



(12)发明专利申请

(10)申请公布号 CN 107075452 A

(43)申请公布日 2017.08.18

(21)申请号 201580052220.1

(22)申请日 2015.07.24

(30)优先权数据

62/028,946 2014.07.25 US

(85)PCT国际申请进入国家阶段日

2017.03.27

(86)PCT国际申请的申请数据

PCT/US2015/041910 2015.07.24

(87)PCT国际申请的公布数据

W02016/014900 EN 2016.01.28

(71)申请人 诺沃吉股份有限公司

地址 美国马萨诸塞州

(72)发明人 A·卡米内尼 E·E·布雷夫诺瓦

(74)专利代理机构 上海专利商标事务所有限公司 31100

代理人 陶启长 杨昀

(51)Int.Cl.

C12N 1/19(2006.01)

C12N 15/81(2006.01)

C12N 15/90(2006.01)

权利要求书4页 说明书33页
序列表30页 附图8页

(54)发明名称

来源于解脂耶氏酵母和解腺嘌呤阿氏酵母的启动子及其使用方法

(57)摘要

公开了来自解腺嘌呤阿氏酵母(*Arxula adeninivorans*)和解脂耶氏酵母(*Yarrowia lipolytica*)的启动子的核苷酸序列,其可以用于驱动细胞中的基因表达。验证启动子、筛选选择的启动子以确定哪些启动子可用于增加含油酵母的脂质生产效率。

1. 一种编码来自解腺嘌呤阿氏酵母 (*Arxula adeninivorans*) 的启动子的核酸, 其中所述启动子是用于下述的启动子: 翻译延伸因子 EF-1 α ; 甘油-3-磷酸脱氢酶; 磷酸丙糖异构酶 1; 果糖-1,6-二磷酸醛缩酶; 磷酸甘油酸变位酶; 丙酮酸激酶; 输出蛋白 EXP1; 核糖体蛋白 S7; 醇脱氢酶; 磷酸甘油酸激酶; 己糖转运蛋白; 一般氨基酸通透酶; 丝氨酸蛋白酶; 异柠檬酸裂解酶; 酰基辅酶 A 氧化酶; ATP-硫酸化酶; 己糖激酶; 3-磷酸甘油酸脱氢酶; 丙酮酸脱氢酶 α 亚基; 丙酮酸脱氢酶 β 亚基; 乌头酸; 烯醇酶; 肌动蛋白; 多药耐药蛋白 (ABC-转运蛋白); 泛素; GTP 酶; 质膜 Na⁺/Pi 共转运蛋白; 丙酮酸脱羧酶; 植酸酶; 或 α -淀粉酶。

2. 如权利要求 1 所述的核酸, 其中所述启动子源自编码下述的基因: TEF1; GPD1; TPI1; FBA1; GPM1; PYK1; EXP1; RPS7; ADH1; PGK1; HXT7; GAP1; XPR2; ICL1; POX; MET3; HXK1; SER3; PDA1; PDB1; ACO1; ENO1; ACT1; MDR1; UBI4; YPT1; PHO89; PDC1; PHY; 或 AMYA。

3. 如权利要求 1 或 2 所述的核酸, 其中:

所述核酸与 SEQ ID NO:5; SEQ ID NO:6; SEQ ID NO:7; SEQ ID NO:8; SEQ ID NO:9; SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; SEQ ID NO:15; SEQ ID NO:35; SEQ ID NO:36; SEQ ID NO:37; SEQ ID NO:38; SEQ ID NO:39; SEQ ID NO:40; SEQ ID NO:41; SEQ ID NO:42; SEQ ID NO:43; SEQ ID NO:44; SEQ ID NO:45; SEQ ID NO:46; SEQ ID NO:47; SEQ ID NO:48; SEQ ID NO:49; SEQ ID NO:50; SEQ ID NO:51; SEQ ID NO:52; 或 SEQ ID NO:53 所示的核苷酸序列具有至少 90% 的序列同源性; 或

所述核酸与 SEQ ID NO:5; SEQ ID NO:6; SEQ ID NO:7; SEQ ID NO:8; SEQ ID NO:9; SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; SEQ ID NO:15; SEQ ID NO:35; SEQ ID NO:36; SEQ ID NO:37; SEQ ID NO:38; SEQ ID NO:39; SEQ ID NO:40; SEQ ID NO:41; SEQ ID NO:42; SEQ ID NO:43; SEQ ID NO:44; SEQ ID NO:45; SEQ ID NO:46; SEQ ID NO:47; SEQ ID NO:48; SEQ ID NO:49; SEQ ID NO:50; SEQ ID NO:51; SEQ ID NO:52; 或 SEQ ID NO:53 的子序列具有至少 90% 序列同源性, 并且所述子序列保留启动子活性。

4. 如权利要求 3 所述的核酸, 其中所述核酸包含 SEQ ID NO:5; SEQ ID NO:6; SEQ ID NO:7; SEQ ID NO:8; SEQ ID NO:9; SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; SEQ ID NO:15; SEQ ID NO:35; SEQ ID NO:36; SEQ ID NO:37; SEQ ID NO:38; SEQ ID NO:39; SEQ ID NO:40; SEQ ID NO:41; SEQ ID NO:42; SEQ ID NO:43; SEQ ID NO:44; SEQ ID NO:45; SEQ ID NO:46; SEQ ID NO:47; SEQ ID NO:48; SEQ ID NO:49; SEQ ID NO:50; SEQ ID NO:51; SEQ ID NO:52; 或 SEQ ID NO:53 的子序列, 并且所述子序列保留启动子活性。

5. 如权利要求 3 所述的核酸, 其中所述核酸包含 SEQ ID NO:5; SEQ ID NO:6; SEQ ID NO:7; SEQ ID NO:8; SEQ ID NO:9; SEQ ID NO:10; SEQ ID NO:11; SEQ ID NO:12; SEQ ID NO:13; SEQ ID NO:14; SEQ ID NO:15; SEQ ID NO:35; SEQ ID NO:36; SEQ ID NO:37; SEQ ID NO:38; SEQ ID NO:39; SEQ ID NO:40; SEQ ID NO:41; SEQ ID NO:42; SEQ ID NO:43; SEQ ID NO:44; SEQ ID NO:45; SEQ ID NO:46; SEQ ID NO:47; SEQ ID NO:48; SEQ ID NO:49; SEQ ID NO:50; SEQ ID NO:51; SEQ ID NO:52; 或 SEQ ID NO:53 所示的核苷酸序列。

6. 如权利要求 1-5 中任一项所述的核酸, 其还包含基因, 其中所述启动子和基因可操作地连接。

7. 一种载体, 其包含权利要求 1-6 中任一项所述的核酸。

8. 如权利要求7所述的载体,其中所述载体是质粒。

9. 一种转化的细胞,其包含权利要求1-6中任一项所述的核酸。

10. 一种包含遗传修饰的转化的细胞,其中所述遗传修饰是用编码启动子的核酸转化,其中所述启动子与SEQ ID NO:5;SEQ ID NO:6;SEQ ID NO:7;SEQ ID NO:8;SEQ ID NO:9;SEQ ID NO:10;SEQ ID NO:11;SEQ ID NO:12;SEQ ID NO:13;SEQ ID NO:14;SEQ ID NO:15;SEQ ID NO:16;SEQ ID NO:17;SEQ ID NO:18;SEQ ID NO:19;SEQ ID NO:20;SEQ ID NO:21;SEQ ID NO:22;SEQ ID NO:23;SEQ ID NO:24;SEQ ID NO:25;SEQ ID NO:26;SEQ ID NO:27;SEQ ID NO:28;SEQ ID NO:29;SEQ ID NO:30;SEQ ID NO:31;SEQ ID NO:32;SEQ ID NO:33;SEQ ID NO:34;SEQ ID NO:35;SEQ ID NO:36;SEQ ID NO:37;SEQ ID NO:38;SEQ ID NO:39;SEQ ID NO:40;SEQ ID NO:41;SEQ ID NO:42;SEQ ID NO:43;SEQ ID NO:44;SEQ ID NO:45;SEQ ID NO:46;SEQ ID NO:47;SEQ ID NO:48;SEQ ID NO:49;SEQ ID NO:50;SEQ ID NO:51;SEQ ID NO:52;或SEQ ID NO:53的子序列具有至少90%的序列同源性,并且所述子序列保留启动子活性。

11. 如权利要求9或10所述的转化的细胞,其中所述细胞选自藻类、细菌、霉菌、真菌、植物和酵母。

12. 如权利要求11所述的转化的细胞,其中所述细胞是酵母。

13. 如权利要求12所述的转化的细胞,其中所述细胞选自:阿氏菌属(*Arxula*)、曲霉属(*Aspegillus*)、奥兰氏菌属(*Aurantiochytrium*)、念珠菌属(*Candida*)、麦角菌属(*Claviceps*)、隐球菌属(*Cryptococcus*)、小坎宁安霉属(*Cunninghamella*)、地霉属(*Geotrichum*)、汉逊酵母属(*Hansenula*)、克鲁维酵母属(*Kluyveromyces*)、柯达酵母(*Kodamaea*)、白冬孢酵母属(*Leucosporidiella*)、油脂酵母属(*Lipomyces*)、被孢霉属(*Mortierella*)、欧格氏霉属(*Ogataea*)、毕赤酵母属(*Pichia*)、原囊藻属(*Prototheca*)、根霉属(*Rhizopus*)、红冬孢酵母属(*Rhodospiridium*)、红酵母属(*Rhodotorula*)、酵母菌属(*Saccharomyces*)、裂殖酵母属(*Schizosaccharomyces*)、银耳属(*Tremella*)、丝孢酵母属(*Trichosporon*)、威克汉姆酵母属(*Wickerhamomyces*)和耶氏酵母属(*Yarrowia*)。

14. 如权利要求13所述的转化的细胞,其中所述细胞选自:黑曲霉(*Aspergillus niger*)、米曲霉(*Aspergillus oryzae*)、土曲霉(*Aspergillus terreus*)、裂殖壶奥兰氏菌(*Aurantiochytrium limacinum*)、产朊假丝酵母(*Candida utilis*)、麦角菌(*Claviceps purpurea*)、浅白隐球菌(*Cryptococcus albidus*)、弯曲隐球菌(*Cryptococcus curvatus*)、拉米隐球菌(*Cryptococcus ramirezgomezianus*)、地生隐球菌(*Cryptococcus terreus*)、威氏隐球菌(*Cryptococcus wieringae*)、刺孢小坎宁安霉(*Cunninghamella echinulata*)、贴梗海棠小坎宁安霉(*Cunninghamella japonica*)、发酵性地霉(*Geotrichum fermentans*)、多形汉森酵母(*Hansenula polymorpha*)、乳酸克鲁维酵母(*Kluyveromyces lactis*)、马克斯克鲁维酵母(*Kluyveromyces marxianus*)、奥默柯达酵母(*Kodamaea ohmeri*)、克雷氏白冬孢酵母(*Leucosporidiella creatinivora*)、产油油脂酵母(*Lipomyces lipofer*)、斯达氏油脂酵母(*Lipomyces starkeyi*)、子囊菌油脂酵母(*Lipomyces tetrasporus*)、深黄被孢霉(*Mortierella isabellina*)、高山被孢霉(*Mortierella alpina*)、马兜铃欧格氏霉(*Ogataea polymorpha*)、塞佛式毕赤酵母(*Pichia ciferrii*)、季氏毕赤酵母(*Pichia guilliermondii*)、巴斯德毕赤酵母(*Pichia*

pastoris)、叶柄毕赤酵母(*Pichia stipites*)、小型原藻(*Prototheca zopfii*)、隐根根霉(*Rhizopus arrhizus*)、巴氏红冬孢酵母(*Rhodospiridium babjevae*)、圆红冬孢酵母(*Rhodospiridium toruloides*)、帕鲁氏红冬孢酵母(*Rhodospiridium paludigenum*)、粘红酵母(*Rhodotorula glutinis*)、胶红酵母(*Rhodotorula mucilaginosa*)、酿酒酵母(*Saccharomyces cerevisiae*)、粟酒裂殖酵母(*Schizosaccharomyces pombe*)、脑状银耳(*Tremella encephala*)、皮状丝孢酵母(*Trichosporon cutaneum*)、发酵性丝孢酵母(*Trichosporon fermentans*)、和塞佛式威克汉姆酵母(*Wickerhamomyces ciferrii*)。

15. 如权利要求13所述的转化的细胞,其中所述细胞是解脂耶氏酵母。

16. 如权利要求13所述的转化的细胞,其中所述细胞是解腺嘌呤阿氏酵母。

17. 一种在细胞中表达基因的方法,包括用编码启动子的核酸转化亲本细胞,其中:

所述启动子与SEQ ID NO:5;SEQ ID NO:6;SEQ ID NO:7;SEQ ID NO:8;SEQ ID NO:9;SEQ ID NO:10;SEQ ID NO:11;SEQ ID NO:12;SEQ ID NO:13;SEQ ID NO:14;SEQ ID NO:15;SEQ ID NO:16;SEQ ID NO:17;SEQ ID NO:18;SEQ ID NO:19;SEQ ID NO:20;SEQ ID NO:21;SEQ ID NO:22;SEQ ID NO:23;SEQ ID NO:24;SEQ ID NO:25;SEQ ID NO:26;SEQ ID NO:27;SEQ ID NO:28;SEQ ID NO:29;SEQ ID NO:30;SEQ ID NO:31;SEQ ID NO:32;SEQ ID NO:33;SEQ ID NO:34;SEQ ID NO:35;SEQ ID NO:36;SEQ ID NO:37;SEQ ID NO:38;SEQ ID NO:39;SEQ ID NO:40;SEQ ID NO:41;SEQ ID NO:42;SEQ ID NO:43;SEQ ID NO:44;SEQ ID NO:45;SEQ ID NO:46;SEQ ID NO:47;SEQ ID NO:48;SEQ ID NO:49;SEQ ID NO:50;SEQ ID NO:51;SEQ ID NO:52;或SEQ ID NO:53的子序列具有至少90%序列同源性;

所述子序列保留启动子活性;并且或:

所述核酸包含所述基因,并且所述基因和所述启动子可操作地连接;或设计所述核酸使得所述启动子在转化亲本细胞后变得可操作地连接到所述基因。

18. 一种在细胞中表达基因的方法,包括用权利要求1-5中任一项所述的核酸转化亲本细胞;

其中:

所述核酸包含所述基因,并且所述基因和所述启动子可操作地连接;或

设计所述核酸使得所述启动子在转化亲本细胞后变得可操作地连接到所述基因。

19. 如权利要求17或18所述的方法,其中所述核酸包含所述基因,并且所述基因和所述启动子可操作地连接。

20. 如权利要求17或18所述的方法,其中设计所述核酸使得所述启动子在转化所述亲本细胞后变得可操作地连接到所述基因。

21. 如权利要求17-20中任一项所述的方法,其中所述细胞是酵母。

22. 如权利要求21所述的方法,其中所述细胞选自:阿氏菌属(*Arxula*)、曲霉属(*Aspegillus*)、奥兰氏菌属(*Aurantiochytrium*)、念珠菌属(*Candida*)、麦角菌属(*Claviceps*)、隐球菌属(*Cryptococcus*)、小坎宁安霉属(*Cunninghamella*)、地霉属(*Geotrichum*)、汉逊酵母属(*Hansenula*)、克鲁维酵母属(*Kluyveromyces*)、柯达酵母(*Kodamaea*)、白冬孢酵母属(*Leucosporidiella*)、油脂酵母属(*Lipomyces*)、被孢霉属(*Mortierella*)、欧格氏霉属(*Ogataea*)、毕赤酵母属(*Pichia*)、原囊藻属(*Prototheca*)、根霉属(*Rhizopus*)、红冬孢酵母属(*Rhodospiridium*)、红酵母属(*Rhodotorula*)、酵母菌属

(*Saccharomyces*)、裂殖酵母属(*Schizosaccharomyces*)、银耳属(*Tremella*)、丝孢酵母属(*Trichosporon*)、威克汉姆酵母属(*Wickerhamomyces*)和耶氏酵母属(*Yarrowia*)。

23. 如权利要求22所述的方法,其中所述细胞选自:黑曲霉(*Aspergillus niger*)、米曲霉(*Aspergillus oryzae*)、土曲霉(*Aspergillus terreus*)、裂殖壶奥兰氏菌(*Aurantiochytrium limacinum*)、产朊假丝酵母(*Candida utilis*)、麦角菌(*Claviceps purpurea*)、浅白隐球酵母(*Cryptococcus albidus*)、弯曲隐球菌(*Cryptococcus curvatus*)、拉米隐球菌(*Cryptococcus ramirezgomezianus*)、地生隐球菌(*Cryptococcus terreus*)、威氏隐球菌(*Cryptococcus wieringae*)、刺孢小坎宁安霉(*Cunninghamella echinulata*)、贴梗海棠小坎宁安霉(*Cunninghamella japonica*)、发酵性地霉(*Geotrichum fermentans*)、多形汉森酵母(*Hansenula polymorpha*)、乳酸克鲁维酵母(*Kluyveromyces lactis*)、马克斯克鲁维酵母(*Kluyveromyces marxianus*)、奥默柯达酵母(*Kodamaea ohmeri*)、克雷氏白冬孢酵母(*Leucosporidiella creatinivora*)、产油油脂酵母(*Lipomyces lipofer*)、斯达氏油脂酵母(*Lipomyces starkeyi*)、子囊菌油脂酵母(*Lipomyces tetrasporus*)、深黄被孢霉(*Mortierella isabellina*)、高山被孢霉(*Mortierella alpina*)、马兜铃欧格氏霉(*Ogataea polymorpha*)、塞佛式毕赤酵母(*Pichia ciferrii*)、季氏毕赤氏酵母(*Pichia guilliermondii*)、巴斯德毕赤酵母(*Pichia pastoris*)、叶柄毕赤酵母(*Pichia stipites*)、小型原藻(*Prototheca zopfii*)、隐根根霉(*Rhizopus arrhizus*)、巴氏红冬孢酵母(*Rhodosporeidium babjevae*)、圆红冬孢酵母(*Rhodosporeidium toruloides*)、帕鲁氏红冬孢酵母(*Rhodosporeidium paludigenum*)、粘红酵母(*Rhodotorula glutinis*)、胶红酵母(*Rhodotorula mucilaginosa*)、酿酒酵母(*Saccharomyces cerevisiae*)、粟酒裂殖酵母(*Schizosaccharomyces pombe*)、脑状银耳(*Tremella encephala*)、皮状丝孢酵母(*Trichosporon cutaneum*)、发酵性丝孢酵母(*Trichosporon fermentans*)、和塞佛式威克汉姆酵母(*Wickerhamomyces ciferrii*)。

24. 如权利要求22所述的方法,其中所述细胞是解脂耶氏酵母。

25. 如权利要求22所述的方法,其中所述细胞是解腺嘌呤阿氏酵母。

来源于解脂耶氏酵母和解腺嘌呤阿氏酵母的启动子及其使用方法

[0001] 相关申请

[0002] 本申请要求于2014年7月25日提交的美国临时专利申请号62/028,946的优先权的权益,其全部内容通过引用并入本文。

[0003] 序列表

[0004] 本申请包含以ASCII格式电子提交的序列表,并通过引用全文纳入本文。所述ASCII副本,于2015年7月16日创建,名为NGX_03425_SL.txt,大小为71,975字节。

背景技术

[0005] 含脂肪酵母例如解脂耶氏酵母 (*Yarrowia lipolytica*) 和解腺嘌呤阿氏酵母 (*Arxula adenivorans*) 可以被工程改造用于脂质的工业生产,脂质是食品和化妆品工业中不可或缺的成分,并且是生物柴油和生物化学工业中的重要前体。可以通过上调或下调调节细胞代谢和脂质途径的基因来提高含油生物体的脂质产量。

[0006] 上调基因的一种方法是使用强组成型启动子控制其表达。例如,可以使用强组成型启动子来上调解脂耶氏酵母 (*Y. lipolytica*) 二酰基甘油酰基转移酶DGA1,并且这种遗传工程改造显著增加生物体的脂质产量和生产力(参见例如Tai和Stephanopoulos, *Metabolic Engineering* 12:1-9 (2013))。

[0007] 选择用于控制基因表达的最佳启动子是遗传工程改造的关键部分,但不同的启动子可能对于不同的应用是最佳的。例如,用于酵母的工业菌株的最佳启动子可以不同于在实验室菌株中最佳的启动子。

[0008] 已经鉴定和验证了一些解脂耶氏酵母 (*Y. lipolytica*) 和解腺嘌呤阿氏酵母 (*A. adenivorans*) (参见例如美国专利号7,259,255(通过引用并入)和7,264,949(通过引用并入);美国专利申请号2012/0289600(通过引用并入),2006/0094102(通过引用并入)和2003/0186376(通过引用并入);Wartmann等,*FEMS Yeast Research* 2:363-69(2002))。然而,两种生物体都含有数百个尚待鉴定的启动子,并且许多这些启动子可用于工程改造酵母和其他生物体。此外,启动子可以在相同物种的不同菌株之间显著变化,并且这种遗传多态性的鉴定和筛选为遗传工程改造提供了更丰富的工具箱。

[0009] 发明概述

[0010] 公开了解腺嘌呤阿氏酵母 (*Arxula adenivorans*) 和解脂耶氏酵母 (*Yarrowia lipolytica*) 的启动子的核苷酸序列,其可以用于驱动细胞中的基因表达。验证这些启动子、筛选选择的启动子以确定哪些启动子可用于增加含油酵母的脂质生产效率。

附图说明

[0011] 图1描绘了pNC303构建体的图谱,其用作模板以扩增包含酿酒酵母 (*Saccharomyces cerevisiae*) 转化酶基因SUC2和TER1终止子的DNA片段。“Sc URA3”表示用于在酵母中选择的酿酒酵母URA3营养缺陷型标记;“2u ori”表示来自2 μ m圆形质粒的酿酒

酵母复制起点;“pMB1ori”表示来自pBR322质粒的大肠杆菌(E.coli)pMB1复制起点;“AmpR”表示用作用氨苄青霉素选择的标记的bla基因;“ScFBA1p”表示酿酒酵母FBA1启动子-822至-1;“hygR(NG4)”表示由GenScript合成的大肠杆菌hygR基因cDNA(SEQ ID NO:2);“ScFBA1t”表示终止后205bp的酿酒酵母FBA1终止子;“Y1TEF1p(PR3)”表示解脂耶氏酵母TEF1启动子-406至+125;“NG102”表示酿酒酵母SUC2基因(SEQ ID NO:1);“Y1CYC1t(TER1)”表示在终止密码子之后300bp的解脂耶氏酵母CYC1终止子。

[0012] 图2描述了在14种不同启动子和相同TER1终止子(终止密码子后300bp的解脂耶氏酵母CYC1终止子)的控制下表达酿酒酵母转化酶基因SUC2的解脂耶氏酵母菌株NS18转化体的转化酶活性。x轴标记对应于表II中的启动子ID。通过二硝基水杨酸(DNS)试验来测量活性。在30℃下在96孔板中的YPD培养基中细胞生长48小时后分析样品。在不同的96孔板中分析2A和2B中的样品。亲本解脂耶氏酵母菌株NS18(“C”)用作每个平板上的阴性对照。

[0013] 图3描述了用于在解脂耶氏酵母菌株NS18和解腺嘌呤阿氏酵母菌株NS252中表达潮霉素抗性基因(hygR,SEQ ID NO:2)的pNC161构建体的图谱。载体pNC161在转化前通过PacI/PmeI限制性消化进行线性化。“pMB1ori”表示来自pBR322质粒的大肠杆菌pMB1复制起点;“AmpR”表示用作用氨苄青霉素选择的标记的bla基因;“Sc URA3”表示用于在酵母中选择的酿酒酵母URA3营养缺陷型标记;“2u ori”表示来自2μm圆形质粒的酿酒酵母复制起点;“ScFBA1p”表示酿酒酵母FBA1启动子-822至-1;“hygR(NG4)”表示由GenScript合成的大肠杆菌hygR基因cDNA(SEQ ID NO:2);“ScFBA1t”表示在终止密码子之后205bp的酿酒酵母FBA1终止子。

[0014] 图4描述了在不同解腺嘌呤阿氏酵母启动子的控制下表达大肠杆菌潮霉素抗性基因(SEQ ID NO:2)的解腺嘌呤阿氏酵母菌株NS2S2转化体的琼脂平板。标记对应于表I中的启动子ID。转化体在含有YPD和300μg/μL潮霉素B的平板上在37℃下生长2天。阴性对照由亲本解腺嘌呤阿氏酵母菌株NS252组成,用水代替DNA。

[0015] 图5描述了在不同解腺嘌呤阿氏酵母启动子的控制下表达大肠杆菌潮霉素抗性基因(SEQ ID NO:2)的解脂耶氏酵母菌株NS18转化体的琼脂平板。标记对应于表I中的启动子ID。转化体在含有YPD和300μg/μL潮霉素B的平板上在37℃下生长2天。阴性对照由亲本解脂耶氏酵母菌株NS18组成,用水代替DNA。

[0016] 图6描述了用于在解脂耶氏酵母菌株NS18中过表达编码二酰基甘油酰基转移酶DGA1(SEQ ID NO:3)的基因的pNC336构建体的图谱。载体pNC336在转化前通过PacI/NotI限制性消化进行线性化。“Sc URA3”表示用于在酵母中选择的酿酒酵母URA3营养缺陷型标记;“2u ori”表示来自2μm圆形质粒的酿酒酵母复制起点;“pMB1ori”表示来自pBR322质粒的大肠杆菌pMB1复制起点;“AmpR”表示用作用氨苄青霉素选择的标记的bla基因;“PR14AaTEF1p”表示解腺嘌呤阿氏酵母TEF1启动子-427至-1(SEQ ID NO:5);NG66(Rt DGA1)表示由GenScript合成的红冬孢酵母DGA1cDNA(SEQ ID NO:3);“Y1CYC1t(TER1)”表示在终止密码子之后300bp的解脂耶氏酵母CYC1终止子;“ScTEF1p”表示酿酒酵母TEF1启动子-412至-1;“NAT”表示用作用诺尔丝菌素选择的标记物的链霉菌属Nat1基因;“ScCYC1t”表示终止密码子后275bp的酿酒酵母CYC1终止子。

[0017] 图7描述了在不同的解腺嘌呤阿氏酵母启动子和相同的TER1终止子(终止密码子之后300bp的解脂耶氏酵母CYC1终止子)的控制下表达红冬孢酵母(Rhodospordium

toruloides) DGA1蛋白的解脂耶氏酵母菌株NS18转化体的脂质试验结果。x轴标记对应于表I中的启动子ID。对于每个构建体,通过实施例7中描述的脂质试验分析12个转化体。在含有脂质生产诱导培养基的96孔板中细胞生长72小时后分析样品。样品“C”描绘了亲本菌株NS18作为对照,误差棒描绘了从三种不同试验获得的一个标准偏差。

[0018] 图8描绘了在不同解脂耶氏酵母启动子和相同TER1终止子(解脂耶氏酵母CYC1终止子,终止密码子后300bp)控制下表达红冬孢酵母DGA1的解脂耶氏酵母菌株NS18转化体的脂质试验结果。x轴标记对应于表II中的启动子ID。对于每个构建体,通过实施例7中描述的脂质试验分析12个转化体。在含有脂质生产诱导培养基的96孔板中细胞生长72小时后分析样品。样品“C”描绘了亲本菌株NS18作为对照,误差棒描绘了从三种不同试验获得的一个标准偏差。

[0019] 图9描述了用于在解腺嘌呤阿氏酵母菌株NS252中过表达编码来自红冬孢酵母的二酰基甘油酰基转移酶DGA1的基因的pNC378构建体的图。载体pNC378在转化前通过PmeI/AscI限制性消化进行线性化。“Sc URA3”表示用于在酵母中选择的酿酒酵母URA3营养缺陷型标记;“2u ori”表示来自2 μ m圆形质粒的酿酒酵母复制起点;“pMB1ori”表示来自pBR322质粒的大肠杆菌pMB1复制起点;“AmpR”表示用作用氨苄青霉素选择的标记的bla基因;“PR26AaPGK1p”表示解腺嘌呤阿氏酵母PGK1启动子-524至-1(SEQ ID NO:14);“PR25AaADH1p”表示解腺嘌呤阿氏酵母ADH1启动子-877至-1(SEQ ID NO:13);“NG66(Rt DGA1)”表示红冬孢酵母DGA1cDNA;“ScFBA1t(TER6)”表示终止密码子后205bp的酿酒酵母终止子;“NAT”表示用作用诺尔丝菌素选择的标记物的链霉菌属(*Streptomyces noursei*) Nat1基因;“AaCYC1t”表示终止密码子后301bp的解腺嘌呤阿氏酵母CYC1终止子。

