025] 有利地,在本发明的方法中,存在于 DNA 构建体中的编码序列编码任选地参与代谢产物生产的酶。

[0026] 而且,本发明提供编码葡糖淀粉酶的 DNA 序列,其包含:

[0027] (a) SEQ ID NO:23 所示的 DNA 序列;

[0028] (b) 能够与 (a) 中 DNA 序列的互补体杂交的 DNA 序列;

[0029] (c) 与 (a) 中 DNA 序列至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、还更优选至少 90% 和最优选至少 95% 同源的 DNA 序列;或者

[0030] (d) 编码葡糖淀粉酶并且与 SEQ ID NO:24 至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、还更优选至少 90% 和最优选至少 95% 同源的 DNA 序列。

[0031] 本发明的又一实施方式提供葡糖淀粉酶,其具有与 SEQ ID NO:24 至少 50%、优选至少 60%、更优选至少 70%、甚至更优选至少 80%、还更优选至少 90% 和最优选至少 95% 同源的 DNA 序列。

附图说明

[0032] 图 1 显示质粒 pENTRY-P6bleTtrpC-Pxeba7flagTgla 的示意图,其为 *R. emersonii* 中启动子测试构建体的基础。启动子测试构建体包含 ble 表达盒,该 ble 表达盒由以下组成:A.nidulans gpdA 启动子 (P6)、ble 编码区 (ble) 和 A.nidulans TrpC 终止子 (TtrpC)、目标启动子 (Px)、EBA7-FLAG 报告子编码区 (eba7flag) 和 A.niger 葡糖淀粉酶终止子。

[0033] 图 2 显示如使用 FLAG- 特异性抗体通过 Western 印迹法检测的 FLAG 标记的 *R. emersonii* β - 葡聚糖酶 CEB 蛋白 (EBA7-FLAG) 的表达,其通过在 *R. emersonii* 培养物的上清液中表达的 5 种不同的 *R. emersonii* 启动子来驱动。泳道 1:CbhI 启动子,100 倍稀释的上清液;泳道 2:CbhI 启动子,未经稀释的上清液;泳道 3 和 4:AXE 启动子;泳道 5:空菌株;泳道 6:空泳道;泳道 7:A.nidulans gpdA 启动子;泳道 8:BG 启动子;泳道 9 和 10:CbhII 启动子;泳道 11:EG 启动子,未经稀释的上清液;泳道 12:EG 启动子,10 倍稀释的上清液;泳道 13:空菌株。

[0034] 图 3 显示质粒 Te pep.bbn 的示意图,其为靶向 RePepA 基因座 (locus) 的 *R. emersonii* 中启动子测试构建体的基础。载体包含靶向 RePepA 基因座的 RePepA ORF 1.5kb 上游的 1500bp 的 5' 侧翼区、lox66 位点、通过 A.nidulans gpdA 启动子驱动的无功能的 ble 编码区 5' 部分 (5' ble)、以及 ccdB 基因。

[0035] 图 4 显示在二分基因靶向方法 (bipartite gene-targeting method) 中结合 pEBA528、pEBA529、pEBA530、pEBA531、pEBA532 和 pEBA533 载体使用的质粒 pEBA1006 的示意图,其目标为在 *Rasamsonia emersonii* 中通过启动子-报告子表达盒替换 RePepA ORF 和起始 ATG 密码子的约 1500 个核苷酸上游。载体包含 ble 编码区的 3' 部分、A.nidulans trpC 终止子、lox71 位点、RePepA ORF 的 2500bp 3' 侧翼区和 pUC19 骨架 (Invitrogen, Breda, The Netherlands)。

[0036] 图 5 显示在二分基因靶向方法中结合 pEBA1006 载体使用的质粒 pEBA528 的示意图,其目标为在 *Rasamsonia emersonii* 中通过启动子-报告子表达盒替换 RePepA ORF 和起始 ATG 密码子的约 1500 个核苷酸上游。载体包含靶向 RePepA 基因座的 RePepA ORF 1.5kb 上游的 1500bp 5' 侧翼区;由 *R. emersonii* 启动子 1、FLAG 标记的 *R. emersonii* 葡糖淀粉酶 (AG-FLAG) 和 A.nidulans amdS 终止子 (TamdS) 组成的启动子-报告子表达盒;lox66 位点;通过 A.nidulans gpdA 启动子 (5' ble) 驱动的 ble 编码区的无功能 5' 部分。在 *R. emersonii* 菌株的转化之前,通过使用限制性酶 NotI 的消化来去除 *E. coli* DNA。

[0037] 图 6 显示质粒 pEBA1001 的示意图。在二分基因靶向方法中结合 pEBA1002 载体使用一部分载体片段,其目标为使在 *Rasamsonia emersonii* 中的 ReKu800RF 缺失。载体包含 2500bp 5' 上游侧翼区、lox66 位点、通过 A.nidulans gpdA 启动子驱动的 ble 编码序列的 5' 部分以及 pUC19 的骨架 (Invitrogen, Breda, The Netherlands)。在 *R. emersonii* 菌株

的转化之前,通过使用限制性酶 NotI 的消化来去除 *E. coli* DNA。

[0038] 图 7 显示质粒 pEBA1002 的示意图。在二分基因靶向方法中结合 pEBA1001 载体使用一部分载体片段,其目标为使在 *Rasamsonia emersonii* 中的 ReKu800RF 缺失。载体包含 ble 编码区的 3' 部分、*A.nidulans* trpC 终止子、lox71 位点、ReKu800RF 的 2500bp 3' 下游侧翼区以及 pUC19 的骨架 (Invitrogen, Breda, The Netherlands)。在 *R. emersonii* 菌株的转化之前,通过使用限制性酶 NotI 的消化来去除 *E. coli* DNA。

[0039] 图 8 显示用于使 *R. emersonii* 的 ReKu80 基因缺失的策略。用于使 ReKu80 缺失的载体包含重叠的无功能的 ble 选择标记物片段 (分开的标记物 (split marker)), 其侧翼是 loxP 位点以及 ReKu80 基因的 5' 和 3' 同源区来用于靶向 (1)。构建体在基因组的 ReKu80 基因座处和在重叠的同源非功能性 ble 选择标记物片段处通过三重同源重组 (X) 来整合 (2) 并替换基因组的 ReKu80 基因拷贝 (3)。随后,通过导致 lox66 和 lox71 位点之间重组的 cre 重组酶的瞬时表达来去除选择标记物,从而以在基因组内剩下的剩余双突变 lox72 位点缺失 ble 基因 (4)。使用该整体策略将 ReKu800RF 从基因组中去除。

[0040] 图 9 显示用于在真菌中瞬时表达 cre 重组酶的质粒 pEBA513 的示意图。pEBA513 是含有 AMA1 区和 CAT 氯霉素抗性基因的 pAMPF21 衍生载体。示出了 cre 重组酶基因 (cre) 表达盒,其含有 *A.niger* glaA 启动子 (Pgla)、cre 重组酶编码区和 niaD 终止子。此外,示出了由 *A.nidulans* gpdA 启动子 (PgpdA)、hygB 编码区和 *P.chrysogenum* penDE 终止子 (TpenDE) 组成的潮霉素抗性盒。

[0041] 图 10 显示如使用 FLAG- 特异性抗体通过 Western 印迹法检测的 FLAG 标记的 *R. emersonii* 葡糖淀粉酶 (AG-FLAG) 的表达,该表达通过在 *R. emersonii* 培养物的上清液中表达的 6 种不同 *R. emersonii* 启动子来驱动。不同的泳道显示在表达以下启动子-报告子表达构建体的转化体的上清液中 AG-FLAG 的表达:泳道 1:pEBA540 (携带 *A.nidulans* gpdA 启动子);泳道 2:pEBA528 (携带 *R. emersonii* 启动子 1);泳道 3:pEBA529 (携带 *R. emersonii* 启动子 2);泳道 4:pEBA530 (携带 *R. emersonii* 启动子 3);泳道 5:pEBA531 (携带 *R. emersonii* 启动子 4);泳道 6:pEBA532 (携带 *R. emersonii* 启动子 5);泳道 7:pEBA533 (携带 *R. emersonii* 启动子 6);以及泳道 8:空菌株。

[0042] 序列表

[0043] SEQ ID NO:1 *R. emersonii* 纤维二糖水解酶 -I 启动子

[0044] SEQ ID NO:2 *R. emersonii* 乙酰木聚糖酯酶启动子

[0045] SEQ ID NO:3 *R. emersonii* 内切葡聚糖酶启动子

[0046] SEQ ID NO:4 *R. emersonii* 纤维二糖水解酶 -II 启动子

[0047] SEQ ID NO:5 *R. emersonii* β - 葡糖苷酶启动子

[0048] SEQ ID NO:6 *A.nidulans* gpdA 启动子

[0049] SEQ ID NO:7 *R. emersonii* RePepA (包括侧翼的基因组序列)

[0050] SEQ ID NO:8 *R. emersonii* RePepA (cDNA)

[0051] SEQ ID NO:9 *R. emersonii* RePepA (蛋白质)

[0052] SEQ ID NO:10 *A.nidulans* gpdA 启动子和 ble 编码区的 5' 部分

[0053] SEQ ID NO:11 ble 编码区的 3' 部分和 *A.nidulans* TrpC 终止子

[0054] SEQ ID NO:12 *R. emersonii* 启动子 1

- [0055] SEQ ID NO:13R. *emersonii* 启动子 2
- [0056] SEQ ID NO:14R. *emersonii* 启动子 3
- [0057] SEQ ID NO:15R. *emersonii* 启动子 4
- [0058] SEQ ID NO:16R. *emersonii* 启动子 5
- [0059] SEQ ID NO:17R. *emersonii* 启动子 6
- [0060] SEQ ID NO:18FLAG 标记的 *R. emersonii* 葡糖淀粉酶 (蛋白质)
- [0061] SEQ ID NO:19FLAG 标记的 *R. emersonii* 葡糖淀粉酶 (DNA, 编码区) 和 *A. nidulans* AmdS 终止子
- [0062] SEQ ID NO:20ReKu80 基因组序列, 具有侧翼的编码区
- [0063] SEQ ID NO:21ReKu80cDNA 序列
- [0064] SEQ ID NO:22ReKu80 蛋白序列
- [0065] SEQ ID NO:23ReGla cDNA 序列
- [0066] SEQ ID NO:24ReGla 蛋白序列
- [0067] 发明详述

[0068] 当今, 基因组学项目使用功能性基因组学方法以确定用于工业和环境应用的新的真菌酶。基于公知的序列来注释基因组 DNA 序列。许多酶在起源微生物进化过程中看起来高度保守。然而, 对于启动子而言, 很难观察到任何保守, 即使在密切相关的物种中, 同一性也普遍少于 5%。因此, 必须开发其他的策略以寻找到新的和有效的启动子。

[0069] 在本发明的上下文中, 启动子 DNA 序列是: 当该启动子 DNA 序列与编码序列可操作相连时, 能够控制该编码序列表达的 DNA 序列。术语“可操作相连 (in operative association)”在本文中被定义为下述结构, 其中启动子 DNA 序列被放置于相对编码序列来说合适的位置, 使得启动子 DNA 序列指导被编码序列编码的产物的生产。

[0070] 术语“编码序列”在本文中被定义为: 当放置在合适的控制序列的控制下时, 被转录为 mRNA, mRNA 被翻译为多肽的核酸序列。编码序列的边界通常通过 ATG 起始密码子 (其通常是 mRNA 5' 端的开放阅读框的开始) 和转录终止子序列 (处于紧邻 mRNA 3' 端开放阅读框的下游) 界定。编码序列可以包括但不限于基因组 DNA, cDNA, 半合成、合成和重组核酸序列。

[0071] 更具体而言, 术语“启动子”在本文中被定义为下述 DNA 序列, 其与 RNA 聚合酶结合, 并指导聚合酶到编码多肽的编码序列的正确下游转录开始位点以起始转录。RNA 聚合酶有效地催化与编码区的合适 DNA 链互补的信使 RNA 的组装。术语“启动子”还将被理解为包括转录为 mRNA 之后用于翻译的 5' 非编码区 (介于启动子和翻译起始之间)、顺式作用转录控制元件 (例如增强子) 以及能与转录因子相互作用的其它核苷酸序列。

[0072] 术语“强启动子”在本文中被定义为: 在合适生长条件下, 在含有作为碳源的 2.4% 葡萄糖或者 2% 纤维素的合适的营养培养基中, 与 *A. nidulans* *gpdA* 启动子相比, 给予报告蛋白更多表达的启动子。合适的报告蛋白的例子为 FLAG 标记的内切葡聚糖酶 (在实施例 2 中描述) 和 FLAG 标记的葡糖淀粉酶 (在实施例 5 中描述)。优选地, 将启动子-报告子构建体的一个拷贝整合至特定基因座, 从而防止由于拷贝数或者整合至基因组的位置造成的表达差异 (在实施例 5 中描述)。用于比较启动子活性的合适的营养培养基和生长条件取决于宿主。例如, 细胞可在实验室或者工业发酵器中通过摇瓶培养、小规

模或者大规模发酵（包括连续、分批、分批补料（fed-batch）或者固态发酵）来培养，其在合适培养基和使得启动子-报告子基因被表达的条件下进行。在包含碳源和氮源以及无机盐的合适的营养培养基中使用本领域已知的工序进行培养（对于丝状真菌宿主而言，参见例如，Bennett, J. W. 和 LaSure, L. 编, *More Gene Manipulations in Fungi*, Academic Press, CA, 1991）。测定 *Rasamsonia* 中启动子活性的具体条件的例子在实施例 2 和实施例 5 中描述。*Trichoderma* 中合适的营养培养基和生长条件的例子在 Zou 等人, 2012. Construction of a cellulase hyper-expression system in *Trichoderma reesei* by promoter and enzyme engineering. *Microb Cell Fact.*, 2012 Feb 8; 11(1): 21 中描述；以及用于 *Aspergillus* 情况下的合适营养培养基和生长条件的例子在 EP 635574 中描述。

[0073] 术语“诱导型启动子”定义为：活性通过存在或者不存在生物或者非生物因素（例如木质纤维素的酶促水解衍生的化合物、金属、温度或者光）诱导的启动子。木质纤维素的酶促水解衍生的化合物的例子为槐糖（sophorose）、龙胆二糖、纤维二糖和木糖。

[0074] 编码序列或者目标基因的过表达是指目标蛋白的表达和 / 或分泌与之前情况相比（例如在连同使能在亲本细胞中表达的编码序列引入启动子之前）是新的或者增加的。

[0075] 能高水平表达的基因（即高表达的基因）在本文中被定义为：在例如诱导条件下，其 mRNA 可占总细胞 mRNA 的至少 0.5% (w/w) 的基因；或者，其基因产物可占总细胞蛋白的至少 1% (w/w) 的基因；或者在分泌的基因产物的情况下，可分泌至至少 0.1g/l 水平的基因（如 EP 357127B 1 所述）。

[0076] 在一个优选的实施方式中，启动子为任何 *Rasamsonia* 启动子。为转录基因所挑选的特定启动子的选择取决于培养基条件，其中启动子应当为有活性的。此外，启动子的强度是启动子选择的标准。启动子的强度取决于宿主菌株和发酵条件。通过在特定发酵条件下生长丝状宿主和通过使用例如微阵列分析、定量 RT-PCR 或者 RNA 测序来定量转录物水平可确定优选的启动子。使用本领域技术人员已知的标准方法能够进行微阵列分析，例如通过在 Kiryu 等人, 2005. Extracting relations between promoter sequences and their strengths from microarray data. *Bioinformatics* 21(7): 1062-1068 中描述的方法。使用本领域技术人员已知的标准方法能够进行 RNA 的测序，例如使用第二代测序技术，例如 Illumina GA2、Roche 454 等，如在 Pareek 等人, 2011 Sequencing technologies and genome sequencing, *J Appl Genetics* 52: 413-435 中所综述。或者，使用 MALDI-TOF 分析、LC-MS 或者 LC/MS-MS，通过蛋白质组学研究能确定感兴趣的启动子，其中根据表达的蛋白量能选择启动子。

[0077] 通过定量转录物，可以评估给定条件下的启动子强度。比较在不同条件下的强度使得可以确定条件特异性诱导型启动子。或者，可识别在不同条件下激活和组成型激活的启动子。

[0078] 在一个优选的实施方式中，本发明的启动子 DNA 序列为以下列表中所示的 DNA 序列：SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 14、SEQ ID NO: 15、SEQ ID NO: 16 或者 SEQ ID NO: 17。

[0079] 根据另一优选的实施方式，本发明的启动子 DNA 序列是能够与以下列表所示的 DNA 序列杂交且仍然保持启动子活性的 DNA 序列：SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID NO: 11、SEQ ID NO: 13、SEQ ID NO: 14、SEQ ID NO: 15、

SEQ ID NO:16 或者 SEQ ID NO:17,。

[0080] 在本发明上下文中,优选通过测量编码序列表达所产生的蛋白质的浓度来确定启动子活性,该编码序列与该启动子可操作相连。或者,通过测量由编码序列编码的蛋白质的酶活性来确定启动子的活性,该编码序列与该启动子可操作相连。根据一个优选的实施方式,通过测量 lacZ 报告基因编码序列的表达(在 Luo(Gene163(1995)127-131) 或者通过测量 FLAG 标记的蛋白例如 FLAG 标记的葡糖淀粉酶(参见实施例)来测定启动子活性(及其强度)。根据另一优选的实施方式,通过使用绿色荧光蛋白作为编码序列来测定启动子的活性(在 Microbiology. 1999Mar ;145(Pt 3):729-34. Santerre Henriksen AL, Even S, Muller C, Punt PJ, van den Hondel CA, Nielsen J. Study)。此外,通过测量在启动子控制下产生的转录物的 mRNA 水平能够测定启动子活性。mRNA 水平能例如通过 Northern 印迹来测量(J. Sambrook, E. F. Fritsch 和 T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 第 2 版, Cold Spring Harbor, N. Y.)。在用于确定启动子活性的所有所描述的测定中,可将该启动子活性与另一启动子活性相比较,这例如通过将相同的报告基因或者编码序列置于不同启动子控制下并在相同条件下测量启动子活性来实现。

[0081] 本发明涵盖在非常低严格条件,优选低严格条件,更优选中严格条件,更优选中-高严格条件,甚至更优选高严格条件,和最优选非常高严格条件下与核酸探针的互补链杂交的(分离的)启动子 DNA 序列,所述核酸探针对应于:

[0082] a. SEQ ID NO:1 或者 SEQ ID NO:2 的第 1 至 1494 位核苷酸,优选第 100 至 1494,更优选 200 至 1494,甚至更优选 300 至 1494,甚至更优选 350 至 1494 以及最优选 360 至 1494 位核苷酸;

[0083] b. SEQ ID NO:2 的第 1 至 1482 位核苷酸,优选第 100 至 1482,更优选 200 至 1482,甚至更优选 300 至 1482,甚至更优选 350 至 1482 以及最优选 360 至 1482 位核苷酸;

[0084] c. SEQ ID NO:3 或者 SEQ ID NO:4 的第 1 至 1503 位核苷酸,优选第 100 至 1503,更优选 200 至 1503,甚至更优选 300 至 1503,甚至更优选 350 至 1503 以及最优选 360 至 1503 位核苷酸;

[0085] d. SEQ ID NO:5 的第 1 至 1979 位核苷酸,优选第 100 至 1979,更优选 200 至 1979,甚至更优选 300 至 1979,甚至更优选 350 至 1979 以及最优选 360 至 1979 位核苷酸;

[0086] e. SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:16 或者 SEQ ID NO:17 的第 1 至 1501 位核苷酸,优选第 100 至 1501,更优选 200 至 1501,甚至更优选 300 至 1501,甚至更优选 350 至 1501 以及最优选 360 至 1501 位核苷酸;或者

[0087] f. SEQ ID NO:15 的第 1 至 651 位核苷酸,优选第 50 至 651,更优选 100 至 651,甚至更优选 150 至 651,甚至更优选 200 至 651 以及最优选 250 至 651 位核苷酸。

[0088] 术语互补链为本领域技术人员所已知,并在 J. Sambrook, E. F. Fritsch 和 T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 第 2 版, Cold Spring Harbor, N. Y 中描述。

[0089] 如本文所使用,术语“杂交”旨在描述用于杂交和洗涤的条件,在该条件下,彼此之间至少约 60%、至少约 70%、至少约 80%、更优选至少约 85%、甚至更优选至少约 90%、更优选至少 95%、更优选至少 98% 或者更优选至少 99% 同源的核苷酸序列通常保持彼此杂交。

[0090] 这些杂交条件的一个优选、非限制性例子为在约 45℃ 下在 6X 氯化钠 / 柠檬酸钠 (SSC) 中杂交,然后在 50℃ 下,优选在 55℃ 下,优选在 60℃ 下以及甚至更优选在 65℃ 下在 1X SSC,0.1% SDS 中洗涤一次或者更多次。

[0091] 高严格条件包括例如在 68℃ 下在 5x SSC/5x Denhardt's 溶液 /1.0% SDS 中杂交和在室温下在 0.2x SSC/0.1% SDS 中洗涤。或者,可在 42℃ 下进行洗涤。

[0092] 技术人员将知道严格和高严格杂交条件适用何种条件。容易在本领域中获得关于这些条件的另外的指导,例如,在 Sambrook 等人,1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N. Y. ; 和 Ausubel 等人 (编辑),1995, *Current Protocols in Molecular Biology*, (John Wiley&Sons, N. Y.) 中。

[0093] 当然,仅与多聚腺苷酸序列 (例如 mRNA 的 3' 末端 poly(A) 段 (tract)) 或者仅与 T (或者 U) 残基的互补性延伸区段 (stretch) 杂交的多核苷酸将不被包括在用于与本发明核酸的一部分特异性杂交的本发明多核苷酸中,因为此类多核苷酸将与含有 poly(A) 延伸区段或者其互补物的任何核酸分子杂交 (例如,几乎任何双链 cDNA 克隆)。

[0094] SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 的子序列 (subsequence) 可以为至少 100 个核苷酸、优选至少 200 个核苷酸、更优选至少 300 个核苷酸、甚至更优选至少 400 个核苷酸以及最优选至少 500 个核苷酸。

[0095] SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 的核酸序列或者其子序列可被用于设计核酸探针,以按照本领域公知的方法来识别和克隆来自不同属或者种的菌株的 DNA 启动子。尤其,此类探针可被用于按照标 Southern 印迹法工序与目标属或者种的基因组或者 cDNA 杂交,以识别和分离其中相应的基因。此类探针可比完整序列短很多,但是长度应当为至少 15 个,优选至少 25 个,更优选至少 35 个核苷酸。此外,此类探针可被用于通过 PCR 来扩增 DNA 启动子。也可以使用更长的探针。可以使用 DNA、RNA 和肽核酸 (PNA) 探针。典型地,对探针进行标记,用于探测相应的基因 (例如,用 ³²P、³³P、³H、³⁵S、生物素或者亲和素或者荧光标记物)。此类探针被涵盖在本发明中。

[0096] 因此,可对从所述其它生物制得的基因组 DNA 或者 cDNA 文库加以筛选,选出与上述探针杂交并编码多肽的 DNA。可通过琼脂糖或者聚丙烯酰胺凝胶电泳或者其它分离技术来分离来自此类其它生物的基因组或者其它 DNA。可将来自文库的 DNA 或者分离的 DNA 转移并固定到硝酸纤维素或者其它合适的载体材料上。为识别出与 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 或者其子序列同源的克隆或者 DNA,可将运载体材料用于 Southern 印迹中。

[0097] 就本发明的目的而言,杂交指:核酸序列与对应于 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 所示核酸序列的互补链的经标记核酸探针在非常低至非常高严格条件下杂交。用例如 X 射线膜来探测在这些条件下与核酸探针杂交的分子。也可使用其它杂交技术,例如,使用荧光来探测以及玻璃片 (glass slide) 和 / 或 DNA 微阵列作为支持物的技术。FEMS Yeast Res. 2003Dec ;4 (3) :259-69 (Daran-Lapujade

P, Daran JM, Kotter P, Petit T, Piper MD, Pronk JT. “Comparative genotyping of the *Saccharomyces cerevisiae* laboratory strains S288C and CEN.PK113-7D using oligonucleotide microarrays”中给出了DNA微阵列杂交探测的例子。此外, PNA微阵列用于杂交的用途在 *Nucleic Acids Res.* 2003 Oct 1; 31(19): e119 (Brandt O, Feldner J, Stephan A, Schroder M, Schnolzer M, Arlinghaus HF, Hoheisel JD, Jacob A. PNA microarrays for hybridisation of unlabelled DNA samples) 中描述。

[0098] 在一个优选的实施方式中, 核酸探针为 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 的核酸序列。在另一优选的实施方式中, 核酸探针为具有以下序列:

[0099] a. SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:16 或者 SEQ ID NO:17 的第 20 至 1480 位核苷酸, 更优选 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:16 或者 SEQ ID NO:17 的第 500 至 1480 位核苷酸, 甚至更优选 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:16 或者 SEQ ID NO:17 的第 800 至 1480 位核苷酸, 以及最优选 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:16 或者 SEQ ID NO:17 的第 900 至 1480 位核苷酸; 或者

[0100] b. SEQ ID NO:15 的第 20 至 651 位核苷酸, 更优选 SEQ ID NO:15 的第 100 至 651 位核苷酸, 甚至更优选 SEQ ID NO:15 的第 200 至 651 位核苷酸, 以及最优选 SEQ ID NO:15 的第 300 至 651 位核苷酸; 或者

[0101] 另一优选探针为紧邻转录起始位点之前的 DNA 序列的部分。

[0102] 对于长度为至少 100 个核苷酸的长探针而言, 非常低至非常高严格条件被定义为: 按照标准 Southern 印迹法工序, 在 42 摄氏度下在以下中进行预杂交和杂交: 5 倍 SSPE、0.3% SDS、200 微克/ml 经剪切和变性的鲑鱼精 DNA, 以及对于非常低和低严格而言的 25% 甲酰胺; 对于中等和中-高严格而言的 35% 甲酰胺; 或者对于高严格和非常高严格而言的 50% 甲酰胺。

[0103] 对于长度为至少 100 个核苷酸的长探针而言, 载体材料被最终洗涤三次, 每次 15 分钟, 其中使用 2 倍 SSC、0.2% SDS, 优选至少在 45 摄氏度 (非常低严格), 更优选在至少 50 摄氏度 (低严格), 更优选在至少 55 摄氏度 (中严格), 更优选在至少 60 摄氏度 (中-高严格), 甚至更优选在至少 65 摄氏度 (高严格) 以及最优选在至少 70 摄氏度 (非常高严格) 下进行。

[0104] 对于长度为大约 15 个核苷酸至大约 70 个核苷酸的短探针而言, 严格条件被定义为: 预杂交、杂交和杂交后洗涤在比根据 Bolton 和 McCarthy (1962, *Proceedings of the National Academy of Sciences USA* 48:1390) 的算式计算得到的 T_m 低 5 摄氏度至 10 摄氏度的温度下, 在 0.9M NaCl、0.09M pH 7.6 的 Tris-HCl、6mM EDTA、0.5% NP-40、1 倍 Denhardt's 溶液、1mM 焦磷酸钠、1mM 磷酸二氢钠、0.1mM ATP 和 0.2mg 酵母 RNA/ml 中, 按照标准 Southern 印迹法工序来进行。

[0105] 对于长度为大约 15 个核苷酸至大约 70 个核苷酸的短探针而言,载体材料如下洗涤:使用 6 倍 SSC 加 0.1% SDS 洗涤一次,15 分钟;用 6 倍 SSC 洗两次,每次 15 分钟,在比计算出的 T_m 低 5 摄氏度至 10 摄氏度下进行。