[0020] 图10描述了解腺嘌呤阿氏酵母菌株NS252转化体的脂质试验结果,该转化体表达在解腺嘌呤阿氏酵母启动子ADH1和TER16终止子(终止密码子后301bp的解腺嘌呤阿氏酵母CYC1终止子)控制下的来自各种宿主生物的不同DGA蛋白。x轴标记对应于表III中的DGA基因。对于每个构建体,通过实施例7和8中描述的脂质试验分析8个转化体。在含有脂质生产诱导培养基的96孔板中细胞生长72小时后分析样品。样品“C”描绘了亲本菌株NS252作为对照,误差棒描绘了从8种不同试验获得的一个标准偏差。

[0021] 图11描述了解腺嘌呤阿氏酵母菌株NS252转化体的脂质试验结果,该转化体表达在解腺嘌呤阿氏酵母启动子ADH1和TER16终止子(终止密码子后301bp的解腺嘌呤阿氏酵母CYC1终止子)控制下的来自各种宿主生物的不同DGA蛋白。x轴标记对应于表III中的DGA基因。对于每个构建体,通过实施例7和8中描述的脂质试验分析8个转化体。在含有脂质生产诱导培养基的96孔板中细胞生长72小时后分析样品。样品“C”描绘了亲本菌株NS252作为对照,误差棒描绘了从8种不同试验获得的一个标准偏差。

[0022] 图12描述了解腺嘌呤阿氏酵母菌株NS252转化体的脂质试验结果,该转化体表达在解腺嘌呤阿氏酵母启动子ADH1和TER16终止子(终止密码子后301bp的解腺嘌呤阿氏酵母CYC1终止子)控制下的来自各种宿主生物的不同DGA蛋白。x轴标记对应于表III中的DGA基因。对于每个构建体,通过实施例7和8中描述的脂质试验分析8个转化体。在含有脂质生产诱导培养基的96孔板中细胞生长72小时后分析样品。样品“C”描绘了亲本菌株NS252作为对照,误差棒描绘了从8种不同试验获得的一个标准偏差。

[0023] 发明详述

[0024] 概述

[0025] 在一些方面,本发明涉及载体,其包含编码衍生自解腺嘌呤阿氏酵母或解脂耶氏酵母的启动子的核苷酸序列,其中所述载体是质粒。在一些方面,本发明涉及载体,其包含编码衍生自解腺嘌呤阿氏酵母或解脂耶氏酵母的启动子的核苷酸序列,其中所述载体是线性DNA片段。

[0026] 在某些方面,本发明涉及包含遗传修饰的转化的细胞,其中所述遗传修饰是用编码衍生自解腺嘌呤阿氏酵母或解脂耶氏酵母的启动子的核酸转化。

[0027] 在其它方面,本发明涉及在细胞中表达基因的方法,包括用编码衍生自解腺嘌呤阿氏酵母或解脂耶氏酵母的启动子的核酸转化亲本细胞。在一些实施方式中,核酸包含基因,并且基因和启动子可操作地连接。在其他实施方式中,设计核酸使得启动子在转化亲本细胞后变得可操作地连接到基因。

[0028] 定义

[0029] 本文使用冠词“一个”和“一种”表示一个或一个以上的(即至少一个)该冠词语法上的宾语。举例而言,“一种元件”表示一个元件或者一个以上的元件。

[0030] 术语“DGAT2”是指编码2型二酰甘油酰基转移酶蛋白的基因,例如编码DGA1蛋白的基因。

[0031] “二酰基甘油酯”,“二酰基甘油”和“甘油二酯”是由甘油和两种脂肪酸组成的酯。

[0032] 术语“二酰基甘油酰基转移酶”和“DGA”是指催化由二酰基甘油形成三酰基甘油酯的任何蛋白质。二酰基甘油酰基转移酶包括1型二酰基甘油酰基转移酶(DGA2)、2型二酰基甘油酰基转移酶(DGA1)和催化上述反应的所有同系物。

[0033] 术语“二酰基甘油酰基转移酶,2型”和“2型二酰基甘油酰基转移酶”是指DGA1和DGA1直向同源物。

[0034] 术语“结构域”是指蛋白质的氨基酸序列的一部分,其能够折叠成独立于蛋白质其余部分的稳定的三维结构。

[0035] “干重”和“干细胞重量”是指在相对不存在水时测定的平均重量。例如,以干重提及油性细胞包括特定成分的指定百分比是指基本上所有的水已经被除去后,基于细胞的重量计算百分比。

[0036] 术语“编码”是指核苷酸序列,其(a)编码氨基酸序列,(b)可以结合蛋白质例如聚合酶或转录因子,(c)调节结合核酸的蛋白质,例如作为转录起始位点,和(d)(a)、(b)和(c)中所述的核苷酸序列的互补序列。例如,核苷酸序列可以编码结合聚合酶的基因,其编码氨基酸序列和/或启动子。DNA和RNA都可以编码基因。DNA和RNA都可以编码蛋白质。

[0037] 术语“内源的”是指存在于天然的、未转化的细胞中的任何物质,即,没有被引入细胞的一切。“内源核酸”是存在于天然的、未转化的细胞中的核酸,例如染色体或从染色体中天然存在的基因转录的mRNA。内源核酸包括内源基因和内源启动子。术语“内源基因”和“内源启动子”是指天然存在于细胞基因组中的核苷酸序列,其没有通过转化或转染引入。

[0038] 术语“外源的”是指引入细胞的任何物质。“外源核酸”是通过细胞膜进入细胞的核酸。外源核酸可以含有先前不存在于细胞的天然基因组中的核苷酸序列和/或已经存在于基因组中但被重新引入基因组的核苷酸序列,例如通过用其他拷贝的核苷酸序列转化。外源核酸包括外源基因和外源启动子。“外源基因”是已经被引入细胞(例如通过转化/转染)

并编码RNA和/或蛋白质的核苷酸序列,外源基因也称为“转基因”。类似地,“外源启动子”是已被引入细胞(例如通过转化/转染)并编码启动子的核苷酸序列。包含外源基因或外源启动子的细胞可以称为重组细胞,其中可以引入额外的外源基因或启动子。外源基因或外源启动子可以来自相对于被转化的细胞的相同物种或不同物种。因此,外源基因可以包括占据细胞基因组中不同于内源基因的位置的基因,或者相对于基因的内源性拷贝处于不同的可操作的连接。类似地,外源启动子可以包括占据细胞基因组中不同于内源启动子位置的启动子或可操作地连接于与内源启动子不同的基因的启动子。外源基因或外源启动子可以以超过一个拷贝存在于细胞中。外源基因或外源启动子可以作为插入基因组(核或质体)或作为游离分子保持在细胞中。

[0039] 术语“表达”是指细胞中核酸或氨基酸序列(例如肽,多肽或蛋白质)的量。基因的增加的表达是指该基因的增加的转录。氨基酸序列、肽、多肽或蛋白质的增加的表达是指增加的编码氨基酸序列、肽、多肽或蛋白质的核酸的翻译。

[0040] 本文使用的术语“基因”可以包括含有内含子的基因组序列,特别是编码涉及特定活性的多肽序列的多核苷酸序列。该术语还包括不衍生自基因组序列的合成核酸。在某些实施方式中,基因缺乏内含子,因为它们基于cDNA和蛋白质序列的已知DNA序列合成。在其它实施方式中,基因是合成的非天然cDNA,其中已基于密码子使用而优化密码子以在解脂耶氏酵母或解腺嘌呤阿氏酵母腺中表达。该术语还可以包括包含上游、下游和/或内含子核苷酸序列的核酸分子,包括启动子。

[0041] 术语“遗传修饰”是指转化的结果。每个转化通过定义(definition)引起遗传修饰。

[0042] 本文所用的术语“同源物”是指(a)相对于在研的未修饰蛋白质具有氨基酸取代、缺失和/或插入的肽、寡肽、多肽、蛋白质和酶,并且具有与它们来源的未修饰的蛋白质相似的生物学和功能活性,和(b)相对于在研的未修饰核酸具有核苷酸取代、缺失和/或插入的核酸,并且具有与它们来源的未修饰的核酸相似的生物学和功能活性。例如,解脂耶氏酵母可以与由相同转录调节子调节的解腺嘌呤阿氏酵母启动子同源。

[0043] 术语“整合的”是指作为插入细胞基因组中(例如插入染色体,包括插入质体基因组中)的插入物而维持在细胞中的核酸。

[0044] “可操作连接”是两个核酸序列(例如控制序列(通常是启动子)和连接的序列(通常是编码蛋白质的序列,也称为编码序列)之间的功能性连接。如果启动子可以介导基因的转录,则其与基因可操作地连接(或“可操作连接”)。

[0045] 术语“天然”是指在转化事件之前的细胞或亲本细胞的组成。

[0046] 术语“核酸”是指任何长度的核苷酸聚合形式,不论是脱氧核糖核苷酸或核糖核苷酸或它们的类似物。多核苷酸可以具有任意的三维结构,且可以执行任何功能。以下是多核苷酸的非限制性例子:基因或基因片段的编码或非编码区域、由连锁分析定义的基因座、外显子、内含子、信使RNA(mRNA)、转移RNA、核糖体RNA、核酶、cDNA、重组多核苷酸、支链多核苷酸、质粒、载体、任意序列的分离DNA、任意序列的分离RNA、核酸探针和引物。多核苷酸可包括修饰的核苷酸,如甲基化的核苷酸和核苷酸类似物。如果存在,对核苷酸结构的修饰可在聚合物的组装之前或之后赋予。多核苷酸可进一步被修饰,如通过与标记组分结合。在本文提供的所有核酸序列中,U核苷酸可与T核苷酸互换。

[0047] 术语“亲本细胞”是指细胞从其传承的每个细胞。细胞的基因组由亲本细胞的基因组和对其基因组的任何后续遗传修饰组成。

[0048] 如本文所用,术语“质粒”是指与生物体的基因组DNA物理分离的环状DNA分子。质粒可以在引入宿主细胞之前线性化(本文称为线性化质粒)。线性化质粒可以不是自我复制的,而是可以整合到生物体的基因组DNA中并与其一起复制。

[0049] “启动子”是指导核酸转录的核酸控制序列。如本文所用,启动子包括靠近转录起始位点的必需的核酸序列。启动子还任选地包括远端增强子或阻抑物元件,其可位于距转录起始位点多达几千个碱基对处。

[0050] “重组”是指由于外源核酸的引入或天然核酸的改变而被修饰的细胞,核酸,蛋白质或载体。因此,例如,重组细胞可以表达在天然(非重组)形式的细胞中没有发现的基因,或者表达与非重组细胞表达的基因不同的天然基因。重组细胞可以包括但不限于编码基因产物或抑制元件例如降低细胞中活性基因产物水平的突变、敲除、反义、干扰RNA(RNAi)或dsRNA的重组核酸。“重组核酸”衍生自最初在体外,通常通过操纵核酸,例如使用聚合酶、连接酶、外切核酸酶和内切核酸酶,或以其它方式在自然界中通常不存在的形式形成的核酸。可以产生重组核酸,例如,将两个或更多个核酸置于可操作连接中。因此,出于本发明的目的,在体外通过将正常状态天然不相连的DNA分子连接形成的分离核酸或表达载体都被认为是重组的。一旦制备重组核酸并引入宿主细胞或生物体中,其可以使用宿主细胞的体内细胞机制进行复制;然而,此类核酸一旦重组产生,尽管随后在细胞内复制,仍然被认为是用于本发明目的的重组。另外,重组核酸是指包含内源核苷酸序列和外源核苷酸序列的核苷酸序列;因此,已经与外源启动子重组的内源基因是重组核酸。“重组蛋白”是使用重组技术,即通过表达重组核酸而制备的蛋白质。

[0051] 术语“调节区”是指影响基因的转录或翻译但不编码氨基酸序列的核苷酸序列。调节区包括启动子,操纵子,增强子和沉默子。

[0052] 术语“子序列”是指在核苷酸序列内发现的小于全长核苷酸序列的连续核苷酸序列。例如,子序列可以由选自427个核苷酸长的SEQ ID NO:5所示的核苷酸序列的100个连续核苷酸组成;可以在427个核苷酸长的序列中发现328个100连续核苷酸长的子序列。由在全长核苷酸序列的3'-末端的100个连续核苷酸组成的子序列是指在该序列中发现的最后100个核苷酸。例如,子序列可以由SEQ ID NO:5的3'-末端的100个连续核苷酸组成,并且该子序列是SEQ ID NO:5的最后100个核苷酸。换句话说,SEQ ID NO:5的3'端的100个连续核苷酸是具有前327个核苷酸缺失的SEQ ID NO:5的核苷酸序列,其是单个子序列。如本文所用,子序列由至少50个核苷酸组成。

[0053] “转化”是指将核酸转移到宿主生物体或宿主生物体的基因组中,导致遗传上稳定的遗传。包含转化的核酸片段的宿主生物体称为“重组”、“转基因”或“转化”生物体。因此,本发明的分离的多核苷酸可以掺入能够引入宿主细胞并在宿主细胞中复制的重组构建体,通常是DNA构建体。这种构建体可以是包含能够在给定宿主细胞中转录和翻译多肽编码序列的复制系统和序列的载体。通常,表达载体包括例如在5'和3'调节序列和可选择标记的转录控制下的一个或多个克隆基因。这样的载体还可以含有启动子调节区(例如,控制诱导型或组成型,环境或发育调节或位置特异性表达的调节区),转录起始起始位点,核糖体结合位点,转录终止位点,和/或多聚腺苷酸化信号。或者,细胞可以用单个遗传元件如启动子

转化,其可以在整合到宿主生物体的基因组中,例如通过同源重组时导致遗传上稳定的遗传。

[0054] 术语“转化的细胞”是指经历转化的细胞。因此,转化的细胞包含亲本的基因组和可遗传的遗传修饰。

[0055] 术语“三酰基甘油酯”,“三酰基甘油”,“甘油三酯”和“TAG”是由甘油和三个脂肪酸组成的酯。

[0056] 术语“载体”是核酸能增殖和/或在生物体、细胞或细胞组分之间转移的方法。载体包括可能或可能不能自主复制或整合到宿主细胞的染色体中的质粒,线性DNA片段,病毒,噬菌体,原病毒,噬菌粒,转座子和人工染色体等。

[0057] 微生物工程改造

[0058] A. 概述

[0059] 外源启动子和基因可以引入许多不同的宿主细胞中。合适的宿主细胞是可广泛存在于真菌家族中的微生物宿主。合适的宿主菌株的实例包括但不限于真菌或酵母物种,如阿氏菌属(Arxula)、曲霉属(Aspegillus)、奥兰氏菌属(Aurantiochytrium)、念珠菌属(Candida)、麦角菌属(Claviceps)、隐球菌属(Cryptococcus)、小坎宁安霉属(Cunninghamella)、汉逊酵母属(Hansenula)、克鲁维酵母属(Kluyveromyces)、白冬孢酵母属(Leucosporidiella)、油脂酵母属(Lipomyces)、被孢霉属(Mortierella)、欧格氏霉属(Ogataea)、毕赤酵母属(Pichia)、原囊藻属(Prototheca)、根霉属(Rhizopus)、红冬孢酵母属(Rhodospidium)、红酵母属(Rhodotorula)、酵母菌属(Saccharomyces)、裂殖酵母属(Schizosaccharomyces)、银耳属(Tremella)、丝孢酵母属(Trichosporon)和耶氏酵母属(Yarrowia)。解脂耶氏酵母和解腺嘌呤阿氏酵母非常适合用作宿主微生物,因为它们可以以三酰基甘油积累大部分的重量。

[0060] 本发明的微生物经遗传工程改造成含有外源启动子,其可以是强启动子或弱启动子。强启动子驱动可操作连接的基因的相当大的转录。弱启动子对于许多应用可能是有价值的。例如,弱启动子可优选用以驱动编码在高浓度下显示毒性的蛋白质的基因的转录或编码针对必需蛋白质的干扰RNA的核苷酸序列的转录。因此,当强启动子产生致死量的蛋白质产物时,优选弱启动子用于表达蛋白质。类似地,当基底水平的靶标是细胞存活所必需的时,优选弱启动子用于表达干扰RNA。

[0061] 微生物表达系统和表达载体是本领域技术人员熟知的。任何这样的表达载体可以用于将即时启动子引入生物体。可以通过转化技术将启动子引入合适的微生物中,以指导可操作连接的基因的表达。例如,可将启动子克隆到合适的质粒中,并用所得质粒转化亲本细胞。该方法可用于驱动与启动子可操作地连接或在转化事件后变得可操作地连接于启动子的基因的表达。所述质粒没有特别限制,只要其能产生对微生物后代可遗传的所需启动子。

[0062] 用于转化合适的宿主细胞的载体或盒是本领域熟知的。通常,载体或盒含有基因,指导相关基因的转录和翻译的序列,包括启动子,选择标记和允许自主复制或染色体整合的序列。合适的载体包含具有启动子和其它转录起始控制的基因的5'区域和控制转录终止的DNA片段的3'区域。优选两个控制区来源于与转化的宿主细胞同源或来自密切相关物种的基因,尽管应理解这种控制区不需要来源于选择作为生产宿主的特定物种的天然基因。

例如,解腺嘌呤阿氏酵母启动子可以用于驱动其他酵母菌种的表达。

[0063] 启动子,cDNA和3'UTR以及载体的其他元件可以通过使用从天然来源分离的片段的克隆技术产生(Green和Sambrook,《分子克隆:实验室手册》(第四版,2012);美国专利号4,683,202;通过引用并入)。或者,可以使用已知方法(Gene 164:49-53(1995))合成地产生元件。

[0064] B. 启动子序列

[0065] 在一些实施方式中,本发明涉及启动子。在一些实施方式中,启动子包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中所示的核苷酸序列。启动子可以包含保守取代、缺失和/或插入,同时仍然起驱动转录的作用。因此,启动子序列可包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高相同性的核苷酸序列。

[0066] 为了确定两个核苷酸序列的百分比同一性,可以为了最佳比较目的而对序列进行比对(例如,可以在第一和第二核苷酸序列的一个或两个中引入间隙用于最佳比对,并且出于比较目的可以不考虑非相同序列)。然后可以比较相应核苷酸位置处的核苷酸。当第一序列中的位置由与第二序列中的相应位置相同的核苷酸占据时,则分子在该位置是相同的(如本文所使用的,核苷酸“同一性”等同于核苷酸“同源性”)。考虑到用于两个序列的最佳比对而需要引入的间隙数目和每个间隙的长度,两个序列之间的百分比同一性是序列共有的相同位置的数目的函数。

[0067] 序列的比较和两个序列之间百分比同一性的确定可以使用数学算法来完成。可用于测定两个核苷酸序列之间同一性的示例性计算机程序包括但不限于BLAST程序套件,例如BLASTN、MEGABLAST和Clustal程序,例如ClustalW,ClustalX和Clustal Omega。

[0068] 当相对于GenBank DNA序列和其他公共数据库中的核苷酸序列评估给定的核苷酸序列时,通常使用BLASTN程序进行序列搜索。使用例如CLUSTAL-W程序进行选定序列的比对以确定两个或多个序列之间的“%同一性”。

[0069] 在整个说明书中用于指包含核苷酸序列和/或由核苷酸序列组成的核酸的缩写是常规的单字母缩写。因此,当包括在核酸中时,天然存在的编码核苷酸缩写如下:腺嘌呤(A),鸟嘌呤(G),胞嘧啶(C),胸腺嘧啶(T)和尿嘧啶(U)。此外,本文呈现的核苷酸序列是5'→3'方向。

[0070] 如本文所使用的,术语“互补”及其衍生物用于通过众所周知的规则来配对核酸,即A与T或U配对,和C与G配对。互补可以是“部分的”或“完整的”。在部分互补中,根据碱基配对规则只有一些核苷酸匹配;而在完全或全部互补中,所有碱基根据配对规则匹配。核酸链之间的互补程度可以对本领域熟知的两条核酸链之间的杂交效率和强度具有显著的影响。杂交的效率和强度取决于检测方法。

[0071] 启动子的全核苷酸序列对于驱动转录不是必需的,并且比启动子的全核苷酸序列

短的序列可以驱动可操作连接的基因的转录。称为核心启动子的启动子的最小部分包括转录起始位点, RNA聚合酶的结合位点和转录因子的结合位点。RNA聚合酶结合到启动子的3'末端。因此, 启动子可包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的3' -端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高相同性的核苷酸序列。

[0072] 另外, 可以组合两个启动子。例如, 结合RNA聚合酶的第一启动子的区域可以与结合一种或多种转录因子的第二启动子的区域组合以产生杂合启动子。因此, 启动子的子序列可以与另一个启动子组合以改变调节可操作地连接的基因的转录的转录因子。因此, 启动子可包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任何位置的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高相同性的核苷酸序列。

[0073] C. 载体和载体组分

[0074] 根据本发明的用于转化微生物的载体可以通过本领域技术人员熟知的已知技术根据本文的公开内容制备。载体通常含有一个或多个基因, 其中每个基因编码所需产物(基因产物)的表达, 并可操作地连接到调节基因表达的一个或多个调控序列(即启动子), 或载体将基因、控制序列或其它核苷酸序列靶向重组细胞中的特定位置。

[0075] 任何核酸载体可以编码启动子。质粒可以是方便的载体, 因为质粒可以在细菌宿主中操作和复制。在一些实施方式中, 线性DNA分子可以是优选的载体, 例如, 在转化前消除质粒核苷酸序列。线性DNA可以从质粒的限制性消化或通过PCR扩增获得。PCR可用于通过扩增质粒DNA、基因组DNA、合成DNA或任何其它模板产生线性DNA载体。例如, PCR可用于从重叠的寡核苷酸片段产生线性DNA载体。合适的载体不限于DNA; 例如, 逆转录病毒载体的RNA可用于转化具有所需启动子的细胞。

[0076] 载体可以包含启动子和基因, 使得启动子和基因可操作地连接。或者, 可以设计载体使得启动子在转化亲本细胞后变得可操作地连接到基因。例如, 含有启动子的第一载体

可以设计为与含有基因的第二载体重组,从而成功的转化和重组事件使得启动子和基因在宿主细胞中可操作地连接。或者,可以将含有启动子的载体设计为与宿主细胞基因组中的基因重组。在该实施方式中,外源启动子取代内源启动子。

[0077] 1. 控制序列

[0078] 控制序列是调节编码序列表达或将基因产物导向细胞内或细胞外特定位置的核酸。调节表达的调控序列包括例如调节编码序列转录的启动子和终止编码序列转录的终止子。另一个控制序列是位于编码多聚腺苷酸化信号的编码序列末端的3'非翻译序列。将基因产物导向特定位置的控制序列包括编码信号肽的那些,将它们所连接的蛋白质导向细胞内或细胞外的特定位置。

[0079] 因此,用于在微生物中表达启动子的示例性载体设计包含与在酵母中有活性的启动子可操作连接的所需基因产物(例如,选择性标记或酶)的编码序列。或者,如果载体不含有与启动子可操作连接的基因,则启动子可以转化到细胞中,使得其在载体整合的位点变得可操作地连接到内源基因。

[0080] 用于表达基因的启动子可以是与该基因天然连接的启动子或不同的启动子。

[0081] 任选包括终止区控制序列,并且如果使用,则选择主要是方便之一,因为终止区是相对可互换的。终止区可以对于转录起始区(启动子)是天然的,对感兴趣的DNA序列可以是天然的,或者可以从另一个来源获得(参见例如Chen和Orozco, *Nucleic Acids Research* 16:8411 (1988))。

[0082] 2. 基因

[0083] 通常,基因包括启动子,编码序列和终止控制序列。当通过重组DNA技术组装时,基因可以称为表达盒,并且可以侧接限制性位点以方便地插入用于将重组基因导入宿主细胞的载体中。表达盒可以侧接来自基因组或其他核酸靶的DNA序列,以利于通过同源重组将表达盒稳定整合到基因组中。或者,载体及其表达盒可以保持未整合(例如附加体),在这种情况下,载体通常包括复制起点,其能够提供载体DNA的复制。

[0084] 载体上存在的常见基因是编码蛋白质的基因,该基因的表达允许含有该蛋白质的重组细胞与不表达该蛋白质的细胞进行区分。这样的基因及其相应的基因产物称为选择性标记或选择标记。在用于转化本发明生物体的转基因构建体中可以使用多种选择标记中的任一种。

[0085] 为了最佳表达重组蛋白,使用产生具有待转化宿主细胞最佳使用的密码子的编码序列是有益的。因此,转基因的正确表达可能需要转基因的密码子使用与其中表达转基因的生物体的特异性密码子偏好相匹配。这种效应背后的精确机制有许多,但包括可用的氨酰基化tRNA库与细胞中合成的蛋白质的适当平衡,以及当满足该需要时更高效地翻译转基因信使RNA (mRNA)。当不优化转基因中的密码子使用时,可用的tRNA库不足以允许转基因mRNA的高效翻译,导致核糖体停滞和终止以及转基因mRNA的可能不稳定性。

[0086] D. 同源重组

[0087] 同源重组可用于用不同的核苷酸序列替代一个核苷酸序列。因此,可以使用同源重组来用外源启动子的全部或部分来替代驱动生物体中的基因表达的内源启动子的全部或部分。另外,同源重组可用于组合含有同源核苷酸序列的两种核酸。

[0088] 同源重组是互补DNA序列比对和交换同源区的能力。例如,可以产生含有与被靶向

的基因组序列(“模板”)同源的序列的转基因DNA(“供体”),并将其引入生物体中以与生物体的基因组序列进行重组。

[0089] 在宿主生物体中进行同源重组的能力对于可以在分子遗传水平上进行的事物具有许多实际意义,并且可用于生产产生所需产物的微生物。根据其本质,同源重组是精确的基因靶向事件;因此,用相同的靶向序列产生的大多数转基因品系在表型方面将是基本相同的,需要筛选少得多的转化事件。同源重组也将基因插入事件靶向宿主染色体,潜在地导致优异的遗传稳定性,即使在没有遗传选择的情况下。

[0090] 因为同源重组是精确的基因靶向事件,所以其可以用于精确修饰目的基因或区域内的任何核苷酸,只要已经鉴定了足够的侧翼区。因此,同源重组可用于修饰影响RNA和/或蛋白质表达的调节序列。其还可以修饰蛋白质编码区,例如通过修饰酶活性例如底物特异性,结合亲和力和 K_m ,因此其可以影响宿主细胞的代谢的所需变化。同源重组提供了操纵宿主基因组的有效手段,导致基因靶向,基因转化,基因缺失,基因重复,基因倒置和交换基因表达调节元件,例如启动子,增强子和3'UTR。因此,同源重组允许用不同的启动子取代生物体中的内源启动子。外源启动子可以提供优于内源启动子的优点;例如,外源启动子可以增加或降低可操作连接的基因的转录,或者外源启动子可允许通过有关内源启动子的不同细胞过程调节转录。

[0091] 同源重组可以通过使用含有内源序列片段的靶向构建体来实现,以“靶向”内源性宿主细胞基因组内的感兴趣的基因或区域。这种靶向序列可以位于感兴趣的基因或区域的上游或下游,或位于感兴趣的基因/区域的侧翼。这种靶向构建体可以作为环状质粒DNA转化到宿主细胞中,任选地包括来自质粒的核苷酸序列;线性化DNA,例如质粒限制性消化;PCR产物,例如重叠寡核苷酸的扩增;或将DNA引入细胞的任何其它方式。在一些情况下,可能有利的是通过用限制酶切割转基因DNA首先暴露转基因DNA(供体DNA)内的同源序列,这可以提高重组效率并减少非特异性重组事件的发生。提高重组效率的其它方法包括使用PCR产生含有与被靶向的基因组序列同源的线性末端的转化转基因DNA。

[0092] E. 转化

[0093] 可以通过任何合适的技术转化细胞,包括例如生物弹射,电穿孔,玻璃珠转化和碳化硅晶须转化。用于将转基因导入微生物的任何方便的技术都可用于本发明。转化可以通过例如D.M.Morrison的方法(Methods in Enzymology 68:326(1979)),通过用氯化钙增加受体细胞对DNA的渗透性的方法(Mandel&Higa,J.Molecular Biology,53:159(1970))等来实现。

[0094] 在产油酵母(例如解脂耶氏酵母)中表达转基因的实例可以在文献中找到(Bordes等人,J.Microbiological Methods,70:493(2007);Chen等人,Applied Microbiology & Biotechnology 48:232(1997))。

[0095] 用于转化微生物的载体可以通过已知技术制备。在一个实施方式中,用于在微生物中表达基因的示例性载体包含编码与启动子可操作连接的蛋白质的基因。或者,如果启动子不与感兴趣基因可操作地连接,则启动子可以转化到细胞中,使得其在载体整合的位点变得可操作地连接到天然基因。另外,微生物可以用两种载体同时转化(参见,例如,Protist 155:381-93(2004))。转化的细胞可以任选地基于其在未转化的细胞不会生长的抗生素或其它选择性标记存在的条件下生长的能力来选择。

[0096] 示例性核酸、细胞和方法

[0097] 1. 衍生自解腺嘌呤阿氏酵母和解脂耶氏酵母的核苷酸序列

[0098] 在一些实施方式中,本发明涉及编码启动子的核酸分子。在一些实施方式中,启动子来源于编码下述的基因:翻译延伸因子EF-1 α ;甘油-3-磷酸脱氢酶;磷酸丙糖异构酶1;果糖-1,6-二磷酸醛缩酶;磷酸甘油酸变位酶;丙酮酸激酶;输出蛋白EXP1;核糖体蛋白S7;醇脱氢酶;磷酸甘油酸激酶;己糖转运蛋白;一般氨基酸通透酶;丝氨酸蛋白酶;异柠檬酸裂解酶;酰基辅酶A氧化酶;ATP-硫酸化酶;己糖激酶;3-磷酸甘油酸脱氢酶;丙酮酸脱氢酶 α 亚基;丙酮酸脱氢酶 β 亚基;乌头酸;烯醇酶;肌动蛋白;多药耐药蛋白(ABC-转运蛋白);泛素;GTP酶;质膜Na⁺/Pi共转运蛋白;丙酮酸脱羧酶;植酸酶;或 α -淀粉酶。在一些实施方式中,启动子源自编码下述的基因:TEF1;GPD1;TPI1;FBA1;GPM1;PYK1;EXP1;RPS7;ADH1;PGK1;HXT7;GAP1;XPR2;ICL1;POX;MET3;HXK1;SER3;PDA1;PDB1;ACO1;ENO1;ACT1;MDR1;UBI4;YPT1;PHO89;PDC1;PHY;或AMYA。

[0099] 在一些实施方式中,启动子来源于编码下述的基因:磷酸甘油酸激酶;己糖激酶;6-磷酸果糖激酶 α 亚基;磷酸丙糖异构酶1;3-磷酸甘油酸脱氢酶;丙酮酸激酶1;丙酮酸脱氢酶 α 亚基;丙酮酸脱氢酶 β 亚基;乌头酸;烯醇酶;肌动蛋白;核肌动蛋白相关蛋白;多药耐药蛋白(ABC-转运蛋白);泛素;参与ER/高尔基体囊泡运输的亲水蛋白;或质膜Na⁺/Pi共转运蛋白。在一些实施方式中,启动子来源于编码下述的基因:PGK1;HXK1;PFK1;TPI1;SER3;PYK1;PDA1;PDB1;ACO1;ENO1;ACT1;ARP4;MDR1;UBI4;SLY1;或PHO89。

[0100] 在一些实施方式中,核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53所示序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在其他实施方式中,核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的子序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方式中,核酸包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中所示的核苷酸序列。在其他实施方式中,核酸包含由SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的子序列组成的核苷酸序列。在某些实施方式中,所述子序列保留启动子活性。在某些实施方式中,子序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、

45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中，子序列保留了全长核苷酸序列的启动子活性。

[0101] 在一些实施方案中，子序列为50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个核苷酸长或更长。在一些实施方案中，子序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在一些实施方案中，子序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。

[0102] 在一些实施方案中，核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方案中，核酸包含由SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、

120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸组成的核苷酸序列。在某些实施方式中，核苷酸序列保留启动子活性。在某些实施方式中，核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中，核苷酸序列保留了全长核苷酸序列的启动子活性。

[0103] 在一些实施方案中，核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方案中，核酸包含由SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸组成的核苷酸序列。在某些实施方式中，核苷酸序列保留启动子活性。在某些实施方式中，核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中，核苷酸序列保留了全长核苷酸序列的启动子活性。

[0104] 2. 包含源自解腺嘌呤阿氏酵母的启动子的载体

[0105] 在一些实施方案中，本发明涉及包含编码来自解腺嘌呤阿氏酵母的启动子的核苷

酸序列的载体,其中所述启动子来源于编码下述的基因:翻译延伸因子EF-1 α ;甘油-3-磷酸脱氢酶;磷酸丙糖异构酶1;果糖-1,6-二磷酸醛缩酶;磷酸甘油酸变位酶;丙酮酸激酶;输出蛋白EXP1;核糖体蛋白S7;醇脱氢酶;磷酸甘油酸激酶;己糖转运蛋白;一般氨基酸通透酶;丝氨酸蛋白酶;异柠檬酸裂解酶;酰基辅酶A氧化酶;ATP-硫酸化酶;己糖激酶;3-磷酸甘油酸脱氢酶;丙酮酸脱氢酶 α 亚基;丙酮酸脱氢酶 β 亚基;乌头酸;烯醇酶;肌动蛋白;多药耐药蛋白(ABC-转运蛋白);泛素;GTP酶;质膜Na⁺/Pi共转运蛋白;丙酮酸脱羧酶;植酸酶;或 α -淀粉酶。

[0106] 在一些实施方案中,载体是质粒。在其他实施方案中,载体是线性DNA分子。

[0107] 在一些实施方案中,载体包含编码来自解腺嘌呤阿氏酵母的启动子的核苷酸序列,其中所述启动子源自编码下述的基因:TEF1;GPD1;TPI1;FBA1;GPM1;PYK1;EXP1;RPS7;ADH1;PGK1;HXT7;GAP1;XPR2;ICL1;POX;MET3;HXX1;SER3;PDA1;PDB1;ACO1;ENO1;ACT1;MDR1;UBI4;YPT1;PHO89;PDC1;PHY;或AMYA。

[0108] 在一些实施方案中,核苷酸序列与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53所示序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在其他实施方案中,核苷酸序列与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的子序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在一些实施方案中,核苷酸序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53所示序列。在其他实施方案中,核苷酸序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的子序列。在某些实施方案中,所述子序列保留启动子活性。在其他实施方式中,子序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,子序列保留了全长核苷酸序列的启动子活性。

[0109] 在一些实施方案中,子序列为50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、

250、255、260、265、270、275、280、285、290、295或300个核苷酸长或更长。在一些实施方案中，子序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在一些实施方案中，子序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。

[0110] 在一些实施方案中，核苷酸序列与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在一些实施方案中，核苷酸序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在某些实施方式中，核苷酸序列保留启动子活性。在某些实施方式中，核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中，核苷酸序列保留了全长核苷酸序列的启动子活性。

[0111] 在一些实施方案中，核苷酸序列与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、

35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在一些实施方案中,核苷酸序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在某些实施方式中,核苷酸序列保留启动子活性。在某些实施方式中,核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,核苷酸序列保留了全长核苷酸序列的启动子活性。

[0112] 在一些实施方案中,载体还包含基因,并且基因和启动子可操作地连接。在其它实施方案中,设计载体使得在用载体转化细胞时启动子与基因可操作地连接。

[0113] 3. 包含源自解脂耶氏酵母的启动子的载体

[0114] 在一些实施方案中,本发明涉及包含编码来自解脂耶氏酵母的启动子的核苷酸序列的载体,其中所述启动子来源于编码下述的基因:磷酸甘油酸激酶;己糖激酶;6-磷酸果糖激酶 α 亚基;磷酸丙糖异构酶1;3-磷酸甘油酸脱氢酶;丙酮酸激酶1;丙酮酸脱氢酶 α 亚基;丙酮酸脱氢酶 β 亚基;乌头酸;烯醇酶;肌动蛋白;核肌动蛋白相关蛋白;多药耐药蛋白(ABC-转运蛋白);泛素;参与ER/高尔基体囊泡运输的亲水蛋白;或质膜Na⁺/Pi共转运蛋白。

[0115] 在一些实施方案中,载体是质粒。在其他实施方案中,载体是线性DNA分子。

[0116] 在一些实施方案中,载体包含编码来自解脂耶氏酵母的启动子的核苷酸序列,其中所述启动子衍生自编码下述的基因:PGK1;HXK1;PFK1;TPI1;SER3;PYK1;PDA1;PDB1;ACO1;ENO1;ACT1;ARP4;MDR1;UBI4;SLY1;或PH089。

[0117] 在一些实施方案中,核苷酸序列与SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34所示序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、

99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在其他实施方案中,核苷酸序列与SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的子序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在一些实施方式中,核苷酸序列包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34所示的序列。在其他实施方式中,核苷酸序列包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的子序列。在某些实施方案中,所述子序列保留启动子活性。在某些实施方式中,子序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,子序列保留了全长核苷酸序列的启动子活性。

[0118] 在某些实施方案中,子序列为50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个核苷酸长或更长。在一些实施方案中,子序列包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在一些实施方案中,子序列包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。

[0119] 在一些实施方案中,核苷酸序列与SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34中任意位置的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、

240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在一些实施方案中,核苷酸序列包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在某些实施方式中,核苷酸序列保留启动子活性。在某些实施方式中,核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,核苷酸序列保留了全长核苷酸序列的启动子活性。

[0120] 在一些实施方案中,核苷酸序列与SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性。在一些实施方案中,核苷酸序列包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在某些实施方式中,核苷酸序列保留启动子活性。在某些实施方式中,核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、

64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,核苷酸序列保留了全长核苷酸序列的启动子活性。

[0121] 4. 包含源自解腺嘌呤阿氏酵母的启动子的转化的细胞和用源自解腺嘌呤阿氏酵母的启动子转化细胞的方法

[0122] 在某些方面,本发明涉及包含遗传修饰的转化的细胞,其中所述遗传修饰是用编码来自解腺嘌呤阿氏酵母的启动子的核酸转化。在一些方面,本发明涉及在细胞中表达基因的方法,包括用编码来自解腺嘌呤阿氏酵母的启动子的核酸转化亲本细胞。在一些实施方案中,核酸包含基因,并且基因和启动子可操作地连接。在其他实施方案中,设计核酸使得启动子在转化亲本细胞后变得可操作地连接到基因。

[0123] 在一些实施方案中,启动子来源于编码下述的基因:翻译延伸因子EF-1 α ;甘油-3-磷酸脱氢酶;磷酸丙糖异构酶1;果糖-1,6-二磷酸醛缩酶;磷酸甘油酸变位酶;丙酮酸激酶;输出蛋白EXP1;核糖体蛋白S7;醇脱氢酶;磷酸甘油酸激酶;己糖转运蛋白;一般氨基酸通透酶;丝氨酸蛋白酶;异柠檬酸裂解酶;酰基辅酶A氧化酶;ATP-硫酸化酶;己糖激酶;3-磷酸甘油酸脱氢酶;丙酮酸脱氢酶 α 亚基;丙酮酸脱氢酶 β 亚基;乌头酸;烯醇酶;肌动蛋白;多药耐药蛋白(ABC-转运蛋白);泛素;GTP酶;质膜Na⁺/Pi共转运蛋白;丙酮酸脱羧酶;植酸酶;或 α -淀粉酶。在一些实施方案中,启动子源自编码下述的基因:TEF1;GPD1;TPI1;FBA1;GPM1;PYK1;EXP1;RPS7;ADH1;PGK1;HXT7;GAP1;XPR2;ICL1;POX;MET3;HXK1;SER3;PDA1;PDB1;ACO1;ENO1;ACT1;MDR1;UBI4;YPT1;PHO89;PDC1;PHY;或AMYA。

[0124] 在一些实施方案中,核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53所示序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在其他实施方案中,核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的子序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方案中,核酸包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53所示的核苷酸序列。在其他实施方案中,核酸包含由SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的子序列组成的核苷酸序列。在某些实施方案中,所述子序列保留启动子活性。在某些实施方式中,子序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、

63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,子序列保留了全长核苷酸序列的启动子活性。

[0125] 在一些实施方案中,子序列为50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个核苷酸长或更长。在一些实施方案中,子序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在一些实施方案中,子序列包含SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。

[0126] 在一些实施方案中,核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方案中,核酸包含由SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸组成的核苷酸序列。在某些实施方式中,核苷酸序列保留启动子活性。在某些实施方式中,核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、

17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,核苷酸序列保留了全长核苷酸序列的启动子活性。

[0127] 在一些实施方案中,核酸包含与SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方案中,核酸包含由SEQ ID NO:5、6、7、8、9、10、11、12、13、14、15、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52或53的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸组成的核苷酸序列。在某些实施方式中,核苷酸序列保留启动子活性。在某些实施方式中,核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,核苷酸序列保留了全长核苷酸序列的启动子活性。

[0128] 5. 包含源自解脂耶氏酵母的启动子的转化的细胞和用源自解脂耶氏酵母的启动子转化细胞的方法

[0129] 在某些方面,本发明涉及包含遗传修饰的转化的细胞,其中所述遗传修饰是用编码来自解脂耶氏酵母的启动子的核酸转化。在一些方面,本发明涉及在细胞中表达基因的方法,包括用编码来自解脂耶氏酵母的启动子的核酸转化亲本细胞。在一些实施方案中,核酸包含基因,并且基因和启动子可操作地连接。在其他实施方案中,设计核酸使得启动子在转化亲本细胞后变得可操作地连接到基因。

[0130] 在一些实施方案中,启动子来源于编码下述的基因:磷酸甘油酸激酶;己糖激酶;6-磷酸果糖激酶 α 亚基;磷酸丙糖异构酶1;3-磷酸甘油酸脱氢酶;丙酮酸激酶1;丙酮酸脱氢酶 α 亚基;丙酮酸脱氢酶 β 亚基;乌头酸;烯醇酶;肌动蛋白;核肌动蛋白相关蛋白;多药耐药蛋白(ABC-转运蛋白);泛素;参与ER/高尔基体囊泡运输的亲水蛋白;或质膜Na⁺/Pi共转运蛋白。在一些实施方案中,启动子来源于编码下述的基因:PGK1;HXK1;PFK1;TPI1;SER3;PYK1;PDA1;PDB1;ACO1;ENO1;ACT1;ARP4;MDR1;UBI4;SLY1;或PHO89。

[0131] 在一些实施方案中,核酸包含与SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34所示序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在其他实施方案中,核酸包含与SEQ IDNO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的子序列具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方式中,核酸包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34所示的核苷酸序列。在其他实施方式中,核酸包含由SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的子序列组成的核苷酸序列。在某些实施方案中,所述子序列保留启动子活性。在某些实施方式中,子序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,子序列保留了全长核苷酸序列的启动子活性。

[0132] 在某些实施方案中,子序列为50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个核苷酸长或更长。在一些实施方案中,子序列包含SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。在一些实施方案中,子序列包含SEQ

ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸。

[0133] 在一些实施方案中,核酸包含与SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34中任意位置的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷酸序列。在一些实施方案中,核酸包含由SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34中任意位置处的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸组成的核苷酸序列。在某些实施方式中,核苷酸序列保留启动子活性。在某些实施方式中,核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,核苷酸序列保留了全长核苷酸序列的启动子活性。

[0134] 在一些实施方案中,核酸包含与SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸具有至少约70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.6%、99.7%、99.8%、99.9%或更高序列同源性的核苷

酸序列。在一些实施方案中,核酸包含由SEQ ID NO:16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33或34的3'端的50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、72、73、74、75、76、77、78、79、80、81、82、83、84、85、86、87、88、89、90、91、92、93、94、95、96、97、98、99、100、105、110、115、120、125、130、135、140、145、150、155、160、165、170、175、180、185、190、195、200、205、210、215、220、225、230、235、240、245、250、255、260、265、270、275、280、285、290、295或300个连续核苷酸组成的核苷酸序列。在某些实施方式中,核苷酸序列保留启动子活性。在一些实施方式中,核苷酸序列保留全长核苷酸序列的至少1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、21%、22%、23%、24%、25%、26%、27%、28%、29%、30%、31%、32%、33%、34%、35%、36%、37%、38%、39%、40%、41%、42%、43%、44%、45%、46%、47%、48%、49%、50%、51%、52%、53%、54%、55%、56%、57%、58%、59%、60%、61%、62%、63%、64%、65%、66%、67%、68%、69%、70%、71%、72%、73%、74%、75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的启动子活性。在某些实施方案中,核苷酸序列保留了全长核苷酸序列的启动子活性。

[0135] 6. 细胞,亲本细胞和转化的细胞的种类

[0136] 细胞可以选自藻类,细菌,霉菌,真菌,植物和酵母。在一些实施方式中,细胞选自下组:阿氏菌属(*Arxula*)、曲霉属(*Aspergillus*)、奥兰氏菌属(*Aurantiochytrium*)、念珠菌属(*Candida*)、麦角菌属(*Claviceps*)、隐球菌属(*Cryptococcus*)、小坎宁安霉属(*Cunninghamella*)、地霉属(*Geotrichum*)、汉逊酵母属(*Hansenula*)、克鲁维酵母属(*Kluyveromyces*)、柯达酵母(*Kodamaea*)、白冬孢酵母属(*Leucosporidiella*)、油脂酵母属(*Lipomyces*)、被孢霉属(*Mortierella*)、欧格氏霉属(*Ogataea*)、毕赤酵母属(*Pichia*)、原囊藻属(*Prototheca*)、根霉属(*Rhizopus*)、红冬孢酵母属(*Rhodospiridium*)、红酵母属(*Rhodotorula*)、酵母菌属(*Saccharomyces*)、裂殖酵母属(*Schizosaccharomyces*)、银耳属(*Tremella*)、丝孢酵母属(*Trichosporon*)、威克汉姆酵母属(*Wickerhamomyces*)和耶氏酵母属(*Yarrowia*)。在某些实施方式中,细胞选自下组:解腺嘌呤阿氏酵母(*Arxula adenivorans*)、黑曲霉(*Aspergillus niger*)、米曲霉(*Aspergillus oryzae*)、土曲霉(*Aspergillus terreus*)、裂殖壶奥兰氏菌(*Aurantiochytrium limacinum*)、产朊假丝酵母(*Candida utilis*)、麦角菌(*Claviceps purpurea*)、浅白隐球菌(*Cryptococcus albidus*)、弯曲隐球菌(*Cryptococcus curvatus*)、拉米隐球菌(*Cryptococcus ramirezgomezianus*)、地生隐球菌(*Cryptococcus terreus*)、威氏隐球菌(*Cryptococcus wieringae*)、刺孢小坎宁安霉(*Cunninghamella echinulata*)、贴梗海棠小坎宁安霉(*Cunninghamella japonica*)、发酵性地霉(*Geotrichum fermentans*)、多形汉森酵母(*Hansenulapolyomorpha*)、乳酸克鲁维酵母(*Kluyveromyces lactis*)、马克斯克鲁维酵母(*Kluyveromyces marxianus*)、奥默柯达酵母(*Kodamaea ohmeri*)、克雷氏白冬孢酵母(*Leucosporidiella creatinivora*)、产油油脂酵母(*Lipomyces lipofer*)、斯达氏油脂酵母(*Lipomyces starkeyi*)、子囊菌油脂酵母(*Lipomyces tetrasporus*)、深黄被孢霉(*Mortierella isabellina*)、高山被孢霉(*Mortierella alpina*)、马兜铃欧格氏霉(*Ogataea polymorpha*)、塞佛式毕赤酵母(*Pichia ciferrii*)、季氏毕赤氏酵母(*Pichia*

guilliermondii)、巴斯德毕赤酵母(*Pichia pastoris*)、叶柄毕赤酵母(*Pichia stipites*)、小型原藻(*Prototheca zopfii*)、隐根根霉(*Rhizopus arrhizus*)、巴氏红冬孢酵母(*Rhodosporidium babjevae*)、圆红冬孢酵母(*Rhodosporidium toruloides*)、帕鲁氏红冬孢酵母(*Rhodosporidium paludigenum*)、粘红酵母(*Rhodotorula glutinis*)、胶红酵母(*Rhodotorula mucilaginosa*)、酿酒酵母(*Saccharomyces cerevisiae*)、粟酒裂殖酵母(*Schizosaccharomyces pombe*)、脑状银耳(*Tremella encephala*)、皮状丝孢酵母(*Trichosporon cutaneum*)、发酵性丝孢酵母(*Trichosporon fermentans*)、塞佛式威克汉姆酵母(*Wickerhamomyces ciferrii*)、和解脂耶氏酵母(*Yarrowia lipolytica*)。因此,细胞可以是解脂耶氏酵母。细胞可以是解腺嘌呤阿氏酵母。

[0137] 通过以下实施例进一步说明本说明书,这些实施例不应被解释为以任何方式构成限制。所有引用文献的内容(包括本申请通篇引用的文献参考文献,授权专利,公开的专利申请和GenBank登录号)通过引用明确地并入本文。当通过引用并入本文的文献中的术语定义与本文所使用的定义冲突时,以本文使用的定义为主。

实施例

[0138] 实施例1:测序解腺嘌呤阿氏酵母基因组和鉴定启动子序列

[0139] 鉴定并筛选解腺嘌呤阿氏酵母启动子。首先,为了获得选定基因的启动子序列,解腺嘌呤阿氏酵母NS252株(ATCC 76597)的基因组进行测序,并通过合成基因组公司(Synthetic Genomics Inc.)(美国加州)注释。

[0140] 基于关于酵母和真菌中常用的启动子的公开数据列举了可能特别用于驱动转录的启动子。例如,鉴定和筛选涉及重要代谢途径例如糖酵解的基因的启动子。可特别用于驱动转录的解腺嘌呤阿氏酵母启动子序列示于SEQ ID NO:5-15和35-53中,并列于下表I中。

[0141] 表I,解腺嘌呤阿氏酵母启动子

[0142]

启动子	启动子编号	相关蛋白质功能	SEQ ID NO
TEF1	PR14	翻译延伸因子 EF-1 α	5
GPD1	PR15	甘油 3 磷酸脱氢酶	6
TPI1	PR16	磷酸丙糖异构酶 1	7
FBA1	PR17	果糖-1,6-二磷酸醛缩酶	8
GPM1	PR18	磷酸甘油酸变位酶	9
PYK1	PR19	丙酮酸激酶	10
EXP1	PR20	输出蛋白	11
RPS7	PR21	核糖体蛋白 S7	12
ADH1	PR25	醇脱氢酶	13
PGK1	PR26	磷酸甘油酸激酶	14
HXT7	PR27	己糖转运蛋白	15
GAP1	PR57	一般氨基酸通透酶	35
XPR2	PR58	丝氨酸蛋白酶	36
ICL1	PR59	异柠檬酸裂解酶	37
POX	PR60	酰基辅酶 A 氧化酶	38
MET3	PR61	ATP-硫酸化酶	39
HXK1	PR62	己糖激酶	40
SER3	PR63	3-磷酸甘油酸脱氢酶	41
PDA1	PR64	丙酮酸脱氢酶 α 亚基	42
PDB1	PR65	丙酮酸脱氢酶 β 亚基	43
ACO1	PR66	乌头酸	44
ENO1	PR67	烯醇酶	45
ACT1	PR68	肌动蛋白	46
MDR1	PR69	多药耐药蛋白 (ABC-转运蛋白)	47
UBI4	PR70	泛素	48
YPT1	PR71	GTP 酶	49
PHO89	PR72	质膜 Na ⁺ /Pi 共转运蛋白	50
PDC1	PR73	丙酮酸脱羧酶	51
PHY	PR74	植酸酶	52
AMYA	PR75	α -淀粉酶	53

[0143] 实施例2:解脂耶氏酵母启动子的鉴定

[0144] 解脂耶氏酵母基因组在KEGG数据库中是公开可获得的,但是每个解脂耶氏酵母启动子的精确序列尚未被鉴定或验证。

[0145] 基于关于酵母和真菌中常用的启动子的公开数据列举了可能特别用于驱动转录的启动子。例如,鉴定和筛选涉及重要代谢途径例如糖酵解的基因的启动子。可特别用于驱动转录的解脂耶氏酵母启动子序列示于SEQ ID NO:16-34中并列于下表II中。

[0146] 表II.解脂耶氏酵母启动子

[0147]

启动子	启动子编号	相关蛋白质功能	SEQ ID NO
PGK1	PR34*, PR54	磷酸甘油酸激酶	16*, 32
HXK1	PR35	己糖激酶	17
PFK1	PR36	6-磷酸果糖激酶 α 亚基	18
TPI1	PR37*, PR55	磷酸丙糖异构酶 1	19*, 33
SER3	PR38	3-磷酸甘油酸脱氢酶	20
PYK1	PR39*, PR56	丙酮酸激酶 1	21*, 34
PDA1	PR40	丙酮酸脱氢酶 α 亚基	22
PDB1	PR41	丙酮酸脱氢酶 β 亚基	23
ACO1	PR42	乌头酸	24
ENO1	PR43	烯醇酶	25
ACT1	PR44	肌动蛋白	26
ARP4	PR45	核肌动蛋白相关蛋白	27
MDR1	PR46	多药耐药蛋白 (ABC-转运蛋白)	28
UBI4	PR47	泛素	29
SLY1	PR49	参与 ER/高尔基体囊泡运输的亲水蛋白	30
PHO89	PR50	质膜 Na ⁺ /Pi 共转运蛋白	31

[0148] *表示启动子和连续转录序列。

[0149] 实施例3:验证解脂耶氏酵母启动子序列并评估其使用转化酶报告物基因的强度

[0150] 使用酿酒酵母转化酶基因SUC2 (SEQ ID NO:1) 作为报告物,在解脂耶氏酵母菌株NS18中筛选选择的解脂耶氏酵母启动子的功能性和强度。转化酶基因既用作选择标记,筛选用于在蔗糖上生长的细胞,又用作用于定量评价启动子强度的报告物。另外,通过实施例4中描述的DNS试验来测量启动子强度。

[0151] 酿酒酵母转化酶基因在解脂耶氏酵母菌株NS18中在十四种不同的解脂耶氏酵母启动子和相同的TER1终止子的控制下表达。使用含有与转化酶基因的5'末端同源的30-35个碱基对的反向引物,从宿主解脂耶氏酵母菌株NS18 (从NRRL#YB-392获得) 的基因组DNA扩增启动子,以允许同源重组启动子和转化酶DNA。从pNC303质粒扩增转化酶核苷酸序列和TER1终止子 (图1)。将每个扩增的启动子的DNA与扩增的转化酶-TER1片段的DNA组合,并使

用Chen等人描述的转化方案转化到NS18菌株中(Applied Microbiology&Biotechnology 48:232-35 (1997))。启动子DNA片段和转化酶-TER1DNA片段在体内组装并随机整合到宿主解脂耶氏酵母菌株NS18的基因组中。

[0152] 将转化体铺板,并在具有2%蔗糖的YNB板上选择,并通过实施例4中所述的DNS试验筛选转化酶活性。就每个启动子分析了几个转化体。DNS试验的结果显示在图2中。大多数启动子在转化体之间显示出显著的集落变异,可能是由于转化酶的整合位点对表达的影响。图2证明所有十四个启动子都允许转化酶表达。对于具有较低表达水平和较低集落数(PR39,PR41,PR43,PR45和PR46)的那些启动子,其转化体在YNB+2%蔗糖选择性平板上生长的事实表明启动子能够实现转化酶的充分转录,以在蔗糖上生长。

[0153] 实施例4:二硝基水杨酸试验

[0154] 将细胞在30℃下在YPD琼脂平板上温育1至2天。来自每个琼脂平板的细胞用于在96孔板的孔中接种300μL培养基。96孔板用多孔盖覆盖并在InforsMultitron ATR振荡器中在30℃,70-90%湿度和900rpm下孵育。

[0155] 将96孔板在3000rpm离心2分钟。在新96孔板中,将50μL上清液加入到150μL含有40mM乙酸钠,pH为4.5-5的50mM蔗糖中,并在30℃温育30-60分钟。

[0156] 在新96孔板中,将30μL蔗糖/上清液混合物加入到60μL DNS试剂(1%二硝基水杨酸,30%酒石酸钠钾,0.4M NaOH)中并用PCR膜覆盖。将板在热循环器中加热至99℃持续5分钟。