[0106] 根据另一优选的实施方式,SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 首先被用于克隆与之可操作相连的天然基因、编码序列或者其部分。这可以用前文定义的 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 或者其子序列来开始,用该序列作为探针。将探针与给定宿主 (*Rasamsonia emersonii* 或者本申请中定义的任何其它宿主) 的 cDNA 或者基因组文库杂交。一旦天然基因或者其部分已被克隆,随后就可以通过如本文所述的杂交实验,用其自身作为探针去克隆来自其它真菌的其同源基因。

[0107] 在本发明的上下文中,同源基因是指与天然基因至少 50% 同源(相同)的基因。优选地,同源基因与天然基因至少 55% 同源,更优选至少 60%,更优选至少 65%,更优选至少 70%,甚至更优选至少 75%,优选约 80%,更优选约 90%,甚至更优选约 95%,甚至更优选约 97%,甚至更优选约 98%,甚至更优选约 99%,以及最优选约 99.5% 同源。

[0108] 同源基因编码序列上游的序列是本发明所涵盖的启动子。或者,使用例如本文所述的比对或者 BLAST 算法,可以用前文定义的 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 或者其子序列去检索基因组数据库来识别对与本发明启动子可操作相连的天然基因、编码序列或者其部分的序列。该识别出的序列随后可被用于识别本申请所定义的任何其它宿主中的直系同源(orthologue)或者同源基因。识别出的直系同源或者同源基因的编码序列上游序列是本发明所涵盖的启动子。

[0109] 根据另一优选的实施方式,本发明的启动子 DNA 序列是(分离的)DNA 序列,其与 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 至少 50% 同源(相同)。优选地,DNA 序列与 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 至少 55% 同源,更优选至少 60%,更优选至少 65%,更优选至少 70%,甚至更优选至少 75%,优选约 80%,更优选约 90%,甚至更优选约 95%,甚至更优选约 97%,甚至更优选约 98%,甚至更优选约 99%,以及最优选约 99.5% 同源。

[0110] 就本发明的目的而言,两条核酸序列之间同源性(同一性)的程度优选通过 BLAST 程序来确定。用于进行 BLAST 分析的软件是公众可通过国家生物信息中心(National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/>) 获得的。BLAST 算法参数 W、T 和 X 确定了比对的灵敏度和速度。BLAST 程序使用缺省值,字长(W)为 11, BLOSUM62 分数矩阵(参见 Henikoff&Henikoff, Proc. Natl. Acad. Sci. USA 89:10915(1989)) 比对(B)为 50,期望(E)为 10, M = 5, N = -4,以及对两条链比较。

[0111] 术语“同源性”、“同一性”或者“同一性百分比”在本文中可互换使用。就本发明的目的而言,在此定义为:为确定两条氨基酸序列或者两条核酸序列的同一性百分比,本着最

佳比较的目的（例如，可以在第一条氨基酸序列或者核苷酸序列上引入缺口，以与第二条氨基酸序列或者核苷酸序列最佳比对）来对序列进行比对。然后对相应氨基酸位置或者核苷酸位置上的氨基酸残基或者核苷酸进行比较。如果第一条序列上某位置的氨基酸残基或者核苷酸与第二条序列上相应位置的相同，那么分子在此位置就是相同的。两条序列间的同一性百分比是序列共有的相同位置的数量的函数（即，同一性% = 相同位置的数量 / 位置（即重叠位置）总数 × 100）。优选地，两条序列长度相同。

[0112] 在另一优选的实施方式中，启动子是 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 的子序列，该子序列仍然具有启动子活性。子序列优选含有至少约 100 个核苷酸，更优选至少约 200 个核苷酸，以及最优选至少约 300 个核苷酸。

[0113] 在另一优选的实施方式中，子序列是由已缺失 5' 和 / 或 3' 端的一个或者更多个核苷酸的 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 涵盖的核酸序列，该 DNA 序列仍然具有启动子活性。

[0114] 在另一优选的实施方式中，启动子序列是“经修整的 (trimmed)”的子序列，即，翻译起始处和 / 或转录起始处上游的序列片段。对启动子进行修整以及功能分析的例子见 Gene. 1994Aug 5;145(2):179-87:the effect of multiple copies of the upstream region on expression of the *Aspergillus niger* glucoamylase-encoding gene. Verdoes JC, Punt PJ, Stouthamer AH, van den Hondel CA) 中所述。

[0115] 在本发明的另一实施方式中，启动子 DNA 序列是 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:11、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 的变体。

[0116] 术语“变体”或者“变体启动子”在本文中被定义为具有下述核苷酸序列的启动子，所述核苷酸序列包含亲本启动子一个或者更多个核苷酸的替换、缺失和 / 或插入，其中变体启动子较之相应的亲本启动子具有更多或者更少的启动子活性。这些替换、缺失和 / 或插入的长度可不同，例如 1-1000 个核苷酸，优选 1-100 个核苷酸，更优选 1-20 个核苷酸，甚至更优选 1-10 个核苷酸，还更优选 1-6 个核苷酸，以及最优选 1-3 个核苷酸，但仍然得到具有启动子活性的生物活性多核苷酸。

[0117] 术语“变体启动子”涵盖天然变体和用本领域公知的方法例如经典诱变、定点诱变和 DNA 改组 (shuffling) 获得的体外产生的变体。变体启动子可以具有一处或者更多处突变。每处突变是独立的核苷酸的替换、缺失和 / 或插入。

[0118] 根据一个优选的实施方式，变体启动子为：较之最初识别出的启动子序列 (SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17) 具有至少一处经修饰的调控位点的启动子。此类调控位点可以被整体去除或者按照上文的解释进行特异性突变。此类启动子变体的调控被修饰从而例如其不再被葡萄糖所诱导。此类启动子变体和如何获得它们的技术的例子在 EP 673429 或者 W094/04673 中描述。

[0119] 启动子变体可以是等位变体。等位变体是指占据相同染色体基因座的基因的两种或者更多种可替换形式中的任意种。等位变化是通过突变天然产生的，其能导致种群内的

多态性。可通过下述获得变体启动子：(a) 在非常低、低、中、中-高、高或者非常高严格条件下，使 DNA 与 (i) SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17；(ii) (i) 的子序列；或者 (iii) (i)、(ii) 的互补链杂交，以及 (b) 从 DNA 分离变体启动子。严格性和洗涤条件如本文所定义。

[0120] 本发明的启动子可以是其序列可提供有接头的启动子，所述接头用于如下目的：引入特定限制位点从而促进启动子序列与编码多肽的核酸序列编码区之间的连接。

[0121] 本文提供的序列信息不应被狭义理解为需要包括被错误识别的碱基。本文公开的特定序列可容易地被用于从优选的丝状真菌（特别是 *Rasamsonia*）分离原始 DNA 序列，再对其进行进一步的序列分析，由此识别出测序错误。

[0122] 除非另有指明，本文中通过对 DNA 分子进行测序测定的所有核苷酸序列都是用自动 DNA 测序仪测定的。因此，如本领域关于通过该自动方法测定的任何 DNA 序列所已知的，本文测定的任何核苷酸序列可能含有一些错误。通过自动方法测定的核苷酸序列典型地与被测 DNA 分子的实际核苷酸序列至少大约 90% 相同，更典型地，至少约 95% 相同至至少大约 99.9% 相同。可以通过包括本领域公知的手动 DNA 测序方法的其它方法对实际序列进行更为精确的测定。

[0123] 本领域技术人员能识别出此类被错误识别的碱基，并且知道如何更正此类错误。

[0124] 本发明涵盖启动子的功能等同物，其典型地含有不改变相关启动子生物功能的突变。术语“功能等同物”也涵盖 *Rasamsonia* DNA 序列的直系同源物。*Rasamsonia* DNA 序列的直系同源物是可从其它生物体、其他真菌物种或者菌株中分离并且具有相似或者相同生物活性的 DNA 序列。

[0125] 本发明的启动子序列可从任何属的微生物中获得。就本发明的目的而言，如本文所使用的与给定来源相关的术语“从……获得”应表示：多肽是该来源产生的或者是被来自该来源的基因所插入的细胞产生的。

[0126] 启动子序列可从真菌来源获得，优选 *Rasamsonia* 菌株，更优选 *Rasamsonia emersonii*。

[0127] *Rasamsonia* 是包括耐热和嗜热的 *Talaromyces* 和 *Geosmithia* 物种的新属 (J. Houbraken 等人，见上)。基于表型、生理和分子数据，Houbraken 等人提议将 *T. emersonii*、*T. byssochlamydoides*、*T. eburneus*、*G. argillacea* 和 *G. cylindrospora* 种改为 *Rasamsonia* gen. nov. *Talaromyces emersonii*、*Penicillium geosmithia emersonii* 和 *Rasamsonia emersonii* 在本文中可交换使用。

[0128] 应理解，对上述物种而言，本发明包括完美的和不完美的状态以及其它分类学等同物，例如，无性型 (anamorph)，而不管它们是以什么种名被知道的。本领域技术人员容易对合适的等同物加以鉴定。这些物种的菌株是公众容易从大量的培养单位获得的，例如，美国典型培养物保藏中心 (American Type Culture Collection (ATCC))、德国微生物和细胞培养物保藏中心 (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ))、荷兰真菌培养物保藏中心 (Centraalbureau Voor Schimmelcultures (CBS)) 和农业研究机构专利培养物保藏中兴北区研究中心 (Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL))。

[0129] 而且,可使用上文提到的探针,从其它来源(包括从自然界(例如,土壤、肥料、水等)分离的微生物)鉴定并获得根据本发明的启动子序列。用于从自然环境分离微生物的技术是本领域公知的。然后可以通过对其它微生物基因组 DNA 文库进行类似筛选来获得核酸序列。一旦用探针探测出了编码启动子的核酸序列,就可以利用本领域普通技术人员已知的技术来分离或者克隆该序列(例如,参见 Sambrook 等人,1989,见上)。

[0130] 在本发明中,启动子 DNA 序列还可以是杂合启动子,其包含本发明的一种或者更多种启动子的一部分;本发明的启动子的一部分以及另外的已知启动子的一部分,例如,一种启动子的前导序列(leader sequence)和来自其它启动子的转录起始位点;或者本发明的一种或者更多种启动子的一部分以及一种或者更多种其它启动子的一部分。其它启动子可以是任何在选用的宿主细胞中显示出转录活性的启动子序列,包括变体的、截短的和杂合的启动子,并且其可以从编码对于宿主细胞来说同源或者异源的、细胞外或者细胞内多肽的基因获得。其它启动子序列可以是对编码多肽的核酸序列来说天然的或者外源的,并且对细胞来说可以是天然的或者外源的。

[0131] 作为一个优选的实施方式,鉴定出的启动子的重要调控子序列可以与其它“基础”启动子融合,从而增强它们的启动子活性(例如,见 Mol Microbiol. 1994May; 12(3):479-90. Regulation of the xylanase-encoding xlnA gene of *Aspergillus tubigenensis*. de Graaff LH, van den Broeck HC, van Ooijen AJ, Visser J. 中所述)。

[0132] 可用于与本发明的启动子一起构建杂合启动子的其它启动子的其他例子包括:从 *A. oryzae* TAKA 淀粉酶、*Rhizomucor miehei* 天冬氨酸蛋白酶、*A. niger* 中性 α -淀粉酶、*A. niger* 酸稳定性 α -淀粉酶、*A. niger* 或者 *Aspergillus awamori* 葡糖淀粉酶(glaA)、*A. niger* gpdA、*A. niger* 葡萄糖氧化酶 goxC、*Rhizomucor miehei* 脂酶、*A. oryzae* 碱性蛋白酶、*A. oryzae* 磷酸丙糖异构酶、*A. nidulans* 乙酰胺酶和 *Fusarium oxysporum* 胰蛋白酶样蛋白酶(WO 96/00787)的基因获得的启动子,以及 NA2-tpi 启动子(来自 *A. niger* 中性 α -淀粉酶和 *A. oryzae* 磷酸丙糖异构酶基因的启动子的杂合体)、*Saccharomyces cerevisiae* 烯醇酶(ENO-1)、*Saccharomyces cerevisiae* 半乳糖激酶(GAL1)、*Saccharomyces cerevisiae* 醇脱氢酶/甘油醛-3-磷酸脱氢酶(ADH2/GAP)和 *Saccharomyces cerevisiae* 3-磷酸甘油酸激酶的基因获得的启动子及其突变的、截短的和杂合的启动子。酵母宿主细胞的一些可使用启动子通过 Romanos 等人,1992, *Yeast* 8:423-488 描述。

[0133] 在本发明中,启动子 DNA 序列还可以是“串联启动子”。“串联启动子”在本文中被定义为两个或者更多个启动子序列,它们每一个都与编码序列可操作相连,并介导编码序列转录成 mRNA。

[0134] 串联启动子包含:本发明的两种或者更多种启动子,或者本发明的一种或者更多种启动子与一种或者更多种其它已知的启动子,例如上文例示的用于构建杂合启动子的那些。串联启动子的两种或者更多种启动子序列可以同时促进核酸序列的转录。或者,串联启动子的一种或者更多种启动子序列可以促进核酸序列在细胞不同生长阶段或者菌丝体不同形态部分的转录。

[0135] 在本发明中,启动子对于编码生物化合物的编码序列可以是外源的和/或启动子对于宿主细胞来说可以是外源的。本发明的变体、杂合或者串联启动子应当被理解为对于编码的编码序列来说是外源的,即使野生型启动子对编码序列或者宿主细胞是天然的。

[0136] 本发明的变体、杂合或者串联启动子的启动子活性为：具有 SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16 或者 SEQ ID NO:17 的启动子的启动子活性的至少约 20%、优选至少约 40%、更优选至少约 60%、更优选至少约 80%、更优选至少约 90%、更优选至少约 100%、甚至更优选至少约 200%、最优选至少约 300%、以及甚至最优选至少约 400%。启动子活性优选如说明书中前述部分所述来测定。

[0137] 本发明还涉及下述 DNA 构建体，该 DNA 构建体包含（至少一个）如上所述的启动子 DNA 序列和与该启动子 DNA 序列可操作相连的编码序列，从而编码序列可以在启动子 DNA 序列的控制下表达。这可在任何合适的宿主细胞中检测。或者，这可在合适的体外表达和 / 或翻译体系中检测。编码序列可从任何原核的、真核的或者其它来源中获得。或者，编码序列可以是合成的或者部分合成的序列。合成基因的密码子使用已经被最优化来匹配宿主细胞物种的密码子使用，从而促进所编码的生物物质的表达和 / 或分泌。优化密码子使用的例子描述于 WO97/11086 中，其中植物多肽的密码子使用被优化为在丝状真菌细胞中表达。优选地，编码序列编码生物化合物。两种或者更多种这些 DNA 构建体可连接以形成新的（串联）DNA 构建体。该新的（串联）构建体可包含两种或者更多种 DNA 构建体，其例如包含（启动子 - 开放阅读框 - 终止子）连接至（启动子 - 开放阅读框 - 终止子），该（启动子 - 开放阅读框 - 终止子）可任选连接下一个（启动子 - 开放阅读框 - 终止子）单元。在例如有 5 个线状排列单元的情况下，DNA 构建体优选包含 5 种不同的启动子，从而防止单元因重组而缺失。优选至少一种启动子为本发明的启动子。

[0138] 或者，编码序列可编码反义 RNA 和 / 或 RNAi (RNA 干扰) 构建体的表达。反义 RNA 表达的实例显示于 Appl Environ Microbiol. 2000Feb ;66(2):775-82. (Characterization of a foldase, protein disulfide isomerase A, in the protein secretory pathway of *Aspergillus niger*. Ngiam C, Jeenes DJ, Punt PJ, Van Den Hondel CA, Archer DB) 或者 (Zrenner R, Willmitzer L, Sonnewald U. Analysis of the expression of potato uridinediphosphate-glucose pyrophosphorylase and its inhibition by antisense RNA. Planta. (1993) ;190(2):247-52) 中。基因表达的完全灭活用于例如灭活控制不期望的代谢途径分支的基因，例如从而提高特定次级代谢产物例如 (β -内酰胺) 抗生素或者类胡萝卜素的产生。完全灭活也用于减少毒素或者非所需化合物的产生 (*Penicillium* 中的黄青霉素 ; *Aspergillus* 中的黄曲霉毒素 ; MacDonald KD 等人, heterokaryon studies and the genetic control of penicillin and chrysogenin production in *Penicillium chrysogenum*. J Gen Microbiol. (1963) 33:375-83)。完全灭活还用于改变生物的形态以改进发酵过程和下游加工。

[0139] 本发明的另一实施方式涉及对宿主细胞的广泛代谢程序重排或者工程改造。引入全新的途径和 / 或修饰非所需途径会提供特别适用于产生特定生物化合物（例如蛋白质或者代谢产物）的细胞。

[0140] 在本发明的方法中，当编码序列编码多肽时，该多肽还可包括融合的或者杂合的多肽，其中另一多肽与该多肽或者其片段的 N- 末端或者 C- 末端融合。融合的多肽通过将编码一个多肽的核酸序列（或者其一部分）与编码另一多肽的核酸序列（或者其一部分）融合来产生。用于产生融合多肽的技术为本领域已知，其包括：连接编码多肽的编码序列，

使得其同框并且使得融合多肽的表达位于相同的启动子和终止子控制下。杂合多肽可包含由至少两个不同多肽获得的部分或者完整多肽序列的组合,其中一个或者更多个所述多肽可以对真菌细胞而言异源。

[0141] 除启动子 DNA 序列外,DNA 构建体可包含一个或者更多个控制序列,启动子 DNA 序列指导编码序列在合适宿主细胞中在与控制序列相容的条件下表达。表达应理解为包括多肽生产中涉及的任何步骤,其包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。一个或者更多个控制序列可以对编码序列或者宿主是天然的。或者,一个或者更多个控制序列可以用对核酸序列而言为外源的一个或者更多个控制序列代替,用于改善编码序列在宿主细胞中的表达。

[0142] “DNA 构建体”在本文中被定义为单链或者双链的核酸分子,其分离自天然存在的基因,或者其经修饰而含有以自然中不会存在的方式组合和并置的核酸区段。当 DNA 构建体含有编码序列和编码序列表达所需的全部控制序列时,术语 DNA 构建体与术语表达盒同义。

[0143] 术语“控制序列”在本文中被定义为包括对编码序列表达必需或者有利的所有成分,包括本发明的启动子。各控制序列可以对编码多肽的核酸序列是天然的或者外源的。这类控制序列包括但不限于前导序列、翻译起始序列(如在 Kozak, 1991, J. Biol. Chem. 266:19867-19870 中所述)、翻译起始编码序列、聚腺苷酸化序列、前肽序列、信号肽序列、上游激活序列、本发明的启动子(包括源自它的变体、片段和杂合以及串联启动子)、转录终止子和翻译终止子。控制序列最少包括转录和翻译终止信号和(部分的)本发明启动子。控制序列可以与用于引入特异限制性位点的接头一起提供,以便于连接控制序列与编码多肽的核酸序列的编码区。

[0144] 控制序列可以是合适的转录终止子序列,即被宿主细胞识别以终止转录的序列。终止子序列与编码多肽的编码序列 3' 末端可操作相连。任何在所选宿主细胞中有功能的终止子可用于本发明。

[0145] 用于丝状真菌宿主细胞的优选终止子得自 *A. oryzae* TAKA 淀粉酶、*A. niger* 葡糖淀粉酶、*A. nidulans* 氨基苯甲酸合酶、*A. niger* α -葡糖苷酶、trpC 基因和 *Fusarium oxysporum* 胰蛋白酶样蛋白酶的基因。

[0146] 控制序列还可为合适的前导序列,即 mRNA 的 5' 非翻译区,其对宿主细胞翻译是重要的。前导序列与编码多肽的核酸序列的 5' 末端可操作相连。任何在所选宿主细胞中有功能的前导序列可用于本发明。

[0147] 用于丝状真菌宿主细胞的优选前导序列由 *A. oryzae* TAKA 淀粉酶、*A. nidulans* 磷酸丙糖异构酶和 *A. niger* glaA 的基因获得。

[0148] 控制序列还可以是聚腺苷酸化序列,该序列与核酸序列 3' 末端可操作相连并且经转录后被宿主细胞识别为向经转录的 mRNA 添加聚腺苷酸残基的信号。在所选宿主细胞中有功能的任何聚腺苷酸化序列可用于本发明。

[0149] 用于丝状真菌宿主细胞的优选的聚腺苷酸化序列由 *A. oryzae* TAKA 淀粉酶、*A. niger* 葡糖淀粉酶、*A. nidulans* 氨基苯甲酸合酶、*Fusarium oxysporum* 胰蛋白酶样蛋白酶和 *A. niger* α -葡糖苷酶获得。

[0150] 控制序列还可以是信号肽编码区,其编码与多肽的氨基末端连接并指导被编码的

多肽进入细胞分泌途径的氨基酸序列。核酸序列编码序列的 5' 端可固在地含有信号肽编码区,该信号肽编码区与编码被分泌多肽的编码区区段按翻译阅读框天然连接。或者,编码序列的 5' 端可含有对编码序列是外源的信号肽编码区。当编码序列天然不含有信号肽编码区时,可需要外源信号肽编码区。或者,外源信号肽编码区可简单地替换天然信号肽编码区,从而增强多肽的分泌。然而,指导所表达的多肽进入所选宿主细胞分泌途径的任何信号肽编码区可用于本发明。

[0151] 用于丝状真菌宿主细胞的有效信号肽编码区为由 *A. oryzae* TAKA 淀粉酶、*A. niger* 中性淀粉酶、*A. ficuum* 植酸酶、*A. niger* 葡糖淀粉酶、*A. niger* 内切木聚糖酶、*Rhizomucor miehei* 天冬氨酸蛋白酶、*Humicola insolens* 纤维素酶和 *Humicola lanuginosa* 脂酶的基因获得的信号肽编码区。

[0152] 用于酵母宿主细胞的有用信号肽由 *Saccharomyces cerevisiae* α -因子和 *Saccharomyces cerevisiae* 转化酶的基因获得。另一些有用的信号肽编码区描述于 Romanos et al., 1992, 见上。

[0153] 控制序列还可以是前肽编码区,其编码位于多肽氨基末端的氨基酸序列。得到的多肽称为前酶 (proenzyme) 或者前多肽 (或者一些情况下为酶原 (zymogen))。前多肽通常是无活性的,并可通过将前肽从前多肽上催化或者自身催化切割转化为成熟的活性多肽。前肽编码区可由 *Bacillus subtilis* 碱性蛋白酶 (aprE)、*Bacillus subtilis* 中性蛋白酶 (nprT)、*Saccharomyces cerevisiae* α -因子、*Rhizomucor miehei* 天冬氨酸蛋白酶、*Myceliophthora thermophila* 漆酶 (WO 95/33836) 和 *A. niger* 内切木聚糖酶 (end1) 的基因获得。

[0154] 当信号肽和前肽区均出现在多肽氨基末端时,前肽区处在与多肽的氨基末端相邻的位置,而信号肽区处在与前肽区的氨基末端相邻的位置。

[0155] 还可期望添加调节序列,其允许相对于宿主细胞的生长调节多肽表达。调节体系的实例为响应化学或者物理刺激 (包括调节化合物的存在) 引起基因的表达被打开或者关闭的那些调节体系。原核体系中的调节体系包括 lac 和 trp 操纵子体系。在酵母中,可使用 ADH2 体系或者 GAL1 体系。在丝状真菌中,可使用 TAKA α -淀粉酶启动子、*A. niger* 葡萄糖淀粉酶启动子、*A. oryzae* 葡萄糖淀粉酶启动子、*A. tubingensis* 内切木聚糖酶 (xlnA) 启动子、*A. niger* 硝酸还原酶 (niaD) 启动子、*Trichoderma reesei* 纤维二糖水解酶启动子和 *A. nidulans* 醇和醛脱氢酶 (分别为 alcA 和 aldA) 启动子 (如 US 5,503,991 中所述) 作为调节序列。调节序列的其它实例为允许基因扩增的那些调节序列。在真核体系中,这些包括二氢叶酸还原酶基因 (其在存在氨甲蝶呤时被扩增) 和金属硫蛋白基因 (其在有重金属时被扩增)。在这些情况下,编码多肽的核酸序列应与调节序列可操作相连。

[0156] 去除 creA 结合位点 (如早先在 EP 673429 中所述的碳分解代谢产物抑制)、改变 pacC 和 areA (用于 pH 和氮调节) 可以是重要的。

[0157] 优选地, DNA 构建体包含本发明的启动子 DNA 序列、与该启动子 DNA 序列可操作相连的编码序列和翻译控制序列,例如:

[0158] - 选自以下序列列表的依照 5' 到 3' 方向的一个翻译终止序列: TAAG、TAGA 和 TAAA, 优选 TAAA, 和 / 或

[0159] - 选自以下序列列表的依照 5' 到 3' 方向的一个翻译起始编码序列: GCTACCCCC;

GCTACCTCC ;GCTACCCTC ;GCTACCTTC ;GCTCCCCC ;GCTCCCTCC ;GCTCCCCTC ;GCTCCCTTC ;GCTGCCCC ;GCTGCCTCC ;GCTGCCCTC ;GCTGCCTTC ;GTTCCCCC ;GTTCCCTCC ;GTTCCCTC 和 GTTCCCTTC, 优选 GCT TCC TTC ;和 / 或

[0160] - 选自以下序列列表的一个转录起始序列 :5'-mwChkyCAAA-3';5'-mwChkyCACA-3' 或者 5'-mwChkyCAAG-3', 采用核苷酸的不确定编码 :m(A/C) ;w(A/T) ;y(C/T) ;k(G/T) ;h(A/C/T), 优选 5'-CACCGTCAAA-3' 或者 5'-CGCAGTCAAG-3'。

[0161] 在本发明上下文中, 术语“翻译起始编码序列”被定义为 DNA 编码序列开放阅读框的起始子或者起始密码子下游紧邻的九个核苷酸。起始子或者起始密码子编码 AA 甲硫氨酸。起始密码子典型为 ATG, 但是也可以是任何功能性起始密码子, 例如 GTG。

[0162] 在本发明上下文中, 术语“翻译终止序列”被定义为从开放阅读框或者核苷酸编码序列 3' 端翻译终止密码子开始并依照 5' 到 3' 方向的三或者四个核苷酸。

[0163] 在本发明上下文中, 术语“翻译起始序列”被定义为编码多肽的 DNA 序列开放阅读框起始子或者起始密码子上游紧邻的十个核苷酸。起始子或者起始密码子编码 AA 甲硫氨酸。起始密码子典型为 ATG, 但是也可以是任何功能性起始密码子, 例如 GTG。本领域公知的是在 RNA 中, 尿嘧啶 U 替换脱氧核苷酸胸腺嘧啶 T。

[0164] 本发明还涉及包含本发明启动子、编码多肽的编码序列和转录和翻译起始子以及终止信号的重组表达载体。

[0165] 上述多种编码和控制序列可以结合在一起, 产生下述重组表达载体, 该重组表达载体可包含一个或者更多个便利的限制位点, 以允许在这类位点处插入或者替换启动子和 / 或编码多肽的编码序列。或者, 可通过例如 Gene. 1989Apr 15 ;77(1):51-9. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR“Site-directed mutagenesis by overlap extension using the polymerase chain reaction”) 中所述的使用 PCR 的序列重叠延伸 (SOE-PCR) 或者通过使用 Gateway™克隆体系 (Invitrogen) 克隆完成编码序列和启动子的融合。或者, 可通过将编码序列或者包含启动子和 / 或编码序列的 DNA 构建体插入适当的表达载体来表达编码序列。创建表达载体时, 编码序列以下述方式位于载体中 :编码序列与本发明启动子和一个或者更多个适当的表达控制序列可操作相连。

[0166] 重组表达载体可以是下述任何载体 (例如, 质粒或者病毒), 该载体可便利地进行重组 DNA 操作并可完成编码序列的表达。对载体的选择典型地会取决于载体与所述载体待被引入的宿主细胞间的相容性。载体可以是线性或者闭环质粒。

[0167] 载体可以是自主复制的载体, 即作为染色体外实体存在的载体, 其复制不依赖于染色体复制, 例如质粒、染色体外元件、微型染色体或者人工染色体。对自主复制而言, 载体可以包含使载体能够在所述宿主细胞中自主复制的复制起点。用于酵母宿主细胞的复制起点实例为 2 微米复制起点, ARS1, ARS4, ARS1 和 CEN3 的组合以及 ARS4 和 CEN6 的组合。复制起点可以是具有下述突变的复制起点, 该突变使宿主细胞中复制起点的功能对于温度敏感 (参见例如 Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75:1433)。丝状真菌中自主维持的克隆载体的实例为包含 AMA1- 序列的克隆载体。AMA1 为自 *A. nidulans* 分离的 6.0-kb 基因组 DNA 片段, 其能够在 *Aspergillus* 中自主维持 (参见例如 Aleksenko 和 Clutterbuck(1997), Fungal Genet. Biol. 21:373-397)。

[0168] 或者, 载体可以是引入宿主细胞时被整合至基因组和与将其整合至的染色体一起

复制的载体。而且,可以使用单个载体或者质粒或者一起包含要引入宿主细胞基因组中的全部 DNA 的两个或者更多个载体或者质粒或者转座子。

[0169] 本发明的载体优选含有一个或者更多个可选择标记物,该可选择标记物允许容易地选择经转化的细胞。宿主可以用至少两个载体共转化,其中一个包含可选择标记物。可选择标记物是下述基因,所述基因的产物提供对抗微生物剂或者病毒的抗性、对重金属的抗性、针对营养缺陷型的原营养等。用于酵母宿主细胞的合适标记物为 ADE2、HIS3、LEU2、LYS2、MET3、TRP1 和 URA3。用于丝状真菌宿主细胞的可选择标记物包括但不限于 amdS(乙酰胺酶)、argB(鸟氨酸氨甲酰基转移酶)、bar(膦丝菌素乙酰转移酶)、hygB(潮霉素磷酸转移酶)、niaD(硝酸还原酶)、pyrG(乳清苷-5'-磷酸盐脱羧酶)、sC(硫酸腺嘌呤基转移酶)、trpC(氨基苯甲酸合酶)及其等效物。也可使用提供针对例如腐草霉素、潮霉素 B 或者 G418 抗性的标记物。在 *Rasamsonia* 细胞中优选使用 ble 和 hygB 选择标记物。

[0170] 就整合至宿主细胞基因组而言,载体可依赖于启动子序列和/或编码多肽的编码序列或者载体的任何其它元件,用于通过同源重组或者非同源重组将载体稳定整合至基因组内。或者,载体可含有用于指导通过同源重组整合至宿主细胞基因组的额外的核酸序列。该额外的核酸序列使得载体能够在染色体中预先确定的靶位点整合至宿主细胞基因组。为了提高在精确位点整合的可能性,整合元件应优选含有足够数量的核酸,例如 30 到 1,500 个碱基对,优选 100 到 1,500 个碱基对,更优选 400 到 1,500 个碱基对,更优选 800 到 1,500 个碱基对并最优选至少 2kb,其与相应的靶序列高度同源以提高同源重组的可能性。整合元件可以是与宿主细胞基因组中靶序列同源的任何序列。而且,整合元件可以是非编码或者编码核酸序列。为了促进靶向的整合,克隆载体优选在转化宿主细胞之前线性化。线性化优选以下述方式进行:克隆载体的至少一端(但是优选两端)侧翼为与靶基因座同源的序列。

[0171] 优选地,克隆载体中与靶基因座同源的整合元件来自高度表达的基因座,即它们来自能够在真菌宿主细胞中高水平表达的基因。能够有高表达水平的基因(即高表达的基因)在本文中定义为下述基因,所述基因的 mRNA 可占据总细胞 mRNA 的至少 0.5% (w/w)(例如在经诱导的条件下),或者,所述基因的基因产物可占据总细胞蛋白质的至少 1% (w/w),或者在经分泌的基因产物情况下,可分泌至少 0.1g/l 的水平(如在 EP 357127B1 中所述)。大量优选的高表达真菌基因以实例的方式给出:来自 *Aspergilli* 或者 *Trichoderma* 的淀粉酶、葡糖淀粉酶、醇脱氢酶、木聚糖酶、甘油醛-磷酸脱氢酶或者纤维二糖水解酶的基因。

[0172] 另一方面,载体可以通过非同源重组整合至宿主细胞基因组。

[0173] 可将编码生物化合物的核酸序列的多于一个拷贝插入宿主细胞中,以促进基因产物的产生。这可优选通过将 DNA 序列的数个拷贝整合进其基因组中,更优选通过将 DNA 序列的整合靶向高表达基因座来完成。或者,这可通过在核酸序列中包含可扩增的可选择标记物基因,使含有经扩增拷贝的可选择标记物基因从而含有额外拷贝的核酸序列的细胞可通过在存在适当可选择剂时培养细胞而被选择。

[0174] 用于连接上述元件以构建本发明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,见上)。

[0175] 本发明还涉及包含与编码序列可操作相连的本发明启动子 DNA 序列的重组宿主

细胞,该宿主细胞有利地用于生产生物化合物。将包含与编码序列可操作相连的本发明启动子的载体引入宿主细胞,使得载体作为染色体整合体或者作为前文所述的自主复制染色体外载体而保持。术语“宿主细胞”涵盖亲本细胞的任何子代,该子代由于复制中发生的突变而与亲本细胞不同。宿主细胞的选择很大程度上取决于编码序列的来源和本发明启动子的来源。技术人员知晓如何选择最合适的宿主细胞。

[0176] 本发明还涉及重组宿主细胞,该宿主细胞包含多于一个本发明的启动子 DNA 序列,各启动子优选地与编码序列可操作相连。这类宿主细胞可有利地用于至少一种生物化合物的重组生产。或者,本发明的重组宿主细胞可包含与本领域已知启动子组合的一个或者更多个本发明的启动子。本领域已知的这类启动子包括但不限于由下述基因获得的启动子:A. tubigensis xlnA、A. oryzae TAKA 淀粉酶、Rhizomucor miehei 天冬氨酸蛋白酶、A. niger 中性 α -淀粉酶、A. niger 酸稳定性 α -淀粉酶、A. niger 或者 A. awamori 葡糖淀粉酶 (glaA)、A. niger 或者 A. awamori 内切木聚糖酶 (xlnA) 或者 β -木糖苷酶 (xlnD)、T. reesei 纤维二糖水解酶 I (CBHI)、R. miehei 脂酶、A. oryzae 碱性蛋白酶、A. oryzae 磷酸丙糖异构酶、A. nidulans 乙酰胺酶、Trichoderma reesei β -葡糖苷酶、Trichoderma reesei 纤维二糖水解酶 I、Trichoderma reesei 纤维二糖水解酶 II、Trichoderma reesei 内切葡聚糖酶 I、Trichoderma reesei 内切葡聚糖酶 II、Trichoderma reesei 内切葡聚糖酶 III、Trichoderma reesei 内切葡聚糖酶 IV、Trichoderma reesei 内切葡聚糖酶 V、Trichoderma reesei 木聚糖酶 I、Trichoderma reesei 木聚糖酶 II、Trichoderma reesei β -木糖苷酶,以及 NA2-tpi 启动子(来自编码 A. niger 中性 α -淀粉酶和 A. oryzae 磷酸丙糖异构酶的多核苷酸的启动子杂合物)及其突变的、截短的和杂合的启动子。启动子的其他例子是描述于 W02006/092396 和 W02005/100573 中的启动子,其通过引用方式并入本文。启动子使用的又一其他例子于 W02008/098933 中描述。诱导型(异源)启动子的实例为醇诱导型启动子 alcA、使用四环素响应启动子的 tet 系统、雌激素响应启动子(Pachlinger 等人(2005),Appl&Environmental Microbiol 672-678)。

[0177] 本发明的宿主细胞和本发明方法中使用的宿主细胞可以是任何宿主细胞。优选地,本发明的宿主细胞为真菌细胞。如本文所使用的“真菌”包括子囊菌亚门(Ascomycota)、担子菌亚门(Basidiomycota)、壶菌门(Chytridiomycota)和接合菌门(Zygomycota)(如由 Hawksworth 等人, Ainsworth and Bisby's Dictionary of The Fungi, 第 8 版, 1995, CAB International, University Press, Cambridge, UK 中所定义)以及卵菌门(Oomycota)(如 Hawksworth 等人, 1995, 见上, 171 页所述)和所有的有丝分裂孢子真菌(Hawksworth 等人, 1995, 见上)。

[0178] 在一个优选的实施方式中,真菌宿主细胞为丝状真菌细胞。“丝状真菌”包括真菌门(Eumycota)和卵菌门亚门的所有丝状形式(如 Hawksworth 等人, 1995, 见上所定义)。丝状真菌表征为由几丁质、纤维素、葡聚糖、壳聚糖、甘露聚糖和其它复杂多糖构成的菌丝体壁。营养生长通过菌丝伸长,并且碳分解代谢为专性需氧。相反,酵母如 Saccharomyces cerevisiae 的营养生长通过单细胞菌体出芽实现且碳分解代谢可以是发酵的。

[0179] 优选地,丝状真菌宿主细胞为以下属的细胞:Acremonium、Agaricus、Aspergillus、Aureobasidium、Chrysosporium、Coprinus、Cryptococcus、Filobasidium、Fusarium、Geosmithia、Hemicola、Magnaporthe、Mucor、Myceliophthora、Neocallimastix、

Neurospora、Paecilomyces、Penicillium、Piromyces、Panerochaete、Pleurotus、Rasamsonia、Schizophyllum、Talaromyces、Thermoascus、Thermomyces、Thielavia、Tolypocladium 或者 Trichoderma。

[0180] 在一个更优选的实施方式中,丝状真菌宿主细胞为 *Humicola grisea* var. *thermoidea*、*Humicola lanuginosa*、*Myceliophthora thermophila*、*Papulaspora thermophila*、*Rasamsonia byssochlamydoides*、*Rasamsonia emersonii*、*Rasamsonia argillacea*、*Rasamsonia eburnean*、*Rasamsonia brevistipitata*、*Rasamsonia cylindrospora*、*Rhizomucor pusillus*、*Rhizomucor miehei*、*Talaromyces bacillisporus*、*Talaromyces leycettanus*、*Talaromyces thermophilus*、*Thermomyces lanuginosus*、*Thermoascus crustaceus*、*Thermoascus thermophilus* *Thermoascus aurantiacus* 或者 *Thielavia terrestris* 细胞。在另一更优选的实施方式中,丝状真菌宿主细胞为 *Aspergillus awamori*、*Aspergillus foetidus*、*Aspergillus japonicus*、*A. nidulans*、*A. niger*、*A. sojae*、*A. oryzae*、*Chrysosporium lucknowense*、*Fusarium bactridioides*、*Fusarium cerealis*、*Fusarium crookwellense*、*Fusarium culmorum*、*Fusarium graminearum*、*Fusarium graminum*、*Fusarium heterosporum*、*Fusarium negundi*、*Fusarium oxysporum*、*Fusarium reticulatum*、*Fusarium roseum*、*Fusarium sambucinum*、*Fusarium sarcochromum*、*Fusarium sporotrichioides*、*Fusarium sulphureum*、*Fusarium torulosum*、*Fusarium trichothecioides* 或者 *Fusarium venenatum* 细胞。在另一更优选的实施方式中,丝状真菌宿主细胞为 *Mucor miehei*、*Myceliophthora thermophila*、*Neurospora crassa*、*Penicillium purpurogenum*、*Penicillium chrysogenum*、*Trichoderma harzianum*、*Trichoderma koningii*、*Trichoderma longibrachiatum*、*Trichoderma reesei* 或者 *Trichoderma viride* 细胞。在一个最优选的实施方式中,丝状真菌宿主细胞为选自以下的种:*Rasamsonia emersonii*、*Aspergillus niger*、*Aspergillus oryzae*、*Aspergillus sojae*、*Myceliophthora thermophila*、*Trichoderma reesei* 或者 *Penicillium chrysogenum*。一个最优选的 *Rasamsonia emersonii* 宿主细胞为 CBS393.64 或者其衍生物。

[0181] 公众可以容易地从多个培养物保藏中心获得丝状真菌的几种菌株,例如美国典型培养物保藏中心 (American Type Culture Collection (ATCC))、德国微生物和细胞培养物保藏中心 (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ))、荷兰真菌培养物保藏中心 (Centraalbureau Voor Schimmelcultures (CBS)) 和农业研究机构专利培养物保藏中兴北区研究中心 (Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL)) *Rasamsonia emersonii* ATCC16479、*Aspergillus niger* CBS 513.88、*Aspergillus oryzae* ATCC20423、IFO 4177、ATCC 1011、ATCC 9576、ATCC14488-14491、ATCC11601、ATCC12892、*P. chrysogenum* CBS 455.95、*Penicillium citrinum* ATCC 38065、*Penicillium chrysogenum* P2、*Acremonium chrysogenum* ATCC 36225 或者 ATCC 48272、*Trichoderma reesei* ATCC 26921 或者 ATCC 56765 或者 ATCC 26921、*Aspergillus sojae* ATCC11906、*Chrysosporium lucknowense* ATCC44006。

[0182] 宿主细胞可以是野生型丝状真菌宿主细胞或者变体、突变体或者经遗传修饰的丝状真菌宿主细胞。

[0183] 可以用本身已知的方式,通过下述方法来转化真菌细胞,该方法涉及原生质体形成、原生质体转化以及细胞壁再生。合适的用于转化 *Rasamsonia* 宿主细胞的工序描述于 W02011\054899 中。合适的用于转化 *Aspergillus* 宿主细胞的工序描述于 EP 238023 以及 Yelton 等人,1984,Proceedings of the National Academy of Sciences USA 81:1470-1474 中。使用 *Agrobacterium tumefaciens* 转化 *Aspergillus* 和其他丝状真菌宿主细胞的合适工序描述于例如 Nat Biotechnol. 1998Sep;16(9):839-42,错误刊载在 Nat Biotechnol 1998Nov;16(11):1074. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. de Groot MJ, Bundock P, Hooykaas PJ, Beijersbergen AG. Unilever Research Laboratory Vlaardingen, The Netherlands. 中。转化 *Fusarium* 种的合适的方法通过 Malardier 等人,1989, Gene 78:147-156 和 W096/00787 描述。可以通过 Becker 和 Guarente, Abelson, J. N. 和 Simon, M. I. 编辑, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito 等人,1983, Journal of Bacteriology 153:163; 以及 Hinnen 等人,1978, Proceedings of the National Academy of Sciences USA 75:1920 描述的工序来转化酵母。

[0184] “生物化合物”可以是任何生物聚合物或者代谢产物。生物化合物可由单个编码序列或者组成生物合成或代谢途径的一系列编码序列编码,或者可以是单一编码序列的产物或者一系列编码序列的产物的直接结果。生物化合物对宿主细胞可以是天然的或者异源的。

[0185] 术语“异源生物化合物”在本文中定义为下述生物化合物,其对给定宿主细胞不是同源的,或者天然生物化合物中进行了结构修饰以改变天然生物化合物。

[0186] 术语“生物聚合物”在本文定义为相同、相似或者不相似的亚单元(单体)的链(或者聚合物)。生物聚合物可以是任何生物聚合物。生物聚合物可以是例如但不限于核酸(例如 RNA)、多胺、多元醇、多肽(或者聚酰胺)或者多糖。

[0187] 根据一个优选的实施方式,生产的生物化合物为多肽。根据一个更优选的实施方式,生产的多肽由 DNA 构建体中存在的编码序列编码,该 DNA 构建体包含与所述编码序列可操作相连的本发明的启动子。多肽可以是具有目的生物活性的任何多肽。术语“多肽”在本文并非意指特定长度的被编码的产物,并因此涵盖肽、寡肽和蛋白质。术语“多肽”还涵盖组合以形成所编码的产物的两个或者更多个多肽。多肽还包括杂合多肽,其包含得自至少两个不同多肽的部分或者全部多肽序列的组合,其中一个或者更多个可对宿主细胞是异源的。多肽还包括上述多肽和杂合多肽的天然存在的等位和工程化的变异。

[0188] 多肽对给定的宿主细胞可以是天然的或者异源的。术语“异源多肽”在本文定义为对给定宿主细胞而言并非天然的多肽。或者,异源多肽为其中进行了修饰以改变天然序列的天然多肽,或者其表达在数量上被改变的天然多肽,该改变是通过重组 DNA 技术操作真菌细胞的结果。例如,可通过以下方式重组生产天然多肽,例如将编码多肽的序列置于本发明启动子控制下,从而增强多肽表达、通过使用信号序列加速目标天然多肽输出细胞外,和提高通常由细胞产生的编码多肽的基因的拷贝数。

[0189] 多肽可以是胶原或者明胶、或者其变体或者杂合体。多肽可以是抗体或者其部分、抗原、凝固因子、酶、激素或者激素变体、受体或者其部分、调节蛋白质、结构蛋白质、受体或

者转运蛋白、分泌过程涉及的蛋白质、折叠过程涉及的蛋白质、伴侣分子 (chaperone)、肽氨基酸转运蛋白、糖基化因子、转录因子、合成肽或者寡肽、细胞内蛋白质。细胞内蛋白质可以是酶,例如蛋白酶、神经酰胺酶、环氧化物水解酶、氨基肽酶、酰基转移酶、醛缩酶、羟化酶、氨基肽酶、脂酶。多肽可以是细胞外分泌的酶。这类酶可以属于以下的组:氧化还原酶、转移酶、水解酶、裂合酶、异构酶、连接酶、过氧化氢酶、纤维素酶、壳多糖酶、角质酶、脱氧核糖核酸酶、葡聚糖酶、酯酶。酶可以是碳水化合物酶,例如纤维素酶例如内切葡聚糖酶、 β -葡聚糖酶、纤维二糖水解酶或者 β -葡萄糖苷酶、半纤维素酶或者胶质水解酶 (pectinolytic enzyme) 例如木聚糖酶、木糖苷酶、甘露聚糖酶、半乳聚糖酶、半乳糖苷酶、果胶甲基酯酶、果胶裂合酶、果胶酸裂合酶 (pectate lyase)、内切多聚半乳糖醛酸酶、外切多聚半乳糖醛酸酶、鼠李半乳糖醛酸酶、阿拉伯聚糖酶、阿拉伯呋喃糖酶、阿拉伯木聚糖水解酶、半乳糖醛酸酶、裂合酶或者淀粉分解酶;水解酶,异构酶或者连接酶,磷酸酶如植酸酶,酯酶如脂酶,蛋白水解酶,氧化还原酶如氧化酶,转移酶或者异构酶。酶可以是植酸酶。酶可以是氨基肽酶、淀粉酶、碳水化合物酶、羧肽酶、内切蛋白酶、金属蛋白酶、丝氨酸蛋白酶过氧化氢酶、壳多糖酶、角质酶、环糊精葡萄糖基转移酶、脱氧核糖核酸酶、酯酶、 α -半乳糖苷酶、 β -半乳糖苷酶、葡糖淀粉酶、 α -葡萄糖苷酶、 β -葡萄糖苷酶、卤素过氧化物酶、蛋白水解酶、转化酶、漆酶、脂酶、甘露糖苷酶、变构水解酶 (mutanase)、氧化酶、胶质水解酶、过氧化物酶、磷脂酶、多酚氧化酶、核糖核酸酶、转谷氨酰胺酶或者葡萄糖氧化酶、己糖氧化酶、单加氧酶。

[0190] 或者,与本发明启动子可操作连接的编码序列可编码细胞内蛋白质,例如伴侣分子或者转录因子。其的一个实例描述于 Appl Microbiol Biotechnol. 1998 Oct; 50(4):447-54 (“Analysis of the role of the gene *bipA*, encoding the major endoplasmic reticulum chaperone protein in the secretion of homologous and heterologous proteins in black *Aspergilli*. Punt PJ, van Gemeren IA, Drint-Kuijvenhoven J, Hessing JG, van Muijlwijk-Harteveld GM, Beijersbergen A, Verrips CT, van den Hondel CA) 中。这可用于例如促进宿主细胞作为蛋白质生产者或者作为代谢产物的效力,如果该编码序列(例如伴侣分子或者转录因子)已知是蛋白质或者代谢产物产生的限制因子。

[0191] 生物化合物可以是多糖。多糖可以是任何多糖,包括但不限于粘多糖(例如肝素和透明质酸)和含氮多糖(例如几丁质)。在一个更优选的选择中,多糖为透明质酸。

[0192] 或者,生物化合物可以是代谢产物。术语“代谢产物”涵盖初级和次级代谢产物;代谢产物可以是任何代谢产物。一个优选的代谢产物为柠檬酸。

[0193] 根据另一优选的实施方式,产生的生物化合物为代谢产物。根据一个更优选的实施方式,DNA 构建体中存在的编码序列编码涉及代谢产物产生的酶,该 DNA 构建体包含与该编码序列可操作连接的本发明的启动子。

[0194] 或者,本发明的 DNA 构建体中可存在若干编码序列。各编码序列可编码涉及引起代谢产物产生的代谢或者生物合成途径的不同的酶。初级代谢产物是细胞的初级或者一般代谢产物,其涉及能量代谢、生长和结构。次级代谢产物是次级代谢的产物(参见例如 R. B. Herbert, *The Biosynthesis of Secondary Metabolites*, Chapman and Hall, New York, 1981)。

[0195] 初级代谢产物可以是但不限于氨基酸、脂肪酸、核苷、核苷酸、糖、甘油三酯或者维

生素。一个优选的初级代谢产物为柠檬酸。

[0196] 次级代谢产物可以是但不限于生物碱、香豆素、类黄酮、聚酮化合物、奎宁、类固醇、肽或者萜。次级代谢产物可以是抗生素、拒食素、引诱素、杀细菌素、杀真菌素、激素、杀虫剂或者杀鼠剂。优选的抗生素为头孢菌素和 β -内酰胺。

[0197] 生物化合物还可以是可选择标记物。可选择标记物是提供针对抗微生物剂或者病毒的抗性、对重金属抗性、对营养缺陷型的原养等的产物。可选择标记物包括但不限于 *amdS* (乙酰胺酶)、*argB* (鸟氨酸氨甲酰基转移酶)、*bar* (膦丝菌素乙酰转移酶)、*hygB* (潮霉素磷酸转移酶)、*niaD* (硝酸还原酶)、*pyrG* (乳清苷 -5'-磷酸盐脱羧酶)、*sC* (硫酸腺嘌呤基转移酶)、*trpC* (氨基苯甲酸合酶)、*ble* (腐草霉素抗性蛋白) 及其等效物。

[0198] 在本发明的生产方法中,细胞使用本领域已知方法在适于生产生物化合物的营养培养基中培养,该生物化合物可以是但不限于多肽或者代谢产物。例如,细胞可在实验室或者工业发酵器中通过摇瓶培养、小规模或者大规模发酵(包括连续、分批、分批补料或者固态发酵)来培养,其在合适培养基和使得编码序列被表达和/或生物化合物被分离的条件下进行。在含有碳和氮源和无机盐的合适的营养培养基中,使用本领域已知工序进行培养。合适的培养基可来自商业供货商或者可通过(例如美国典型培养物保藏中心的产品目录中)公开的成分制备。如果生物化合物分泌进营养培养基,则该生物化合物可直接从培养基中回收。如果生物化合物(其可以是但不限于多肽或者代谢产物)不分泌,那么其可从细胞裂解物中回收。

[0199] 可通过本领域已知方法回收所得生物化合物,其可以是但不限于多肽或者代谢产物。例如,可通过常规工序从营养培养基中回收多肽或者代谢产物,该常规工序包括但不限于离心、过滤、提取、喷雾干燥、蒸发或者沉淀。

[0200] 多肽可通过本领域已知的多种方法纯化,所述方法包括但不限于色谱法(例如离子交换、亲和、疏水、色谱聚焦和尺寸排阻)、电泳工序(例如制备型等电聚焦)、差异溶解度(例如硫酸铵沉淀)、SDS-PAGE 或者提取(参见例如 Protein Purification, J.-C. Janson 和 Lars Ryden 编辑, VCH Publishers, New York, 1989)。多肽可使用本领域已知的对于多肽特异的方法检测。这些检测方法可包括特异抗体的使用、酶产物的形成或者酶底物的消失。

[0201] 本发明还涉及用于改变编码多肽的编码序列的表达的 DNA 构建体,该多肽对于真菌宿主细胞是内源的。该构建体可含有改变内源基因表达所需的最少数量的成分。

[0202] 在一个实施方式中,核酸构建体优选含有:(a) 靶向序列;(b) 本发明的启动子 DNA 序列,(c) 外显子;以及(d) 剪接供体位点。将核酸构建体引入细胞时,构建体通过同源重组在内源基因位点处整合至细胞基因组内。靶向序列指导元件(a)-(d) 进入内源基因的整合,使得元件(b)-(d) 与内源基因可操作相连。

[0203] 在另一实施方式中,核酸构建体含有:(a) 靶向序列;(b) 本发明的启动子 DNA 序列;(c) 外显子;(d) 剪接供体位点;(e) 内含子;以及(f) 剪接受体位点,其中靶向序列指导元件(a)-(f) 的整合,使得元件(b)-(f) 与内源基因可操作相连。然而,构建体可含有额外的成分,如可选择标记物。可使用的可选择标记物先前已描述。

[0204] 在两种实施方式中,这些成分的引入引起新转录单位的产生,其中内源基因的表达被改变。事实上,新的转录单位是通过靶向构建体引入的序列和内源基因的融合产物。在

一个内源基因被改变的实施方式中,基因被激活。在该实施方式中,使用同源重组替换、破坏或者失活调节区,该调节区通常通过调节序列的插入与亲本细胞的内源基因相关联,这引起与相应的亲本细胞中的现象相比所述基因以更高水平表达。

[0205] 靶向序列可以在内源基因内、紧邻该基因、在上游基因内或者在内源基因上游和与内源基因有段距离。可使用一个或者更多个靶向序列。例如环形质粒或者 DNA 片段优选使用单个靶向序列,而线性质粒或者 DNA 片段优选使用两个靶向序列。

[0206] 构建体还含有内源基因的一个或者更多个外显子。外显子定义为下述 DNA 序列,该 DNA 序列被拷贝为 RNA 并存在于成熟的 mRNA 分子中,使得外显子序列与内源基因的编码区同框。外显子可任选地包含这样的 DNA,该 DNA 编码一个或者更多个氨基酸和 / 或部分地编码氨基酸。或者,外显子含有对应于 5' 非编码区的 DNA。当一个或者更多个外源外显子编码一个或者更多个氨基酸和 / 或氨基酸的一部分时,设计核酸构建体使得在转录和剪接时,内源基因的编码区是同框的,从而来自第二个外显子的 mRNA 部分的合适读码框是不变的。构建体的剪接供体位点指导一个外显子剪接到另一外显子。典型地,第一个外显子位于第二个外显子的 5', 并且与第一个外显子 3' 侧重合并位于其侧翼的剪接供体位点识别位于第二个外显子 5' 侧翼的第二个外显子侧翼的剪接受体位点。剪接受体位点(类似于剪接供体位点)是指导一个外显子剪接到另一个外显子的序列。与剪接供体位点一同作用的剪接装置使用剪接受体位点来促成内含子的去除。