[0157] 然后将70μL混合物转移至康宁96孔透明平底板中,在SpectraMax M2分光光度计(分子装置公司(Molecular Devices))上监测540nm处的吸光度。

[0158] 实施例5:利用hygR报告物基因验证解腺嘌呤阿氏酵母启动子序列

[0159] 实施例3和4中所述的转化酶报告物试验不适用于解腺嘌呤阿氏酵母菌株NS252,因为该菌株具有在蔗糖上生长的天然能力。因此,大肠杆菌hygR基因(SEQ ID NO:2)用作解腺嘌呤阿氏酵母中的报告物并且用作通过潮霉素B(HYG)选择的转化选择标记。在十一个选择的启动子和相同的终止子的控制下,hygR基因在解脂耶氏酵母和解腺嘌呤阿氏酵母中表达(图4和5)。图3显示了使用来自酿酒酵母的FBA1启动子(SEQ ID NO:4)作为示例,用于在解脂耶氏酵母和解腺嘌呤阿氏酵母中过表达hygR基因的表达构建体pNC161的图谱。FBA1启动子也用作阳性对照,因为它可以驱动解脂耶氏酵母和解腺嘌呤阿氏酵母中的hygR表达。所有hygR表达构建体与pNC161相同,除了启动子序列。用水转化细胞作为阴性对照。

[0160] 在通过PacI/PmeI限制性消化转化之前将表达构建体线性化。每个线性表达构建体包括hygR基因的表达盒和不同的启动子。将表达构建体随机整合到解脂耶氏酵母菌株NS18和解腺嘌呤阿氏酵母菌株NS252的基因组中,使用Chen等人所述的转化方案(Applied Microbiology&Biotechnology 48:232-35 (1997))。

[0161] 在具有300μg/mL HYG的YPD平板上选择转化体,并基于在平板上生长的菌落的大小筛选启动子强度。每个转化体的YPD+HYG平板的照片示于图4和5中。解腺嘌呤阿氏酵母的转化效率远低于解脂耶氏酵母,可能是因为转化方案针对解脂耶氏酵母优化而不是解腺嘌呤阿氏酵母。转化体的数目在不同构建体之间不同,可能是由于在不同转化期间使用的DNA量略有不同,尽管启动子的强度可能导致这种变化。图4和5仍然证明所有11种启动子在解脂耶氏酵母和解腺嘌呤阿氏酵母中有功能。

[0162] 不同的解腺嘌呤阿氏酵母启动子的解腺嘌呤阿氏酵母转化体的菌落大小没有显著变化,表明当与hygR报告物连接时天然的解腺嘌呤阿氏酵母启动子具有相似效率。同时,解脂耶氏酵母菌落的大小显著变化。该数据可以表明不同的解腺嘌呤阿氏酵母启动子与解腺嘌呤阿氏酵母调节因子类似地相互作用,并与解脂耶氏酵母调节因子不同地相互作用。

[0163] 在解腺嘌呤阿氏酵母和解脂耶氏酵母中筛选的每个启动子能够在解腺嘌呤阿氏酵母和解脂耶氏酵母中驱动基因表达,这表明SEQ ID NO:6-53中鉴定的所有启动子在所有酵母中都是有功能的。

[0164] 实施例6:使用DGA2作为报告物评估解腺嘌呤阿氏酵母和解脂耶氏酵母启动子序列的强度

[0165] 选择通过实施例3-5中所述的转化酶和hygR试验评估的最有效的启动子,用于在解脂耶氏酵母中使用二酰基甘油酰基转移酶DGA1作为报告物进一步定量测试。DGA1蛋白催化三酰基甘油(TAG)合成的最后步骤,因此DGA1是脂质合成途径中的关键组分。DGA1在解脂耶氏酵母中的过表达显著增加其脂质生产效率。因此,在DGA1试验中启动子的强度与脂质生产效率相关。

[0166] 编码来自红冬孢酵母菌的DGA1的基因(SEQ ID NO:3)在解脂耶氏酵母中在十二个选定的启动子和相同的终止子的控制下表达。图6显示了作为示例的表达构建体pNC336的图谱;该构建体用于用来自解腺嘌呤阿氏酵母的TEF1启动子(SEQ ID NO:5)过表达DGA1。所有其他DGA1表达构建体与pNC336相同,除了它们的启动子序列。

[0167] 在通过PacI/NotI限制性消化转化之前将表达构建体线性化。每个线性表达构建体包括编码DGA1的基因的表达盒和用于通过诺尔丝菌素(NAT)进行选择的作用标记物的Nat1基因的表达盒。将表达构建体随机整合到解脂耶氏酵母菌株NS18的基因组中,使用Chen等人所述的转化方案(*Applied Microbiology&Biotechnology* 48:232-35(1997))。在具有500 μ g/mL NAT的YPD平板上选择转化体,并通过实施例7中所述的荧光染色脂质试验来筛选累积脂质的能力。

[0168] 使用实施例7中描述的荧光染色脂质试验分析12个转化体的每个表达构建体(图7和8)。大多数构建体在转化体之间显示出显著的集落变异,这可能是由于在仅获得功能性Nat1盒的一些转化体中缺乏功能性DGA1表达盒,或者DGA1表达盒整合位点对DGA1表达的负效应。然而,图7和8证明所有十二种启动子增加解脂耶氏酵母的脂质含量,这证实了每种启动子用于增加脂质产生的功能性并再次确认它们驱动基因表达的功能性。

[0169] 实施例7:脂质荧光试验

[0170] 将含有0.5g/L尿素,1.5g/L酵母提取物,0.85g/L酪蛋白氨基酸,1.7g LYNB(无氨基酸和硫酸铵),100g/L葡萄糖和5.11g/L邻苯二甲酸氢钾(25mM)的过滤灭菌的培养基装入高压灭菌的多孔板的每个孔中。针对24孔板,每孔使用1.5mL培养基,针对96孔板,每孔使用300 μ l培养基。或者,使用酵母培养物在高压灭菌的250mL烧瓶中接种50ml灭菌的培养基。使用在30 $^{\circ}$ C下在YPD琼脂平板上温育1-2天的酵母菌株接种多壁板的每个孔。

[0171] 多孔板用多孔盖覆盖并在Infors Multitron ATR振荡器中在30 $^{\circ}$ C,70-90%湿度和900rpm下孵育。或者,用铝箔覆盖烧瓶,并在New Brunswick Scientific烧瓶中于30 $^{\circ}$ C,70-90%湿度和900rpm下孵育。96小时后,将20 μ L 100%乙醇加入到在分析微量培养板中的20 μ L细胞中,并在4 $^{\circ}$ C下孵育30分钟。然后将20 μ L细胞/乙醇混合物加入到Costar 96孔、黑

色、透明底板中的80 μ L预混合的溶液中并覆盖有透明密封物,该预混合的溶液含有50 μ L1M碘化钾,1mM Bodipy493/503,0.5 μ L 100%DMSO,1.5 μ L 60%PEG 4000和27 μ L水。30 $^{\circ}$ C下用SpectraMax M2分子光度计(分子装置公司)动力学试验监测Bodipy荧光,并通过用荧光除以600nm处的吸光度来标准化。

[0172] 实施例8:解腺嘌呤阿氏酵母启动子增加酵母中的脂质生产

[0173] 选择通过实施例5中所述的hygR试验评估的启动子,以筛选来自解腺嘌呤阿氏酵母中的各种生物体的编码二酰基甘油酰基转移酶(DGA)的基因,从而增加脂质产生。DGA蛋白催化三酰基甘油(TAG)合成的最后步骤,因此DGA是脂质合成途径中的关键组分。

[0174] 编码来自各种宿主生物体的DGA1,DGA2和DGA3的基因在解腺嘌呤阿氏酵母菌株NS252中在解腺嘌呤阿氏酵母ADH1启动子(SEQ ID NO:13)和CYC1终止子控制下表达,所述宿主生物体例如解腺嘌呤阿氏酵母、解脂耶氏酵母、红冬孢酵母菌、斯达氏油脂酵母、土曲霉、麦角菌、裂殖壶奥兰氏菌、球毛壳菌(*Chaetomium globosum*)、牧草红酵母(*Rhodotorula graminis*)、花药黑粉菌(*Microbotryum violaceum*)、禾柄锈菌(*Puccinia graminis*)、密粘褶菌(*Gloeophyllum trabeum*)、倒卵形红冬孢酵母(*Rhodosporidium diobovatum*)、三角褐指藻(*Phaeodactylum tricornutum*)、蛾幼虫被虫草菌(*Ophiocordyceps sinensis*)、绿木霉(*Trichoderma virens*)、蓖麻(*Ricinus communis*)、和落花生(*Arachis hypogaea*)。图9显示了表达构建体pNC378的图谱作为示例。该构建体用于用来自解腺嘌呤阿氏酵母的启动子ADH1(SEQ ID NO:13)过表达红冬孢酵母DGA1。所有其他DGA表达构建体与pNC378相同,除了DGA序列。使用解腺嘌呤阿氏酵母PGK1启动子(SEQ ID NO:14)驱动所有构建体中选择标记物NAT的表达。

[0175] 表III.用解腺嘌呤阿氏酵母ADH1启动子筛选DGA的列表

基因	基因号	供体生物体
DGA2	NG168	解腺嘌呤阿氏酵母
DGA1	NG167	解腺嘌呤阿氏酵母
DGA1	NG15	解脂耶氏酵母
DGA1	NG66	红冬孢酵母
DGA1	NG69	斯达氏油脂酵母
DGA1	NG70	土曲霉
DGA1	NG71	麦角菌
DGA1	NG72	裂殖壶奥兰氏菌
DGA2	NG16	解脂耶氏酵母
DGA2	NG109	红冬孢酵母
DGA2	NG110	斯达氏油脂酵母
DGA2	NG111	土曲霉
[0176] DGA2	NG112	麦角菌
DGA2	NG113	球毛壳菌
DGA1	NG286	牧草红酵母
DGA1	NG287	花药黑粉菌
DGA1	NG288	禾柄锈菌
DGA1	NG289	密粘褶菌
DGA1	NG290	倒卵形红冬孢酵母
DGA1	NG293	三角褐指藻
DGA2	NG295	三角褐指藻
DGA2	NG297	蛾幼虫被虫草菌
DGA2	NG298	绿木霉
DGA3	NG299	蓖麻
DGA3	NG300	落花生

[0177] 在通过PmeI/AscI限制性消化转化之前将表达构建体线性化。每个线性表达构建体包括编码DGA的基因的表达盒和用于通过诺尔丝菌素 (NAT) 进行选择的作用标记物的Nat1基因的表达盒。将表达构建体随机整合到解腺嘌呤阿氏酵母菌株NS252的基因组中。简言之,将5mL YPD培养基用来自过夜集落的NS252在YPD平板上接种,并在37℃下孵育16-24小时。接下来,将2.5mL的过夜培养物用于在250mL摇瓶中接种22.5mL的YPD培养基。在37℃下3-4小时后,将培养物在3000rpm离心3分钟。弃去上清液,用水洗涤细胞,离心,弃去上清液。

[0178] 为了使细胞具有感受态,将2mL的100mM LiAc和40 μ L的2M DTT加入到细胞沉淀中并在37 $^{\circ}$ C下孵育1小时。将细胞溶液以10,000rpm离心10秒,弃去上清液。先用水洗涤沉淀,然后用冷的1M山梨醇洗涤沉淀。将洗涤的沉淀重悬于2mL冷的1M山梨醇中并置于冰上。将40 μ L细胞-山梨糖醇溶液和5 μ L消化的构建体加入预冷的0.2cm电穿孔比色皿内。将细胞在25uF,200欧姆和1.5kV下电穿孔,时间常数为约4.9-5.0ms。在1mL YPD中在37 $^{\circ}$ C下回收细胞过夜。将100 μ L-500 μ L回收的培养物在具有50 μ g/mL NAT的YPD平板上铺板。

[0179] 使用实施例7中描述的荧光染色脂质试验分析每个表达构建体的八个转化体。大多数构建体在转化体之间显示出显著的集落变异,这可能是由于在仅获得功能性Nat1盒的一些转化体中缺乏功能性DGA表达盒,或者DGA表达盒整合位点对DGA表达的负效应。然而,图10、11和12证明两个解腺嘌呤阿氏酵母启动子ADH1和PGK1可用作构建可行的表达盒的工具。

[0180] 参考文献的引用

[0181] 本文引用的所有专利,公开的专利申请和其他文献通过引用并入本文。

[0182] 等同形式

[0183] 本领域技术人员应了解或能够确定采用不超过常规实验即可获得本文所述的本发明具体实施方式的许多等同形式。这类等同形式应包含在所附权利要求书的范围内。

序列表

<110> 诺沃吉股份有限公司(NOVOGY, INC.)

<120> 来源于解脂耶氏酵母和解腺嘌呤阿氏酵母的启动子及其使用方法

<130> NGX-034.25

<140>

<141>

<150> 62/028,946

<151> 2014-07-25

<160> 53

<170> PatentIn version 3.5

<210> 1

<211> 1599

<212> DNA

<213> 酿酒酵母(*Saccharomyces cerevisiae*)

<400> 1

```

atgcttttgc aagctttcct tttccttttg gctggttttg cagccaaaat atctgcatca 60
atgacaaacg aaactagcga tagacctttg gtccacttca cacccaacaa gggctggatg 120
aatgacccaa atgggtttgt gtacgatgaa aaagatgcca aatggcatct gtactttcaa 180
tacaacccaa atgacaccgt atggggtacg ccattgtttt ggggccatgc tacttccgat 240
gatttgacta attgggaaga tcaaccatt gctatcgctc ccaagcgtaa cgattcaggt 300
gctttctctg gctccatggt ggttgattac aacaacacga gtgggttttt caatgatact 360
attgatccaa gacaaagatg cgttgcgatt tggacttata aactcctga aagtgaagag 420
caatacatta gctattctct tgatgggtgggt tacactttta ctgaatacca aaagaacct 480
gttttagctg ccaactccac tcaattcaga gatccaaagg tgttctggta tgaaccttct 540
caaaaatgga ttatgacggc tgccaaatca caagactaca aaattgaaat ttactcctct 600
gatgacttga agtcctggaa gctagaatct gcatttgcca atgaaggttt cttaggctac 660
caatacgaat gtccaggttt gattgaagtc ccaactgagc aagatccttc caaatcttat 720
tgggtcatgt ttatttctat caaccagggt gcacctgctg gcggttcctt caaccaatat 780
tttgttggat ccttcaatgg tactcatttt gaagcgtttg acaatcaatc tagagtggta 840
gattttggta aggactacta tgccttgeaa actttettea aactgacce aacctacggt 900
tcagcattag gtattgcctg ggettcaaac tgggagtaca gtgcctttgt cccaactaac 960
ccatggagat catccatgct tttgggtccgc aagttttctt tgaacactga atatcaagct 1020
aatccagaga ctgaattgat caatttgaaa gccgaaccaa tattgaacat tagtaatgct 1080
gggccctggg ctcgttttgc tactaacaca actetaacta aggccaatc ttacaatgct 1140
gatttgagca actcgactgg taccettagag tttgagttgg tttacgctgt taacaccaca 1200
caaaccatat ccaaatecgt ctttgccgac ttateacttt ggttcaaggg tttagaagat 1260
cctgaagaat atttgagaat gggttttgaa gtcagtgett cttccttctt tttggaccgt 1320
ggtaactcta aggtcaagtt tgcaaggag aacctatatt tcacaaacag aatgtctgtc 1380

```

aacaaccaac cattcaagtc tgagaacgac ctaagttact ataaagtgta cggcctactg 1440
 gatcaaaaaca tcttggattt gtacttcaac gatggagatg tggtttctac aaatacctac 1500
 ttcatgacca ccggtaacgc tctaggatct gtgaacatga ccactgggtg cgataatttg 1560
 ttctacattg acaagttcca agtaagggaa gtaaaatag 1599

<210> 2

<211> 1026

<212> DNA

<213> 大肠杆菌(*Escherichia coli*)

<400> 2

atgaagaagc ccgagctgac cgctacctct gttgagaagt tcttgattga gaagtttgat 60
 tccgtttccg acctgatgca gctgtccgag ggcgaggagt ctcgagcctt ctcttttgac 120
 gtgggcccggac gaggttacgt tctgagagtg aactcgtgtg ccgacggcct ctacaaggat 180
 cgatacgtct accgacaactt tgettctgce getctgccc teectgaggt tctcgacatt 240
 ggcgagttct ctgagtcctt cacctaetgc atctctcgac gagctcaggg agtcaccctg 300
 caggacctcc ctgagactga getgcctget gtctctcage ctggttctga ggccatggac 360
 gctatcgtct ctgctgatct gtcccagacc tcgggtttcg gcccctttgg acctcaggga 420
 attggacagt acaccacttg gcgagaetct atctgtgcta ttgccgatec tcacgtctac 480
 cattggcaga ccgttatgga cgatactgtg tcggcttctg tcgctcagge tctggacgag 540
 ctgatgctct gggcccaggga ttgccccgag gttcgacacc tgggtgatgc tgacttcggt 600
 tccaacaacg ttctcaccga caacggcccga atcactgccg tgattgactg gtccgaggct 660
 atgtttggcg actcgcagta cgaggtggcc aacatcttct tttggcgacc ctggctggct 720
 tgtatggagc agcagaccgg atacttcgag cgacgacatc ctgagctcgc tggatccctt 780
 cgactgcgag cttacatgct ccgaattggt ctggaccage tctaccagtc gctggtggat 840
 ggcaactttg acgatgctgc ctgggctcag ggacgatgtg acgcatcgt gcgatctggc 900
 gctggaaccg tcggacgaac tcagattgcc cgacgatccg ctgctgtctg gaccgacgga 960
 tgcgtggagg tcctggctga ttcgggtaac cgacgaccct ctactcgacc tcgagctaag 1020
 gagtaa 1026

<210> 3

<211> 1047

<212> DNA

<213> 红冬孢酵母(*Rhodospiridium toruloides*)

<400> 3

atgggccagc aggcgacgcc cgaggageta tacacacget cagagatctc caagatcaag 60
 ttgcaccct ttggcgtccc gcggtcgcgc cggetgcaga cttctctcgt ctttgcttgg 120
 acgacggcac tgccatcct actcggegtc ttctctctec tetgctcgtt cccaccgctc 180
 tggccggctg tcattgecta cctcactggt gtctttttea ttgaccagge gccgattcac 240
 ggtggacggg cgcagtcttg getgcggaag agtcggatat gggctctggt tgcaggatac 300
 tatcccgtea gcttgatcaa gagegccgac ttgccgctg accggaagta cgtctttggc 360
 taccaccgc acggcgtcat aggcattggc gccatcgcca acttcgcgac cgacgcaacc 420

ggcttctcga cactcttccc cggettgaac cctcacctcc tcacctcca aagcaacttc 480
 aagctcccgc tctaccgca gttgctgctc gctctcgca tatgctcctc ctcatgaag 540
 agctgtcaga acattctgcg acaaggctct ggctcgctc tcactatcgt cgtcgggtggc 600
 gccgccgaga gcttgagtgc gcatccccga accgccgatc ttacgctcaa gcgacgaaaa 660
 ggcttcatca aactcgcgat ccggcaaggc gccgacctg tgcccgtctt ttcggttcggc 720
 gagaacgaca tctttggcca gctgcgaaac gagcgaggaa cgcggtgta caagttgcag 780
 aagcgtttcc aaggcgtgtt tggttcacc ctccctctct tctacggccg gggactcttc 840
 aactacaacg tcggattgat gccgtatcgc catccgatcg tctctgtcgt cggtcgacca 900
 atctcggtag agcagaagga ccaccggacc acggcggacc tcgaagaagt tcaggcgcgg 960
 tatatcgcag aactcaagcg gatctgggaa gaatacaagg acgcctacgc caaaagtcgc 1020
 acgcgggagc tcaatattat cgctga 1047

<210> 4

<211> 822

<212> DNA

<213> 酿酒酵母 (*Saccharomyces cerevisiae*)

<400> 4

gatccaactg gcaccgctgg cttgaacaac aataccagcc ttccaacttc tgtaaataac 60
 ggcggtagc cagtgccacc agtaccgta ctttcggta tacctcttt ccccatgttt 120
 ccaatgccct tcatgcctcc aacggctact atcacaate ctcatcaage tgacgcaage 180
 cctaagaaat gaataacaat actgacagta ctaaataatt gcctacttgg cttcacatac 240
 gttgcatacg tcgatataga taataatgat aatgacagca ggattatcgt aatacgtaat 300
 agttgaaaat ctcaaaaatg tgtgggtcat tacgtaaata atgataggaa tgggattctt 360
 ctatttttcc tttttccatt ctacgagccg tcgggaaaac gtggcatcct ctctttcggg 420
 ctcaattgga gtcacgctgc cgtgagcatc ctctctttcc atatctaaca actgagcacg 480
 taaccaatgg aaaagcatga gcttagcgtt gctccaaaa agtattggat ggtaataacc 540
 atttgtctgt tctcttctga cttgactcc tcaaaaaaaaa aaaatctaca atcaacagat 600
 cgcttcaatt acgcccctac aaaaactttt ttctttcttc ttgcccacg ttaaatttta 660
 tccctcatgt tgtctaacgg atttctgcac ttgatttatt ataaaaagac aaagacataa 720
 tacttctcta tcaatttcag ttattgttct tcttgcgtt atttctctgt tcttcttttt 780
 cttttgtcat atataacat aaccaagtaa tacatattea aa 822

<210> 5

<211> 427

<212> DNA

<213> 解腺嘌呤阿氏酵母 (*Arxula adeninivorans*)

<400> 5

catggctcac ttgcggtcac cgettgcgat aagcgcagat taccacaaag gtcttagtag 60
 cttgaagggt gaaaacttga ggtttacaag ggccccaaaa ctcaattgca gccactaaaa 120
 tgagcattca atctataatc agtccatagt caacaagagc gctcaaaatt gatacagttt 180
 agtgaatctt gctcgagatg agcgggcat agttgctttt ggggagccct aagtgttacg 240

tgcggcgcgc gggatgttcc cctattagcc aaaggccgac egggtaacce ctcgagaaaa 300
 aaaaattttt cgccgctaatt ctgggtgttat ataaagetcc cccctgctett ggattttttc 360
 cttgtcaact cacaccgga atcgaaggca tttcattctg agtagttctc aaaaaacata 420
 atcaaca 427

<210> 6

<211> 910

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 6

cctggtcgtc ctctcttcc tctgggtgtg gatcctgcaa tacctcagcg atccggttcc 60
 atgcaattgc ctgcgctcgt gtcaactggg actcaggcca ttgattgggt tagccagaat 120
 aggtgcgaga gaagaacatc aaaatctgct gccagggtcc tactgcgctc eggcggtcgc 180
 tttcatccat gtacgccatc aatggetett ccggttccatt cttataacc cgcctcgcgt 240
 cgtagatgat cggttggaac cctatcgatt taagggttag ttcggcagac atgaacagcg 300
 actccatggc gttccaaatt cgtccccgcc attccaggte ttggccgctg ctgttaacat 360
 gactgggctt ttcgaccaga gccgcaaggc cgatgggggt aaaatctgcc acgtactttc 420
 cggccgcca cttctctata aactgggttg gtgccatggc aatccacacg cgtcgtcaag 480
 ctaatgtccc ttccacatta ctgcggcttt gcaatgtgag gttcgggtaca attacatcat 540
 acgccgcaac tacaccagca acctaaaga actagtccga atctgtccag aaaccaattg 600
 tcagcaaaaca atcagacaca catacatggt ttgaccacac aacaccaca ccattatgac 660
 cagtcacat tgcgtcctac aaggatcatg taccatgact gcggtggtat tttgtttgcc 720
 atttgtcata cttcattat gctgcaacgt tagacggctg tgctgatctc cgtggtgaca 780
 ccacaatagg ccacgtttat ccgttggtcc gctcattact acacccttg tgccctgtgt 840
 ttggtgttgt catgccttta attcagtatc tgaggccact ttaacggaat caccctgag 900
 aatcgcaatc 910

<210> 7

<211> 499

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 7

cgttccgttc actttcccgt cgcgacaaa ttgacttctg ttgctatatt ttaactctc 60
 cgggactggc tcgtaagccg cacgcgcttg ttatgacagg gteaagctgc cccaacaag 120
 ctttcaaggc acgccgttat gacgaattgg atgacgatta tgatcaaac ccgggtcaat 180
 cggccgcccc agagaccctt tttcgacgca ttgaaaatt caaactccc tagcctagcc 240
 gaccgcatt cggggagtcc gcgaaaagt cccggaacag cccatacggg ggcctaccgc 300
 ctcaattgct cggtaatcac cgtataaagc caataagggt acagagctgt tttttgtgac 360
 tgatagttcg gttgatacaa gaaggaaaga aaaaaatc acaatggtga gtagaatttg 420
 cgagacgaca gtatggcaat tgattgtgac gtaacattt ccgtaggctc gaaagttctt 480
 tgtcggagga aacttcaag 499

<210> 8

<211> 882

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 8

ttgctgccat accacagtcc acctggtaca tttctacgct gttccaaggg gaataaccga 60
gccgcatgat aaccgacccg atcgcaagct caacaaattg ccgtacgggc atacgcgacg 120
agaatttctc ggagacctgg gactacggaa acgggtgtcg gatttcagtc cattggcaac 180
tcaaaccgac aattgacaag actcaattgc tggaagacaa tgccaaaggt ataaccacg 240
ttaccgcct cactggctac cgggtccgcg atacgagtc cttgtcatgc accgtggtea 300
gggtccattg tacggtttga atttgcggtt gtcagggcg agccgaacaa aagtcgtggc 360
acgagaataa tcgtgceggg gtacacttc ccatactcg tgtatataag tatccatccc 420
tactctgttt ccatcacctc ttgtacggg gaatacacia aaggtaagtc aattgttggg 480
acctctgtag tatgacgcat taggetagge tgTTTTTTTT tcaaacggtt tcaccggcat 540
caccgcaggg tcagccttag gggccaccgt tgcaaggtac tgtttagtg gtcattgtg 600
tgacggttgc agggcaggaa ttgacccta tctgaggcaa agacgtcatt ggccccgcaa 660
ccaacacac cagcccctat tcagccatt gtctgatta gttttggcac aattgcaatt 720
ggctccta at gcagagagaa ccctgcaaat gttgcttgat tggtcgcccc actacagata 780
gtgatgtagt ggtgaaccac ccaacattgg tgtgatatat ataaaagccc ttgtttgtct 840
attgtgtcat tctttcttga accaaaaaag actaattcca ga 882

<210> 9

<211> 741

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 9

ccaaacgcga atccgcccga tgtagaagct ccaaacagcg gtttgattc agaactagag 60
gtagtagtag tggcaggagt agtggatcca gccgaagcgc caaacagtcc ccccgaagga 120
ttgcccggcg gcgtcgtcga ccaaaaactg aaccgcctt gtttaggagt ggagacccg 180
ctgttctgtc caaacattgt tgcaactggtc aaacctgaag gttggtaa acagtagaatg 240
tgTTTTTCGCA cgtgccgtcc acatgaacag agtcagatga cagtcagatg acatacggct 300
ataaaaagcgt cataaatcac ctgactaccg catgactacc acgcgataat cacctgacta 360
caaccggaag attcatcacc tcatacccct cggcagatgc cgaaagtccg ggcaattatc 420
gtgataatca tgacgccccaa ttgggacca attgcgagag caacacaaa cgacacgcag 480
tgtcaccgca attgggctc tggtgccact ggtgctggcg ctgacgttgc ctgtcggaca 540
aggaccaccg cctccaatc tgccaccag cgccccact tgatacgtt gataagcctt 600
tgttggggccg ccagccgcgc acccgttgtg ttttcagaac tgtacctcag ggtggtgagg 660
ctgcaagggc caagcagtat atatagggcc cttcggaacc atgggatgtg attagttgaa 720
cagacagcag ctgattgaat a 741

<210> 10

<211> 983

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 10

gaacacttca ggcacccaca cacataccaa ctgagacaca aacacaaaca cagacaaaca 60
 caatcgtatg acctgacagc atagtgcatt atctatgctc cttcccagc tatacacgtc 120
 acaatctcag ctaatagacc taataggggtg acagttgcac tttcccctt gatgtctcat 180
 ccagaccctc gtcattcccat ccgccaate cttcccagtg cattccaacg ccctctcaat 240
 ggcaaggcca acctctgagc cattgacct taaccageca tagtttactc aactggccga 300
 ccgccgtccc tttaccatt gccatacgea atattccaat ggcctagagg ggctgtacgg 360
 cccattgtcc attgtccatt gttactgctg gtatTTTTAT ctcaacgtcc caaaaccgt 420
 tgtggacagc cgcgccgggt gtataggccg ccgatcgcag cctctccgga agcggcgcag 480
 agcaaaaacga gcccagtcgg gagtcaaatt ccgtccttg tatgaattag tccggcaca 540
 agagcccaaa cggggttacc atgagatgcg gaacggcggc aatcagttca gaggctacag 600
 tcgtccccta atcgccatcg gatactgccc attgtcgttg tactttacag ttttacaagc 660
 atagcgataa gcccgaagcc aaccaactcat caacgagctt aacctgttgg gtcgcgtaag 720
 cgagcggggg gatgctgagt caggacaaca tttgggttgg tcagccgcgg cccgtagggt 780
 agtcgtcgtc gaggtcacgg ctgagtaagc ggcaccaatt aatccgtttg ttacatccgt 840
 atgggtgggt gctTTTTTTT acctgtacgg tttggaacaa caaaaattg gcagggaccc 900
 atTTTTTTTt cTTTTccctt ataagcacct cctaggtccc ttttagtagc attcgaattt 960
 ttgacacaca caaaaaaggt acc 983

<210> 11

<211> 930

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 11

gggtgcggca tttagacagc aacaaagact gcgatcggcg atcagcacta ccgttccctg 60
 taaatggat caacaaagag cgttccaatt cgtcgttcc agcgcggta cccgtaattt 120
 tctacgtaaa agtgccecca gcaactgtgc gccatgtct aacctaatca taggctcagt 180
 gggaaccacc agaccacct ccgactgtgt ctgactgtct ctgactctct ctggccccag 240
 aacggctacc gcggagaaaag ggtaatcggg actttgttct gatgggttgc atgtttgttt 300
 tgtcccaatg gggttagtgc ggcaggtacg gcaggtgaca ggatggcacc gtctcacaag 360
 ggaacgcagt ggaagatgag ttttgggggg aattagacag agaatgggc aatttgggtg 420
 actagggagc agtccatgtg tatctagcag tetccattta gtggcctatg tttttctta 480
 tttctTTTTT gtcaaaaagg agcatttacg taaccatcta caaaaaaag aattactaaa 540
 atgacacaaa ccgggggggag ccgggatgcc gctcacaggg tacgcagcgt ttgtgcaatt 600
 caataaccac caacaatagg agaatatatt aacaaagcat acaacagatg tatcccctt 660
 ggctttgtgc atcgcactgt acctttaatg tttgtttga cagtcctcag acgcaaccgg 720
 attgtcccga gtctttgtga tcaaaccgcc tcattgtgca tetatttccc attcgggctt 780

gtttgcttat ttcccaaaag caatcccca gggatatataa aggcgcaacg acccgcaccg 840
 acggggaact gataaactaa gtacagttgt tttcacggtt accggattga ttaatctttt 900
 ttttaactaa aaactactag tacaacaaac 930

<210> 12

<211> 602

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 12

cgaggcagta acctcccgtt gtcgtcagta attggggccg aagccgagag aattgacgac 60
 ggggtgacta ccgaatggag gcaggaacc tgttcgttt tttccatgt atacgagggc 120
 aaaggtcgga ccatcgtaca cactgagaat gaggaaaaga taattgaatg ggaaaaggca 180
 gatactttct gcgttcccag ctgggcaaag tttcggcaca ttgctgaggg aaacgctgac 240
 ggcccagctt atttgtttag tttttcagac aagccattac tagatagctt ggctttttac 300
 cgagcaaata gcgtatagca atacatteta tatttttttc gagttaaagg tactgataag 360
 ataagggatc cgtcacccat tttttgactt gacaccacga ctgggagcgg agagccgcac 420
 aacggttttg tatggggcac agcgaaaggg agggagggaa aaaatgaaa aatgtgagc 480
 cgcattagcc ctaagcagtc acacgcggac ccacgattac tcctctcca tcgcagcacc 540
 atacagagta aggacgattc aaactgtcaa agtgttcgac tgcccaactg aagactcaca 600
 tc 602

<210> 13

<211> 877

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 13

tgcgtcggaa cgggatatgc attcccctag tttcgccga gtgcagaatc aggcggtttc 60
 tttgcaccac accacatacg gaggatgacg ggcattattg atgttgaata gtaacctgat 120
 cgtgactagt atgacggaac ccaacagcaa cagccgaccg tttgtgagcg tttttcggc 180
 cggtcaggcg agtttttccg gctgccaat ggtccttccg taccctttac cctgtacgct 240
 gtacctgcca cggataggcc gtgctccacc tgctcactat ggtgggtgcg gggaaaacaa 300
 caggcaggct caattgctct gcaaatgggt tgaggggggtg attgatgtca ctggtacacc 360
 aacaggggaa tgctcggcgt tgattttggg ccacctttt tgtttgccag agcttgtctc 420
 tattgtcaaa tttaacggtc tgcaactggt gcccaaatg ggacaatgat ccgatgcctg 480
 catagacacc ctgcttgagg gtgcgatcgc cctaatacga ggcaaacca gttttccaat 540
 tgaccttcaa ttgacgagcg gttgttgcca caggggactg gactgctacc tgttttagagt 600
 tcaaatccgt caccagcat tgaaagtttt tccccgatt ggatgattgc aatgccgcta 660
 acccgtcat ccgccaaagt tcatagtecc acctgectc gacttatcgg accacatggg 720
 gctcccttat gcgcgcgat atggcgettg attgcttttt ggtcaacggt tgggacaaat 780
 ttctttgtt aaggcggacc cgccagcaga tacgaagta taaatagggc tcactttcac 840
 catcttgtec attcaattgc aagactcaaa agtaata 877

<210> 14

<211> 524

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 14

```
cccagcccga cttttaacct caatagctag ctacgcaaca gacagttaaa gctacgtact 60
caactatata ttccattgac aattgacaat tacaactggt tcttctcctg catcgttctc 120
atcctcattg gcttatctcc tgttatcaat taattataat aatatagtag ttctgaacta 180
attacgtgat cgcacgcagt acggetgacg cgtattattg gaccaacaaa ccctaaaaat 240
tgtttcatcc aattgaacag ttcacgcaac cgtgattgtg ccaaaaaggc attgccggcc 300
tcaagtaggc gcccattgta cgactactgc ggtctaggcg ctcccgtatc cctcaatcgt 360
ggcccttttc cggcttaccg gctgagtcag ccccgcccaa caaaaaagc acaccacaag 420
ttcgacatgg tccaggggca cggetgcagg gttgcggtat aaatacagtc accatttcca 480
ccgcacctcc gtgctttggt tttcaattgg caacctataa caca 524
```

<210> 15

<211> 668

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 15

```
caggctgcat gtatgcacgg tttttccggc agcaatgctg ccgcctcct tgggagtaac 60
atgaacactc acaccaatgt gtggtccaaa aaactgctga cattagttgc aactccggat 120
ctttttgcca acggtttgcc ccggcaccgc ccatggggcc ccagtatacg gggttaactt 180
acagtcccac tggaaagcctt tgttaccga tggcaatgct gcaccggacc gtgagccgta 240
catgagacgc acgcaaaact ttgtagccga ggcaaggac aacaattggg ttgaatagag 300
ccaggggaga gctcagggtc cctcgggttt tgaattatcc ctgaacctct agcaaaaggt 360
tccaaactaa gggttgcgct taactgtacg ctttttgatt tccggcctgt gaattgttgc 420
cccattaggg actcaaaccg tctaccggca cctcccggcc gagggccgct ctgtgccgaa 480
aaagcaatgt gagattccct tgtggcaggc tgggatttgc tataatTTTT ttttttatgg 540
ttggttgaat atataaagaa gcagaggccc ccaatagaat tgtcatcagt cactttttta 600
cactgaacta acaatcatac attaccatta attgatagac attggagaag gaaaagtact 660
aactctaa 668
```

<210> 16

<211> 1366

<212> DNA

<213> 解脂耶氏酵母(*Yarrowia lipolytica*)

<400> 16

```
tggcagacag tgacgagtca tacattctcc gtataatata gtgtatgtcc agacgatagt 60
cgtactcgtc ctcgttactg taactactgt gcgagtactc gtgcatgtat cgtaggtatt 120
gtatgttcga gtacatacac atacgatacc aaacactgcc cactgttctg tcatgttaga 180
```

tcatggccaa tccacgtgac ttgcatgcag gtttggcatt gaatattcag cgtggctact 240
 acaagtagta catactgtat caatacgatt glacatacgg tactcaccet ttgctacagt 300
 atgtacatac aagggcgcac atggcagaat accatgggag aattggcccc catggagttc 360
 agatgagccc taacaacgcc cctgttcggc ttcagaagca attggctttt ggaaattatt 420
 tggcgagtga acaatggcgt gtatggagcc gtattcgtgc tggtgcttgt tgaatcagcc 480
 cattgcgcga aattgttggc tctcacaact caacggcttc ttttaccctg tcgtgacgag 540
 acgctactgt agcgcttgtc ggtcggacca caccaaaact gggcctgtat tgcattgtac 600
 tcagatgtaa gcaccaagag ctgggatcca cgtgatcgcc cccacacaag acgcgtccat 660
 ctgtctattg ctcatctccc ccggcgtctc ccgatctctt ccgacgaaaa ttagcacatt 720
 tcacacgcat ctcaagtcag tttggaggca gtagggcgag ggtagaggtc tggatagga 780
 aaacgagtg ggaaccttat tatttggttg ggcacatecc aaagacctcc acgtttcgaa 840
 atcagttgtg ttctttttct ttgaacttca cgatatttcg tttattcagg tgagtacca 900
 cgaaacgcag ccaattggtt ccaattgagt cctagggagg tgacacaaac acacagcgac 960
 acagagacgg acacacaggg ctccgtctgt ggtgagagat aactagtaa cactactgg 1020
 ggcgcccatt gccgtgagcg agagtgtcag caaagtagta tatggagcta tgcacaaatg 1080
 ctaaggcaaa ttgggatgca cgggtgtgca atgggataac gcaagttgtg ttgctacgcc 1140
 actggtgact ctcttgtctg gtgatgttgt cttggtgcag tgttggggtt gagctcttgt 1200
 gctgttgggt ccgtctgtgt ggatgatttg acctatttct gtgtcaagtc cacatacaaa 1260
 caggatctat caatgccacg gaccagtcac aacctgcc a cgtccgcctc tgcgacctc 1320
 tacctcctct tcgactgcac atgttccctc attetaacta actcag 1366

<210> 17

<211> 1000

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 17

gggagagcat ggagcagaaa cggttcgtg cttcaagttc gactacaagt gcacagtgat 60
 gttgcaacac agtcaccact gctgagtc ctctggctgc ctatcaggat gtactctcac 120
 acatctcacg ttcggctcac ctctctttg tcaaggcata aagttcttag accattgtga 180
 ctacgcagtt tgctccgaaa agatgcatga tcccaccac ttgcgcctgg aaccggtgga 240
 cgtgtgctac cagcaggacg tgtatggcac gegatttatt gttgagtacc aaaatagtac 300
 ttgtagaacg tattccaatt cacctttggc ttaccggtt ttgtgacctg agctactgta 360
 ctctctgtac ttccaatcta ggctatttcg ccacttatag caaggcaatg gtatcaacgg 420
 tgcgtcgtaa ctgcggcagt atgcggagag caggcgtatc tatgacgtcg gtcggccccg 480
 gggcagtggg gaatcaggtc atgtgtttgt ttctttgcca tttttgctt gtccaatgag 540
 ctgagtcaga agttcactaa gcatcgatct ttatggaaca ccacgggtac ttagctatg 600
 gtgacgtaat tgtagacta cgtagecac cagactacaa gtccatacat ccagacagag 660
 agtgctaaaa agaaaaatg gggcagcata gacatcgga accatgggag atgatagcta 720
 cccaccacac ccaaacagtc agggtagcgt acgtacctgt agtgtgtact taaccagctc 780
 gggtagcagtc tcgtgcaaaa cctccgatcc actcttctg gtcactctac tttatctggc 840

agtaactctg gtccccatac cttgctgtca gactctccgc atttaacctg cacaacceta 900
 attcggcctc acacactctc caaatacaga taaaacacaa aggtttcgtt caccctatac 960
 tccgaatcaa cgctacctac tgcatttctc taccgcaaca 1000

<210> 18

<211> 971

<212> DNA

<213> 解脂耶氏酵母(*Yarrowia lipolytica*)

<400> 18

gaggtagcggc ctcgttcagg ctatacggaa aagttttctt ttgacgtttt tttgagtgga 60
 tttccacgag ccatttaggt catggttgga tatecaggtea tggccttggt acgagaagca 120
 tcttgtcgac tgattagctt ggactcatgt tatctgecte ggtagcaact ccagtgcgaa 180
 caagcacact gtactgctgc tcacatgcgg tttcaaatgc acggggagac gcccagtgcc 240
 aatgtcgcca catttccagg tcgtcgagtt gaccttttcc gcacaattga gtccacattg 300
 tctacttggc cctggtaact cactcgtacc ggtacctttt gatccgatgg ttacttttta 360
 tgttctatatt tacattagcg tggaaataga ccatgccate tttggcacc cggaaaaact 420
 tgatccaata gattgttgg gtggagctag tgactggcgg caattggaga gcttctagaa 480
 gacgaaccag gagcccgata ggaactccgt tggcgtgagt cggcccccca gtagcaattc 540
 gaatcacgtg acgtggagtt ttccctcgcc cgcgttctg gattgtcccg gtgtgacgag 600
 gccgactgga tttgatcacc caaccccaca cgacgcataa tgtaaattga tcatcataca 660
 gtacatgccc gagtctaatt attggctggc tcacggggac cgggagcgtc cagtgggccc 720
 tatggcggtg gttcaaacac ctacagatcag ctactatcg gctgagacaa tccttaactg 780
 ttggtcgggt gccgtttttg cctgctccta acagctcgca ccgactctaa aaaacctata 840
 cgacctggcc ggcgtaactt tgagtgtcgt caaactctga tatatatata gagagacgta 900
 tcccaacagt tgatagtcga caaacgcaaa acagacggac actgaacccc ccgcgcttca 960
 aaacaccgac a 971

<210> 19

<211> 1076

<212> DNA

<213> 解脂耶氏酵母(*Yarrowia lipolytica*)

<400> 19

gttggattta gttagaaatt agttgactgg aaaagtcacc tgggggttca tttctggtgt 60
 tacaagaatg gaagaacatt gagatgtagt ttagtagatg gagaagactt gagttctaaa 120
 caaaagagct gaaatcatac ccttcagtag tagtatagtc ctgttateac agcatcaatt 180
 acccccgtcc aagtaagttg attgggattt ttgtttacag atacagtaat atacttgact 240
 atttctttac aggtgactca gaaagtgcac gttggaaatg agccacagac caagacaaga 300
 tatgacaaaa ttgcaactatt cgatgcagaa ttcgacgggtg tttccattgg tgttatgaca 360
 ttcatctgca ttcatacaaa aaagtcttgg tagtggtact tttgcgttat tacctccgat 420
 atctacgcac cccccaaccc ccttgetaca gtaaagagtg tgagtctact gtacatgctt 480
 actaaaccac ctactgtaca gcgaaacccc tcagcaaaat cacacaatca gctcattaca 540

acacacccaa tgacctcacc acaaattcta tacgcetttt gacgccatta ttacagtagc 600
 ttgcaacgcc gttgtcttag gttccatfff tagtgeteta ttacctact taaccctgat 660
 aggcagatca ggccatggca ctaagtgtag agctagaggt tgatatcgcc acgagtgctc 720
 catcagggtc aggggtgggt tagaaataca gtccgtgcgc actcaaaagg cgtccgggtt 780
 agggcatccg ataatatcgc ctggactcgg cgccatattc tcgacttctg ggcgcgttgt 840
 attcatctcc tccgcttccc aaccttcca cccgtttctc catccaacc aatagaatag 900
 ggtaacctta ttcgggacac tttcgtcata catagtcaga tatacaagca atgtcactct 960
 ccttcgtact cgtacataca acacaactac attcaaaatg gtgagtatt gaggagacaa 1020
 ttggccggcg gtcaattgag cgacacaaaa ccccgcgcgc gcacgcacta acacag 1076

<210> 20

<211> 997

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 20

cccttcgta cctctctgcc ccttctggac aggtcaatga tagactcaga gcgacacaca 60
 tgtctgacgt accatgttag accttgtatt gacctggacg aatgtgtgtg aggagtgagg 120
 caggccaaga cgaaccacgg tctttatata tgcccacgga gtgacacggt ctgtgtcgtc 180
 accgcagctc cactcaccac ccgcatcatg atcgtccaac cagaaccac tccccagttt 240
 cgaccaaca ccattctcaa ctgtaagtat gattaccaca gtgatactcg cccagtgcgg 300
 cactcgtact gtagccactc cactgcaaca tccgtatcgt attgcaccgc cccgattcac 360
 ctgcttcctt caagccttca accacgtact getccacctc ctaccgttga gccactcggg 420
 atcggccaga gtcatgtctt agggtttggc tgcagttgtg gcgtaaacta tggagaaggc 480
 gacggaaacg agagcgcctac cggtagcgac ttggcgacac gtctggctcg ggaagggggc 540
 cgttgcagag accaagactt ccgtcacgtg accgctgttt ggtcaattct aacgcagtta 600
 ttttccgtct gattcgtga tacgagtact cgcttgcgtg agatgactca gaccaagaca 660
 agagaagggg aaataaaaaa aacttccaaa aaaaacttcc aaaaaaaaaa aatcaaaatt 720
 tgacaaacct tttctgcctg ggaccaggaa ctttgtgagt ccattgaggg agtttagccac 780
 ccatcagcca cagccacagt ttggacaaga agtaaaagtg gatataatta tgttatggag 840
 accatgtagt gttgtgggag ggaggggttt ttttgttgtt tttggctgag taatcaacag 900
 caagtggcgt atatcgtata tctatcgtga ctcagactat tcaccgcttg tatggtgcta 960
 tctcgacttg tgcttagtct caggtacacg tgattgc 997

<210> 21

<211> 1691

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 21

tgcaaccagt ctccgtgggt tgcagcatac attgttcccg cctctccttg tcttgttggg 60
 aggccgatgt cgctgactgt atgtaccgtt tttttgtac cgtagtacet gcaggccttg 120
 gtattttcca actacagtac atacaggtct tagagtgtctg attggagata gatatgaatg 180

gagtgtacga gtggaacaa agcgggtag atatgtgtac ttgtacatct gtgatattgg 240
 tagtattgac aagcggtagt catttcagtg catcgccgtg ccctttctac tatccccttg 300
 cgccatcaat ctcccccttc atcaatccac ctctggcagc tcttctagaa gaccttttta 360
 cagtctccca attttatcgt ctagtacgg cagacctgt aagcagatat gtatcatgag 420
 tcacgatagc tggacagacc aatggcatgc gggcaaataa ctcccacaga cgctctccct 480
 ccggcgcaca aagcctcgtg ctctgaacac gccccagttg atttgacagc tctcaacatt 540
 cgtgtgaact tttttagcgg gaaaaagtaa catgacgttg accgtgcggg gctacatgta 600
 gcagctgggt gtgctaacta cggatacatg cctacaacc ccacaagtca agaccattgc 660
 gacgcggaaa caggagcccg caaaaagagga gaaaaacaac ggcgagactc gggggcggag 720
 tgggtcacgt gactttcctt tttcccctca cctggcccgc tccgtccata tctctgtcgt 780
 acaagacaat attgtcgcaa cgcaaaaggt ccataaatta ctgggtagac gcaactctat 840
 ttgaaggcaa cctaccgttt gcttttagtg ttttggtttt gttaccatat ccaaaaaaaaa 900
 accatatact caaaaaattcc getgcaccat ctcttctctt ctccatcaac taccctgcg 960
 gagaaattca caccacagtt acaatgattt acaccgcaa ttcgtcccct tccaccaacc 1020
 tgcagtggct cagtaccctg aacacggatg acattcccac caagaactac cgaaagtctg 1080
 ccatcattgg tactatcggg gagtatactt atccacagac cagacgccga ttgcgcggtt 1140
 tggtgcaaca ttcgacgagc ccacaagagg taggcgtcac aggataacgg acccgctcat 1200
 gtgaacatgt ggcgagccca ttgtaccctg gtcgectgcc cccaagtcca ttcgccgaat 1260
 gcgctcaaac gctggctcgg tctccgcctc aggcctcagt aaaaacggca aactaacagc 1320
 aggtcccaac acaaaactccg ccgagatgat ctccaagctt cgacaaggtg agtaaccata 1380
 atgcgacagc agtgtgcgcc gtgaccgat tcgcggtggc cacgtctatc tgctccttct 1440
 ttcgttacc caattggcac cgtgcctta ttttttgget ttggtttccc gggtttgtcc 1500
 aatacacggc tcatgcgcat gcacattttt tccggtcgga taaaccaac gaactctaag 1560
 tgacaaacat gaaatgaaaa cgacgcaagt ggtaaggcg ctaatggtga cgttcatgac 1620
 gttgccagtc tgggtgccctt atcgatgacg tatggacca tgtgtctatc atgccgcaat 1680
 actaaccaca g 1691

<210> 22

<211> 996

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 22

tccagactac ttgccacaaa tgcagcgagc tgcacattga tgcgttcatg caagctacaa 60
 gtacgagtaa tttgacgtat tgggcaette aaggcagtct ttcgaaatgg ccaatctggg 120
 agctcgtca ccctccgaga taactgttgg gcacaccagc aggtctcagc aacggttgaa 180
 aatgggctct cagttcaact aatgatccaa gaaaatacaa gtacgatggt gtgattggtc 240
 ggactacttg tagacgacac tagccaaagc gaaaaggcac ccaccctatc tgaatgctga 300
 gctgtgttca gcccactc ggaatgetga actgttgtaa gtcgatagcc gatagatata 360
 tatcgtagca aacacaagtt gttgactcaa acgcattgac aaggaagtac agatccgaga 420
 aattgtgccg tgcaactgc tcccaggcac ggtctcaatt ggggtatata ctctgtatag 480

agtaagccag gttggccccc caccacagag aaatgcacca accagtcggc gagctcaaca 540
 gccgtatggg agcctctcgt ttgatgatg tgtgacagga ggtgtatatt ggggctactg 600
 ggtgaaaata aaaacgcgag agagaatata ggggtttcag cgaaatccca gtggagagac 660
 cgaatcatag tattataact atgacagtat cgtgcgctct cctctttcat cacttctctc 720
 caatatgcgt accatctatc accactctct tgcttagcct ccctccctcc ctctctctct 780
 gttagacccc cacacgctca acagtactca atatccgcgc agaaaaataa ggttggtggg 840
 acatttctcc ggtgtgagcg attagtgggt tgtggtggtg cccatagcag acctaaaatt 900
 gactctctc acttgacaac acacagatag aacctgcaac ttcaccaag tggcgctttt 960
 ctcaatccag tccgtgtgaa aacaagacaa ttgacc 996

<210> 23

<211> 985

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 23

gaccaaccta aattagtcg ggtggacgtg tactagaac gttgtaatac caaggtagtt 60
 gcgcttgttt tgaccaaaaa tgtgtgacaa aacattgcca gtgtatccag tctgggaatt 120
 gagtcgttct ataaccgtca tttccaactcc acttccgcgc aacgcgctgc tgagcactcg 180
 gaaaattagc tcgaaaagtt tttccgggtt atgtgaccg ccagcaggtt aggctctatt 240
 ctgttgggaa ataactctca accgcccctc cgagctaaat ctctcaactac aagactcttc 300
 atcgacaaac gatatctgag atctttttcg gtcccacagc aacaagccac aaacatgtct 360
 gccatcaagc aattgaaccg tctggccgcc accgccaaga cctctgttct caagcccgcc 420
 tccaagcaga ttctgctgcc cactgctggc cagcaggetg ccatccgaat gatgggccag 480
 acccgagctg cctcaaccga aggcgccgcc actaacgtga gtatTTTTTg tgtgaacgac 540
 acgatatata cacgacggcc gtgcgcttgc ggettcgca tgcgccctga atggggcaac 600
 tcgagcgttg tgtaacgggt gttcatcaac agcaaacagt gcttttcgga cttagacat 660
 ggcagaagaa gcaaacacgg ttatagcgag agagatcaca atggagtgac gagctttcag 720
 tgatatttgc caccagtcaa attttcagca actcctgaaa cgcaccatt ttatcatcat 780
 tgtgacgcgg acattcagcc tcatgctgta actgcactcc gtgtctggag ctgccggagt 840
 tatcaaagct gtaggtgcca attgtgaaat agcgattcgg cgattcagcc gttgattgag 900
 ttctgctat gtcgcaattg tacaatgctc tgtacactc caacaacatc aacatgacaa 960
 cacccaaaac aacgtactaa cctag 985

<210> 24

<211> 999

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 24

ggtgagtcgt gtcgctaaaa ggtttgcaat gggctcccc aaaggctttt ggtggtttgt 60
 agggcggtga aaaaattgtc cattttaggg ccaagattta gacgtgtcga gatggggagg 120
 ttttgaaca cgccgaatcg catcgacacg actcccctcc gctgaacca caacctgcc 180

ggtcacatga catggctcct gcacttcgga tacggaagcc eggatccttt atgctctacc 240
 cccgagttgt acctgtccaa tagaacaaga gtcaattggc ettactcgca tgcaactcaa 300
 acttgggccc gggttgagag gtacagttga caacgtgaaa ataagagggg ggggaggtta 360
 aggcctcagg ggcgaatttg agagcactta tatagacaaa tccgcaccga agtgacaaca 420
 tggacaatgt gacacgtaga tacacgccgg atccagctgt ccacacacat ttatcccga 480
 aaatagcccc catcacatgc acgtctcgta aaaaaaaaaag agctgcgggc caaaggacca 540
 ataagtgccg aggaatgtta agccaaaaga acaacgacga tcgccagaca gttttagtgg 600
 gagcagcagc agcagaggcc gtgcaacggc aggagagaga ggtctggcga aaaggaggag 660
 acgggggtgtt aattgatttg cggattttcc gccagccac aaaaatggcc tattttggcg 720
 ggtttaacgg cgccccctcc aattaatecg aacccccgtt accacgcagc ctacactatg 780
 tactgttgac aacaccccat gacggtagtc tccggagccg agccggactt gtgtttaaaa 840
 tcggcacgat tttgttcaga ggttaggggt caccctggct aatagattgg cgctgattgg 900
 cccgacacaaa cccaaaatgg gcaactctgca gtgtttataa aacctctccg aggcccacga 960
 ttcaactttc tcctttccgc tctaacacca catatcaca 999

<210> 25

<211> 999

<212> DNA

<213> 解脂耶氏酵母(*Yarrowia lipolytica*)

<400> 25

atggtgcgtg gaggctttgg catcctttct acttgtagt gctatagtac ttgcagtcca 60
 agcaaacatg agtatgtgct tgtatgtact gaaacccgct tacgtaata ttttagagtg 120
 tggaaactatg ggatgagtgc tcattcgata ctatgttgte acccgatttg ccgtttgcga 180
 ggtaagacac attcggtggt tcaggcggct acttgtatgt agcatccacg ttcatgtttt 240
 gtggatcaga ttaatggat ggatatgcac ggggcgtttc cccgtaacg tgtaggcagt 300
 ccagtgcac ccagacagct gagctctcta tagccgtgcg tgtgcggctca tatcacgcta 360
 cacttagcta cagaataaag ctcggtagcg ccaacagcgt tgacaaatag ctcaaggcg 420
 tggagcacag ggtttaggag gtttaaatgg gcgagaaggc gcgtagatgt agtcttcctc 480
 ggtcccatcg gtaatcacgt gtgtgccgat ttgcaagacg aaaagccacg agaataaacc 540
 gggagagggg atggaagtcc ccgaacagca accagcctt gccctcgtgg acataacctt 600
 tcaactgcca gaactctaag cgtcaccacg gtatacaagc gcacgtagaa gattgtggaa 660
 gtcgtgttgg agactgttga tttgggcggg ggaggggggt atttgagagc aagtttgaga 720
 tttgtgccat tgagggggag gttattgtgg ccatgcagtc ggatttgccg tcacgggacc 780
 gcaacatgct tttcattgca gtcctcaac tatecatete acctcecca atggctttta 840
 actttcgaat gacgaaagca ccccccttg tacagatgac tatttgggac caatccaata 900
 gcgcaattgg gtttgcacga tgtataaaaag gagcaatccc ccactagtta taaagtcaca 960
 agtatctcag tataaccgct taaccacaca tttateacc 999