[0207] 用于改变给定 DNA 序列表达的一个优选策略包括缺失给定 DNA 序列和 / 或通过经修饰的启动子 DNA 序列(例如本发明的启动子)代替给定 DNA 序列的内源启动子序列。

[0208] 替代地或者与其它提到的技术相组合,可使用基于在 *E. coli* 中粘粒的体内重组的技术,如在 *A rapid method for efficient gene replacement in the filamentous fungus A. nidulans* (2000) Chaverroche, M-K., Ghico, J-M. 和 d' Enfert C; *Nucleic acids Research*, vol 28, no 22 中所述。该技术可应用于其它丝状真菌,例如 *R. emersonii*。

[0209] 本文描述和要求保护的发明并不限于本文公开的特定实施方式的范围,因为这些实施方式旨在说明本发明的多个方面。任何等效的实施方式旨在包括在本发明范围内。事实上,除本文示出和描述的以外,通过前述说明,本发明的多种变型对本领域技术人员而言是明显的。这类变型也旨在落入所附的权利要求书范围内。在冲突的情况下,以包括定义的本公开内容为准。

[0210] 本发明通过以下实施例进一步描述,这些实施例不应当解释为限制本发明的范围。

实施例

[0211] 应当理解,尽管示出本发明的优选实施方式,但是这些实施例仅通过示例方式给出。根据上述讨论和这些实施例,本领域技术人员可确定本发明的必要特征,并且在没有违背本发明精神和范围下,本领域技术人员可得到本发明的各种变化和修改以使其适应各种用途和条件。因此除本文示出和描述的以外,由前述说明书,本发明的多种修改对本领域技术人员而言是明显的。这类修改也旨在落入所附权利要求书的范围内。

[0212] 试验信息

[0213] 菌株

[0214] 本文所使用的 *Rasamsonia emersonii* (*R. emersonii*) 菌株来自 ATCC16479, 将其用作野生型菌株。ATCC16479 之前也称为 *Talaromyces emersonii* 和 *Penicillium geosmithia emersonii*。在使用名称 *Rasamsonia emersonii* 时, 也表示 *Talaromyces emersonii*。 *R. emersonii* ATCC16479 的其他菌株名称为 CBS393.64、IF031232 和 IMI116815。

[0215] *Rasamsonia* (*Talaromyces*) *emersonii* 菌株 TEC-142 于 2009 年 7 月 1 日保藏在荷兰真菌培养物保藏中心 (CENTRAAL BUREAU VOOR SCHIMMELCULTURES), 荷兰乌特勒支 NL-3508 AD 乌普萨兰 8 号, 邮箱 85167 (Uppsalalaan 8, P. O. Box 85167, NL-3508 AD Utrecht), 其登录号为 CBS 124902。TEC-142S 是 TEC-142 的单一分离菌株。

[0216] 分子生物技术

[0217] 在这些菌株中, 使用技术人员已知的分子生物技术 (参见: Sambrook&Russell, *Molecular Cloning: A Laboratory Manual*, 第 3 版, CSHL Press, Cold Spring Harbor, NY, 2001), 如下所述将若干种基因过表达, 并将其他基因下调。用于基因过表达的表达载体和用于下调的破坏载体、转化、标记物使用 and 选择培养基的一般设计的实例可在例如 W0199846772、W0199932617、W02001121779、W02005095624、EP 635574B 和 W02005100573 中找到。

[0218] 培养基和溶液

[0219] 土豆右旋糖琼脂, PDA (Fluka, Cat. No. 70139)

[0220]

土豆提取物	4 g/l
右旋糖	20 g/l
细菌琼脂	15 g/l
pH	5.4
水	调节至一升
消毒	在 120°C 下 20 分钟

[0221] *Rasamsonia* 琼脂培养基

[0222]

盐部分 no.3	15 g
纤维素(3%)	30 g
细菌蛋白胨	7.5 g
谷物面粉	15 g
KH ₂ PO ₄	5 g
CaCl ₂ ·2aq	1 g
细菌琼脂	20 g
pH	6.0
水	调节至一升
消毒	在 120℃下 20 分钟

[0223] 盐部分组成

[0224] “盐部分 no. 3”适用 W098/37179, 表 1 的公开。与该表组成差异为 CaCl₂·2aq 1.0g/l, KCl 1.8g/L, 柠檬酸 1aq 0.45g/L(螯合剂)。

[0225] Rasamsonia 的摇瓶培养基

[0226] Rasamsonia培养基 1

[0227]

葡萄糖	20 g/L
酵母提取物(Difco)	20 g/L

[0228]

Clerol FBA3107 (AF)	4 滴/L
pH	6.0
消毒	在 120℃下 20 分钟

[0229] Rasamsonia培养基 2

[0230]

盐部分 no.3	15 g
纤维素	20 g
细菌蛋白胨	4 g
谷物面粉	7.5 g
KH_2PO_4	10 g
$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$	0.5 g
Clerol FBA3107 (AF)	0.4 ml
pH	5
水	调节至一升
消毒	在 120°C 下 20 分钟

[0231] Rasamsonia培养基 3

[0232]

盐部分 no.3	15 g
葡萄糖	24 g
谷物面粉	7.5 g
KH_2PO_4	10 g
$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$	0.5 g
Clerol FBA3107 (AF)	0.4 ml
pH	5
水	调节至一升
消毒	在 120°C 下 20 分钟

[0233] Rasamsonia的孢子批次制备

[0234] 使菌株从储备液在 40°C 下直径 10cm 的 Petri 培养皿中 Rasamsonia 琼脂培养基上生长 5-7 天中。对于 MTP 发酵,使菌株在含有 Rasamsonia 琼脂培养基的 96 孔板中生长。在 -80°C 下,将菌株储备液 10% 甘油中保存。

[0235] 染色体 DNA 分离

[0236] 使菌株在 42°C 及 250rpm 下 YGG 培养基 (每升: 8g KCl, 16g 葡萄糖, H_2O , 20ml 的 10% 酵母提取物, 10ml 的 100× 青霉素 / 链霉素, 6.66g YNB+ 氨基酸, 1.5g 柠檬酸和 6g K_2HPO_4) 中生长 16 小时,然后使用 DNeasy 植物小试剂盒 (Qiagen, Hilden, 德国) 来分离染色体 DNA。

[0237] 蛋白分析

[0238] 通过加入 228 μl TCA- 丙酮 (1.2g 三氯酸、9ml 丙酮、1ml 的 H_2O) 来沉淀在 65 μl 上清液中的蛋白质。在 -20°C 下沉淀 3 小时之后,在 4°C 下将样品在 eppendorf 离心机上以 14,000rpm 离心 10 分钟,然后使用丙酮洗涤沉淀物。将干燥的沉淀物溶解在

1× 样品缓冲剂 (25 μl 的 LDS 样品缓冲剂 (Invitrogen, Breda, 荷兰), 10 μl 的还原剂 (Invitrogen, Breda, 荷兰), 65 μl 的 H₂O) 中。

[0239] 在 NuPAGE 4-12% Bis-Tris 凝胶 (Invitrogen, Breda, 荷兰) 上在还原条件下分离蛋白样品, 并如所示染色。根据制造商说明书, 使用 InstantBlue (Expedeon, Cambridge, 英国)、SimplyBlue safestain (Invitrogen, Breda, 荷兰) 或者 Sypro Ruby (Invitrogen, Breda, 荷兰) 来染色凝胶。

[0240] 对于 Western 印迹法而言, 将蛋白转移至硝酸纤维素。使用含有 3% 脱脂乳的 TBST (含有 0.1% Tween 40 的 Tris 缓冲盐水) 来封闭硝酸纤维素滤膜, 并使用抗 -FLAG M2 抗体 (Sigma, Zwijndrecht, 荷兰) 孵育 16 小时。使用 TBST 洗涤该印迹两次, 每次 10 分钟, 然后使用辣根过氧化物酶缀合的兔抗小鼠抗体 (DAKO, Glostrup, 丹麦) 染色 1 小时。在使用 TBST 洗涤印迹 5 次 (每次 10 分钟) 之后, 使用 SuperSignal (Pierce, Rockford, 美国) 使得蛋白可视化。任选地, 可使用 ChemiDoc 系统 (Biorad, Veenendaal, 荷兰) 定量 Western 印迹。

[0241] 实施例 1

[0242] 构建包含与编码序列可操作相连的本发明启动子的 DNA 构建体

[0243] 该实施例描述了构建包含本发明的启动子的表达构建体, 该启动子与编码序列可操作相连。此处使用的编码序列或者报告构建体为 FLAG 标记的 *R. emersonii* 内切葡聚糖酶基因 (EBA7-FLAG)。EBA7-FLAG 用作报告酶以使得能够通过使用 FLAG 标记的特异性抗体的 Western 印迹法来检测重组蛋白。将构建体随机整合至基因组内。

[0244] 根据常规克隆工序来构建载体 pENTRY-P6bleTtrpC-Pxeba7flagTgla。载体包含 ble 表达盒, 其由 *A. nidulans* gpdA 启动子 (P6)、ble 编码区 (ble) 和 *A. nidulans* TrpC 终止子 (TtrpC)、目标启动子 (Px)、EBA7-FLAG 报告编码区 (eba7flag) 以及 *A. niger* 葡糖淀粉酶终止子组成 (图 1)。将五种 *R. emersonii* 启动子克隆至载体内: *R. emersonii* 纤维二糖水解酶 -I (PcbhI, SEQ ID NO:1)、*R. emersonii* 乙酰木聚糖酯酶 (Pace, SEQ ID NO:2)、*R. emersonii* 内切葡聚糖酶 (Peg, SEQ ID NO:3)、*R. emersonii* 纤维二糖水解酶 -II (PcbhII, SEQ ID NO:4) 以及 *R. emersonii* β-葡萄糖苷酶 (Pbg, SEQ ID NO:5)。此外, 将 *A. nidulans* gpdA 启动子克隆至载体内以比较上述启动子与 *A. nidulans* gpdA 启动子 (PgpdA, SEQ ID NO:6) 的活性。由 W02011\054899 中描述的载体 pGBFINEBA7 中获得 EBA7-FLAG 报告基因盒。测试在 *R. emersonii* 中 5 种启动子构建体的表达。

[0245] 实施例 2

[0246] 在 *Rasamsonia emersonii* 中启动子 - 报告子构建体的表达

[0247] 使用如之前 W02011\054899 所述的方法将实施例 1 描述的 pENTRY-P6bleTtrpC-Pxeba7flagTgla 启动子构建体用于转化 *R. emersonii* 菌株 TEC-142S。根据如 W02011\054899 所述的工序, 在腐草霉素培养基上选择转化体和将菌落纯化, 然后测试。生成含启动子 - 报告子构建体的转化体的孢子批次。

[0248] 转化体在 45 °C 下在 *Rasamsonia* 培养基 1 的摇瓶中生长, 在振荡培养箱中以 250rpm 孵育 24 小时, 然后将该预培养物用于接种含有作为 C 源的纤维素的 *Rasamsonia* 培养基 2 (约 10% 接种)。40 小时之后取样, 然后使用 TCA- 丙酮沉淀蛋白, 并且在 SDS-PAGE 上分离。通过使用 FLAG 标记的特异性抗体的 Western 印迹法来检测重组 FLAG 标记的葡糖

淀粉酶。

[0249] Western 印迹的结果显示在图 2 中。所有 6 种启动子-报告子构建体的转化体上清液显示在 Western 印迹上特定 EBA7-FLAG 蛋白带。在含有 2% 纤维素的培养基中的测试条件下,所有 5 种 *R. emersonii* (半) 纤维素启动子能够驱动 AG-FLAG 报告基因的表达,并且与 *A. nidulans* *gpdA* 启动子相比,其显示更强的表达(比较泳道 1-5 和 8-12 与泳道 7)。

[0250] 实施例 3

[0251] 构建包含与编码序列可操作相连的本发明启动子的 DNA 构建体,其靶向基因组

[0252] 该实施例描述了构建包含本发明的启动子的表达构建体,该启动子与编码序列可操作相连。在此使用的编码序列或者报告构建体为 FLAG 标记的 *R. emersonii* 葡糖淀粉酶基因。葡糖淀粉酶-FLAG 用作报告酶以能够通过葡糖淀粉酶活性测量或者通过使用 FLAG 标记的特异性抗体的 Western 印迹法来测量本发明启动子的活性。将启动子-报告子表达盒靶向整合至 *pepA* 基因座内。

[0253] 为了使启动子-报告子构建体靶向至 *pepA* 基因座内,将表达载体克隆用于靶向。测序和分析 *Rasamsonia emersonii* 菌株 CBS393.64 的基因组 DNA。鉴定具有作为蛋白酶 *pepA* 注释的经翻译蛋白的基因。包含 ORF 和约 3000bp 的 5' 区以及 2500bp 的 3' 侧翼区的基因组序列的 *Rasamsonia emersonii* *pepA* (RePepA) 序列、cDNA 和蛋白质序列分别在序列表 7、8 和 9 中示出。

[0254] 根据用于靶向至 RePepA 基因座内的常规克隆工序来构建两种载体。该两种载体的插入片段可一起应用于所谓“二分基因靶向 (bipartite gene-targeting)”方法中 (Nielsen 等人, 2006, 43:54-64)。该方法使用选择标记物的两种非功能性 DNA 片段,其与基因靶序列重叠至一起(二分方法的进一步细节也参见 WO 2008113847)。当正确同源重组时,通过在同源靶基因座处整合,选择标记物变成有功能的。也如在 WO 2008113847 中详细所述,设计和构建两种不同的缺失载体 (Te *pep. bbn* 和 pEBA1006),从而能够提供用于二分基因靶向的两种重叠的 DNA 分子。第一个载体 Te *pep. bbn* (如在图 3 中总体布置图) 包含靶向 RePepA 基因座的 RePepA ORF 约 1.5kb 上游的 1500bp 5' 侧翼区 (ORF 和约 1500bp 的 RePepA 启动子)、*lox66* 位点、通过 *A. nidulans* *gpdA* 启动子驱动的 *ble* 编码区的非功能性 5' 部分 (缺失在 *ble* 的 3' 端处编码序列的最后 104 个碱基的 P*gpdA*-*ble* 序列, SEQ ID NO:10)。为了使能在 *E. coli* 中有效克隆启动子-报告子盒,将 *ccdB* 基因插入 5' RePepA 侧翼区和 *lox66* 位点之间。第二个 pEBA1006 载体 (如在图 4 中总体布置图) 包含 *ble* 编码区的非功能性 3' 部分和 *A. nidulans* *trpC* 终止子 (缺失在 *ble* 的 5' 端处编码序列的最前面的 12 个碱基的 *ble*-T*trpC* 序列, SEQ ID NO:11)、*lox71* 位点以及用于靶向 RePepA 基因座的 RePepA ORF 的 2500bp 3' 侧翼区。当同源重组时,第一和第二非功能性片段变成有功能的,从而得到功能性 *ble* 盒。RePepA 上游和下游基因侧翼区靶向在预定 RePepA 基因组基因座处二分片段的同源重组。

[0255] 根据常规克隆工序,通过启动子-报告子盒来替换在载体 Te *pep. bbn* 中 *ccdB* 基因。将六种 *R. emersonii* 启动子 (通过 SEQ ID NO:12、13、14、15、16 和 17 表示) 克隆至具有 *A. nidulans* *amdS* 终止子的 FLAG 标记的 *R. emersonii* 葡糖淀粉酶编码区 (AG-FLAG) 的上游,从而分别生成构建体 pEBA528、pEBA529、pEBA530、pEBA531、pEBA532 和 pEBA533。此外,将通过 SEQ ID NO:6 表示的 *A. nidulans* *gpdA* 启动子克隆至在 Te *pep. bbn* 中具有

A.nidulans amdS 终止子的 FLAG 标记的 *R.emersonii* 葡糖淀粉酶编码区 (AG-FLAG) 的上游,从而生成构建体 pEBA540。AG-FLAG 的氨基酸序列和具有 *A.nidulans* amdS 终止子的 AG-FLAG 编码区分别通过 SEQ ID NO:18 和 19 表示。pEBA528 的示意图显示在图 5 中,其代表 pEBA529、pEBA530、pEBA531、pEBA532、pEBA533 和 pEBA540。

[0256] 实施例 4:在 *Rasamsonia emersonii* 中灭活 ReKu80 基因以改善基因靶向

[0257] 克隆 ReKu80 缺失构建体

[0258] 测序和分析 *Rasamsonia emersonii* 菌株 CBS393.64 的基因组 DNA。识别 *Rasamsonia emersonii* Ku80 基因 (ReKu80)。包含 ORF 和约 2500bp 的 5' 区和 2500bp 的 3' 侧翼区的基因组序列、cDNA 和蛋白序列的 ReKu80 的序列分别显示在序列表 20、21 和 22 中。

[0259] 根据常规克隆工序来克隆 ReKu80 的两种替换载体, pEBA1001 和 pEBA1002 (参见图 6 和 7)。两载体的插入片段可一起应用于如在实施例 3 中所述的所谓“二分基因靶向”方法。pEBA1001 载体包含靶向 ReKu80 基因座的 ReKu80ORF 的 2500bp 5' 侧翼区、lox66 位点以及通过 *A.nidulans* *gpdA* 启动子驱动的 *ble* 编码区的 5' 部分 (图 6)。pEBA1002 载体包含 *ble* 编码区的 3' 部分、*A.nidulans* *trpC* 终止子、lox71 位点以及靶向 ReKu80 基因座的 ReKu80ORF 的 2500bp 3' 侧翼区 (图 7)。

[0260] 使 *Rasamsonia emersonii* 中 ReKu80 缺失

[0261] 使用如之前在 W02011\054899 中所述的方法分离缺失构建体 pEBA1001 和 pEBA1002 的线性 DNA,并将其用于转化 *Rasamsonia emersonii* 菌株 TEC-142S。这些线性 DNA 可整合至在 ReKu80 基因座处的基因组内,从而如在图 8 中所示通过 *ble* 基因替代 ReKu80 基因。根据如 W02011\054899 所述的工序,在腐草霉素培养基上选择转化体和将菌落纯化,然后测试。使用在缺失盒的 *gpdA* 启动子中引物和针对在 5' 靶向区正上游的基因组序列的引物,通过 PCR 来诊断生长菌落在 ReKu80 基因座处的整合。从约 250 个转化体的库中,4 种菌株显示出去除基因组 ReKu80 基因。

[0262] 克隆编码 cre 重组酶的瞬时表达质粒 pEBA513

[0263] 通过 DNA2.0 (Menlo Park, 美国) 来构建 pEBA513,并且 pEBA513 含有以下成分:由 *A.niger* *glaA* 启动子、编码 cre- 重组酶 (AAY56380) 的 ORF 和 *A.nidulans* *niaD* 终止子组成的表达盒;由 *A.nidulans* *gpdA* 启动子、编码潮霉素 B 抗性蛋白的 ORF 和 *P.chrysogenum* *penDE* 终止子 (Genbank:M31454.1,核苷酸 1750-2219) 组成的表达盒;含有 AMA1 区以及 CAT 氯霉素抗性基因的来源于 pAMPF21 的载体。图 9 示出 pEBA513 的图谱。

[0264] 通过瞬时表达 cre 重组酶对腐草霉素抗性 ReKu80 缺失菌株的标记物去除

[0265] 随后,使用 pEBA513 来转化 3 种候选 ReKu80 敲除菌株,从而通过 cre 重组酶的瞬时表达来去除 *ble* 选择标记物。在含有 50 μ g/ml 的潮霉素 B 的再生培养基上的覆盖层中铺板 pEBA513 转化体。潮霉素抗性转化体在含有 50 μ g/ml 的潮霉素 B 的 PDA 上生长,从而允许表达 cre 重组酶。将单个菌落铺板在非选择性 *Rasamsonia* 琼脂培养基上,从而得到纯的孢子批次。通过使转化体在具有和不具有 10 μ g/ml 的腐草霉素的培养基上生长来表型检测 *ble* 标记物的去除。在使用 pEBA513 转化 (有 cre 重组酶) 之后大部分 (>90%) 转化体对腐草霉素敏感,这表明去除基于 pEBA1001 和 pEBA1002 的 *ble* 标记物。随后,通过使转化体在具有和不具有 50 μ g/ml 的潮霉素的培养基上生长来表型判断在 *ble* 阴性菌株中

pEBA513 构建体的去除。由于 pEBA51 质粒的自发丢失,约 50%的转化体缺乏潮霉素抗性。

[0266] 通过 Southern 分析和 PCR 来检测候选无标记物的敲除菌株的 ReKu80 基因缺失。获得无标记物的 ReKu80 缺失菌株,并将代表菌株用于启动子-报告子构建体的靶向整合(实施例 5)。

[0267] 实施例 5

[0268] 通过启动子-报告子盒替换在 *Rasamsonia emersonii* 中 RePepA 基因

[0269] 使用如之前在 W02011\054899 中所述的方法,分离缺失构建体 Tepep. bbn 和 pEBA1006 的线性 DNA,并将其用于转化在实施例 4 中获得的 ReKu80 缺失菌株。这些线性 DNA 可整合至在 RePepA 基因座处的基因组内,从而如在实施例 4 中对于 ReKu80 基因所述,通过 ble 基因替换 RePepA 基因,不同在于不仅使 RePepA ORF 缺失,也使起始密码子约 1500nt 的上游缺失,从而也去除 RePepA 启动子。根据如 W02011\054899 所述的工序,在腐草霉素培养基上选择转化体和将菌落纯化,然后测试。使用在 ble 盒的 AmdS 终止子中的引物和针对 3' 靶向区正下游基因组序列的引物,通过 PCR 来分离转化体的染色体 DNA 以确定在 RePepA 基因座处正确的整合。生成转化体的孢子批次,该转化体显示 RePepA 基因座缺失。

[0270] 使转化体在 45°C 下在 *Rasamsonia* 培养基 1 的摇瓶中以 250rpm 在振荡培养箱中生长 24-48 小时,并将该预培养物用于接种含有作为 C 源的葡萄糖的 *Rasamsonia* 培养基 3 (约 10% 接种)。在 24 小时之后取样,然后使用 TCA- 丙酮来沉淀蛋白,并且在 SDS-PAGE 上分离。通过使用 FLAG 标记的特异性抗体的 Western 印迹法来检测重组 FLAG 标记的葡糖淀粉酶。

[0271] Western 印迹的结果显示在图 10 中。所有 7 种启动子-报告子构建体的转化体的上清液在 Western 印迹上显示特异性 AG-FLAG 蛋白带。在含有作为 C 源的 2.4% 葡萄糖的培养基中在检测条件下,所有 6 种 *R. emersonii* 启动子能够驱动 AG-FLAG 报告基因的表达,与 *A. nidulans* gpdA 启动子相比,它们中一些显示更强的表达(比较泳道 2-7 与泳道 1)。

[0001]

序列表

<110> 帝斯曼知识产权资产管理有限公司

<120> 用于在细胞中表达基因的启动子

<130> 28987-WO-PCT

<160> 24

<170> PatentIn version 3.5

<210> 1

<211> 1494

<212> DNA

<213> *Rasamsonia emersonii*

<400> 1

atgcgcactt cgtgtagaga ataccgtcgt caattggctg tgttacgcag cttacgccag	60
cagctgggett ggacgcata gcattcagcc cgcagcgtg ctaatttttag cgcggaaccg	120
gactagcgcg tgacgtgtta aaccctaaat cccgccagtc ctaatctggt tacagctagg	180
gttttagggtc tagggctgaa tactagggtt tctgagggt aggctgaggg ttagggctga	240
gtttgggtcac tgggcacgc tgccattagc tttctctttt ctccattagg aaaacagtgt	300
ggccatccct aatctcaatg acggtgagtt tgaaaagtta ccgaactgca gtagacgate	360
tttgcacctc aaaaatcaga atgccaggt ggcacacgc ttggaggtga tgatcgggca	420
tgatgacctc ctctgtcagc aatctggcat gaacatcgtc aggacacggt agagagagtg	480
gaggacgacg aaaatctgaa cagctcgcgg ttaacatcgc atatecgcta gcgcggttga	540
gaacgacacc catcggaatg tcgggaacte caaggtgaca gtgctggatg ggtcacggt	600
gtcttcaaag gccaaagtgc ggtcttatgc ctatcgcgtt gagaatgtct tccgaacac	660
catcagcttt ggaaggggt caccgagagc ctccaacctt attagatggt ctgattctg	720
agttacagaa ttaacgttgc gctgattggt ttacatgac gagceatgta acgacgcgt	780
gctggaaaag catagacaga catgccgtt tatccggatc acaggctggt caattgctct	840
gcgtgtggca ctccctctgt aacctcatga tgggtctatc gggttatgta acagcacatt	900
atctcagatg aagcatgcag gtaagaaaca ttattctctc agccaagcat ctgcgcgagt	960
ggcaagctgg accaggtccc cgcattggca ttgacggaat ttagtcgcac cccgattgc	1020
gggatcgaca atctttacta cgcctggcgt gtttgggtccc tgagacgaac catcctgcca	1080
agctacccct tcttcagctt aaacagtcgt ccaatcgtca ctggctgttg agcagcatcc	1140
cccgcategg cacttgactt taatttcccg atgtccctta tctgagccag acatgcggag	1200

[0002]

tgtcgggaga tgagegatat cctcagactg actgaccagc aagcaagaac gatcgatcac	1260
tctggcgatt tccccacacg gggagtctcg cacttcagga acaatgcacc ggcaactcaa	1320
caatggcate gcttcccagg aategatgcc caattcggga tggagatgtc gatcagatgg	1380
tgtataaaaag ggcaggatga atccagggtc caggccatct gccatcactc agactccaac	1440
actcgatcag cagcttcgaa agtgggtcttt ttgctatcat ccaccgtcaa aatg	1494

<210> 2

<211> 1482

<212> DNA

<213> *Rasamsonia emersonii*

<400> 2

atcgtgctgt cggatgtcta ccccatcgcc gtcaatacaa cctgggtctgt ccaatatcac	60
acggtatgca acaccacgta cggatgtctgt ggctgcgacg actgcaaggg cgactttgaa	120
gacatctcga cgcgtctcga tctgttcacc cagtaccagg actggattgg aggcccgccc	180
aagacctctt ggggcgtccc gcaggccttt ggggggtggag agttctgggc gcggaatcca	240
acgccggacg aggaagtgac gatgaacctg ctcagcatca accacaacgc gaaggggacg	300
gtcatgtgga gctacccgac ggagccggat ctcgaggacg tgaccagctc gctgagcaag	360
gtccttgetg cgccgaacgt gacaagcttt cttctaggag cccccaccac cgcgttgaag	420
acggcagggtg cagatcgaat cgacgetgcc gggtggcagg tcggaaagca gatgetcgtc	480
agcattttgt atctggagta catcgattcc gattctcagg tcacgatctc tttgccacg	540
ccggccagct cagtcacca agttttgtgg ggtactggct ggactgtgtc agggggcaat	600
ctcgtgaaga atggcctgaa gggcttagaa gcagatttac tcattgtggg attggagtaa	660
tctagtcata tacatgetta gtttacttgg ctaatatctg tggetatcaa taaatctgtt	720
ttccggaaact ttaattctac tttatcaatt agcctaaaat catttagaca tatgcttttt	780
ttcccctcat gcaatgccta taacctttcc ccgcagcaaa gccaaagcact atcaacaata	840
tccccgattc tccgcgataa tacataccaa aatatgaaaac ggcatttttag ccaattgacg	900
atctacccta ctaaacgita attggttacc tacataagta cctggccaca taatttagcc	960
agtcgctgta aactacctag gtatactagg tacgtctacc tggagtggag tgatgggctt	1020
tttttttcaa gggggacatc tgactaacct ttaggtcaac gacctagcta ggcagtgagg	1080
tacgttaagc ggggaaccgg gtgtaacgcg aatccgggga agcctccgtg gacatggata	1140
tagcttgtac agtacagctg tggagatata ccgcatattt ttttaaagat taattacaac	1200