<210> 26

<211> 999

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 26

tgaccaacct tgtttggtag atggggggga agcgaaccgg caatattcca caatgtgctg 60
gcatttactt gtgctggcaa aagaggcaca aagaatactt gtagtcggag cactcactg 120
tcccacaaat agctccccgc tgtcaatctc tctgcaccg cctgctcaca tggatgctaa 180
gccgcactag gtcgcatata tggctctgca ctaaaaatta ggggtcaacc acagtgcggt 240
atTTTTtagat tgcaccaag cagcgagtaa gcaaaaatac gcctaccggg gtccgatatt 300
attcaggagg tgccattaga ggagggcaga tgagagtcgg atatcggaga tattaccgag 360
gctataatta ccccatccac gcctttcacc cctcccactc tctccctcac cgcacacca 420
cccaccactt tcaaaaatata ccgcaacatt gacataatct ccggtacagt ggtagcacc 480
gagaggacc caaaaagctt gggggagata gaggtaggt tttttttgtc agtcaaatcg 540
tatatgccaa tacacacaca cacacacaca cacacacaca gtttctgaca taacagtata 600
ttggaagggg gtgtgcttgg caaagacagg agaagacggt gctgttagag ggcaatccag 660
acgggctaga gctctgtaac tttcggatcg atttcaatc ctctagaata ccaaatacca 720
gtggttaagc ggctcattta ccagtcetaa taccctctc accagccacc tcccctatt 780
cctcggcagt gcttttttac ctttgagatg tggccttgtc tccgttactt cccaaccgtg 840
agtgtgtgt ggtgtgctgg acagtgcgac ataactaacc ctaaccaga cgagccagcg 900
caccacaatt ttgtgttgc caactctac ttttctctc tctccatcg gtatttcatc 960
gacaaatctc tttgctacca acaaccacac aaattaaaa 999

<210> 27

<211> 452

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 27

tgtgtgtttg gtcgaggttt tttctgttc agtacagacc ttgtgtggtg aggaacagca 60
atagcaaggg tggcttttga ttgggtgcag gtgcccttac cctgttggga gttttgtcta 120
ggtgcctggg atggaggaca atgttttgtc actgtcaaga cgggatattg tggggatttg 180
agaaatataat ttgatcagcc ggtctcgaag attatattcg cgctttcgcc tttgaaattg 240
ctccttttgt tgccgtttcg aactgtagtc tegtgetact gagtctcatg ttaatttttg 300
tttcggcctc gacttaatta actctaacca atgttatttt cgtgcattaa cgaaactcga 360
acgcacgatac agtcacactc tccaccatca aatateacgt acagactggt acccataca 420
tactaccatc tgaagacgac acaccacca tc 452

<210> 28

<211> 1000

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 28

ctcgaataag gcactattta ggaccagacc acaccccgcg gatgtcaagc cgaaccttgt 60
tgcataaaga taatactagt caagtggggt gtcgaccga tgagagaata aaccgattgc 120

aacggttttt atttcattcg cttcttccag cagacactct tggtttttett cctcacagct 180
 ttccgccatt atcagctgcg tgtatcgtga gtatattggg agtgagagat gccctcacga 240
 taagacaaca gctatagtac aaatgttaac acagatgtca gatcaagcgc cgccaaactc 300
 gcccgggaaca cgggtaccag gggagatcgg tccccaacaa tcttcccagc aagttcccat 360
 ggcttatacc atcccaggaa caacaagtac agctctagat gaggagatct cggagtaccg 420
 tgacaccaac cgaacgacca agaccctgg gatagacgag ctgacacca cagcgtttta 480
 tgacaatcgt ggtgttaggc atgagcatag gggaatatct gaagagatga agaaggagct 540
 caagagacag gagagacgcc agcacgagat gttgcaacag aagcagcttg agctgagaca 600
 acaggaagcc ctacaccaac accaaatgct tatcattgag cagcagaagc aagatcagat 660
 tattcaacag cagaaacagc tgcaacaact gcagcggcag caacaggaag aggtggctag 720
 acaacagcag ctgcagcaac agcagcaact gtaccagtac tatcagcaac agcaacagca 780
 gcaacagcag tatgccgcac acatgttaca attcgagcaa caaaggcggg agcagatgcg 840
 acaacttcag ttggcccagt accagcate tcaggetggt cagacacatc atcaagatgt 900
 ttctcateta accccgtctg ttcccgtacc tgcagccgta acacagctc ctgcctccgt 960
 agcacgtacg gcatcagtct cagacatggt ggtacctct 1000

<210> 29

<211> 1000

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 29

gtagatacac ggtaagtaca tactatatct atagatgata catttttttt ttataccgac 60
 cgcccaagcc acacggcacc ttaattaaac ggccactttg acatgagacc gagctacaaa 120
 ccagtcgact acaagtactg tcaaagagtc gaaatttgtg gagtcgggag ttataatgt 180
 ccatccaaga acaccctcat ttctgctcg tcttgtgttt cagtagctaa tttcacatgt 240
 aaaacggcgg tcttgatcca cctgtctta actccggtcg gactttgctg ccataacgtt 300
 cggacacgca actctttccc aaatccaact tacagcatct tacctaataca cacctgccct 360
 cacattagc accaacctaa acccaagctc aaccgtcgtc gactcagccc cgaagaagca 420
 ggtactcgtg caaatatata acgaacagtt taacggcggc ccgaaaaag attcggctcgt 480
 cacgtgacct acctccacc taagccggtc cttcaccct cacttttct cactgttctc 540
 acttttctca ccccactgt ggtctataca aactctacga tgacacacaa tggcagaaaa 600
 gtgcctctgc atacacgate caataaaacg gtcagtacac gcaacttagt gagggggagg 660
 ggttacatcc agcaggtggt gctaatgtta cggcagcttt tcagtagtgt gctcgatatt 720
 tcagcccccg ttggaccgcg aaaagcaetc tacaactcgtc ttctagtatg ttcggctcgtg 780
 tcccacgcac ttagttgca taagcgetaa tcatgtttt ctttgtctgt gcggtggcga 840
 ttcggaacat aatcaactgta agcggcgcac gttgaacctt attttgcctt tgagcccaca 900
 catggataac acctcatata taacctgtcc cctccgetaa ctctcttget tctctacaac 960
 ataacctgtt gaaccacaaa acacctaatc aacaacaac 1000

<210> 30

<211> 989

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 30

```

gtcttagtgg gactggaagg agtatcagtc tcaactggta actgtactgg ctagaccccc 60
gaaagggatg gctgtgtgct tgtggttcat tgggtgcggt gtggtgtcta caactcgtgt 120
tgccagacct ggacaagggc atttgtgaat gtgacggtac tcgtaggttc accagagatg 180
gtgtcgaacg acacatgatg agagtggaag ctctttggat gccatcgaca tcacgtgaac 240
ctgtctgata gtccatcgct ggttttagg acgcgtttga aggttccgac ttgacgttgt 300
tggtatgatg cacgagtaca ggcgattgta aggtggtcga gcgtgtttta atgtacaggt 360
ggaagtaatt gtacttgtat cagggeectt tgcagctcgt cttgtgttgt tcgcatcaaa 420
tgacactcgt cttgtacagt acagtctcca tgaactgctc cagattatgt atccaaaaca 480
ggggttgat acttgcagag tacaagcaca ggcataccta tgtacaagcc tctttatata 540
tttaagagta caagtaaacc tactcgeact tgaacttgea cgggcgagat gtatggctgc 600
agaaaacctg tcggcagccc tccgtctctc acatacgaac atgactgact tgcactcttc 660
acctgttcag caagtttcat actgcaactg tccaaatagg taaatcact tggectccta 720
tttgggacag ggtaagggcg tccagaagag gacaccagtg aaattacata atacaagctg 780
cagtacttgt ccgatacgac ctgtctcgaa acagccgttt ggagcagcgc atttcttgc 840
caatcaattc cctgactact ctcaactctc cccaacacgg tgettttctc ccattctggt 900
cacatgactg acacgtcca cctaacctta tetccaaaga ccacgacata cgcactctct 960
cttcagagga gtttcggaag tctagccca 989

```

<210> 31

<211> 940

<212> DNA

<213> 解脂耶氏酵母 (*Yarrowia lipolytica*)

<400> 31

```

aaggcgagcg aacggctttg tccagtggtc aattttcaag tcaatttttg gctaaaaaaaa 60
agaccaaatt gcagccatcc aaactggtea ctactcgacc aatatggccg atattttcaat 120
ccacatcgaa ccagtaaate agaatgaacc accagatcaa tgaagaacaa caaaatcaaa 180
cgaaaaactc cctcgggccc gcatgctccc gccaaatcga caaatctct tctccatag 240
gcgacattga ccccatgcaa tatecgtgac atttgtaaat aagatctgaa ctttaatta 300
tcatactttg gtggtgtatg gtgcgtggtc cacgtggggt aggggaataa aaaaattgga 360
acaaattggg aaatatataa aaattgaaa ataaatggaa aataaaaaaa acgtggatct 420
ttc gatgaat aaaaatcagg ctaatcccag acaagatcg ggagtcttct tcctgagcc 480
aacgtcatcc tgactaatga aaacatcaaa taaaataaat ctgacaccta aactaacca 540
ctttatttgg gccaatgaga cggetgaaag tccgcacgtt gtggggggaa atggacaaag 600
tttattttaa acgtgaaaag ttggggggaa aaaacaaaa aatacgaaaa tgtagccctg 660
atcggtcaca gcccaattat cccctcgaaa aaaatcccct ccaaatcccc atttttctac 720
cgccatttct gtccatactt ttcgataacc ctaaaaaagg tcactctatca gtctaaatct 780
tgtattaacc tcgaagacta accgtaactt agactaatgc taacgttaaa atacaactct 840

```

aaatattaac cgacatcaaa ccccgaaaag aatatataat cgtgaggcca tcctgaggat 900
 tttgtctcca tcgaattcga ccaccacaaa ctctctaca 940

<210> 32

<211> 709

<212> DNA

<213> 解脂耶氏酵母(*Yarrowia lipolytica*)

<400> 32

tggcagacag tgacgagtca tacattctcc gtataatatac gtgtatgtcc agacgatagt 60
 cgtactcgtta ctcgttactg taactactgt gcgagtactc gtgcatgtat cgtaggtatt 120
 gtatgttcga gtacatacac atacgatacc aaacactgcc cactgttctg tcatgttaga 180
 tcatggccaa tccacgtgac ttgcatgcag gtttggcatt gaatattcag cgtggctact 240
 acaagtagta catactgtat caatacgatt gtacatacgg tactcaccct ttgctacagt 300
 atgtacatac aagggcgcac atggcagaat accatgggag aattggcccc catggagttc 360
 agatgagccc taacaacgcc cctgttcgge ttcagaagca attggctttt ggaaattatt 420
 tggcgagtga acaatggcgt gtatggagcc gtattcgtgc tgggtcettgt tgaatcagcc 480
 cattgcgcga aattgttggc tctcacaact caacggctctc ttttaccctg tcgtgacgag 540
 acgctactgt agcgcttgc ggtcggacca caccaaaact gggcctgtat tgcattgtac 600
 tcagatgtaa gcaccaagag ctgggatcca cgtgatcgcc cccacacaag acgcgtccat 660
 ctgtctattg ctcatctccc ccggcgetct ccgatctett ccgacgaaa 709

<210> 33

<211> 997

<212> DNA

<213> 解脂耶氏酵母(*Yarrowia lipolytica*)

<400> 33

gttggattta gttagaaatt agttgactgg aaaagtcacc tgggggttca tttctggtgt 60
 tacaagaatg gaagaacatt gagatgtagt ttagtagatg gagaagactt gagttctaaa 120
 caaaagagct gaaatcatat ccttcagtag tagtatagtc ctgttatcac agcatcaatt 180
 acccccgtcc aagtaagttg attgggattt ttgtttacag atacagtaat atacttgact 240
 atttctttac aggtgactca gaaagtgcatt gttggaaatg agccacagac caagacaaga 300
 tatgacaaaa ttgcactatt cgatgcagaa ttcgacgggtg tttccattgg tgttatgaca 360
 ttcatctgca ttcatacaaa aaagtcttgg tagtggtaact tttgcgttat tacctccgat 420
 atctacgcac cccccaaccc ccttcttaca gtaaagagtg tgagtctact gtacatgctt 480
 actaaaccac ctactgtaca gcgaaacccc tcagcaaaat cacacaatca gctcattaca 540
 acacacccaa tgacctcacc acaaatteta tacgcctttt gacgccatta ttacagtagc 600
 ttgcaacgcc gttgtcttag gttccatttt tagtgeteta ttacctact taaccctgat 660
 aggcatatca ggccatggca ctaagtgtag agctagaggt tgatategcc acgagtgetc 720
 catcagggtt aggggtgggt tagaaataca gtccgtgcgc actcaaaagg cgtccgggtt 780
 agggcatccg ataataatgc ctggactcgg cgccatattc tcgacttctg ggcgcgttgt 840
 attcatctcc tccgcttccc aacacttcca cccgtttctc catcccaacc aatagaatag 900

ggtaacctta ttcgggacac tttcgtcata catagtcaga tatacaagca atgtcactct 960
ccttcgtact cgtacataca acacaactac attcaaa 997

<210> 34

<211> 983

<212> DNA

<213> 解脂耶氏酵母(*Yarrowia lipolytica*)

<400> 34

tgcaaccagt ctccgtggtg tgcagcatac attgttcccg cctctccttg tcttgttgga 60
aggccgatgt cgctgactgt atgtaccgtt tttttgtac cgtagtacat gcagggcttg 120
gtattttcca actacagtac atacaggtct tagagtgctg attggagata gatatgaatg 180
gagtgtacga gtggaaacaa agcgggtagg atatgtgtac ttgtacatct gtgatattgg 240
tagtattgac aagecgtagt catttcagtg catcgecgtg cctttctac tatecccttg 300
cgccatcaat ctcccccttc atcaatccac ctctggcage tcttctagaa gaccttttta 360
cagtctccca attttatcgt ctagtgacgg cagacctgtg aagcagatat gtatcatgag 420
tcacgatagc tggacagacc aatggcatgc gggcaataa ctcccacaga cgctctccct 480
ccggcgcaca aagcctcgtg ctctgaacac gccccagttg atttgacagc tctcaacatt 540
cgtgtgaact tttttagcgg gaaaaagtaa catgacgttg accgtgcggg gctacatgta 600
gcagctgggt gtgctaacta cggatacatg cctacaacce ccacaagtca agaccattgc 660
gacgcggaaa caggagcccg caaaaagagga gaaaaacaac ggcgagactc gggggcggag 720
tgggtcacgt gactttcctt tttcccctca cctggcccgc tccgtccata tctctgtcgt 780
acaagacaat attgtcgaac cgcaaaaggt ccataaatta ctgggtagac gcaactctat 840
ttgaaggcaa cctaccgttt gcttttagtg ttttggtttt gttaccatat ccaaaaaaaaa 900
accatatatc caaaaattcc gctgcacat ctcttctctt ctccatcaac taccctgcg 960
gagaaattca caccacagtt aca 983

<210> 35

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 35

tgcaacctca ggctcagggt cccctgtctc actgtctctat ccaccatcca ctgttccacc 60
ccctcttaga cctcagccag acgccgcagc gggcaagcag cccgggttta cagagcgctg 120
cgggcatcgg catgatgcga cagggcctcg atgagcgggg atactggacc agaccacgga 180
ataaatcctt cggaaaagtg cgttttttga aattggccga cccggcgaat caggccaggt 240
caaatcccgc ccccgttcc ccacaattga ccgatcctga acatgcacaa tctatgacaa 300
tggtcgcgat caaatcgcct tgcaatagca cttagcggtc gaggtgtcta accctgtcga 360
ggtttgtgac cgtaacttc ttgcaagagc gaaggatgca aggcgctcct tectgaatag 420
gcaattgagc cccatgtcgt gaggettaaa gcgtgcttct tgccgaatcc gaaacaacg 480
ccgccgatga tatgacaaaa gccacaacaaa taccgctgg agcgataacg taagggttg 540
gggtatcaac ggacgcggca aacaagcctg tgaacccttt gcgagccatg gtttggcctt 600

agtttttgtc tcccgctatg gttacattgg ctctcgcgatg ctatggtacc tcattctcacc 660
 gaaaattttt caagagggcg ataatggctg tctcgggcaa cggtttgcac acggctacgt 720
 cggttctcgg cctatgattg gctctggctt tatctctatc cgcccacaca tacttcaaaa 780
 ggaaattgag actatgcaaa aagcaattct ggggtgctgga gtgctgtatg acgattccat 840
 aagattttgc cgggtcgtat cgaataaaaa ccctctttt cccccattg tcaccagatt 900
 cctgttgtgt ttttttaata atctcctttt caaccgctt gttggtggtt tgaaaatata 960
 cccatttttt ctaatttaat ttgctctttg ttagcgtaaa 1000

<210> 36

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 36

caacttgtgt agtagacaaa gtgtaaaaga aagcaatttg cgacttttagc gctgctctgg 60
 cacgtgtata cccggtcaga gtgatgcaat tgagtgagec tggcatggag attatgaccg 120
 ggcccatcgg attccgagtt ttttgatccc ggetccaact tcattgctca tcgacccta 180
 ctgtattgaa ctgacgacca acagggccag tttctccaac caaacagtg cagtctaatt 240
 agtttghtaat tggcaacttt agccttagtc tctctgaaga gttctacccc aattccccct 300
 ggaccacccc agaacccatg ttgaccagga tagcgcgcga tgcaggggcc acgtgaagca 360
 gcgcgataag attgataaatt gataatggtg cgggtgcatgg ccagaggcag agcgacggtg 420
 ctgaacacac aactggcgca acattggtgt atatgactgc cggggcactg tatccgtgtt 480
 gacacggtgt gctcacggtt gctagcaaaag ttagggttta atcggtattt aatggtaggt 540
 gttgagttgg ttgagttggg atgagcctca ggatcgccgc acagggtat acgctcacac 600
 gagcaacgcg acaaatgacg taacctgag ggtaaatatg agctctgtgg acgctcgttc 660
 ttgttgcaaa cgttctgaga gaacactcac ggtgtagcga tcgaagcgcg cgtgggttgt 720
 tatacctgtg tccagcgtc ctggcagtc acttttgata tcagtgtgtt ccgtgcccc 780
 gcttcttacc tgagccgcac cgcttatccc gacacaagaa aactataaag aaggctggac 840
 ccccagattg ctcatcatct tgccacagga actctgagat acctgtggat atacagcttt 900
 ctcaggctca gactgcgcgt tttctgtttt attttccctt ttagatcga ctggattgat 960
 tcctagttga tttcattttt attccgtttg tctgaacaca 1000

<210> 37

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 37

agctaggtea agcgacgcct gttagcgata acgacctga aatatctacg cgtgggcccgt 60
 gtgtcgtaac tgtacagtga cgttacgacc agacaatagt ggtggagggg tagccagtgg 120
 gaatggagct tgagcgagag aaaaatgaca tcaccgaaaa aaaggcgggt agggttttgt 180
 tactgggggag acgcgcgtgc gccccgtggt gtgcggcgtg gggctcggca gtgccgacc 240
 atttcacca tggaaatcgtc tagacaggca aaatggcgtg agcgctgcc ggagatacta 300

aagtttgcag cgaaagaagg agaacaaacg cacgaaccaa atcagagcca aattggccag 360
 gtggcaaagc caacgggcaa gtccacgggc aattgcattg ccettgcccc tctttggccc 420
 atactcggac atggtcggga tagaattgtg aagaacgata agctttagtt aaaactgagt 480
 cattccctca tcggctaacg tgatggagge acgtgattct ccggggggtt ttcgctcggg 540
 caggctcggc cgaccgtcgg acggcacggc gcggttaattg tccggcccc ttgtgagtgt 600
 cacctaccct gcagggccca ggcaattagt caatcccag gacagatgga cgagaggta 660
 ggcggtatgt tgagaggatg ttggccattg tgtagaatat aaaggagact aaaaaattgc 720
 gagaatTTTT ccgagtagaa ccatgtaact tttgtctgtc caaatcggta catttccgtg 780
 tctttgtttg gaaaagctgt ctctcttcc ctccctaagc ccgaatctgg ggtgcagacg 840
 ataacccag accacgagge tgcctcggcc ctccgateat tgacagaaca agaatgaatc 900
 acctgaaaat ttggtctata taaagggcc cateccctct ccatgttcca tcattaatca 960
 accaattggt ttttaagtta ttgacattat aaaaacaaaa 1000

<210> 38

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 38

gccgcgggtg tattttcaat ccaataattc acagttctga gcgtttgtaa tagcatctcc 60
 cgataacttc aggcattcatg ccacagatca gcaacccgag tacacacacg tgaccagtag 120
 gcacgtgaca tccccatt tcggcatttg cgatcgttca tgtgccagca tatgaccaca 180
 gagcttgtga tagtttagct ccatcaggtg attttattag aattatcaac ctctggagtg 240
 gtcagagatg gcaccagggg cacccgaagt gtagtgggtc gtgcagacat ccaatgtccg 300
 aagggttat tgacccttct gccatagtgt gcaagtagag ccgacgagat cggtcagca 360
 ccgctttgtc aattaatTTT tcccttgta aaaaggetgc ttgccattgt ctcgacaaat 420
 cgactgaaaa gtggcccgat ttggatctcg acaatcattt gcaatcattt ggagaggcca 480
 cagttgtctg cgggtggcatt gtcattgtccc cctgttgcta tgtgtgccag tgactcgctc 540
 cgcttgcatt ttagttcccc attcataccc cgtaacccc gggcgtttcc ccagatttcc 600
 tcggcaccgc tcaccgaagc ccttaacccc ccgagtgcgg aaaagtcggt attctcggaa 660
 ggcatataga gaattatgaa ataaaaagag gacaataaag cacgccgat acagagcgag 720
 cggtagccaa cctctaccg tcttgtecca ttctctagea teatttctcc gtccgtacct 780
 tcaccaatc ctacctcccg gacttgtctt acgcggttcc catcggcgag cgcagccgca 840
 cactttcacg agccgaggtc cacccccctt cttctttttt gggaccacac acttccccca 900
 cattgcacat ataaagctcc cgaatcagcc atcatacgaac ttctcacaag agcctttggc 960
 cgtttctatt ttatcacaaa accttcgata atataacaca 1000

<210> 39

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 39

attgggtgtg gacaaagctg ctagccccga gcccgaggag gatgaacagg aggattctga 60
 caagcgtgag tatcccatga tggagaccct ccctcaccct cgattcaatg ctgctacatg 120
 cgtagttgat gacactctat ttatctttgg aggcacctat gaggatggcg agcgggagat 180
 ttatctcaat tccatgtatg cagttgatct aggccgtctg gatggtgtta ggggtttctg 240
 ggaggatcta cgggagctgg agcaggccgg ctacagacgat gaagacgatg acgacgatga 300
 agatgatgac gaagaggacg atgatggtga agatggagag gatcacgacg aggatcagga 360
 tcaagtcgaa gccgaggacg aaaaggacaa tcaagaagag gaggaggaag ctgaaaagag 420
 cgacatgacc attccagatc ctcgacctg gctgectcat cccaagccat tcgaatcgct 480
 ccgagcattc taccagcga cgggacctca attcctggaa tgggccctgt ccaaccatcg 540
 ggacgctaga ggaaaggact tgaagcgaat tgcattttaa ctgagcgaag accgatggtg 600
 ggagcgcaga gaggaggctt gtatctccga ggaccagttt gaagagatgg gcggagtcgg 660
 tgaggtcatt gaaaaggacg ctctagaaa agcacgacga taaatagact aatccatcta 720
 tcggtatcag gctatgaaac tatcaatctg tcaaaatctg tcaacatata agctactaat 780
 cctacgaagc ctacactacc aatcetaate ctatcaatcc tatcagccta tcaagctata 840
 aactaccaac ccatcaacct accatcetaa caaacctate aacctataa cctatcaacc 900
 tatcaaccta tcaatcctat caacctgtca acctaccaac ccaccagcct ataaacctg 960
 tatgtgttgc tccgcaatcc ccggtggccc gcagattaat 1000

<210> 40

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 40

taagtcttgt atctgttacg acgctcccag tctccgccct tgtcgtatgag cagtttgacc 60
 gcctccagct cctgggccac aaacacctg tcgtcgaaaa agaagccaat acggatgatg 120
 gttagcgaaa tgtcaatctt gactccggac gagcccgtga gctcaaacgc ctttcgcagc 180
 gtcgaaatgg cctgctcgcg gtcgccaatc ttagcgtagt actctccgag cttgaccag 240
 ttttctacaa tcgcaagctc ttctctctcc tctctgect cggcaatctt cttctgcagc 300
 tctctacct gctgctggtt ctctttctta agctctctgt acaacgactc gtcccactcc 360
 agcactcctg gcagaccctc ggtgtgaagg tactgatata atggggccag cttttccttt 420
 ttgatttctg tcatgagcgt ctttttagcc tggtcatgct gcggtttgag aaacggcgtc 480
 ttcagcacia agatgcattg cgccagatta taatcgggca ctcggtcaat tggagtgcgc 540
 gctccttctg tcaacccat cctacctgtc tatttactec agcagtgtgt gttagtggca 600
 actgggaagt gtcgctggtt ttggtgtcga tgggtgcagc gtgccgtatg agccaccact 660
 agccacaatc tcccgccggt gtggcggtgc tcgctctatt tatacagcaa atgtgcaaca 720
 caactgtagt tttgtttaa ttctgccaat tgcacaacia attcacagaa aaattcacia 780
 gaatgttcta ctaacgtage agtacccttg gccaaagtaa cgtatcgate gategcaate 840
 ctgatctcaa tcggtcccaa ttctggatcc cttttacct agtctctctc cctgctggtc 900
 ccctactacc agcgtaaaca aggcggaaga ccctgcgttc ctctgcggtg gagcaaacct 960
 ctctctgtca ctttcacttt tttcactagc agcttgtaca 1000

<210> 41

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 41

gtctgagttt ggtcagattt tcaaaaacc atcaaaggag ttcttccaga aggcagagct 60
 tcgagctgcc agagcgacat ggccaagat gtcccacatt cacaaccgtg tggccatcga 120
 gttggcttta gtaaaggcaa ttcacaagct tcgtgccctg attgtatctc agagcgtcca 180
 tgagcctggc agttctctac aagtacatgc tgctaatgac gaaggcacc tagcacctat 240
 tcgccgtcgc cattcttcga ccaagcttca ccatagacga caacggctccg atggaatggc 300
 cgtgaaatac ttggctcgca gacattcgtc acagtacttt ggcaactgagg gccctggctc 360
 cgctgcgcta tctcgtaaaa agagtccggc cgggcttacc caggctcata ctctacgcc 420
 ttcactgacc aacagcgta gtgtaggggg cagtccaagg caccgtcgtc tcaactactag 480
 ctctagacag tctcaggag accatttggg aatgttctct caaatcctc cgctagaacg 540
 tatctctacc ggctgaccgc aacggtcttc attcatggca attagacagc tttaaattat 600
 ttagaactac aaactaccaa tgcattgctt acgacctta cgacctctac gaccgttaac 660
 aaccgtaaca accttggcgc taattatcac agtctatcac agtctattac agtccatcac 720
 agttcatgct gtattcatct ataaccttcc atgacttccc tcgtccctgt cgaaggccat 780
 cgaacttgcc cgtagttatt aatttggcgc tcatcatcaa gctgcatgac ccccgaccgc 840
 gcacgccccg gccgaccaac catcaaaggc gataagaatg agtcaaaaag gactaaatat 900
 tccgatcac gtaatcggcg cagtataaaa ctgagctcat ccgcatattt ctaggcactg 960
 aaaattccaa agacttttcc aactctaate aaaaacaaaa 1000

<210> 42

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 42

tcagaatggt atcgacgagg ccaacaaggc cacccaatct taatgatcta cgattggact 60
 ttgtacgaca tagggatgac gatttttaga ttagtaatat ataaccgaag acaataaaga 120
 tatttgtgga ttctattaac aaactacta aaagaatagg atgatacga gcaattgagg 180
 tccaatgct tactggagcc tggggaaaaa tgccagtaag gtgccagcat ggcaggggtt 240
 tgcggtgggt cggtaggcgc cgtttggaca ggggtcaggt acagcggaaa gctgaccatt 300
 caacgcaaac ctaataactg gaatttttgt agttttatte tacatgttca attgctgggt 360
 ttactcaaat tctgaacct ggcagcgtt gtctacaggt cctaaagtcc ctacagctcc 420
 gtgtatgcag cttgtcaaca ggtgtgacga gcactacacg tttcagcaca attgcgttcg 480
 taacagattt ttccaagget tactagcctg tcaactattat tctaccggcc aaattataca 540
 ctttcaagca attacttita taattgcaac tctactttgc aattgttaat tgtccacgac 600
 cgtcgatgac atgggtccct aatgcgtggg ggccgcgcgc acggctggga ggactcgaca 660
 taataaatta ttgcaacaaa gccaaatcaa ttaggtgagg gctgcaacgc attggcaacg 720

agtgaccgta tctgaccaat gtccaatctg cctactgaaa gctgceattg cgctcgtatac 780
 ccctgatttg tgacatatca gccattgcct ccttaattgt catgctcata tactctttct 840
 aactaaata aacccccctca cggggaacgc cggcaaccg cagcataacc cgagtaacgc 900
 tccaacaaa tttggcacgg cccggtagat accggaaaaa ggctcggaaa aaaatctaaa 960
 taactttgca actgaccctt caaaggttga acagtacatc 1000

<210> 43

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 43

ggagaagatg tgggatatta ttggtcttgt aggagcccga ttagggtatg attgggaccg 60
 acaaggaagg attgtctaga ctagtctagc ctagtccaga ctaggctctgt accattacga 120
 gtcgagaact gcactctgat cttgtgetat gtacgtgtga tgtaaataa tgacgaacaa 180
 tatgacgcag acgtggatgt taatctttgg atggacacat ttatatgatg gtggaatggt 240
 ggtcgttgtg aacagtattt aacaaccaga ttcccacact caacttaata caaggactca 300
 atggctctaa atagagctga ataagtacaa ggcattgtta cttatacca attgagctat 360
 ccaattgagt tatatcaacc gtttgacgat ccataattct cagtctgtc tacctcgaat 420
 aactggaact actggtcca attgaccggc ccagccagtg ccagacagta ccaattagtc 480
 caaccactcc catatcacca attgaacaaa tccaattccc ctaccaccgt tacctgtaac 540
 tcacccatt tcaatttgc tgtccagett atccagetta tccatccggg attccgtttc 600
 ctttctcact gctgttgac ccccactett tcctaacac actatttact ctagtacaca 660
 actaattata atactattct cacctcact ccattcctcc tcaactaattg ccaactgaacc 720
 tgccacaacc accgcaccgt accatactaa ttatcctggc caatttcgcc agccaattcc 780
 atccacttgt ctgcaatgtt tacatcgta ctttcctac acgcttctc gaccgggct 840
 ttgccagcgt ccagcgggt tccaactag tgcacggcag acccgggtag ggcccccaac 900
 taatgatacc taccgggcca ctctgaaaaa aagacgccgt tcgagccgga tttttccgtt 960
 gtatttggtc agaacttttt ttctactcc tgtattaaca 1000

<210> 44

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 44

tatattgaat tgatacctaa tatacaatag attgtcctg ggacattaca cgtagacgtt 60
 gaattgtcaa ctacagtatc gtcaacagga agaacattct gtatgccca attgccatta 120
 ccaatcctgg tattcaatc cctgtcccet getgtctctt gctgtctctg tggctctctat 180
 tctagaata cactggccga gagttcgget cagtgcctgc tcgtgatact cggtagcaag 240
 cctaaattgt ccccgatgg ttcgattcca actggaatca ttttctggag taaaatctc 300
 ggcccacgac aataatccgg gtgacgtcat gtaccctag gagggcaaac gccggcgttt 360
 cgcaacaaag cagccccaga aggccccgtt tgaagcgcca gaaccgtctc ccagcgagac 420

tacaaccgt actacgtatc taccggtttt gtagcgattt ccagegtcaa tttcatgtcc 480
 ttttttcat ctccagcttc tccagttcac tctccagecc ttctgttcat ctcttactc 540
 cgaatcgggg gatttttggc aagggttgtc cgattcgtcg gtcgggctgc ggcttgggtg 600
 cccattaacg tgaccgaatg ccgactccc cccgattgac gaaacaaagg aaagcaataa 660
 ctggggtaag aggagattgg gtcgcaatga ttgcacgagc ccggacggta gcgcaattga 720
 gcaccattgt cggcggtcga ctgcctgggc ttctgggatg cctgcaaagc ccggcagcat 780
 ctccgaccaa ttaccgtagt gaattttgtg cagcattttt taccattaac ccgatgcccc 840
 aatcggcccc gacacctgcg tttgtgtaat tgcgagccca tgattggttg ccccgacag 900
 gcgtggcttt ccggcccaaa agtatataac aaactgcaat cgcaaattcg catttttttt 960
 tccgtcgtct agttgcaagt ccaattcgcg gagatttacc 1000

<210> 45

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 45

tgcaggccg ctaatagaac agtgggetca tttgggcggc tcaagccgca ttaccactgt 60
 ggcctcacgg ggcttacggg gctcctgcgg ctctcgggc gcacaaccgt gtatatattt 120
 ccgctggatt ccacgccac ggtagtctaa tccatgtaac gggttgctaa attgtctaga 180
 attgctaaac ttgctaaatt gctaaacttc tgaacgctaa aactgctaaa ttgcttactt 240
 ctactactgc cattaacact ctggetattg cttatcccta tacctacctg ttcttcgctt 300
 ttctatagct attttcacac tgcccattgg tttcccattg gttagaacce gagggtcccc 360
 atgccggcag ccgtacacce tggcgtcttt gtccaaaact gggccgtatc gcggtcagac 420
 aacaggccat tctcgggtgg tatgagagac ggactaatgc gctagtaaca tccggtctat 480
 accattgagc gcctgagtaa ccacaattgc gtgactaatt ctgtttgcat tcggttaacc 540
 cctctctgct ctgatactaa tctgacggc gcggcgcaat tatcgtgttt gttgggcgtc 600
 ccctgtccga gatttgaagg tcccgataat tatcgtcggc aaaaaccgtt actataatgc 660
 atttgacgga cccaaatgat gagttggcaa ccgtttgcaa tcacaatgac cccaaatcct 720
 gatggaaaat gccttgaaag gtacatttcc acatttagtc cactcccccc cttggtcttg 780
 ttgagcggcc cactgcgtct cattcaatgc tgattggtc tttttgacca aacggtggta 840
 tattatctaa cgcaccatca cacaaggcg aggetagttg ctacatggca tcatgctgca 900
 gatgatatat atataaagcc ctctcccacc cgcaggacca acaagaaaaa gtttcacacc 960
 aaagtccgtg tatctttttt gtccaaaaaa aataacaaca 1000

<210> 46

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 46

aaaaggggag acgtcagcca tcctgttagt gtcagtttgc cctacatttg cgcgtccctg 60
 tgtcctttta tattcctctc tcctgaagcc gaaaaaagta attgcaaact accatgcggt 120

ggggacatga tggcagataa tcaccgatga tgattategc acaccgtgat tagcggctca 180
 tgtcccatga tgtggcgcta cctgcccga gcgccgaaa acctaccgca gcagctagtt 240
 tccccaggct gccacatgaa acgaggagaa atagcaatcc cttggccgcg ggaccagttg 300
 ggggccagct gggggccatt gaggtgtcat tgaagtgtca ttggcttgcc atagaatcta 360
 tccatagtag agaacgtcca cttttgttc ttggatatgc ataagcgact ccagggtggg 420
 taaggattat ccatcttcta tcttggcaca taggtagaag tccgcattct tgccgagtag 480
 ccgacaatat atccttaage tccacaattg actttcagat tagaggttta cccaagtagt 540
 accaaggagt accaagtagt accaaatagt accaactagc agttgtgaac tcatataact 600
 gtttcatttg gtggatggaa atcgtcaata gcggagtcc atagaacggt tgtataatac 660
 ggaagggaca cactttgttg gttccattcc aattgtgcta gccaagcaat agtcggattg 720
 cctgcagggt aaagttagtc acgggtacag atcccagtt cagcttcgag ggagtagcct 780
 cgtggcagtt gtccacgagc atcaatggat caagccacat ggttttcagt tctcaatttc 840
 aaagaaacca tcgcatagca tcgaacttagt ccaattcttg agctcttggg tgcgcatctc 900
 ggtcgtggtc agggactggg aaaaaatgcg ccgcatacac agcggcgtgc ggccattacc 960
 ttcacgcgca gagtgcggtt tgtgtttgtca cgaatgacgg 1000

<210> 47

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adeninivorans)

<400> 47

agcaatccaa acagtcacgt ggccgttgtc aagtgaggac tgcccgtgag tgccccgcc 60
 atggatgtgt cattatcacg tgactctgac aaccaagcca attgccccg tgtctcacac 120
 tcacattcca gcaactgggc gccgatggag tgttacgagc ggtgagtcac cagatgtgtc 180
 aactacgtac gagaacaata cacttgatca ttctccgttc cctgacgtg ccccttgcca 240
 tggatgata actaaaggat ggtgcccga acttttcctt tcttctcaa acggaaagga 300
 gtgtttcggg tacgggagcg cgcgcagact ccggtccgga gtttgacaag actcaggggc 360
 ttctgacagg ccttattgtg aagaaccag cacttttctc cagtaactat cctcacagga 420
 tgccatacac gtagattagt accaatttac cctcagtagc ttgctcattg agcaaacttt 480
 ccaattcaat ctagaatgat gtccggcgat tctcgcata acgggtaccg gcgatctccc 540
 tgcccccac gtgcgcctct tggacgttcg gactccgaa tatccactgt tttgccttgc 600
 ctgtgggtgcg gaggatgagt aaccagtggg tacaattggc tccagtttgc catcatcatg 660
 tagataagaa tagaagcaaa ctggacagct gtagtcgcca cactagaca gttgcaattg 720
 ccaactcacg gttctataca ccaaaccacg gtctggttct gcccctttat ttgaccgttg 780
 tcgttggtc ttgtctcaa caaagctcgc ctacctcga tacgagtag catgcgctc 840
 acttttttaa atacgaaaaa gaaattcttg ggcaaatac gaaaagaat tattgggctt 900
 ttctccccg ccgatcaacg cagtgatctt gcgaagacga tatataaca gccaaagatc 960
 cccgaatcat aaacttttct atccgcgaat tagtgctgaa 1000

<210> 48

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 48

ctaattcaag cccactgttg ctaatctctc gacaaagcgt tgagaaactg tcagaggatg 60
 ctcaaacggg tgtggatcca tcggtattca ggggtaacat tgtcatctct ggtacaccag 120
 catacaagga agacgaatgg aactatata caattgccgg acagcggtag cggctcctgg 180
 gcccttgtcg tcggtgcaac atggtatgtg tgaatggcca aggagagatc aattcggaac 240
 cctatTTTTgc actacatcgt accagaaaga cccaaggcaa actactattc ggtcagcaca 300
 tgactcttga tcaacctact gattcactga accctgcaga agctacaatt aaagtaggcc 360
 aattgttcac tcctatctga gacagttcac ctgcagecgt gcaaactgtc aacgagggcc 420
 gaatgatatg gaaataatga ttatgccgtt atgactgtaa tatgaatgaa aaaatTTTcc 480
 ttatgcatta ttaaagaccc aaaataaaca ttctgcccc tgatttacag gtttatccgg 540
 aaggaccggg tcaaagaaaa gTTTTccatg cgtaaaaata atattctgcg tgggggggtcg 600
 gctcccgaact gtggccctat caatagtgcg getgaagagc ttacagacca agctTTTTtag 660
 ctccggacaa atgaatttgg taacaagcat acaatTTTgt tagaagtatt gcgcttcttt 720
 ggtaatTTTT tagtatcttt agtagtcttt atccaattta tgttcattta tactttgact 780
 tggccccctc gttatcttaa cgggtgccagg aactatcgt gcattatcgg accggatacg 840
 gccgataaag cgggtcaatg tcacagttac cgattgetta cataaaagtg gcgcggcgaa 900
 ccgtctagaa tgggtggcgag tatataagga ggccatagcc tagctctgga cacatcat 960
 aaacaactac aaactTTTtac atttacacgt cgcactacc 1000

<210> 49

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 49

acggcggat cgcagcttt gttgacgaca aggctctgcg atggttggca gtcaactttg 60
 cataccacga ccttctggcg tcttcggcgt gctcccga cactcacttt ccatccgcag 120
 aatacgatea cgatcatgaag catggctacg gtctggatgc tctcacgggc tgctgccagc 180
 ctctgttcaa gattctggcc gagatttccg agctcgcct caagtggcag cgagtggacg 240
 atgcgtcctt ggaaaagctc cgaatggtec aggtcccgct tagcgtctg gagcagaagc 300
 ttgaatcttg tcacctgat cctctagaca tgatctcct ttctccccag cagcttgacc 360
 tccaattgat tctatTTTgac accgtcaaga ttaccgccag gctccacctg cgcagtcgg 420
 ttctgcgtct caatgctgcc tccctgaca tgcaatgect tgtcaaacag ctcaacaaga 480
 acttggagct ggtgctgggt acccaggtcg aagggttggc ggtgttccct ctgtttgtgg 540
 ctggtatcca ttgcgtgacc acctcagaca gagagctcat caccaaagc attgatgact 600
 actactctcg caacctggcc cgcaacattt ctagagcaaa agatctcatg gaggaggtgt 660
 ggtctcttga tgatcacgga tctcgtcacg ttgattggta ccgaatcatc caggctagag 720
 gatgggatat ctgtTTTgccc taacagctaa cacgtaacga cttatgacta ctaactgcat 780
 atcaactatc aacaatacta tccttattca atcaactata ctatTTTTat taagatcatc 840

tactatcctt attcaaatca tctatcaact atcctcaaat ttcgtctgta tgtgatccat 900
gcacgtgacc tttaccctg accacatccc gtgacaatac acgtgaacag ttgtgccaac 960
tcagcaccaa atcccccttc gagcttaacc gacgacagca 1000

<210> 50

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 50

aatgattgat acaccttggt acgacctgac tgcgtggtgg gcaagtaaac tggaacttga 60
tatatgctg ccgtttatct gtcataagcc aatcgteaac cacacaatca aatcaaaaac 120
tactgctagc atggcgaacc taaatgggca tcaatggaaa ttatacaaac agtacgagat 180
gaaaacagtc agctatgtca tgggtgtgata gttaccaggt tcattttctg atttcccttg 240
ccagttctgt gcgcctgect cattggattt gactcttttt ggcatcatgc tcacctctgg 300
tgatacacga gctagactgc tgaaagaagt atcagcacag ccaacgaggt tgcagcaaat 360
agggcatacc tgttatcgcc gaccaggeat tatcgaccac cagctatttg cgtctcatgc 420
atcccatttc ctgatgaage tgtgtcccgt cgattacgcc tatecttctt tgccaagcct 480
taccagggac ccaatcatat cgggacceta cgcaacgtga atccggggta ggatatcgag 540
ctcccgaacg ttgaacaaaa ttttaacggt ggtgggagat cacagatcag cgacaccact 600
ataatctgca gtcgcaacca tcacagacct ccgtgaagt atataagaat gctccagaaa 660
gactatggca ggctcgtttt tcccagtgca agagctattt cgggcgagct tctagcggct 720
cccattgtca gaccttaatt gtgctccatt taggcagtg gaggtgccaa gattagtgtt 780
tgaggattct ccctgttgcc aagtctctaa agaagataga cagtgttaag ctactgagct 840
tggaacttga ctaccaatg agaaggatga gccaacccac ctgatgagta ggtatcaggt 900
aacggttgac catagaacga gtcaattgtg gaaatataaa aaggagcca aattggattg 960
attcaccaag aatccaataa aaaaaagaag tactgaaaa 1000

<210> 51

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(*Arxula adenivorans*)

<400> 51

agctcgttcc acccccttcc ccctgtctc caccetaacc ctccggteat actagcacca 60
ctaccgaatg agagtagcac catgtatcat aataaccgag ccaggcgac acaacattga 120
ccgaacaata tcaatatcga ggtacaataa ctgcgtgtct gtgaggccag attacatgag 180
tctgcacgtt tgtgaccgat atcaggcggc ggccgataag ggcaagtga atttcacgtg 240
gaccgtctca cgtgaacacg ggatggcggc agcaatcgtt ggcccaccgt actggccaag 300
caggcccaac aataaagaaa ttcagtggaa aaaccagac caggggacgc agegcacccc 360
tgtaaccgcc cggcacgccc ggccgattg agaccaccgc agagtttttc cggcacagtt 420
tttccggcct ggggtgacct ttgagcgcgc cggaatggcc cgtatcacc tactccgaca 480
gaaccgggtg cggcgagctg aggcggtggg acgattgcgg cggcctgcgg cgcatttcgg 540

gaccgccttc cttgttatga tacgattgcg gcaccgtgag gcgttctga tggttccgag 600
 attcagcgca accttgatgc aacaagtaat caattcgag ccagaatggt ggcaatttgg 660
 tgagcaatag taaaaaaca gtagaatata ggtgtaggaa aacgtagaca gtaggctttt 720
 tgggtccctt tagccattgt aactaaatag ctggacctgc aggacaaaga ccctgtacac 780
 ggaacaattt aagcccttag ctgtaccac aggcatcccc caccgtttta aggacgtcg 840
 caactaacgc ctaaaccgaa caaggacctg gaaagtcgta cgtctaatac ggcaaagtgg 900
 gctataaaaag ggggcgctac tgccaacca atgagttcat ccgatcacca ttgacagttg 960
 tcaattaaca atacacatcc atcttgacc ctaaacaata 1000

<210> 52

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 52

tctaccatca ggaaactgga ggggcgttct cagtacgaca aggccgagag gtatcgtact 60
 ctttggaac tgctcctagg atcgatteta ctctttgtg tctatgcat tactaatttt 120
 ttatttttta tgccgaaga accgacctg gattccagea gcacttgga gtctcgttgg 180
 ttattctgg aagaatttcc taatctggtt tacttcgtt acttttagct tattgcctac 240
 atttggcggc ccaatactaa cgacgtcagg ttcatgtcgt ccaagattgc ccaggatgag 300
 aatgaagttc aagagtttga aattggatct ctccgagagt ctatggacga gtaagagata 360
 ttaaggaatt gaaaaagggc aagaaaagag cgatgagcgt agaaattgcg tagaaattgt 420
 agcagtatca atacccttac catcacctaa agcaaacc aaagatcccc gtgaatctcc 480
 gggacctgag tagatggtaa tacagaatac tggcagaata ctgcactcag aagaactctg 540
 gaagaactct ggaagcagtc taacggacct cagtttggct cttgaacatt cacgtgactg 600
 gaaacttaac atcacgtgac ctctgccagt ctggattgaa atagggtgaa aataaaaaat 660
 cagtacacaa tgagagttg gccgagtggt ctatggcgtc agatttaggt aaaccctaaa 720
 gtgaattctc tgatatcttc ggatgcgca gttcgaatct cgtagctctc attatctttt 780
 ttactccctt tccgtttcgg actaaccacg gatacctttt ccaagcaatt tgcgatccaa 840
 ttatttttgt tcttttaatt aaatttagtt tcattcatct ccggtcccc ttgatagatg 900
 aacgtccgta tttaccgta agccgcataa ccgccaggaa agccccgatc tgtcaacctt 960
 ggcattact acgtttcgtt tataactctc gctcgtttta 1000

<210> 53

<211> 1000

<212> DNA

<213> 解腺嘌呤阿氏酵母(Arxula adenivorans)

<400> 53

gccagtgca ttgtccttgt cattetagger gtggcctttt tcattgcatt tgtacttgta 60
 gaacgagctg tggacacccc tctagtaccg gttcgcaagt ttaactaa tatggccaga 120
 gtgctcgtt gtgtggcctt tggatggggc acttttggt tctggattta ctaccttgg 180
 cagattatgg aatacctgcg acacaactcc ccattgttgg cttcagctca gttctctcca 240

gctgccgcga tgggtgccat tgctgcaatt gctactggat acctcatgtc aaagctacat 300
cctttccgag tgctggcaat ttccctgttg gcgttctctg tegettcaat taccaccgcc 360
acggcgcctg taaaccaaac gttctgggct cagacgtttg tatcaatctt agtagcttct 420
tggggtatgg acatgaactt cctgctgctg acccttatct tatcagagac cgtgcccagg 480
gaacagcagg gaattgccgc ctcttttagtg gccactgttg tcaattattc aatctcccta 540
agcctgggag ttgcaggtag tatcattgag caggtagctc caggtttggg ccctaattca 600
tatttaaagg gcgtccgaag cgcctatat ttctgcattg gcctctctgc cgccggcctt 660
cttgctgctc tctatgggtg catcagagac gacattcttg ctaaccatgg gaaatcctct 720
aacgacgaag aaaagaatac tgcttgaat gcttttttaa tagaattttg ctcttatttg 780
tcctatttaa tctatatttc atgtacgaat cgatttctaa tettaacacc gcggagattc 840
ttttgttatt actaaatcag gaaaagatgc acggagaact cggcccagat tggatttgat 900
ggcatctcgg tccgagttaa acgtggggta atcttttagc ggggaaagtt ataaaacccc 960
taciaagccc aggatttgat aattcacatt tgacaacaca 1000