[0003]

agccgtacat atacgtcttc atcegtctgt ccggaagagc tagctagggt acattttctgc	1260
ttttccagtg gggactgtca aatatgtctt tggtagggga aaaaaaaac aaagctcttt	1320
ctetcccagag ttccctacat ataaggatgc agtgcaaccg gegatctctt ccattctaaa	1380
tcacggcaat tctctgcact tcccaggtag gaagcatctc tgacaattat ttgcecttcc	1440
ttctaagaga aaggggagaga gagaacaaaa aaattcaaaa tg	1482

<210> 3

<211> 1503

<212> DNA

<213> *Rasamsonia emersonii*

<400> 3

actactgcaa gggcgagacg gttgtggctg ttgcgacgat gcagagtgac ccgatcatgg	60
tcaagtgtgc ggagctgatg cggcggggga acatgccgtc gaagaaggag attgaggatg	120
gggtggatgt gctggctgtg ggcgtgccca agagcgttgc tatataagaa gccagcgaatg	180
tatatacgca gatatttttg gtacatgtta gacatgttat actctatacc atctgtctat	240
gaattcgata tcttgccctc gaataataacc ctactatctg tattatatag tgtctaagge	300
attcaatatg aactcttctt cgatataatg cattaattgc tagctgttag gaaacaaaact	360
gtactgctag atagatgctt agcaactate tcagagaacc gcacctgggc aaaggtattg	420
ttaacgcagt caattatttg cagtgtttga gggcttgcca tctggetgtg aaagcctgca	480
gctacctcag acgatgaaga cgatggaaat ccatcaaaca taggaaggat gtcttcaaag	540
gcttgttata ttataaaagg actatattct gttcttctgt tcttaggtag ttacctacgt	600
acagtacgta attcttccca ttaatagggt cgttagcatg ccaaatgact gcacatcccc	660
agccattggc aggaggaatc tctcccataa ttaattaatt acatetaatc taagtctaaa	720
aatcccacct gattatcaac caaagagatt aatctaaacg aatctaaact ccgcagccca	780
gccagacca gaccagacca agatcgatcc gacactccgc gccacggcgc tccaccatca	840
cgaacaacac atcgctttag tccgtttcaa gttggtctat ttcgaggta atcagaggga	900
tccgcgggat gtccgccccg aaaccctgca taatcttaat taacctcttt tttgggttca	960
ttcttgtttc tcttcttttt acctttagggt atagatgecc tggatcctcg gaatgttgtt	1020
tgtttatatt actactgtaa tatgtgatat tgtattgcat tgaattgact cggatgatcg	1080
gattccagta aatacggagt aaacgactac gtctcttgag caatattcca caaatcaatt	1140
gttgaagttg aagacgtctg tactatccac actgttccat acgctcgatc gacgcggaga	1200

[0004]

accggctaaa ttaagatcga ttgttgccca gatgatcatg cagccagatg aaacgtccat	1260
caattcatcg cttttgtggc ctgatgtcgg agtcgtcatg aggtgacgcg tttttctttt	1320
tcctgatgat tatagccaga atgcggggaa gctgtcttgt ctaacagtct acgagatgga	1380
agatgggcta tataaaegg tctgtttcct ttcagecagg ttcttttttt tctgttctga	1440
ttcagtcagt cagatagaca gacaaagtca agctgatccc agttectcca ttectgcaac	1500
atg	1503

<210> 4

<211> 1503

<212> DNA

<213> *Rasamsonia emersonii*

<400> 4

caatcgatta cgactgacct gacgtttgtt tatagtcctg agaagcttgt gtggcgggac	60
tgtttctgc ctgaatatcg ggggcgcgat ctcaaggcta gattgaggaa gtgtctcgtc	120
tgaactagat ttgtacatc tcaggaggag atgctcagtg gaatgttctt gaggcaggct	180
ctcttggtt ttgteagtac aaaccaaact catagagaat catetaaaat atagcacatg	240
atactgtggc cattgcaatt ggctcttgtt atgcagetga gcaatactgg catagcctgc	300
agacaattgt ttattgaatg acagcacage aagagctgca cgaattgacc aatcgtttg	360
ttgttttag tgcactgitt ctgaaatgac gaaatgaacc tatectaata accccgcatt	420
caggggctag gaacaaggac aggtagactg attcaatcaa tggtttggtc caggttctag	480
gcatgcttcc atttatccgg tgacccctca catacageac agtagtactg tgcaaggta	540
caaggtgtct agcaagtacg actacacaat ccctaataca agctctgtat gtagcgtcat	600
atcctcatat ctgacgagat atcaagttgc ctattcgcta ttcaactage ctgtaaatcc	660
cactacccaaa tactaataat taatacgaaa cttctgtctc accggagtcg gacattcccc	720
gtggggatta tccagaagga tcateggcaa ttactgcaact tgceagcaag caagctta	780
taagctagga acgaaagcat gcagggtgtat ccgacattcc gcagtagttt aatttagtta	840
aatttaagta gcctgcgacc ccgcatgtta ggggatgctg atgcaacggt gcctgtgcct	900
agttagtga ttgttttagct agatacggga ataggggaat catetaataa ttagcctatg	960
tcatgtgagg attggttctt atggtcttca catgccattg atgttcttgg cagtgggatg	1020
aaatggtgtc cttgagatag tcgaagatta ggcaattgga ttgtatgttg tttctcgcgt	1080
agatgttcta gtcatataga atagggatca atatgctcag cgtcagctac ctcaggctga	1140

[0005]

tgtaaagcac gaatatatta ggatagttca actaggaaac tctaattgtac ccagcagtag	1200
caattgaaga aatccgacat tcccattgtc atcgaaatcg ggcagaaatg tgcaaagaga	1260
gacgattcgg taggcgtgat cgaactgcaa gcttcaatcc atagettagt tccagaccat	1320
gtcgtttcca caaagctega tccaagttag atgaaagcta ctageagetc cgacactccg	1380
cgagacagct ccatecatcca ttataaatat ggccaataac cctgcagaac agcttcaate	1440
tgcagcctgc ttgaccgtca agaacagaca tcataatata aatttcagag agaattcaaa	1500
atg	1503

<210> 5

<211> 1979

<212> DNA

<213> *Rasamsonia emersonii*

<400> 5

caggactcga atttacacta taacactggt gtactgtaca acaatgtgct ctgatccatt	60
ggtaaattaa ctcgatctcg atctttctct cactgcaaaa ccaagctcat gtctgcgtta	120
aacgtctttc cagaagggtt ccttttgaat ccgaaacgaa ggtagcgacg tgcggtggcg	180
accaatecat ctccaatat ggttatgate gtacgagatc gtcaggggaa ttgtcaactg	240
aatgccaatg atagaaaaca ggttttctgg tgttctetta gaagacacce tctcaccttc	300
acagcaattc ccatecacat tcaageggac atccagegaa tgetggcgat aacgttgcta	360
atcgaattgt gtaattgtgg caatccgcat gcggcggaagg cattaattac tccgtcgcca	420
cgaggagggg gaaatacgtt aaggacagtc aggtatagct gcctccgtac tgccctgccta	480
aaacggcatc cccgttgccg atccccggg gtacggccag acctgcata ggaatgacgc	540
tgccggcagt tgggccagag cagccttgcc tctgcgtctg gagtatggag taacagagta	600
cggcatatta cagagtactt ggttccagtg ctaagaatag gcatcagtaa taaatatttg	660
actttgtctt gctacaagtt gcagtggtaa acacgggaac tgggtgtctg gaaaaccggg	720
acgccatagg aatcagggtc cggggcagac ggtgacggaa ggetcgggct cgggctgtct	780
ctcgggttga cttgccagca tcccggctgc ctgcagcctg gatcaaggga aaacaatcag	840
gcgcattctc cgtgagatcg agacaccccg aggacgttat ttgccagcag gcttccgaaa	900
caacaatgcc cacaacaaac actcgggcaa tcgatcgggg ttgttattga tagaagctcg	960
tggcaaagcc tggaaggcg acctgactgc cgactaagcc gtctctggaa tctacggtag	1020
catttcagcg caatgcgggg gggttggccg acctatggaa gagagagaac ttggaagaaa	1080

[0006]

agtaagtggc tttgagctgc tggcaagtec tttgtaccg ttacacctgt cagtactgcg	1140
cggfagaggt tgcctcgctg ccattgactc gggagtgatg catgaattaa ttattagaca	1200
gcgaggggact gtcattaata attaatatit tgcgatgga ggattgccac gaatttagcc	1260
tgcaataatt actccgtgct gggcgttgtt ccgtacctc tgettttgte agccctagcg	1320
ccggcgccct cccagggcaa cgcaagcccc tgctagtgcg ggggaagccg gagcccgga	1380
cagccaaccc aggaggaagg tacggcccag aggtacggtg ctgcctgccc aggtttctgc	1440
agctgcatgg cgtgtgctgg atagcaatit cctctgctc ccagaagaca gccaaaccaca	1500
cgatgcgtcg tctgcagctc ccaggaaacc cagtggtag gtgttaccct tgactcttg	1560
acagactcgc ctgtccatc gacaacagga acttgcccag tagccgcctt ccgagtctct	1620
gcgtctccac cgcgtgtcca cgcccgatc gcccgtatga tctcagacce gtccaccgtg	1680
tgttgtaatc ccttgctgat cgtacgccg ccatggggct cccgttctcg atctttttcg	1740
ggcagacgcc attccggtec aagactcctg ctaccggcct tactagtggc ccagggccac	1800
tggggaccga cctcatitc ccgtgtcggt gcttggtgcc attgttgtec tgcccgttct	1860
ttttataaaa agagcccgtc gtctccctc ctccaaggc ctgaaagagc atgtcagctt	1920
tcctctgtgc ctgacttctt gtgttacctg gtccctctcg ccgctcacc gtcaaaatg	1979

<210> 6

<211> 803

<212> DNA

<213> *Aspergillus niger*

<400> 6

ccagacagct ctggcggctc tgaggtgcag tggatgatta ttaatccggg accggccgcc	60
cctccgcccc gaagtggaaa ggctgggtgtg cccctcggtg accaagaatc tattgcatca	120
tcggagaata tggagcttca tgaatcacc ggcagtaage gaaggagaat gtgaagccag	180
gggtgtatag ccgtcggcga aatagcatgc cattaacctg ggtacagaag tccaattgct	240
tccgatctgg taaaagattc acgagatagt accttctccg aagtaggtag agcgagtacc	300
cggcgcgtaa gtccctaat tggeccatcc ggcattctga ggcgtccaa atatcgtgcc	360
tctctgctt tgcccgggtg atgaaaccgg aaaggccgct caggagctgg ccagcggcgc	420
agaccgggaa cacaagctgg cagtcgacct atccgggtgt ctgcactcga cctgctgagg	480
tccctcagtc cctggtaggc agctttgccc cgtctgtccg cccggtgtgt cggcgggggt	540
gacaaggctg ttgcgtcagt ccaacatttg ttgcatatt ttctgtctct cccaccagc	600

[0007]

tgctcttttc ttttctcttt cttttcccat cttcagtata ttcattctcc catccaagaa	660
cttttatctc ccctaagtaa gtactttgct acatccatac tccatccttc ccateccctta	720
ttcctttgaa cctttcagtt cgagctttcc cacttcacgc cagcttgact aacagctacc	780
ccgcttgagc caccgtcaaa atg	803

<210> 7

<211> 6493

<212> DNA

<213> *Rasamsonia emersonii*

<400> 7

ggagttgctt gcccgggcag gatatcggtt ggccggcgtg ttggacttga ttgcgaagca	60
gtagagacag gtattctgct gatagcatgg atttatatta tcaataagca gaggcctaaa	120
gacattcaac cagaagaaac gtcttcatg caacagcaag ctagaacata tacatatgag	180
aacaagcata gtgccaatca agcagaccta tactgtactc tattacaaca aactactatc	240
acggtagtcg tacactggtc aatgaaataa tgtgagtaaa aatgateatg attctatgac	300
agaacagcta gtacgcttga tttattgggg tataaatttt actttattta ggtgggtaag	360
agccagcaga ctagatatac agctcagata taatataatt aatagtcaca gaaaaaaata	420
aataaaaaata aaaatagcaa gatecatgat atggtataca caaaaaata ataatacataa	480
atcacacaat tccatcctct ccaaaaccac ctagccagct cctaccaaac gatacatact	540
cagtccaagc aaateccccg ttccgtcctc ccgtccgtcc gatcagtcgc gaataccgac	600
caaaaaaaaaa aagagacaaa atccaaatca cgggttcatt cacateccca caggataccc	660
atggatcagt cgtcctactt actgtgttac agattagaac agaaaattag gtttttacac	720
aactcagggt gggtgcattg cattgcattg tgctgtggag ttagttaact tagttgtact	780
ccatccagtt catacgcagt acattattgg gcatttgacc ccatcagaca agatatctaa	840
ggataagagt agaaaattagg taataatagt caaaagagaa gaagagatac caagggaact	900
agatactaac aaccaatcag gatatgccac agtgtggaac agaattggaag cagacaggat	960
caacataaact ggaaataacc ttttctttct ttcttcgtac agcatcttgg caggaagtaa	1020
cttgatattg ttaattaatg tccatgtcca tgtccatgtc tttgttatgt cttgtcttgt	1080
ctttcttgtg tgctacaagt acagtgtaac agattcatac ccgetgaaac agacataaca	1140
ttcgacacat ggaatacggg gataaagaaa taaattatac tatatacgga atgatatcaa	1200
taaataatccg tgttgtactc cttattaaga aagagtggcg cttggcgcgt ctacgtgcat	1260

[0008]

ggactggtac tacatatttg attacttcga tttttaataa caacaaccta gtagacgtat	1320
gatatgtcatg tgaaactttc gattgcgtgc tttcttgtct acttgtegcac ttgttacaat	1380
cttgtegaat attaataata ataatccatc gcactgacat cttggcaagt actccgtaca	1440
tcaggttaca tacatactga ttctctaaag ctagataacg aataggattc tcgcacagac	1500
agtatgtgtc tcttgtctgt cagatgataa gcagatgaac aaagaaagta taactgctta	1560
ctacctacat gccgacattt agtcgattcc tttcggagaa tttattatgg attattaata	1620
gcataccccc ggattggcag aaggggtaaa aggtccgact agacaaggat atccatacag	1680
tacataccgt tgatacagat cgaatcacat gcatactgct gatgggtgtga tgaatccttg	1740
aattagacaa tcatccagac ctgtctggac agagatcctg gcactgaaca atccactcat	1800
tgctatetat eggtactctg tacctgtttc agctgaagct tgccaatcgc agactgceat	1860
ctgcaactga tcagcgccag gatgcaggtc atgatacccc agcgttgttc ccgagggtgc	1920
attgcttaaa cgcgttaacc agtgtgctaa acgtgctaaa cgtgctaaat gctaaactgc	1980
tgatgctatg cagctgcac gccgaatctg gagaatgcag atcacctgcc gacggcgggc	2040
tccgggcacg tgcacggggg accccgtagg acagaaacgt ccatcgagag tacggagtac	2100
ggagtattac aagacctgt ccacagacc ctgtccatcg tcattgcaa gatctctcat	2160
tgtttgetgt tteatgctc gatcaccagt ggacagcaat gcccgtgaa cagcaagccg	2220
catgctggtc cgtgtcttgt ccgtgtgcc atgtagtatt getaacgaga cccagaatgg	2280
catcaatgac gttgcgatg acagaatgag ggggatcatc agtacgtctg ctatcaggat	2340
gattatccta cggagtattt actcagctga agacaggaac aagatcgtct gatggatgag	2400
gcccacggcc agccagcaca gactccgtac tcttcagtct tctggatttg accgttcgac	2460
ggegcctccg acgtagcatc tegetageet gatccttggc tgegcctate gtcggetcat	2520
gcccctgttg atgacgggga agtggagcgg cgccgcgata aggttgcett gctaatttag	2580
cgctgcacg ctccagccaa aaagaccaat attgaggtcg atcgtctccc ctggetccgt	2640
gctgctggcc tgcgategcc ggcgcgatea taccctgcaa tcacgccgcc agcctatac	2700
agaccatgcg gtccttgca cactctggag ctcgagctct cctgactgcc gtcggggcgt	2760
caatgcgtcc ggagcctccg acgagggcct ctgctcctcg tctgtcctac tggagcttgt	2820
ccgtcagacg tcgcatcctg agccgtgtgc tgatatgcc atggetctga cgtgatcgac	2880
tgcgagcggc cggcgaggct ataagaagcc gcaacttgct gctcgaagta ccgtctccca	2940

[0009]

tccatcgatc agacagtcag cagtcctcac tcagtcagtc ctcagtcgtc cttcaccacc	3000
atgggtctgt ccaaagcett cgtgtctgca ctctcgctgt gctccgccgt cgcctgtggc	3060
gccccgaccg ggccagctcc caacgtgcag ttctccctga agcaggtcgc ggtgccccgg	3120
accaagcttc gtgcgcccc agctgccgac tacgcgcgcg ctctggccaa gtatggcgt	3180
ccaattccgt cgtctgtgcg gacggccgcg tccggcacgc agagcggctc tggggccaac	3240
acgcccgtcg ccggcgacag ctgttatctc acgcccgtta ccateggcca gacacgctg	3300
aacctggact ttgacacggg ctctgcggat ctgtaagtgt cccaactctc gcaagaacaa	3360
gaacggagca gctgactcgt ccagctgggt cttctccaac gagacgccct ccagcgagcg	3420
cggcaaccac gccatctaca agcccagctc gacggccaag aagctgaacg gctacacctg	3480
gagcatctcg tacggcgacg gcagctcggc cgggcgcgac gtctaccagg acagcgtctc	3540
ggtgggcggc gtcaacgctt ccaaccaggc ggtcgaggcc gccaccaagg tcagctccga	3600
gttcacgcag gagccgggcg acggcttget gggcctggcc ttcagcagca tcaacaccgt	3660
caagcccaag ccgcagacga cttctctcga cacgggtcaag tcctcgctcg ccaagccgt	3720
gttcgcctgc accctcaagc acaacgagcc cggcagctac gactttggct acatcgacag	3780
ctccaagtac aagggcagea tccagtacac ccaggctcgac aactcgagc gcttctggca	3840
gttcacggcc gacggetact cgattggcgg cagcagcgcc agcgggtggct ccatttctgg	3900
cattgctggt aagaactccc cctacatcag agttatctag atgctgattt cgcagacacc	3960
ggcaccaccc tcctcttget cgacgaccag atcgtcaacg agtaactacca gcaggtccag	4020
ggcgcgcaga acgaccagaa cgccggcggc tacaccttcc cgtgcgacgc gcagctgccc	4080
gagctgagct tcaccategg ccagtacacc gccaccgtgc cgcccgagta cctcaacttc	4140
cagcccgtgt cgcagggcag ccagacctgc ttggcggtc tgcagtccaa ccagggcatt	4200
ggttctctca tcttcggcga cgtcttctc aagagccagt acgtcgtctt tgactcggac	4260
ggctctcagc tgggctttgc tgcctaggcg tagaccagtc gtctccagc ccaggttggt	4320
tggtaggaga tgatttttcg atcgatcgat tatcatggtg attgatagga tatgtgcatg	4380
agcagttgcc tgtacataca tacataatga ttatttgaat caattagtta tgatcaatct	4440
cgaatatatt ttcagtgaaa tacgtacatg gtcatagcat aacgatatac tccgtttct	4500
tcaggtagct agtaaataa cacaattcca tcgttctccc ggtccgtcag gtccaggaag	4560
gctttgtctc cgatcgtccc gtccgggata ctctcgctgg tctcgtgata gcgctccctc	4620
atcgagtaat caaccacett ggccggcttt ccattccgca cgcgcgcgcg ctctcgcac	4680

[0010]

cggttcagca cgaccaaatt cgcccactgc gccagcacga cggccaccag cgccacgaag	4740
atggccagac aagcccgcac gcccggccga tacgccggcg cgtccttctt gctgaagagc	4800
agcgggccga cgatgttgcc cgccgagetg gccgcgttgt acaggctcat cagegccgat	4860
ttcttegttg tgccgccgtt gttgccacg atccacgtca cgatcagtgg gttgccgccg	4920
aagagaaacg cgagcaggta gtagcctacc aggagggagg gttcgactga gttttgcttc	4980
gtgctgttat tactgcgtgg cacggcgtag agaattgccg gccccgcgac taccggcagc	5040
atgaagccgg ccagcacgac gcccttcac cgcgcccgct gccccagata gctccccgcc	5100
aggatgacca gcagctgcag cgcgccaaac ggcatgttga gcagactcgt cgtgtacgcg	5160
tcgtacecca ggccgttgag gatcagcggg ccgaacgtgt tgetcacgct ggccgccgacg	5220
ttcagcagca tcgcatgcc gatccagagg taggttttgg gctcgagcgc tgcctcgacg	5280
acgtgccgga tcttgaactc gcggctccct gtgcccgctt ggttcgcgcg cagccgctcg	5340
atggcctgcg ctttttccgt ctccgtcagg aaccgcgctg aggggatgtc gttgtctagt	5400
ttccagtaga tgaacggcac tgagatgatg gtcaggaggc cgacggtgag gaagatactg	5460
cgccagtcgt cageagtcag tcagtcagcg ttcggggtag aaggggagta catctgccat	5520
ggcctcagaa caggcgaetc gatatggccc aaccctacg acagggccgc cgcgatgaca	5580
gtcgccgcgc cgttggtact gtaccaggcc gcaatgcgca gcggtgctc ggccgcggcg	5640
taccactggc tggatgatgac gctgaacagc gccagacagg cggcctcgaa caggccgagg	5700
aagaagcgcg cggccatcag ggaggcgaaa ctgcgacagg cggccatgge ggctggggcg	5760
acgccccagc ccagacacag cgcgggcac caggcggcgat gccgcacgcg cacgatcagc	5820
cacgacgaga acggctgccg gacgagctgg gcgatgggcg cgatcgacce cagcagcgag	5880
tactggttgc ccgtcaggtg cgtgtcggec tgcaagccga aggtggcccc gtacccgagc	5940
accgaattgt ccaggatctg caggaagtac acccacacga ggatggccag gatgaecggg	6000
tctgtcttgc gccggatgcg cttgtgtgct gcgtccgtga gtgggattct ctgttgccg	6060
atcaggcgga gcgcgtgtc gccgtggaag gcgggttgct cttcttcatg ggtgacggtc	6120
ggtttggatg ccattgtagc gattactaga tgtaatcaag ttgtaatggg agacaaacga	6180
ccaagttctc tctcgacgtt ttataccggc ttatatgtct gttcagcagc attgcaagtc	6240
aagtaatgac atcggaattc ctccggttcc ccgcattgcg cggcgatcat cggctggcac	6300
tagcagtata gctagctcag agtccgtatt actggattct attgcattgc gctgattgca	6360

[0011]

gacgttgact gacagcagga gctttgactc tattaccccc acgttcggc aattccccgc	6420
gtgctcgggc ctctatgcac ccccaegtgg gggaacattc cagagtatgc aggacagtagt	6480
atgcagcatg gat	6493

<210> 8

<211> 1194

<212> DNA

<213> *Rasamsonia emersonii*

<400> 8

atgggtctgt ccaaagcett cgtgtctgca ctctegctgt gctccgccgt cgccgtggcc	60
gccccgaccg ggccagctcc caacgtgcag ttctccctga agcaggctgc ggtgccccgg	120
accaagcctc gtgcgcccc agctgccgac tacgcgcgcg ctctggccaa gtatggcgt	180
ccaattccgt cgtctgtgcg gacggccgcg tccggcacgc agagcggtc tggggccaac	240
acgcccgtcg ccggcgacag cttgtatctc acgcccgtta ccacggcca gagcacgtg	300
aacctggact ttgacacggg ctctgaggat ctctgggtct tctccaacga gacgcctcc	360
agcgagcgcg gcaaccacgc catctacaag ccagctcga cggccaagaa gctgaacggc	420
tacacctgga gcattctgta cggcgacggc agctcgcccg gcggcgacgt ctaccaggac	480
agcgtctcgg tgggcggcgt caacgcctcc aaccaggcgg tcgaggccgc caccaaggtc	540
agetccgagt tcacgcagga gccgggcgac ggcttctggt gcctggcctt cagcagcctc	600
aacaccgtca agcccaagcc gcagacgacc ttcttcgaca cggtaagtc ctgcctcgcc	660
aagccgctgt tcgccgtcac cctcaagcac aacgagcccg gcagctacga ctttggctac	720
atcgacagct ccaagtacaa gggcagcctc cagtacacce aggtcgacaa ctgcagggc	780
ttctggcagt tcacggcga cggtactctg attggcggca gcagcggcag cgggtgctcc	840
atttctggca ttgtgacac cggcaccacc ctctctctgc tcgacgacca gatcgtaac	900
gagtaactacc agcaggcca gggcgcgcag aacgaccaga acgccggcgg ctacaccttc	960
ccgtgcgacg cgcagctgcc cgagctgagc ttaccatcg gccagtacac cgccaccgtg	1020
ccggccgagt acctcaactt ccagcccgtg tcgcagggca gccagacctg ctccggcgg	1080
ctgcagtcca accagggcct tggtctctcc atcttcggcg acgtcttctt caagagccag	1140
tacgtctctt ttgactcgga cggctctcag ctgggctttg ctgtcaggc gtag	1194

<210> 9

<211> 397

<212> PRT

[0012]

<213> Rasamsonia emersonii

<400> 9

Met Gly Leu Ser Lys Ala Phe Val Ser Ala Leu Ser Leu Cys Ser Ala
1 5 10 15

Val Ala Val Ala Ala Pro Thr Gly Pro Ala Pro Asn Val Gln Phe Ser
20 25 30

Leu Lys Gln Val Ala Val Pro Arg Thr Lys Pro Arg Ala Pro Pro Ala
35 40 45

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

Ser Val Arg Thr Ala Ala Ser Gly Thr Gln Ser Gly Ser Ala Ala Asn
65 70 75 80

Thr Pro Val Ala Gly Asp Ser Leu Tyr Leu Thr Pro Val Thr Ile Gly
85 90 95

Gln Ser Thr Leu Asn Leu Asp Phe Asp Thr Gly Ser Ala Asp Leu Trp
100 105 110

Val Phe Ser Asn Glu Thr Pro Ser Ser Glu Arg Gly Asn His Ala Ile
115 120 125

Tyr Lys Pro Ser Ser Thr Ala Lys Lys Leu Asn Gly Tyr Thr Trp Ser
130 135 140

Ile Ser Tyr Gly Asp Gly Ser Ser Ala Gly Gly Asp Val Tyr Gln Asp
145 150 155 160

Ser Val Ser Val Gly Gly Val Asn Ala Ser Asn Gln Ala Val Glu Ala
165 170 175

Ala Thr Lys Val Ser Ser Glu Phe Thr Gln Glu Pro Gly Asp Gly Leu
180 185 190

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

Thr Thr Phe Phe Asp Thr Val Lys Ser Ser Leu Ala Lys Pro Leu Phe

[0013]

210	215	220
Ala Val Thr Leu Lys His Asn Glu Pro Gly Ser Tyr Asp Phe Gly Tyr 225	230	235 240
Ile Asp Ser Ser Lys Tyr Lys Gly Ser Ile Gln Tyr Thr Gln Val Asp 245	250	255
Asn Ser Gln Gly Phe Trp Gln Phe Thr Ala Asp Gly Tyr Ser Ile Gly 260	265	270
Gly Ser Ser Gly Ser Gly Gly Ser Ile Ser Gly Ile Ala Asp Thr Gly 275	280	285
Thr Thr Leu Leu Leu Leu Asp Asp Gln Ile Val Asn Glu Tyr Tyr Gln 290	295	300
Gln Val Gln Gly Ala Gln Asn Asp Gln Asn Ala Gly Gly Tyr Thr Phe 305	310	315 320
Pro Cys Asp Ala Gln Leu Pro Glu Leu Ser Phe Thr Ile Gly Gln Tyr 325	330	335
Thr Ala Thr Val Pro Ala Glu Tyr Leu Asn Phe Gln Pro Val Ser Gln 340	345	350
Gly Ser Gln Thr Cys Phe Gly Gly Leu Gln Ser Asn Gln Gly Ile Gly 355	360	365
Phe Ser Ile Phe Gly Asp Val Phe Leu Lys Ser Gln Tyr Val Val Phe 370	375	380
Asp Ser Asp Gly Pro Gln Leu Gly Phe Ala Ala Gln Ala 385	390	395

<210> 10

<211> 1066

<212> DNA

<213> Aspergillus nidulans

<400> 10

agacagctct ggcggtctctg aggtgcagtg gatgattatt aatccgggac cgccgccecc 60

tccgccccga agtggaagg ctggtgtgcc cctcgttgac caagaatcta ttgcatcatc 120

[0014]

ggagaatatg gagcttcac gaatcaccgg cagtaagcga aggagaatgt gaagccaggg	180
gtgtatagcc gtcggcgaaa tagcatgcc a ttaacctagg tacagaagtc caattgcttc	240
cgatctggta aaagattcac gagatagtag cttctccgaa gtaggtagag cgagtaccgc	300
gcgcgtaagc tccctaattg gcccatccgg catctgtagg gcgtccaaat atcgtgcctc	360
tcctgctttg cccgggtgtat gaaaccggaa aggccgctca ggagctggcc agcggcgag	420
accgggaaca caagctggca gtgcacccat ccggtgctct gcactcgacc tgcgtaggtc	480
cctcagtcct tggtaggcag ctttgccccg tctgtccgcc cgggtgtgtc gcgggggtga	540
caaggtcgtt gcgtcagtc aacatttgtt gccatatttt cctgctctcc ccaccagctg	600
ctcttttctt ttctttttct ttteccatct tcagtatatt catcttccca tccaagaacc	660
tttatttccc ctaagtaagt actttgctac atccatactc catcttccc atccettatt	720
cctttgaacc ttccagttcg agctttccca cttcctcgca gcttgactaa cagctacccc	780
gcttgagcag acatcaccat ggccaagttg accagtgccg ttcgggtgct caccgcgcgc	840
gacgtcgccg gagcggctga gttctggacc gaccggctcg gttctcccg ggacttcgtg	900
gaggacgact tcgccggtgt ggtccgggac gacgtgaccc tgttcacag cgcgggtccag	960
gaccaggtgg tgccggacaa caccctggcc tgggtgtggg tgcgcggcct ggacgagctg	1020
tacgccgagt ggtcggaggt cgtgtccacg aactteggg acgcct	1066

<210> 11
 <211> 1065
 <212> DNA
 <213> 人工序列

<220>
 <223> 人工序列

<220>
 <221> misc feature
 <222> (685)..(685)
 <223> n为a, c, g或t

<220>
 <221> misc feature
 <222> (893)..(893)
 <223> n为a, c, g或t

<220>
 <221> misc feature
 <222> (984)..(984)
 <223> n为a, c, g或t

[0015]

<400> 11
accagtgccg ttccggtgct caccgcgcgc gacgtegccg gagcggtcga gttctggacc 60
gaccggctcg gttctcccg ggacttcgtg gaggacgact tcgccggtgt ggtccgggac 120
gacgtgaccc tggtcatcag cgcggtccag gaccagggtg tgccggacaa caccctggcc 180
tgggtgtggg tgcgcggcct ggaagagctg tacgccaggt ggtcggaggt cgtgtccacg 240
aacttccggg acgectccgg gccggccatg accgagatcg gcgagcagcc gtggggggcg 300
gagttcgecc tgcgcgaccc ggccggcaac tgcgtgcaact tcgtggccga ggagcaggac 360
tgaccgacgc cgaccaacac cgccggctcg acggcggccc acgggtccca ggagcttgag 420
atccacttaa cgttactgaa atcatcaaac agcttgacga atctggatat aagatcgttg 480
gtgtcgatgt cagctccgga gttgagacaa atggtgttca ggatctcgat aagatacggt 540
catttgtcca agcagcaaag agtgccttct agtgatttaa tagctccatg tcaacaagaa 600
taaaacgctg ttccgggttt acctcttcca gatacagctc atctgcaatg cattaatgca 660
ttgactgcaa cctagtaacg ccttncaggc tccggcgaag agaagaatag cttagcagag 720
ctattttcat ttccgggaga cgagatcaag cagatcaacg gtcgtcaaga gacctacgag 780
actgaggaat ccgctcttgg ctccacgcga ctatatattt gtctctaatg gtactttgac 840
atgctcctct tctttactct gatagcttga ctatgaaaat tccgtcacca gcnctgggt 900
tcgcaaagat aattgcatgt ttcttccttg aactctcaag cctacaggac acacattcat 960
cgtaggtata aacctcgaaa tcanttecta ctaagatggt atacaatagt aacctgcat 1020
ggttgccatg tgaatgctcc gtaacaccca atacgccggc cgccc 1065

<210> 12
<211> 1501
<212> DNA
<213> *Rasamsonia emersonii*

<400> 12
ctgcgcggga gggcagtcac ggtaacgcc tccatctatt cccagagcg agaattcggc 60
ccgacgggac cctggagtcg aactcttgct cgaagtcggg atcccgatct atacctgca 120
ccagtggggt cgagccatga tgatgacgag atgcctcagg ttgtaatgag ccgccgatag 180
cacatggccc tggcatcgac gccacgttga ggaacgaacc tgacgagaca cgcatacacg 240
cacgttgata cgcattgac acaggaagat ataccttggg ggtctaatct gcctgccac 300
ttcagatcga ggcgggggagc tgatgtcacc gctgacagga atgaggatc tcgtaaataa 360
ggccttaaaa atgcatctga caaaaaatca ataatagacc tagagatgaa tggatgtatt 420

[0016]

tagccatgtc gatccaacga ctcttttgat cgagtctacg aactgacaaa tcatgggtcca	480
tatgtggtag ttttagtgaga actgaggttc tgtcgcccaa ggtgttggtc ttttttttta	540
tccggctaaa ttagcgagaa ctgtagtccc tcagtgtcca attctcgacg aattctctcc	600
catectccgg tgtaeaaagt agacaggatt taggataggc aagaetgata ttcgcgattg	660
cgccatctct tcagatgtcg attggtgggt tggaggatgg tcttagccct cgccttccgc	720
gtgccitttg atattcggcc gatactgtcg ggtgagaagt tgaggetttt tctgtgatg	780
gatatccaag ttatgagata tatccggagg gaattgagag gaatgatcag gaactgggtt	840
cagagggcac cagagggtga acctttttaa cagcattcga ttactggcat atacaacgcc	900
gtcttaaatg gtatttagag tactgccatg ctcccgaaac ggaaaagggc aaacaccctc	960
ccgactccac tcttgcgga ctcgaccgag tccacgcctc tccccgggg ctcgcaatga	1020
cgccatccgt ctattcgctc gagatggctt ctgatgtgca gattcccctt ccagaaccat	1080
gggtgtcgtg caccgcaaaa agggaatgcg ggacaccac cggcggccgg atcgcgagtg	1140
acaagaaccg cccacgccc gaagtctga tcgccgtcca atgggagccg ttctgccgcg	1200
ggcggtaggc gattgcgtga tcgccgagcg caggaggagc gcgattcacg acacaccttc	1260
ggggttcgtt gcccagtga ggttgggaac cgggtggagag aggggaggag gggctctgcg	1320
aggggtatca gtctcgag cttcttttat ttcaggggc gcgtcctgtt ctggtccaga	1380
gttgittctt totctcaac agettcgtcc gcagtcaatt cetctgcat cctctctc	1440
gaatcagaca accaccgtcc ttgtcaaaag tttctatacc ccctecggca ccgtcaaaat	1500
g	1501

<210> 13

<211> 1501

<212> DNA

<213> *Rasamsonia emersonii*

<400> 13

ccaattattc aaatcttcaa cctgggtect ttgcgactag actcggcaet ataccacact	60
gagtcgaatc tccgtggac gatttttttt cttagaagaa aaaaagcttc agaggctaag	120
gattaggett ccgtacggac tccatgccct atcagacaga gactccgcaa ctcattctgat	180
ctcttcgatg gagggaaaat ctgctgtttt tcgcaaattt ccaaccacc agaaacacc	240
agaactgtca cgaactcac gtcctacggt ctttgtgtga gtataactga ttatattcac	300
agatgggtta ctcaatgcag gtacaaactt tgatcgcgct ctttttggga ccgtctatca	360

[0017]

accggttcg aaaaatggct cgactagcca atctgacagg aaattgcgat gttgcaaccg	420
tgtatacggga gtctcttgta caacctctgc caacctctgc cacctcggta catgtacgga	480
gtagctcccc gcagccgcga ttggatgcat taaagtgggt caaccgcagt ggcttgcagt	540
ccgctgcacg agtccgtatg caataattct tgacacacac gagtgcacat aataatagga	600
aagcagacaa actttgagct gaaggctgtc gagcttggca aattgcagga tctggctagt	660
ttcgaagtcg acttcgcgog cgcagcagta ttgcattatt gagtgtgacc tgctgcgtgg	720
gattagcgtc gcaccggccg aaagctagtc tcatccaagg ctgagcctga gcgctaatta	780
ccccggatca gccaaagcct aatggatcta atgaggtgcc tectccagca ttcgccctgc	840
atgggtgcggc gacccctctc tccacgtcca ataattgctg ttgcgcctgt cgaaccctgc	900
caccgcatct ttgccgtttt actccgagat ctgaaaagcc tgetgtggat ggcagttegc	960
aatatgcact ctcaatcagg tctgtagcat cttttaacta ttattctatt actaattgct	1020
tctggaagge ttgtggggtg tggtttgtca tcaagttggc tccctgagcg ccgcgttgca	1080
atctccacgc gcggttgtac ggagtatatt catgcggatc cccggggcag agccgtagtg	1140
catgtgacac taatcgatca tccgctcaat tggatcctgg atttcgaccc tggtttgaac	1200
atatccaatg atcttccagg gacgaaccga ccgggtcatg ctttgttacc tacgtacgga	1260
gtagcggcct ggggtgatgg tccggaaggt ctgctaaaag gagatcgagt ataccccccg	1320
gggtccgtct gagacttata aagggtctct tgcaactctc cgcccgactt tttcttcat	1380
tcgacagcca tcaactgttt atctggtega ttctgcagac ttgcccaagg agcaaagagc	1440
atcttcatac gcgcateate catctccagc tttctctctc caaacatata ccgtcaaaat	1500
g	1501

<210> 14

<211> 1501

<212> DNA

<213> *Rasamsonia emersonii*

<400> 14

cagtaagtac tccccgtacg gggtagaatt tagcaattta cattacgtga tggaattgtg	60
gatgcagtca gggattggat gtagtagtat attaagcggt aaactaactg tattgtttaa	120
ttaggcacta aggtatgtac ctaggcctta tgctgtaagg tagggtttta gaagaaagaa	180
tgaaacgggtg cgtatagagt atggacacgg atttgtgcag atatgttatg tcaaagatca	240
tgcaaagcca agggatggac aggctggtat gtattgaccg aatcagcatg ctgtccttca	300

[0018]

cagcagtcag tctaccgagg gtattgactt ggcagtagcc ctaaattcta ttattgtgcc	360
atttgcacaa actttaagtg tgtctgccat tctgaagagg aataatacaa gttaaaacga	420
gaatagaata acaaggctat tattatcagg agaaagggtc tagatcaaac aaaatccttg	480
caagcaatta atttacctag gtattttatt ttattttctt caactagata gtttatgcat	540
gtgategate cgggtcaattat tatgcattca gatacatgct ttccacctga ctagcagaac	600
cctactccgt acgtagaccc etcagaaaca tggcgtgtgg cgttgcatga tttgatggcg	660
caacgctaata cgtttccacc ccaatggggtt ggccatctgt cttgcatgga gttgcctgat	720
agggtggtgac atgcgctage tgtttagtag tcccaacaga gtgtagacca tggttaatga	780
tactatactg gctggatgac gagaattatg cgattctgac ctctagcctt gagctgccgg	840
aacaactaac caaaggaaat tattgatgag cagacgagat atcatcacac gagattcagg	900
gaaatttcag gtaaatacca gccitggcagc agcacagtgg gacaacgtgt ctggtcagcg	960
aatcttgcca tataacctcg atggatggcg tgcagaccct cagccttgge ggcagctcgt	1020
tactgggaag gtttgatcca agtgagacat tgctatgctg gatcgagaca agtcgctaac	1080
catcatccag gatgcccttc atgcgactga ctcagcacga ttgactggc gagcacatgc	1140
cagccctctt ctcagccgag gttgtccgaa ctgccgaaat gaagctgcgc tgtttgtctg	1200
ccaccgattg ggttgatttt tgggctgaet ttctgggat gaccctgat ccatgtcgat	1260
atcgcgatgt ctgcgccgga tggcgaatca gggcgacacc taggggggga gtgcagtctt	1320
tgctcgtggc tttgtggtat ataaagggtg tgaaaccegc tgtccacgat cgatcgatcg	1380
tccagttega tcgtccagtc tgagaatata ccacaggccg ttctctgaac tatcttgatc	1440
acgtctgttc cgcacgctcg ctcatgacac aatcagtacc agactctcca ccgtcaaaat	1500
g	1501

<210> 15

<211> 651

<212> DNA

<213> *Rasamsonia emersonii*

<400> 15

catctgccat gaaaatggct tctgggagat attagtgcga tgtatataat tcaaactcttg	60
agttacttgc gagcatgtac tgtacatacc taggcaccag gatccttcat tgatcaatca	120
cgtattatct tategataac gagcctgcaa atctcggttc gctegaactga ttacggagt	180
caagagaaaa tactctgate taggccaaga agtcctttatt atatactccc aaggcagget	240

[0019]

tcacaccate ttectegcca tacttgacgt atttctacgg agtatattgt agaaaagcac	300
gtgacagtgt gcaaaaatgg aaattaatgt tctgagagta gtactccgta gtgatgattg	360
gccatgttca tctaaateac gtgagagaat accgatgtcc gtctagectc ttgcgggtat	420
agtagatagg gcacgagact taaacctaa acctaaagcc ctactctcgc gtaccgcact	480
ccaactccgc agacggccaa ccaaccaaac gctctgccgc gaaacctcc aatttggctt	540
gcacatcgat cgaacttgac cgatccgacc gacaaacgac gaacgacgac gacaacttcg	600
ctcccagcaa ccacagtcgc caacaccegc agaccacca ccgtcaaaat g	651

<210> 16

<211> 1501

<212> DNA

<213> *Rasamsonia emersonii*

<400> 16

ctgatattat gccaaaggctc ttattgagca acttcttta cctacgttaa gttgcagtta	60
agcgatagtc ctagccctct aaggtaatat acggagatac tccggagtcc gattgctttg	120
cagaacaacta tgaagaggat gcattgcaaac ccattgtctta agggcgcaat ccgtgccatt	180
atcccatagg ggttttactt tgtctccctt gggggaaaac tgtccaaaat accgtgtttt	240
ctatatccgt cattgttgat tgagaagcct actactgact ctttgatcag tcgaaattgg	300
atacatgtat ttcccagttc tgccatgttt tacaaaagtg gcaggatcct acaaccgatt	360
taccttactg atcgaagcga aagcactaat tcgaacaagt ttctacccc tcatccagga	420
cgatataact ctaattcctc ccccaaattc gtccctcgca ccgtgcagg aactctgcag	480
ataactggat agcctccctt ccacgtgggg gctaagttcc gactactccg tagctacatc	540
ttgctagaca tggatcattc cgaacccacc tatagtattt cgagaacggt acctaaact	600
ggaactaatc acgtcttggc agaggaatta aagttggggc tcgttaagat ctggctctct	660
taatcgagat aaggtaacat cggcgatgat ccaagatttg ataccaaact atggagtaca	720
tacgtcagc ctcattccgat cgaacccaaa tttttagtatt tcaattagg tagctgatat	780
tttggttaga tcattctaga attacaaaat aggccagtat attgctattt cagcagtget	840
ggactatitt taagtectac ccaaattaaa gtccacgtgc ctttgggacc cgttgcctc	900
gggaaaacct cagaagtaca gacgactctg ctgattcctt aggcattac ttgtactcca	960
taatagcctt totatttttg aggtctttga ggcgtttttg attcgaataa tgtgatattt	1020
tcccgtgcca agctcgccaa cagtccctac ggactgcttt gataaagtcg gcaccgcggc	1080

[0020]

atgatgccct acatecgcaa tgcaggaacc aaattggaga ctttccaatg aaaagtctcg	1140
aaattagctt cagggaactag ttagttcttc cagataggat tcgcagacc tgatctgaet	1200
tgactatagt tgcctaaac cagecacatt gctaccacga ggtcaaagac atcaaattag	1260
tacggaacat cgctctatac gacatcacac tgtaaccttt tttggttcaa cgatggcacc	1320
ggttagggaa aacatgaggc ttcttagaca tgttctcgat tttttttctt gggtatataa	1380
ttgccatgga tacataatgc ccatggagtc cagaatttcg ccagagcagt tgcaactttt	1440
tgctcattct atagttcttt ctttctcgtg ttattattga attcattgca ccgtcaaaaat	1500
g	1501

<210> 17

<211> 1501

<212> DNA

<213> *Rasamsonia emersonii*

<400> 17

cggtcattta tttctatga tggccatagc taaagtacag gtagtcttga ctaagtttac	60
actgccttcg ccagaatatt cttaacactg gcacgtagag gagactcgtt gttcacgacg	120
tcattcatttg caatgaagag ttcttcacat ggcattatat tgtttaccag ttgcagcgac	180
tccgcacctg aatttcagg cattctcacc ggagatctaa caatggcaat tgtagaagc	240
atccatggat atgcgcgctc ccgcaacgtt gcagccttgg tgccggagca gacggagaag	300
gaaagtaaaa tatattaggg agcagagtcg catgtgaggg gaacgtgatc ttatgtatca	360
tttttcggcg aacaaacgtg atgggccaac atgatcgttc gagttagctc gactggacaa	420
gctttagccc aggaataag gtacggagta aatgcgaatc cgacgcata tcggcaacta	480
gcaccgacac tatgeaacc tggacgacgc catcatacct ccctttactc ttacggcgt	540
actttggaca ggggacgggc ttttatcatt tctccaactc caataattgg cagactaagt	600
cttgcgactc ttgggggata ctctgtctgg atctgaagcc ggatccagcc gggtcgagct	660
gtgcgacacg cgacgcacag gcatcaccag gcaggtcact tgatgctggt caccaggac	720
aagagaggca cattaccagg tcattgctgc aacatgagcg tgagcacgga ctgcgagaga	780
ttgtcccagc cctggagggtg agcttggttg ccaatcaact gcggggagga ctcatgacaa	840
tcgacggtcc acggatgatg gccatgcaga ccagggtgca gcatcaaatg cgattctgct	900
gtgacgccgg acggaaatga gatgacctg gcatgcgact ccgggcaccc ccttgcagg	960
gcctcgtctt cagccccgca tccgacttg ctgctatcca tggatcgca caggcatccg	1020

[0021]

```

cgcgggccgga ctggccaact tcccaattct gcccctgccg cgcgcacggt gtcgctagcc 1080
ccaggcttgt cgcgcacatct ccatgaacta gcccgaacgc ctgacgagct actgggacga 1140
ggtagacattg agcctctacc atgggtctga tgataccctc catttacgag cctgacgtgg 1200
cccggcaatc gtcacgcaa tgcctcacc tcggggcatt tcccaggcca acgtcgcagt 1260
cctctgtcgc tgtttcctgc agtcgattgg cagcctttg ggaattgccc cgtgtctgcc 1320
atttctcggg cttecccttc cgtggctgac tggtgatgg tagacgatag aaggtaatat 1380
tccacgatgg gcacggtggg agattcaaaa acgtgcacat ccccgagcgt cactcgggaa 1440
agagcctctc tccagactct cccgttcgcc ctgtctttcc ttaacgacca cgtcaaaaat 1500
g 1501

```

<210> 18
 <211> 625
 <212> PRT
 <213> *Rasamsonia emersonii*

<400> 18

```

Met Ala Ser Leu Val Ala Gly Ala Leu Cys Val Leu Gly Leu Thr Pro
1           5           10           15

```

```

Ala Ala Phe Ala Arg Ala Pro Val Ala Ala Arg Ala Ser Gly Ser Leu
          20           25           30

```

```

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

```

```

Asn Asn Ile Gly Pro Asn Gly Ala Asp Val Ala Gly Ala Ser Ala Gly
          50           55           60

```

```

Ile Val Val Ala Ser Pro Ser Arg Ser Asp Pro Asp Tyr Phe Tyr Ser
65           70           75           80

```

```

Trp Thr Arg Asp Ala Ala Leu Thr Ala Lys Tyr Leu Val Asp Ala Phe
          85           90           95

```

```

Ile Ala Gly Asn Lys Asp Leu Glu Gln Thr Ile Gln Glu Tyr Ile Ser
          100          105          110

```

```

Ala Gln Ala Gln Val Gln Thr Ile Ser Asn Pro Ser Gly Asp Leu Ser
          115          120          125

```

[0022]

Thr	Gly	Gly	Leu	Gly	Glu	Pro	Lys	Phe	Asn	Val	Asn	Glu	Thr	Ala	Phe	130	135	140
Thr	Gly	Pro	Trp	Gly	Arg	Pro	Gln	Arg	Asp	Gly	Pro	Ala	Leu	Arg	Ala	145	150	155
Thr	Ala	Leu	Ile	Ala	Tyr	Ala	Asn	Tyr	Leu	Ile	Asp	Asn	Gly	Gln	Ala	165	170	175
Ser	Thr	Ala	Asp	Glu	Ile	Ile	Trp	Pro	Ile	Val	Gln	Asn	Asp	Leu	Ser	180	185	190
Tyr	Val	Thr	Gln	Tyr	Trp	Asn	Ser	Ser	Thr	Phe	Asp	Leu	Trp	Glu	Glu	195	200	205
Val	Glu	Gly	Ser	Ser	Phe	Phe	Thr	Thr	Ala	Val	Gln	His	Arg	Ala	Leu	210	215	220
Val	Glu	Gly	Asn	Ala	Leu	Ala	Thr	Arg	Leu	Asn	His	Thr	Cys	Pro	Asn	225	230	235
Cys	Val	Ser	Gln	Ala	Pro	Gln	Val	Leu	Cys	Phe	Leu	Gln	Ser	Tyr	Trp	245	250	255
Thr	Gly	Ser	Tyr	Val	Leu	Ala	Asn	Phe	Gly	Gly	Ser	Gly	Arg	Ser	Gly	260	265	270
Lys	Asp	Val	Asn	Ser	Ile	Leu	Gly	Ser	Ile	His	Thr	Phe	Asp	Pro	Ala	275	280	285
Gly	Gly	Cys	Asp	Asp	Ser	Thr	Phe	Gln	Pro	Cys	Ser	Ala	Arg	Ala	Leu	290	295	300
Ala	Asn	His	Lys	Val	Val	Thr	Asp	Ser	Phe	Arg	Ser	Val	Tyr	Ala	Val	305	310	315
Asn	Ser	Gly	Ile	Ala	Glu	Gly	Ser	Ala	Val	Ala	Val	Gly	Arg	Tyr	Pro	325	330	335
Glu	Asp	Val	Tyr	Gln	Gly	Gly	Asn	Pro	Trp	Tyr	Leu	Ala	Thr	Ala	Ala	340	345	350

[0023]

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

[0024]

Glu Tyr Lys Phe Phe Lys Lys Glu Thr Asp Gly Thr Ile Val Trp Glu
580 585 590

Asp Asp Pro Asn Arg Ser Tyr Thr Val Pro Ala Tyr Cys Gly Gln Thr
595 600 605

Thr Ala Ile Leu Asp Asp Ser Trp Gln Asp Tyr Lys Asp Asp Asp Asp
610 615 620

Lys
625

<210> 19
<211> 2208
<212> DNA
<213> Aspergillus nidulans

<400> 19
atggcctctc tggctcgtgg tgcctctctgc gtccttggat tgactcccg cgccttcgct 60
cgcgcccccg ttgctgctcg tgcctccggc agccttgact cgttctggc tactgaaace 120
cccattgctc tgcagggtgt tctcaacaac attggtecca acggtgccga tgcgctgggt 180
gccagcgctg gtattgttgt tgcctcccc tcccgctccg acctgaacta ctctactec 240
tggactcgtg acgcggcggt gactgccaag taactgggtg atgccttcac tgettgtaac 300
aaggatctcg agcagaccat ccaggagtac atctctgccc aggcgcaggt ccagaccate 360
tccaaccctt ccggtgacct cagcactggc ggcttgggtg agcccaagtt caacgtcaac 420
gaaaccgctt tcaccggacc ttggggacgc cccagcgtg atggacctgc tctgcgcgcc 480
accgcgctga ttgcctacgc caactacctc atcgacaacg gccaggcctc cactgccgac 540
gagatcatct ggcccattgt gcagaacgac ctctctacg tcaccagta ctggaactct 600
tegacctteg atctctggga agaagtegag ggttcgtcgt tcttcaccac tgcgtecag 660
caecgtgctc tgggtggagg taacgcgctg gccacccgce tgaaccacac ctgcccacac 720
tgtgtctccc aggtcccca ggctctttgc ttcctccagt cctactggac tggcagctac 780
gtccttgcca actteggtgg ctctggccgt tccggcaagg atgtcaactc cactctcggt 840
tccatccaca ccttegacct tctgggtgga tgcgatgaca gcacttcca gccttgctcg 900
gcccgtgctt tggccaacca caaggctgct accgactctt tccgctcggt ttacgcgctc 960
aactccggca ttgctgaggg ttctgcgctc gctgttggac gttaccccca ggatgtctac 1020
cagggtggaa acccctggta cttggcgact gctgctgctg ctgagcagct ctacgacgce 1080

[0025]

atctaccagt ggaacaagat eggcagcate tccatcaccg atgtctctct tgcctttctt	1140
caggacatct acccctccgc cgccgttggt acctacaact ccggcagcag caccttcaac	1200
gacatcatct ctgctgttca gaacctagcc gatggctacc tctccatcat cgagaagtae	1260
acccctccg atggtctcct caecgagcag ttctcccgct ccgaecgtac tcccttgctg	1320
gtttctggtc tgacctggtc ctacgectcc ctctgaccg ctgctgctg ccgtcagtc	1380
atcgtccccg ccagctgggg tgagctctct gccagctccg tgcctgctgt ctgctctgcc	1440
acctcggcga ctggtcctta cagcactgcc accaacaactg cgtggccctc ctctggtcc	1500
ggtcccagea ccaccacctc cgttcttgc accaccccca cttctgttgc cgtcacctc	1560
gacgaaattg tgtctaccac ctacggcgag actatctacc ttgttggtc cctccctgag	1620
cttggttaact ggagccctag ctctgccatc cccctccgtg ccgatgcta cactcgtcg	1680
aacccctct ggtaectgac cctcaacctt cctgctggta cctcttctga gtacaagttc	1740
ttcaagaagg agactgacgg caccatctg tgggaggatg accccaaccg cagctacacc	1800
gttctgctt actgcccga gaccaccgc atcctggacg acagctggca ggactacaag	1860
gacgatgacg acaagtaaac taataagtgt cagatagcaa ttgcacaag aaatcaatac	1920
cagcaactgt aaataagcgc tgaagtgaac atgccatgct acgaaagagc agaaaaaac	1980
ctgccgtaga accgaagaga tatgacacgc ttcctctct caaaggaaga atccctcag	2040
ggttgcttt ccagtctaga caegtataac ggcacaagtg tctctacca aatgggttat	2100
atctcaaattg tgatetaagg atggaaagcc cagaatattg gctgggttga tggctgctc	2160
gagtgcagtc tcatgctgcc acaggtgaet ctggatggcc ccatacca	2208