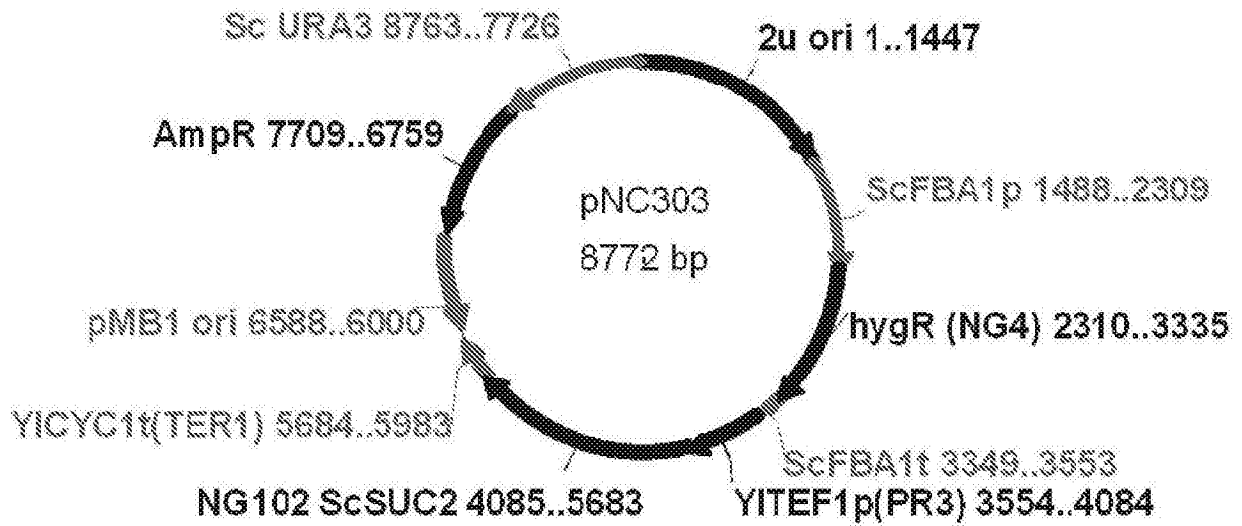


图1

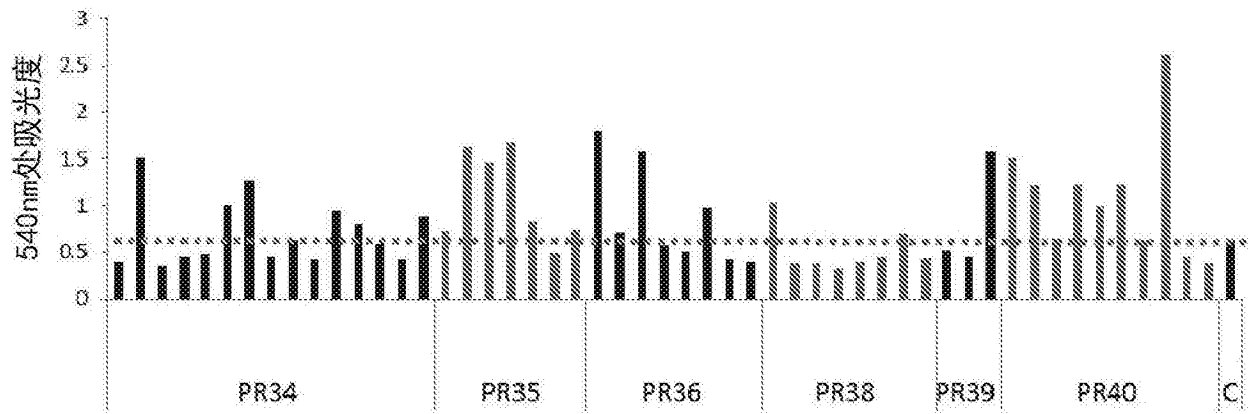


图2A

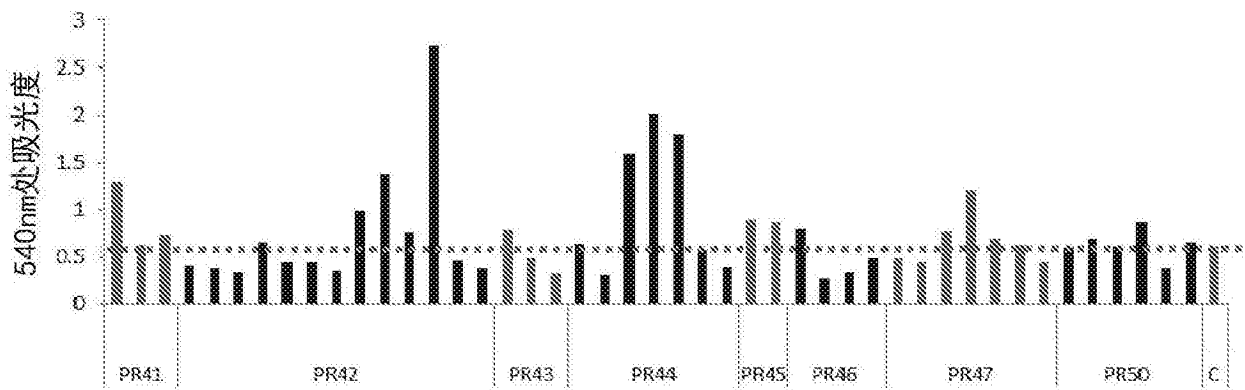


图2B

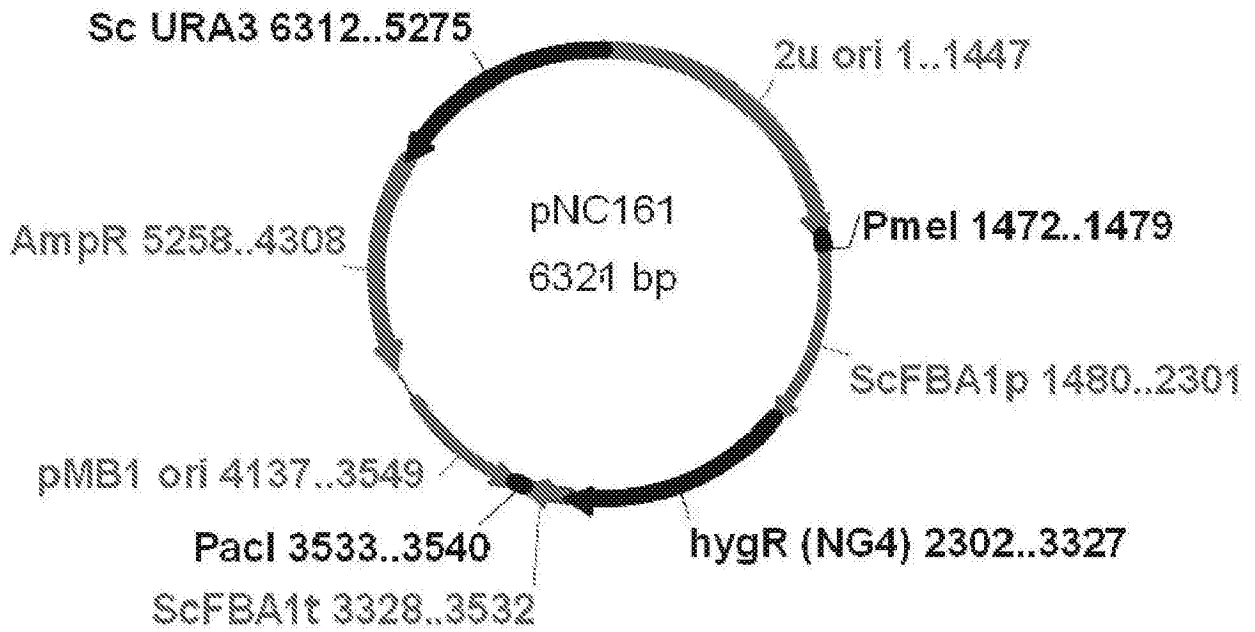


图3

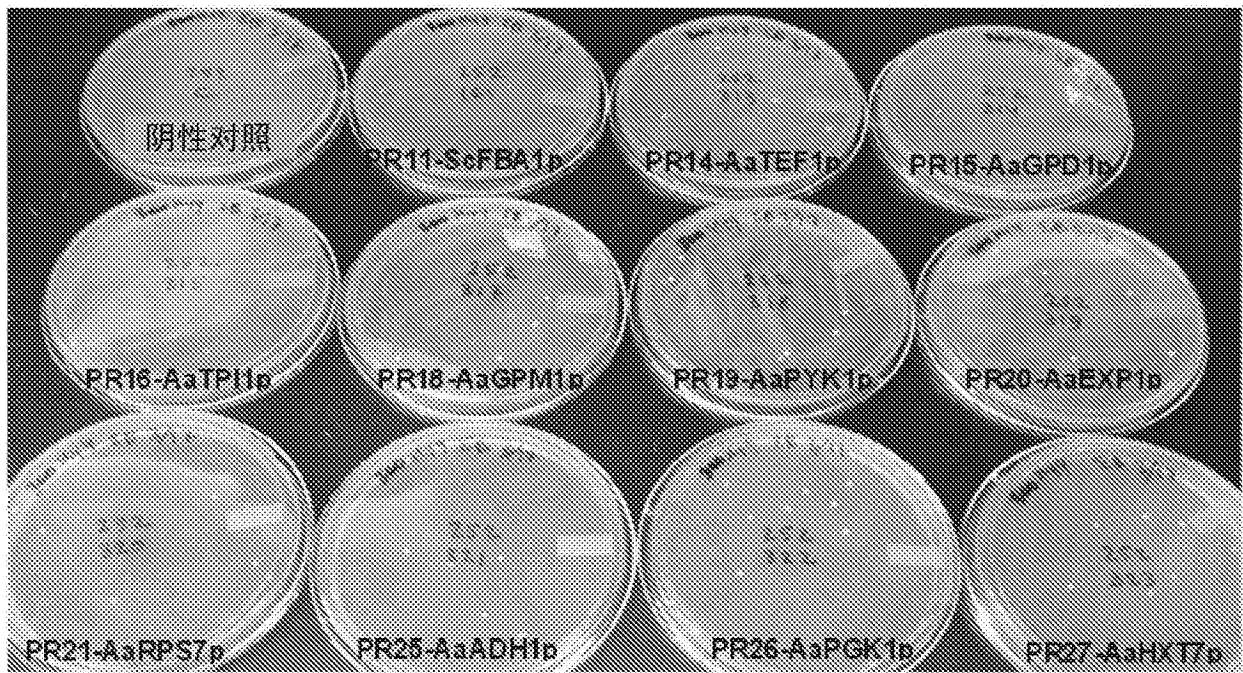


图4

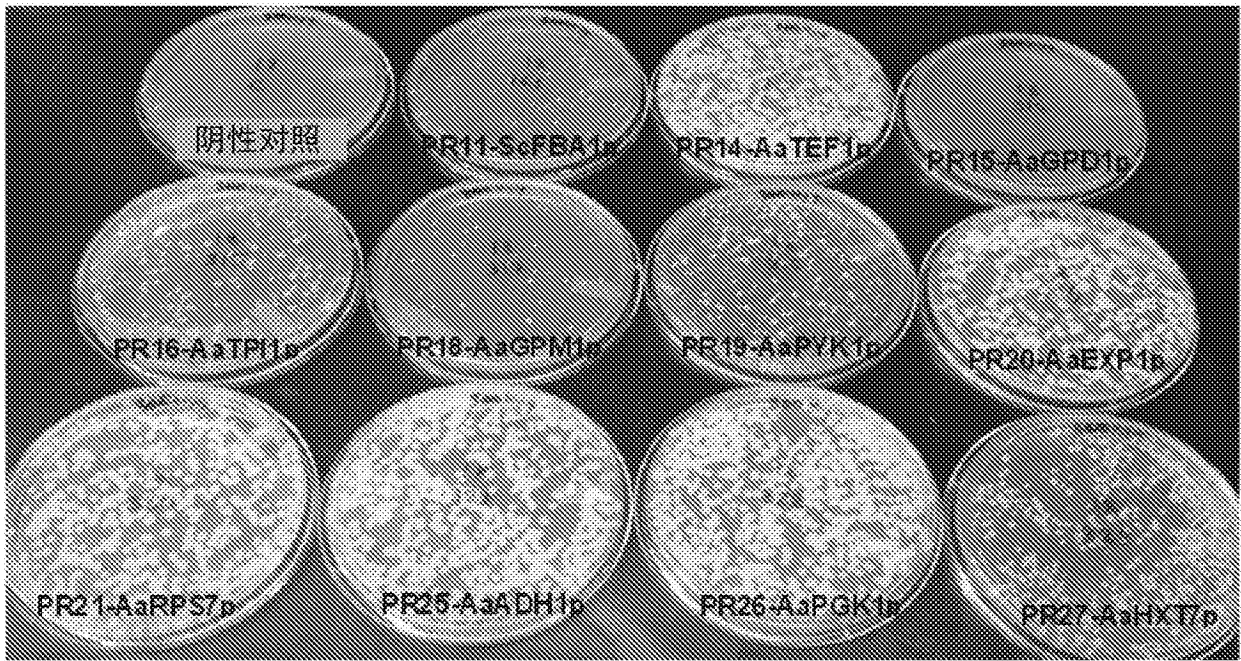


图5

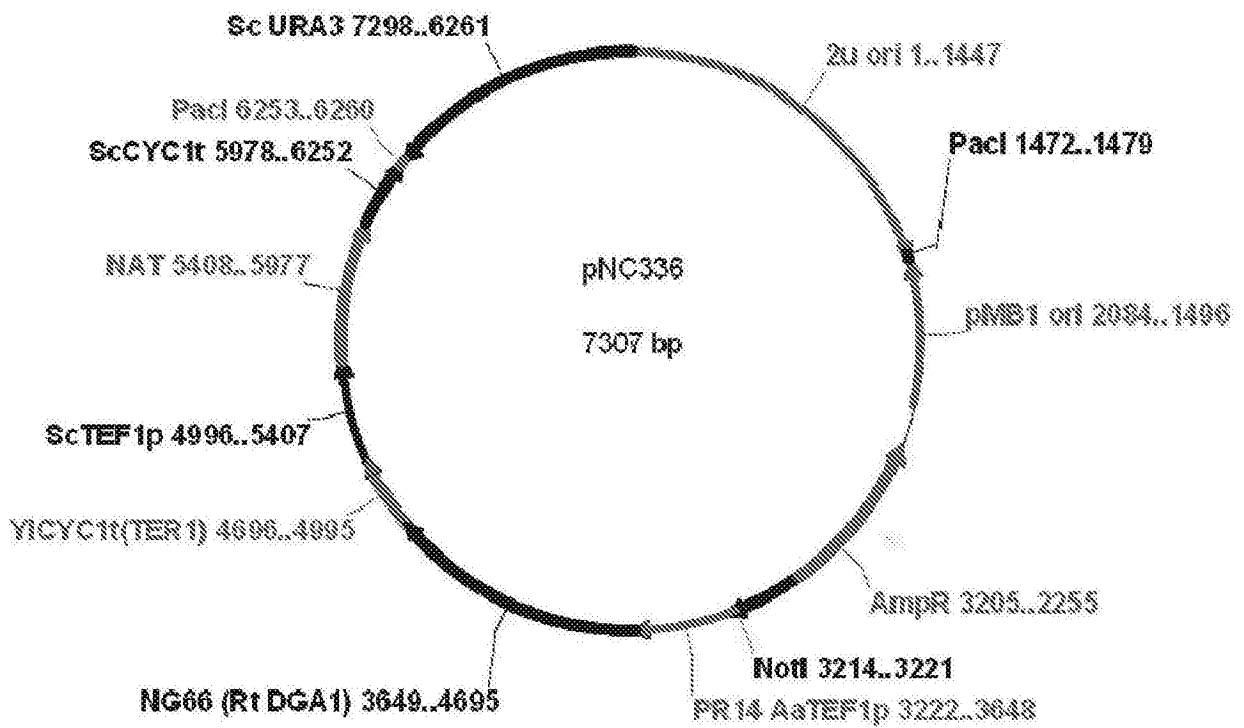


图6

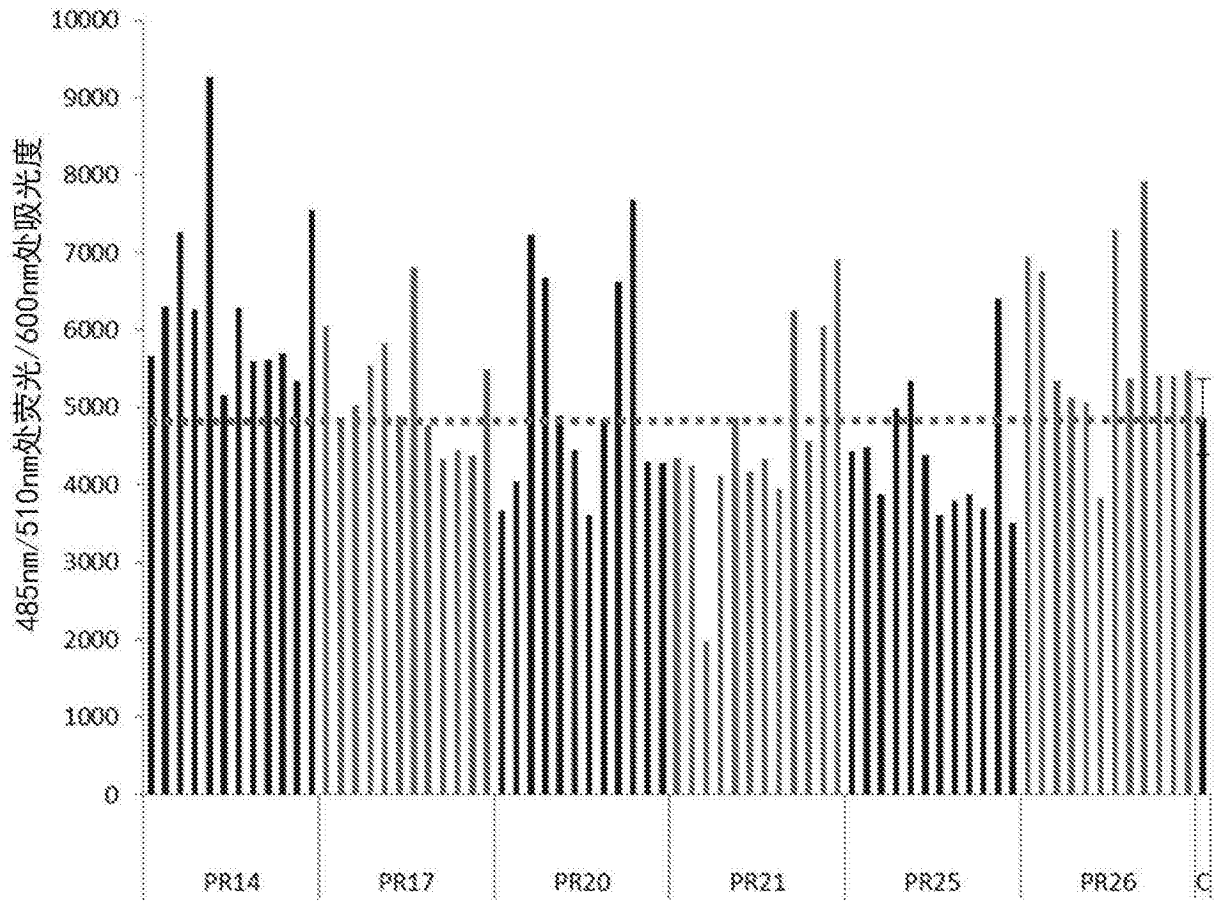


图7

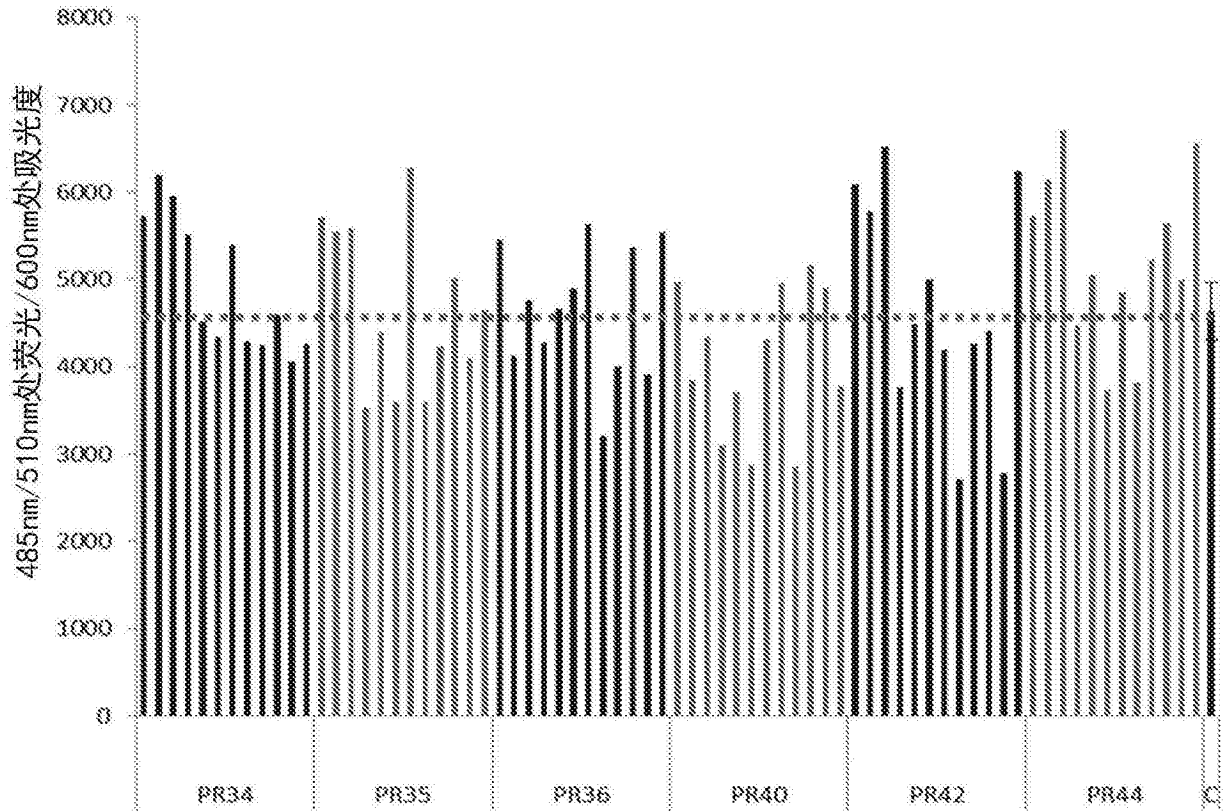


图8

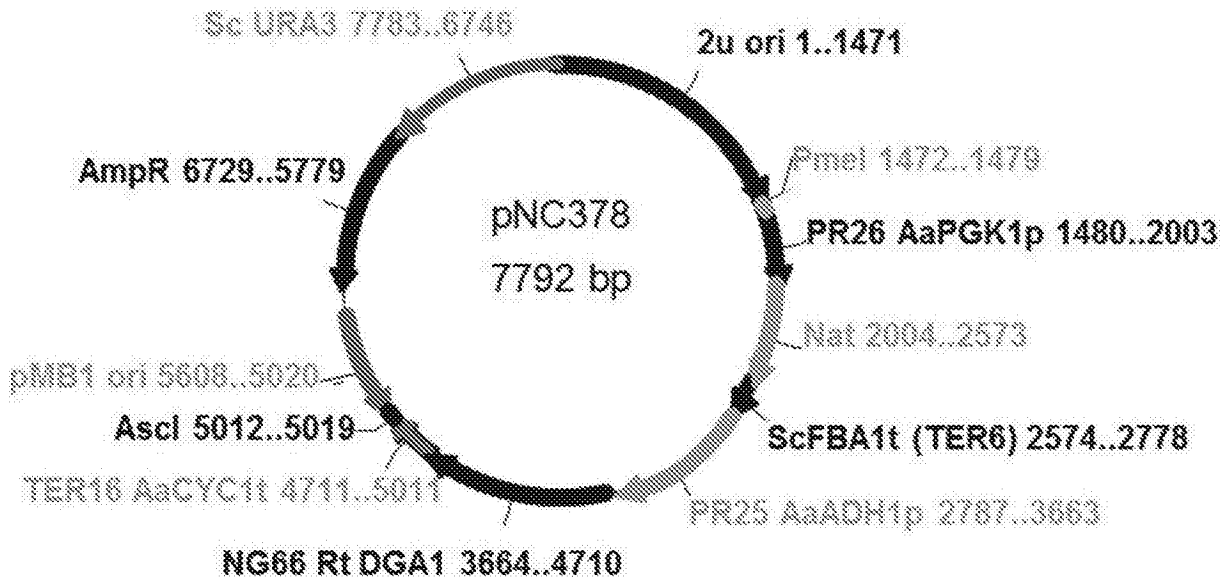


图9

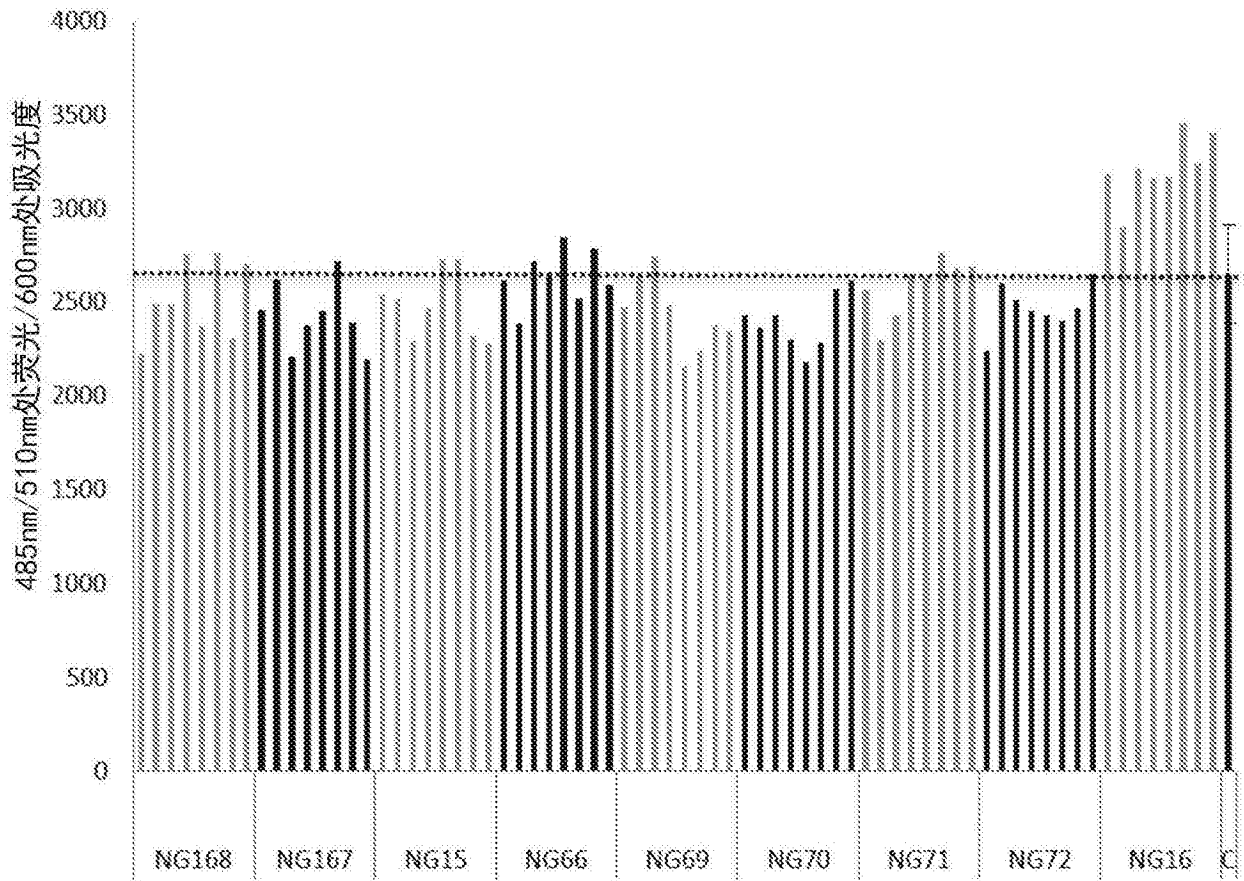


图10

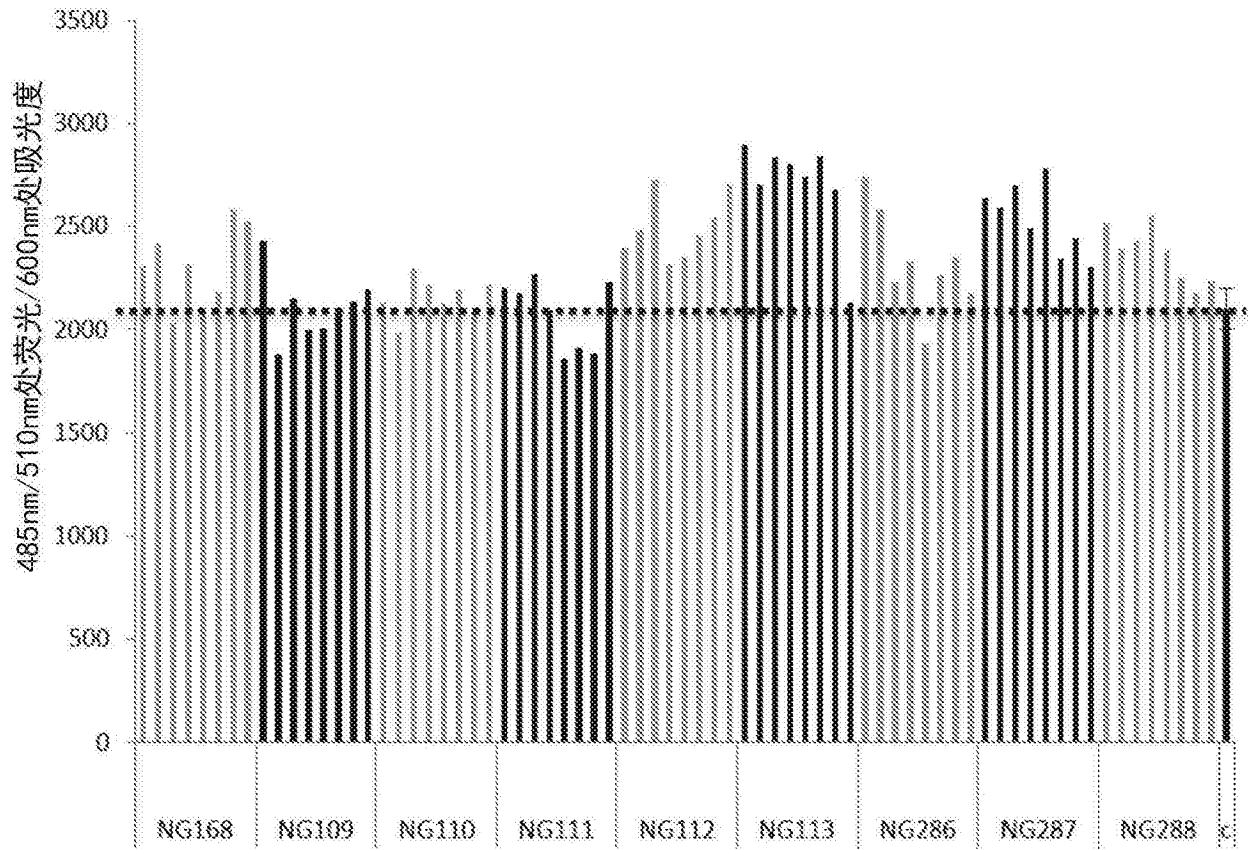


图11

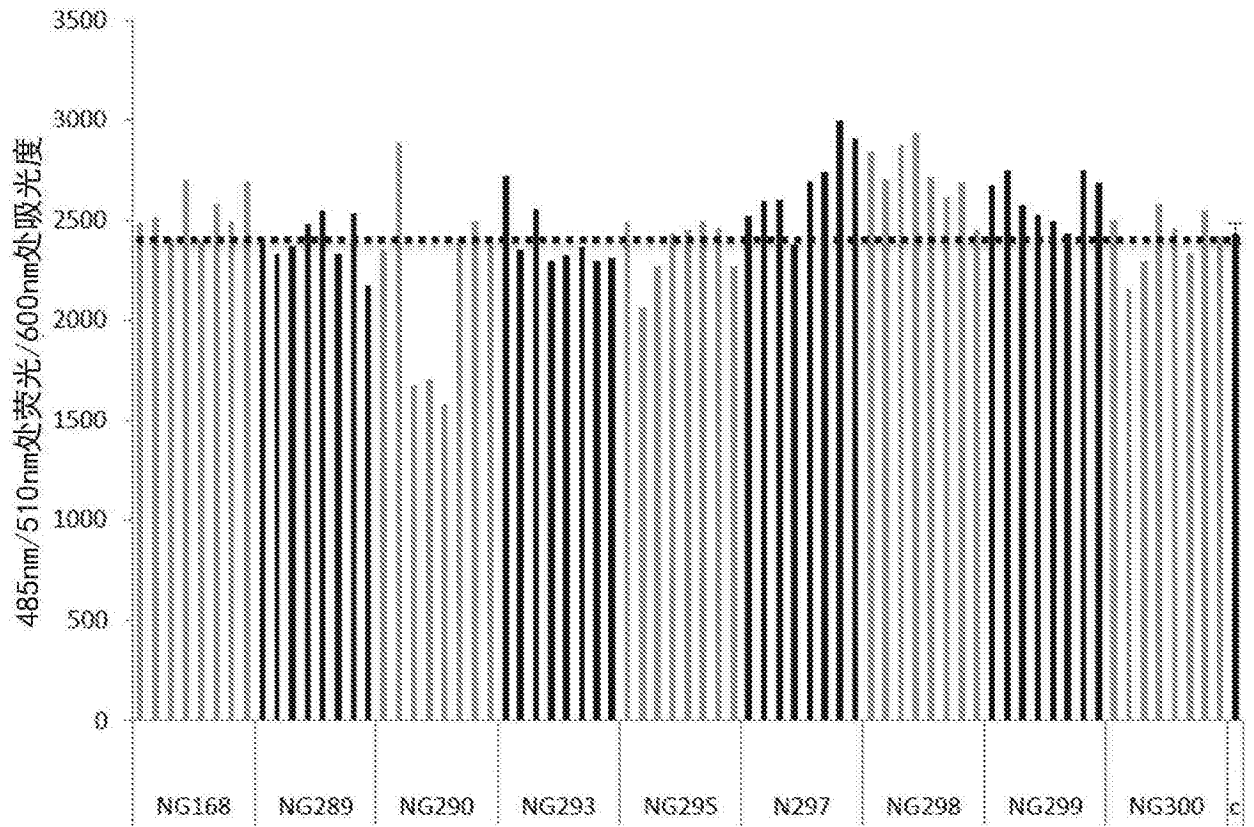


图12