<210> 20

<211> 7890

<212> DNA

<213> *Rasamsonia emersonii*

<400> 20

ggagcctggc tcagcatgct caeggactgc aggatgaaca cgcgtctccg aaggtgcaac	60
cggggagaag ataaccaccg tcttgaggg aagaacatct gtcctcatc aaacagaage	120
agtctcagtc tcggtctcgg tctcagtcct gtcctctata gtcgtcacag ttgggacgg	180
aaacgtggcc gaccagtccc tggctcttga ccaggtgtta ttctctatc gcggcggcgc	240
gatcatgtcg agcgcctcac caatcttcc gaggacatag cccatctga accgcccctg	300
gaggcgtctg atgaagttgt tggcgggatg aacggcggtg aaagggatg catggaagat	360

[0026]

gccgaggatg gccatggcca tggagaacat gaacaggccg aaggggtcca tggatgatgat	420
gattcgatcg atatcagaca gctgaatatg tattagcttt ccaaatttct tctatcatga	480
aattcgacag gaaatagaaa aagaaaggat gctggctttg attaaagaga gtagctgggt	540
cttatecaaa gagattaaaa agtgtcttaa acaaagacaa gacaagacaa gcatcctct	600
gatgcatgag ttctcagata tatataatac aaaggatgga ttactgaac agttcgttga	660
cttctgaaga acaaaggctg gtctctgcat gcccgagaat agtaacaaca ataaagagca	720
aatacaacgt tgtgtacagc aagcgaatga cctgccttga atgagtctgg cagatttatt	780
ccccatcta ctaccttcaa gtacctcatc agatggccag cagaagggtc aagtgggtat	840
atcttctcat tcgaagaact tagtgtttag tttctgagc agcaatatag ctagtgtcaa	900
gttagaaaag agtataaaga accgtttccg caacaccagc taccctcaaa gaaaggaatt	960
gaacaaactg ggaactaccg attactacca agctggaggg cataggcaat gaacgccagg	1020
aattggtaaa gactgaagat aggaaggat ggtgaatact gcatagtga atgtagacct	1080
tcagtatcat aaggatcatc cattcattat agacgcta ataaagtatt tctgaaaaaa	1140
aatgttggag aggagatcaa gtctttttat tcaacggctt tcaaagacta aaagctggaa	1200
aaggcaggct atcatcatga tacttagcat agcataccat gtttgcgttt atcattcat	1260
attcaggaca acctctgcc caggaccacg ccataggcaa ctgtgccag gagaccgcc	1320
aggttgacca gcgtcgagat tccatgcagc ttggcaaaact tttgttgag ctcggtcatc	1380
tccttggagt gaggaggcgg gtcataactc ttcttgccgt cgategatte tgccagagat	1440
catatatatc agcccgtaac tatacagtgt actcgtgcgt actcggcagc agccaaaccg	1500
gtaggtagta ggcaggtaac aaccttgctg ccacctctgc ttgataactc ccacaacctt	1560
cggggtcaga tagagcagg tggtcagtc cgaaacgaag aegatcga aa gcgcagcaa	1620
caccgtcaga egatttctt cgagaagcac ccccgcgagg ctagagggcc cagtteccag	1680
gaccgttctg gctccgggt aggtcagcgc tgcgacaacg gggagagcgc tctgcagggt	1740
gaagtagatc gggaacaggc tattctggag cgtcgagaac tgetgacggg gaagcgtcct	1800
gaacgcgacg gttccgcaa cgaaggctta catatataac ggaatgttgt aggagctttt	1860
gatgagttat ctatctcctt gaccaattga ccagaaacaa aactcaggat gtcaaacctg	1920
atagatctcg actcccagaa gggtgccgta gctgattgaa gtgtgatgtc agccgatcta	1980
ttagtagcag ccgtagtact ggtgacgcac cttagtatgt ggaaggggcc caggatagac	2040

[0027]

atcctgtctct gttgtactga tatggaaaca cctcgtgact ggaacagaac tatatgggat	2100
tatacttaga cagatacccc actgactggg aattcagagg gaagagtaag ttgtgttatg	2160
ctacgggtag gttagagaag ctgtcaagct tgggtctccc gagctaacgc tagctgcatg	2220
tggggcatgt tcttatctcc acggcccgtc caaacctaga tctgttcca acaaagcaca	2280
aatatctata cacacggcct tttccgtaag gcccacgcac cttcccgcgc tcatgtgcac	2340
tegegtctgc cgcgcctcaa aaaggaaata tcacgcgtct gcttgaggc gctccttagt	2400
catagaaaga aacgcattca cgccatgcag tgatttatct atctgacatt tcttctctct	2460
tcgttgacgc aggagggaca gctgacatct cttttgcaaa atggctgaca aggaggccac	2520
cgtctacgtg atcgatgtgg gaaagtccat ggggaggcgc cgccatggac ggccgggtatc	2580
tgacctggaa tgggcaatgc aatatgtctg ggacaagatt acgacaaccg tatgtctgaca	2640
cttgatccgg tctcctggaa attaaattcc tgcgttgaga actgacatat cttctgttag	2700
gttgccacgg ggccgaaaac ggctacaatt ggagtggctg ggctgaggac agatggtgag	2760
atcttaccgt gcccgaaatca ggtaaataatg atttactgat gtatctggac agaaacatcg	2820
aacgacttgc aggatgatga cagctattcg cacatctctg tctttcagga aattggacag	2880
tatgtgcctc agctgacact gatgactagt gacttttctc cgcataact aaataaatca	2940
ctgccagggt cctcatgcct gatctgcgaa aactgcgcga cctgatcaag cctagcaaca	3000
ctgatgaagg agatggtgag ttttgcctgt atcttcggac tcatttgatt tgatattgag	3060
acctatctac ctatagctat cctctccctt gtcgtcgcga tccagatgat caccacttat	3120
acaaaaaagc tgaagtatcg acggaaaatc attctcgtga cgaacgggga aggatccatg	3180
agtaccgatg gtcttgatga gatcgtgaaa aagctcaagt ccgatagcat tgaattggtg	3240
gtcttgatg tttttcactt ctctttgact tttcttggtg ctggtatgca aaatggctaa	3300
actggtttcg ttgcagggtt gttgactttg atgatcctga atttggtgtc aaagaggagg	3360
acaagaatcc agcaaaagta ttcaatgttt tttttttagc aggttggaag agttgctgat	3420
tegatctgcc gcaggctgag aatgaagcgg tctcagagg tctcgttgat tcttgcgcgc	3480
gagtctacgg gacattacaa caggccatat tggagctgga cacaccgcgt gtgaagggtg	3540
ttcgtggaat accctccttt agaggagagc tccgactggg gaacctgaa gagtattcgt	3600
ctgcccttcg tatcccagtc gaaagatact accgaactta tgttgccaag ccgccgacag	3660
cgagctectt tgctctacga tetgacgtcg cagctggtea agagggtgca gagaatgcac	3720
tgacaagcgt ccgaaacgca cggacatatc acgtcagtga tgagtccgca ccaggaggca	3780

[0028]

agagagacgt ggagcgagaa gatctcgcca agggctacga gtatgggaga accgcggtgc	3840
acattagtga gtccgatgag aatatcacca aactccagac gaaccctggt ctggaaatca	3900
tcggttcat tcagagtac catgtatgtt tctcgtcaag ggtatctcat ctgaaccgtg	3960
attaacctag gatccagtac gaccgataca tgcacatgtc taccagcaat gtcataattg	4020
cacagaaaagc aaacgaaaag gcgatccttg ctctttcattc tttcattcac gccttgttcg	4080
agttggactg ttatgctgtg gccagacttg ttaccaagga caacaagccc cactcatcg	4140
tattactggc accatctatt gaagcagact ttgaatgtct tctagaagtc cagctccctt	4200
ttgctgaaga tgttcggtcg taccgtttcc ctcccttgga caagggtggtc actgtctctg	4260
gaaagacagt caaagagcac cgacatctcc caagtgaaga attgctgaat gcgatgagca	4320
aatacgtcga cagcatggag ctcgctcgaca aggatgaaaa cgggtgagtc atcacaggga	4380
aaccgtcatg ctgctcatct caagtatact gacaactcca cagagaacca gttgacagcc	4440
tggtcccgag actggaggat tegtactctc cactgctgca caggatcgag caagctatcc	4500
ggtggcgtgc catccatcca aacgagcctc ttccgcccc ttctgagaag ttgacgcagc	4560
tgtaacgacc gccagcagat ctgcaagcgc gcgcaagaa atacctggat cgggtcattg	4620
ccgccgccga tgtgaagaaa ggtctgtcaa cttctacgtt ccccagaat gcctctgact	4680
aaaaaatget gcacagttcc accaaaagca aaaggctgca ageggaatcg cgaagccgac	4740
aaacccctat cgggtcttga cgttgacgag ctctctctgc gegagaagcg cgccaagatc	4800
tcagccaaca acgcatccc cgagttcaaa cagtcgctgg tcaacgccga gaccatcgac	4860
gccgtccgtg acgcagtcag ccagatggaa agcatcatcg agaaccacat ccgaagcagc	4920
tttgagacg ccaactacga ccgcgtgac gaggagctgg gtgtcctccg cgaggagctg	4980
atgcctacg aagagccgga tetetacaac gacttctgc ggaggetgaa ggacaagatc	5040
ctcaatgagg agctgggcgg agacagacga gagctgtggt ggetcgtcag gaggcaacgg	5100
gtcggcttga tagacaagaa ggcgtcgga cgggttgaag ttactgaaca ggaagccagg	5160
gaggtaagta agcagataca ttattccttt agttccatta aacgagctgc atgatgagct	5220
gacttttgtt cactagttca tgacctcgaa ataaaatagt ccattattgc tatgtatgtc	5280
aaggcgctg gccgtagtag tcttaacatg ctgatgctgt gaatcaaagc gccagatgaa	5340
caataataga aataatacca cttggtagct gtctccattc tcacagatag acaacgttaa	5400
agaaaagaaa aacgtaaaaa gagggatatat gtggtctagt aacgccgcaa ggaaaaaaaa	5460

[0029]

actcatacgt tagtttcgaa cgcaaattctc aaaatcgagc acttcgagta aatactctgt	5520
cgtatcgttt cgcctcagga tatcttcccg agccttctct ttccgatatc gattttccgt	5580
tgtaatctag ttattattac tccagttagt aaatgcacga cgggcagtat tgtaaataat	5640
gaaatcagea gcgagagtac gaacatgtcc acatcctcat cggtttcccg gagcaactcg	5700
ttctggatct ccagctcatt gttaatggcg atccccagct cttctgtcgc agcgacgatt	5760
ttcatcaact cctccacgct cgggtcctgc tcttccatcg tctgttctg cagctggagc	5820
acgcccgtgt tgtecgagtc ccgcgtcttg tccgtttcct tgcccaggac tcgtccagaa	5880
cggggtttgg cgctccccac cagggcgtcc ttgtcctgca tcgaagcgac agcgttgctc	5940
agcttgcctt tcgtcaccat cgcgttggtc agattctcca atccgtcctt ctctttcttg	6000
gcgctcgcca tgagatcctt cctccgacgg atctctccct cgcctaacct gctgccgccc	6060
catccagaag acttgctgct caggttcttc agccttctct caagagcgcc gatcatcgac	6120
cctgctttca ccaagctgct tttagcctgt gccgagctct cgtgttggtt ctgtggagtc	6180
gtggcctgat cagctctcgt cagatgcagc ctctgtctgt gtaagtgcgc ctteatctcc	6240
cggtagcaat ccagccacag gaactggatcc gtgatcggtc cgcgcctgg cgcgcccggt	6300
tccgtgatgg acttggtgaag ctccgatgcg gcagagctat ccgacagggc ttgggagggg	6360
aggtttagaa aggacctcca gacgcttggt tgacgccatc gcgggtcctc gctctcgttg	6420
atggccccga ggtatcttcc caggccttgc cgcgcctctt cgcgcagagt ctcttccag	6480
ttcgtgttgg aaaaccagga ctccccgggc agagcgacgg gtggttgggc gccaacctgg	6540
cggactagtg cgtcatggaa cgatgcaaat tctgaatagc gttcttggaac aacgaacgac	6600
cgtagaggca gccggatggg gatgttgtat agcgtatacg gaactgggagc gtccgcgatg	6660
gtggctgtcg ggatggaaat ttcgacattc ggggccatga ttatagttca gacgggaaaa	6720
agaacaaaaa aaagagcagg cccttggtat cgaccaggaa gcataattcc cgcgccttct	6780
cttgcggtat ctctgtcgtt gcagagttgg ttgcagagta gtggagtcgg ccggcgggtg	6840
gaaactcccg caatgacgca ggcgccccat ctctttctgc caccgccgat ctgtggctta	6900
gttctttctt gtcaagactc gactccacca tcgcgactcc aggcagcacg aatcgcacga	6960
ttgccgaaaa actacaccgt actaggggaa ggcctaatta atctattacc ctagctaaaa	7020
atggggttgt caaacttacc atatagccgt gcgacccgcc cttggaggtc actagatcca	7080
acctgcgcac ggcttggtta cggttgatgg gagctaaaaa tagaacgaaa gatatactgg	7140
cggctccgtc ccgcgtctat ccacaatcca aaactcgtat gcagagttat ctacaggteg	7200

[0030]

atccaatcat gagtcttttg tgacatgtcg ttgaatacat ggtctcaatc gagtctgccg	7260
ttcttacaatg accatctctca ceaagatcaa tgtcccgtga ttcgactgtc agccaagata	7320
cgtctcacct ggccccatct ctactgtcga caacgtctgc ctatactgta ggtgatcaga	7380
atacgcagtc cgggggagtc tactcgcgat ggggtgggtc atacgtcggc tcctcgtcga	7440
cgttgctctct gggtcctctg gagagcgtca atatagaagg gagaagaaag ttgctcttga	7500
tctatatcca tggcttcatg ggtgaagaag cgagcttcca caagttccct gctcatgtcc	7560
ataaccttgt caccattgct ctggccgagt cgcacgttgt gtattcgaag gtatatcctc	7620
gatacaaatc ccgccgagca atggacattg cacgtgatga tttcagtcga tgggtgcgtt	7680
gcagactggc atatctctct ttagagatca tcctagaaaag aaacgcataa tactaagtgt	7740
cgaatagget atcaccgcat gagtcggaag atacagatgt gatcetaetc ggccacagcc	7800
tgggtgggat cctagccgca gaggttgcgc tgctcccatc agccctggg agcaaggaga	7860
tcttcgagca tcgtatcttg ggactcatca	7890

<210> 21

<211> 2145

<212> DNA

<213> *Rasamsonia emersonii*

<400> 21

atggctgaca aggaggccac cgtctacgtg atcgatgtgg gaaagtccat ggggaggcgc	60
cgccatggac ggccggtatc tgacctggaa tgggcaatgc aatatgtctg ggacaagatt	120
acgacaaccg ttgccacggg gcggaaaacg gctacaattg gagtggtcgg gctgaggaca	180
gatgaaacat cgaacgactt gcaggatgat gacagctatt cgcacatctc tgtctttcag	240
gaaattggac aggtctcat gcctgatctg cgaaaactgc gcgacctgat caagcctagc	300
aacctgatg aaggagatgc tatctctctc cttgtcgtcg cgatccagat gatcaccact	360
tataccaaaa agctgaagta tcgacggaaa atcattctcg tgaagaacgg ggaaggatcc	420
atgagtaccg atggtcttga tgagatcgtg aaaaagctca agtccgatag cattgaattg	480
gtggctcttg gtgttgactt tgatgatcct gaatttggtg tcaaagagga ggacaagaat	540
ccagcaaaaag ctgagaatga agcggctctc agaggtctcg ttgattcctg cgacggagtc	600
tacgggacat tacaacaggc catattggag ctggacacac cgcgtgtgaa ggttgttcgt	660
ggaataccct cctttagagg agagctccga ctggggaacc ctgaagagta ttcgtctgcc	720
cctcgtatcc cagtegaaag ataactaccga acttatgttg ccaagccgcc gacagcgagc	780

[0031]

```

tcctttgtcc tacgatctga cgetgcagct ggtcaagagg gtgcagagaa tgcactgaca      840
agcgtccgaa acgcacggac atatcacgtc agtgatgagt ccgcaccagg aggcaagaga      900
gacgtggagc gagaagatct cgccaagggc tacgagtatg ggagaaccgc ggtgcacatt      960
agtgagtccg atgagaatat caccaaactc cagacgaacc ctggctctgga aatcatcggc     1020
ttcattcaga gtgaccatta cgaccgatac atgcacatgt ctaccagcaa tgtcataatt     1080
gcacagaaaag caaacgaaaa ggcgatcctt gctctttcat ctttcattca cgccttggtc     1140
gagttggact gttatgctgt ggccagactt gttaccaagg acaacaagcc cccactcatc     1200
gtattactgg caccatctat tgaagcagac tttgaatgtc ttctagaagt ccagtccect     1260
tttgctgaag atgttcggtc gtaccgttcc cctcccttgg acaagggtgt cactgtctct     1320
ggaaagacag tcaaagagca cgcacatctc ccaagtgcg aattgctgaa tgcgatgagc     1380
aaatacgtcg acagcatgga gctcgtcgac aaggatgaaa acggagaacc agttgacagc     1440
ctggetccca gactggagga ttcgtaactct ccactgctgc acaggatcga gcaagctatc     1500
cgggtggcgtg ccatccatcc aaacgagcct ctccgcctcc cttctgagaa gttgacgcag     1560
ctgtcacgac cgccagcaga tctgcaagcg cgcgcgaaga aatacctgga tcgggtcatt     1620
gccgccgccg atgtgaagaa agttccacca aaagcaaaag gtcgcaagcg gaatcgcgaa     1680
gccgacaaac ccctatcggt tcttgacgtt gacgagetcc ttgctcgca gaagcgcgcc     1740
aagatctcag ccaacaacgc catccccgag ttcaaacagt cgttggtcaa cgccgagacc     1800
atcgacgccg tccgtgacgc agtcagccag atggaaagca tcatcgagaa ccacatccga     1860
agcagctttg gagaegccaa ctacgaccgc gtgatcgagg agctgggtgt cctccgcgag     1920
gagctgatcg cctacgaaga gccgatctc tacaacgact tcctgcggag gctgaaggac     1980
aagatcctca atgaggagct gggcggagac agacgagagc tgtggtggct cgtcaggagg     2040
caacgggtcg gtctgataga caagaaggcg tcggaacggg ttgaagttac tgaacaggaa     2100
gccagggagt ccattattgc tatctgtctc cattctcaca gatag                       2145

```

<210> 22

<211> 714

<212> PRT

<213> Rasamsonia emersonii

<400> 22

```

Met Ala Asp Lys Glu Ala Thr Val Tyr Val Ile Asp Val Gly Lys Ser
1           5           10           15

```

[0032]

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

[0033]

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

[0034]

Ser Met Glu Leu Val Asp Lys Asp Glu Asn Gly Glu Pro Val Asp Ser
 465 470 475 480

Leu Ala Pro Arg Leu Glu Asp Ser Tyr Ser Pro Leu Leu His Arg Ile
 485 490 495

Glu Gln Ala Ile Arg Trp Arg Ala Ile His Pro Asn Glu Pro Leu Pro
 500 505 510

Pro Pro Ser Glu Lys Leu Thr Gln Leu Ser Arg Pro Pro Ala Asp Leu
 515 520 525

Gln Ala Arg Ala Lys Lys Tyr Leu Asp Arg Val Ile Ala Ala Ala Asp
 530 535 540

Val Lys Lys Val Pro Pro Lys Ala Lys Gly Arg Lys Arg Asn Arg Glu
 545 550 555 560

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

Glu Lys Arg Ala Lys Ile Ser Ala Asn Asn Ala Ile Pro Glu Phe Lys
 580 585 590

Gln Ser Leu Val Asn Ala Glu Thr Ile Asp Ala Val Arg Asp Ala Val
 595 600 605

Ser Gln Met Glu Ser Ile Ile Glu Asn His Ile Arg Ser Ser Phe Gly
 610 615 620

Asp Ala Asn Tyr Asp Arg Val Ile Glu Glu Leu Gly Val Leu Arg Glu
 625 630 635 640

Glu Leu Ile Ala Tyr Glu Glu Pro Asp Leu Tyr Asn Asp Phe Leu Arg
 645 650 655

Arg Leu Lys Asp Lys Ile Leu Asn Glu Glu Leu Gly Gly Asp Arg Arg
 660 665 670

Glu Leu Trp Trp Leu Val Arg Arg Gln Arg Val Gly Leu Ile Asp Lys
 675 680 685

Lys Ala Ser Glu Arg Val Glu Val Thr Glu Gln Glu Ala Arg Glu Ser

[0035]

690	695	700	
Ile Ile Ala Ile Cys Leu His Ser His Arg			
705	710		
<210> 23			
<211> 1854			
<212> DNA			
<213> <i>Rasamsonia emersonii</i>			
<400> 23			
atggcgtecc tegtgtctgg cgtctctctgt gtctctgggccc tgacgcctgc tgcatttgcg		60	
cgagcgcccc ttgcagcccc agccagcgggt tccctggact cttttctcgc aaccgagacg		120	
cccattgccc tccaaggcgt gctgaacaac atcgggcccga atggcgcgga tgtggcaggg		180	
gcaagcgccc ggatcgtggg cgcagtcgc agcaggagcg acccgatta tttctactcc		240	
tggacacgcg acgcagctct caccggccaaa tacctcgtcg acgccttcac cgcgggcaac		300	
aaggacctgg agcagaccat ccaggagtac atcagcgcg aggcgcaagt gcaaaccatc		360	
tccaatccgt cgggggatit atccaccggg ggcctaggag agcccaagtt caatgtgaac		420	
gagacggctt ttaccgggcc ctggggctgt ccacagaggg acggaccagc gttgagagcg		480	
acggccctga ttgcatatgc gaactatctc atcgacaacg gccaggettc gaetgccgat		540	
gagatcatct ggccgatitgt ccagaatgat ctgtcctacg tcaccagta ctggaactca		600	
tccacctttg acctctggga agaagtagaa ggatcctcat tcttcacaac cgcctgtcag		660	
caccgcgctt tggtegaagg caatgcactg gcaacaaggc tgaaccacac gtgccccaac		720	
tgcgtctctc aggcgccaca ggtcctgtgt tttctgcagt catactggac cggatcatat		780	
gttctggcca actttgggtg cagcggctgc tccggcaagg acgtgaattc gattctgggc		840	
agcatccaca ccttcgatcc cgcgggaggg tgtgacgaet cgaccttcca gccgtgttcg		900	
gcccgtgcct tggcaaatca caaggtgggc acagactcgt tccggagcgt ctatgcggtc		960	
aactcgggca tgcagaggg atctgcctg gcagtcggcc gctaccccga ggatgtttac		1020	
caggcgggga acccctggta cctggccaca gcagcggccg cggagcagct ttatgacgcc		1080	
atctaccagt ggaataagat cggctcgatc agtateaccg acgttagttt ggcccttttc		1140	
caggatatct acccttctgc cgcggtgggc acctataact ctggctcctc gactttcaac		1200	
gacatcatct cggccgtcca gacgtatgct gatggatata tgagcattat cgagaaatat		1260	
actccctcag acggtctctc caccgaacaa ttctcccggtt cggatggcac tccgctctct		1320	

[0036]

```

gcttctggcc tgacttggtc gtaagcctcc ctctttaccg ctgttgcccg cagacagtcc 1380
atcgteccctg cgteectgggg cgaaagctcc gcgagcagcg tccctgcegt ctgttctgcc 1440
acctctgcca cgggaccata cageacgget accaataccg cetggccaag ttcttggtct 1500
ggccccctcaa ccaccaccag cgttccatgt accactecca cctctgtggc tgtaaccttt 1560
gacgaaatcg tcagcaccac ttacggggag acaatctatc tggecggtc gatccceag 1620
ctggggcaact ggccccgtc cagcgcgate cccctccgcg cggacgcgta caccagcagc 1680
aaccgctttt ggtacgtgac cctcaacctg cccgctggca ccagcttcga gtacaagttc 1740
ttcaagaaag agacggacgg gaccatcgtc tgggaggacg acccgaaccg gtcgtacacg 1800
gtcccagcgt actgtgggca gactactgcc attcttgacg atagttggca gtga 1854

```

<210> 24

<211> 617

<212> PRT

<213> *Rasamsonia emersonii*

<400> 24

```

Met Ala Ser Leu Val Ala Gly Ala Leu Cys Val Leu Gly Leu Thr Pro
1           5           10           15

```

```

Ala Ala Phe Ala Arg Ala Pro Val Ala Ala Arg Ala Ser Gly Ser Leu
          20           25           30

```

```

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

```

```

Asn Asn Ile Gly Pro Asn Gly Ala Asp Val Ala Gly Ala Ser Ala Gly
          50           55           60

```

```

Ile Val Val Ala Ser Pro Ser Arg Ser Asp Pro Asp Tyr Phe Tyr Ser
65           70           75           80

```

```

Trp Thr Arg Asp Ala Ala Leu Thr Ala Lys Tyr Leu Val Asp Ala Phe
          85           90           95

```

```

Ile Ala Gly Asn Lys Asp Leu Glu Gln Thr Ile Gln Glu Tyr Ile Ser
          100          105          110

```

```

Ala Gln Ala Gln Val Gln Thr Ile Ser Asn Pro Ser Gly Asp Leu Ser
          115          120          125

```

[0037]

Thr	Gly	Gly	Leu	Gly	Glu	Pro	Lys	Phe	Asn	Val	Asn	Glu	Thr	Ala	Phe	130	135	140
Thr	Gly	Pro	Trp	Gly	Arg	Pro	Gln	Arg	Asp	Gly	Pro	Ala	Leu	Arg	Ala	145	150	155
Thr	Ala	Leu	Ile	Ala	Tyr	Ala	Asn	Tyr	Leu	Ile	Asp	Asn	Gly	Gln	Ala	165	170	175
Ser	Thr	Ala	Asp	Glu	Ile	Ile	Trp	Pro	Ile	Val	Gln	Asn	Asp	Leu	Ser	180	185	190
Tyr	Val	Thr	Gln	Tyr	Trp	Asn	Ser	Ser	Thr	Phe	Asp	Leu	Trp	Glu	Glu	195	200	205
Val	Glu	Gly	Ser	Ser	Phe	Phe	Thr	Thr	Ala	Val	Gln	His	Arg	Ala	Leu	210	215	220
Val	Glu	Gly	Asn	Ala	Leu	Ala	Thr	Arg	Leu	Asn	His	Thr	Cys	Pro	Asn	225	230	235
Cys	Val	Ser	Gln	Ala	Pro	Gln	Val	Leu	Cys	Phe	Leu	Gln	Ser	Tyr	Trp	245	250	255
Thr	Gly	Ser	Tyr	Val	Leu	Ala	Asn	Phe	Gly	Gly	Ser	Gly	Arg	Ser	Gly	260	265	270
Lys	Asp	Val	Asn	Ser	Ile	Leu	Gly	Ser	Ile	His	Thr	Phe	Asp	Pro	Ala	275	280	285
Gly	Gly	Cys	Asp	Asp	Ser	Thr	Phe	Gln	Pro	Cys	Ser	Ala	Arg	Ala	Leu	290	295	300
Ala	Asn	His	Lys	Val	Val	Thr	Asp	Ser	Phe	Arg	Ser	Val	Tyr	Ala	Val	305	310	315
Asn	Ser	Gly	Ile	Ala	Glu	Gly	Ser	Ala	Val	Ala	Val	Gly	Arg	Tyr	Pro	325	330	335
Glu	Asp	Val	Tyr	Gln	Gly	Gly	Asn	Pro	Trp	Tyr	Leu	Ala	Thr	Ala	Ala	340	345	350

[0038]

Ala Ala Glu Gln Leu Tyr Asp Ala Ile Tyr Gln Trp Asn Lys Ile Gly
 355 360 365

Ser Ile Ser Ile Thr Asp Val Ser Leu Ala Phe Phe Gln Asp Ile Tyr
 370 375 380

Pro Ser Ala Ala Val Gly Thr Tyr Asn Ser Gly Ser Ser Thr Phe Asn
 385 390 395 400

Asp Ile Ile Ser Ala Val Gln Thr Tyr Ala Asp Gly Tyr Leu Ser Ile
 405 410 415

Ile Glu Lys Tyr Thr Pro Ser Asp Gly Ser Leu Thr Glu Gln Phe Ser
 420 425 430

Arg Ser Asp Gly Thr Pro Leu Ser Ala Ser Gly Leu Thr Trp Ser Tyr
 435 440 445

Ala Ser Leu Leu Thr Ala Ala Ala Arg Arg Gln Ser Ile Val Pro Ala
 450 455 460

Ser Trp Gly Glu Ser Ser Ala Ser Ser Val Pro Ala Val Cys Ser Ala
 465 470 475 480

Thr Ser Ala Thr Gly Pro Tyr Ser Thr Ala Thr Asn Thr Ala Trp Pro
 485 490 495

Ser Ser Gly Ser Gly Pro Ser Thr Thr Thr Ser Val Pro Cys Thr Thr
 500 505 510

Pro Thr Ser Val Ala Val Thr Phe Asp Glu Ile Val Ser Thr Thr Tyr
 515 520 525

Gly Glu Thr Ile Tyr Leu Ala Gly Ser Ile Pro Glu Leu Gly Asn Trp
 530 535 540

Ser Pro Ser Ser Ala Ile Pro Leu Arg Ala Asp Ala Tyr Thr Ser Ser
 545 550 555 560

Asn Pro Leu Trp Tyr Val Thr Leu Asn Leu Pro Ala Gly Thr Ser Phe
 565 570 575

Glu Tyr Lys Phe Phe Lys Lys Glu Thr Asp Gly Thr Ile Val Trp Glu

[0039]

580	585	590
Asp Asp Pro Asn Arg Ser Tyr Thr Val Pro Ala Tyr Cys Gly Gln Thr		
595	600	605
Thr Ala Ile Leu Asp Asp Ser Trp Gln		
610	615	

pENTRY-P6bleTtrpC-Pxeba7flagTgla

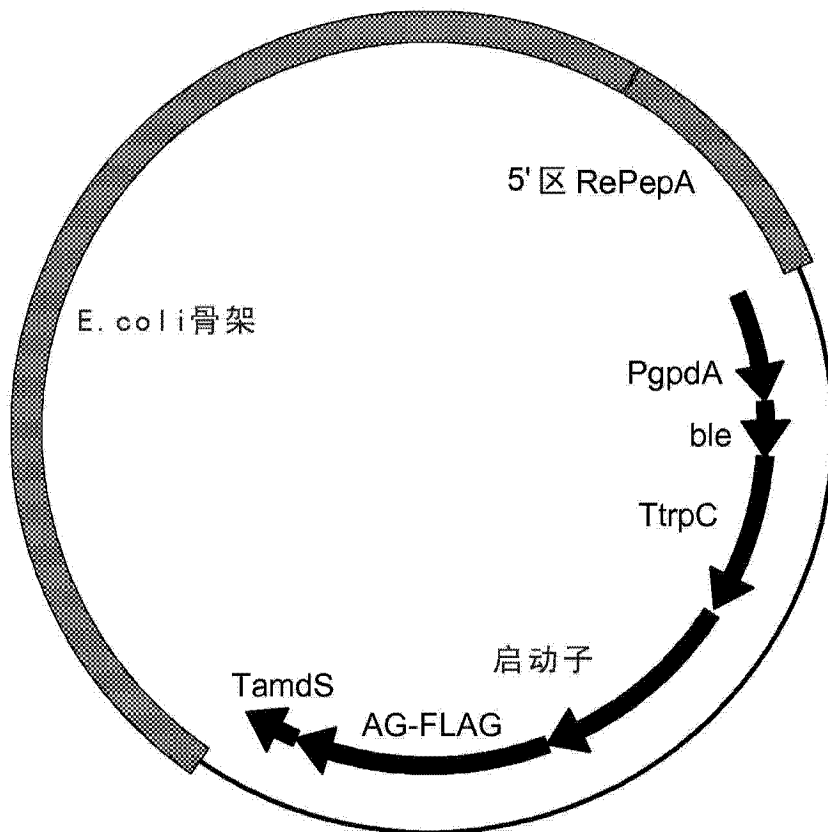


图 1

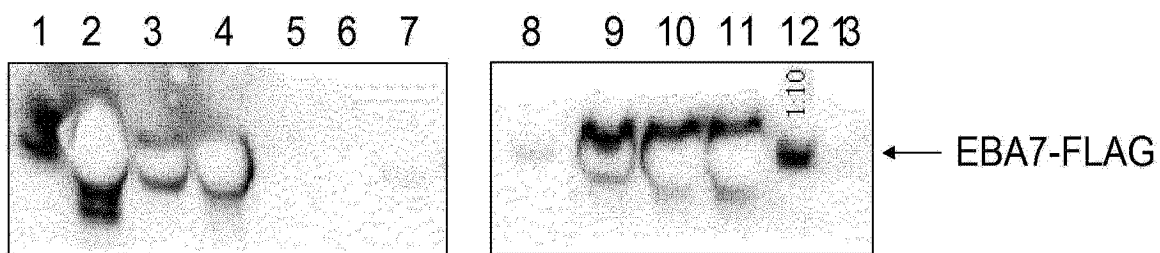


图 2

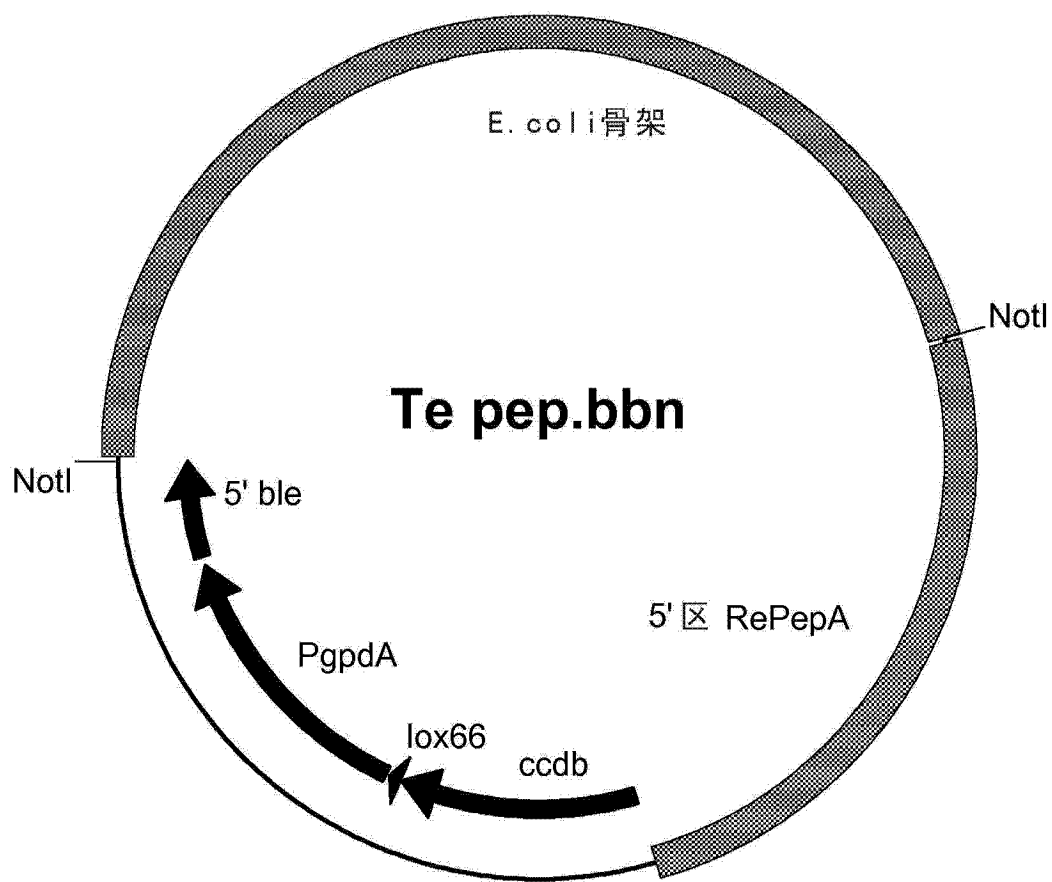


图 3

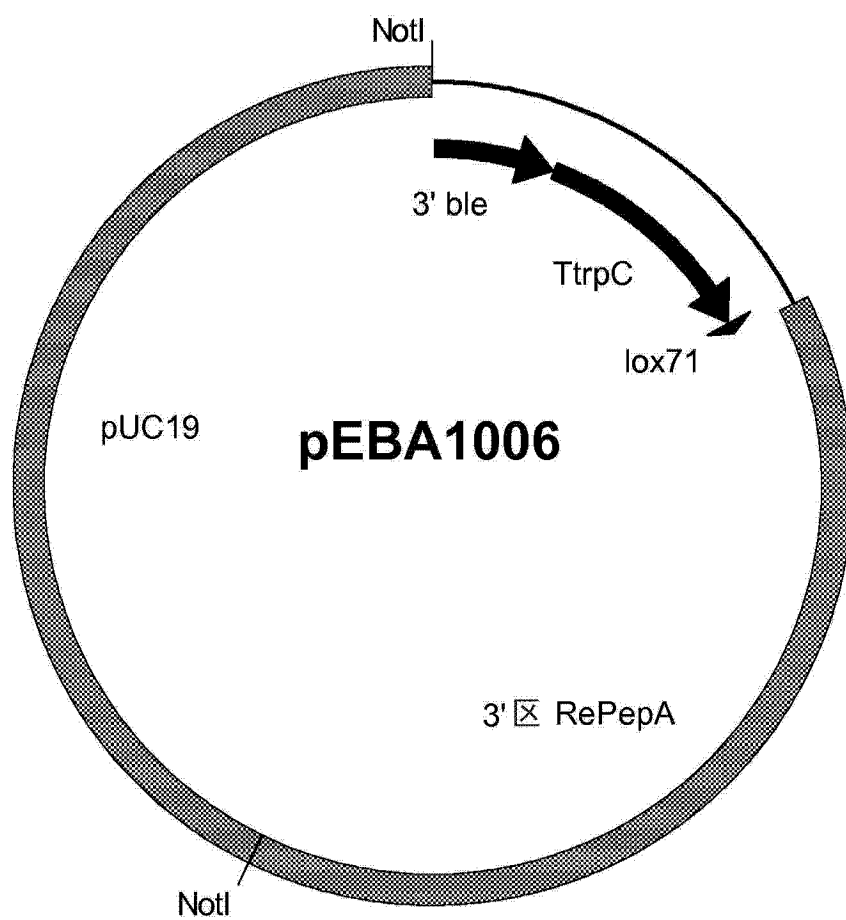


图 4

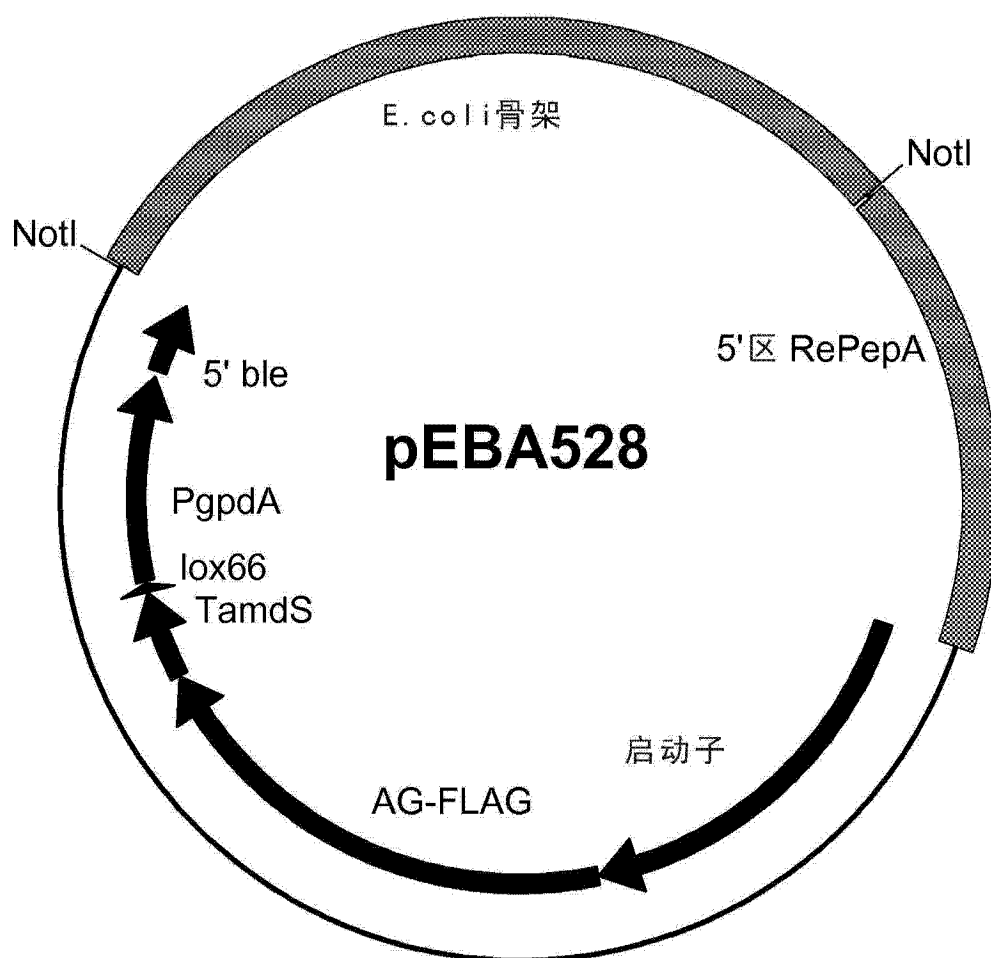


图 5

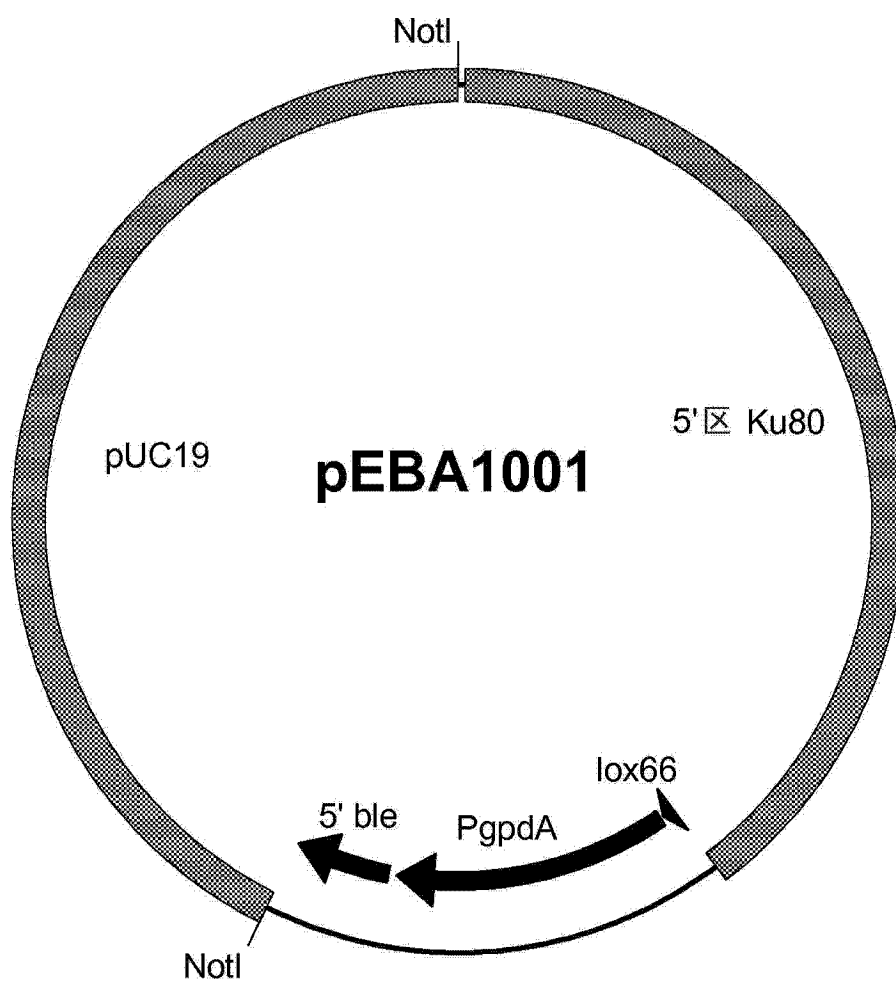


图 6

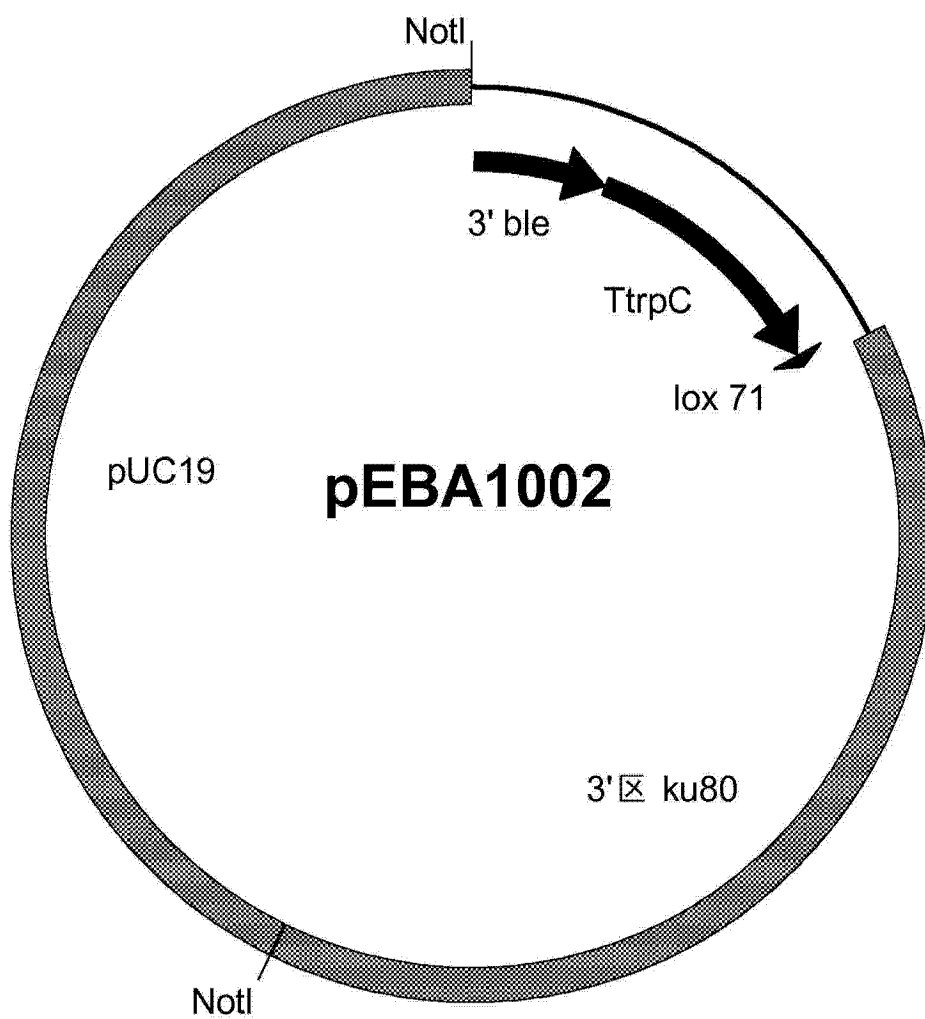


图 7

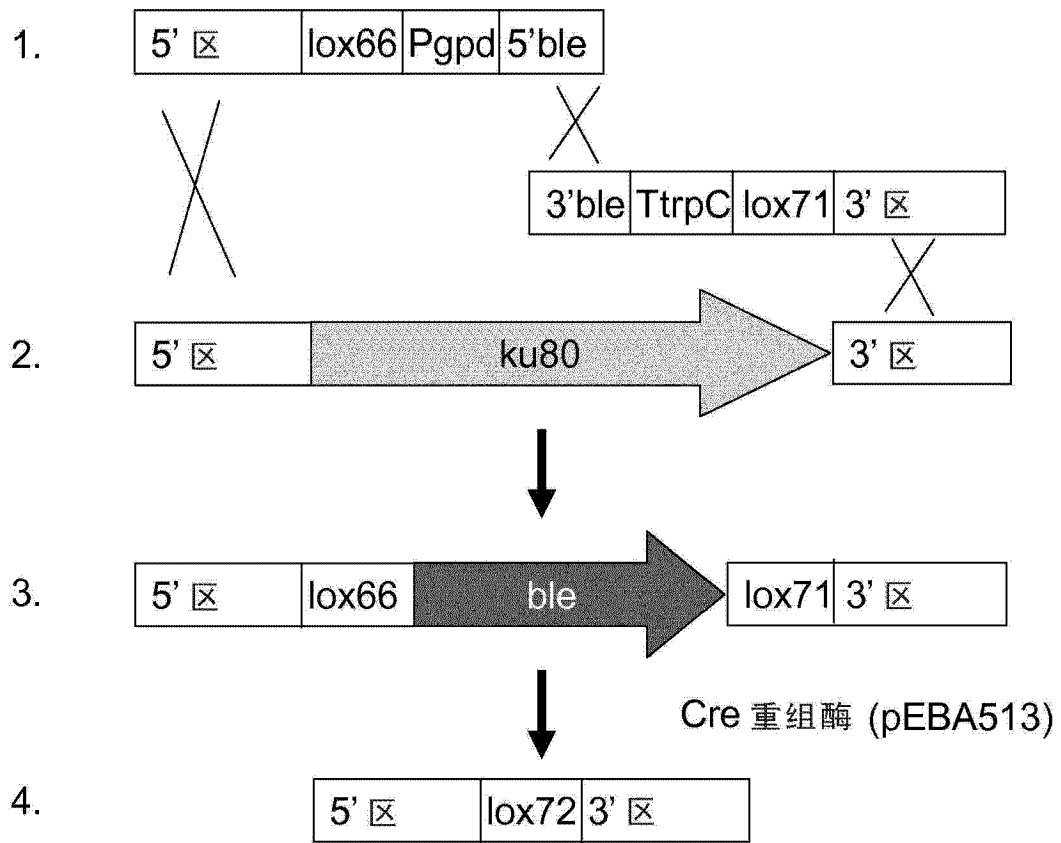


图 8

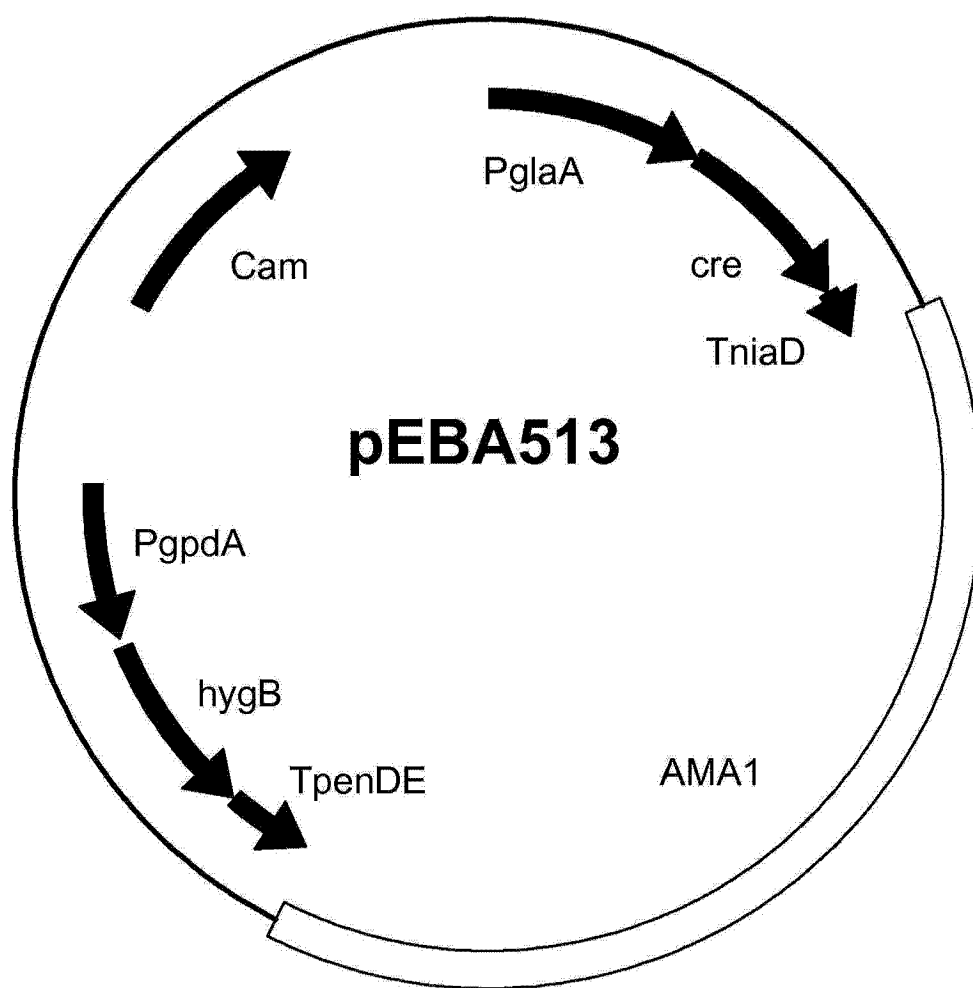


图 9

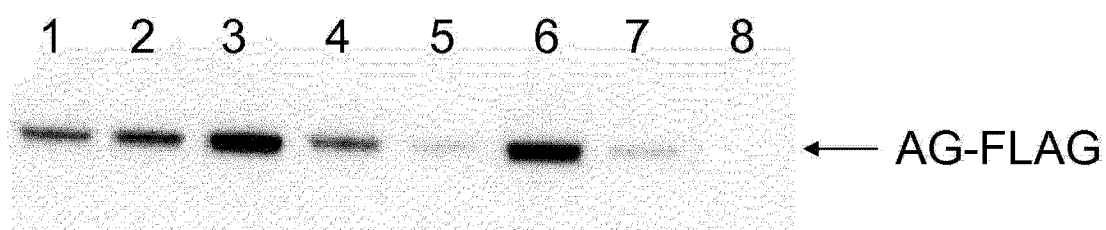


图 10