



(12)发明专利

(10)授权公告号 CN 105026560 B

(45)授权公告日 2020.09.18

(21)申请号 201480013264.9

(51)Int.Cl.

(22)申请日 2014.03.10

C12N 9/92(2006.01)

(65)同一申请的已公布的文献号

C12P 7/10(2006.01)

申请公布号 CN 105026560 A

C12P 7/16(2006.01)

(43)申请公布日 2015.11.04

C12P 7/18(2006.01)

(30)优先权数据

(56)对比文件

13/792321 2013.03.11 US

CN 101260394 A, 2008.09.10

(85)PCT国际申请进入国家阶段日

Earl,A.等.xylose isomerase

2015.09.09

[Lachnospiraceae oral taxon 107 str.

(86)PCT国际申请的申请数据

F0167].《GenBank: EGG90672.1》.2011,

PCT/US2014/022358 2014.03.10

Durkin,A.S.等.xylose isomerase

(87)PCT国际申请的公布数据

[Lachnoanaerobaculum sp. ICM7].《GenBank:

W02014/164410 EN 2014.10.09

EJP20197.1》.2012,

(73)专利权人 纳幕尔杜邦公司

Weinstock,G.等.xylose isomerase

地址 美国特拉华州.威尔明顿

[Lachnospiraceae bacterium oral taxon 082

(72)发明人 Z.陈 K.J.凯利 R.W.叶

str. F0431].《GenBank: EH053553.1》.2012,

(74)专利代理机构 中国专利代理(香港)有限公

审查员 徐俊

司 72001

权利要求书1页 说明书14页

代理人 唐华东 杨思捷

序列表35页

(54)发明名称

酵母细胞中的细菌木糖异构酶活性

(57)摘要

在酵母细胞中,特定多肽被鉴定为能够提供木糖异构酶活性的细菌木糖异构酶。所述木糖异构酶活性能够完成木糖利用途径,使得酵母在发酵过程中能够利用木糖,诸如在生物质水解产物中的木糖。

1. 一种重组酵母细胞,其包含编码多肽的异源核酸分子,所述多肽具有木糖异构酶活性并且其氨基酸序列选自SEQ ID NO:1、3和5。
2. 根据权利要求1所述的重组酵母细胞,其还包含完整的木糖利用途径并具有在作为唯一碳源的木糖上生长的能力。
3. 根据权利要求2所述的重组酵母细胞,其还包含目标化合物,其中所述目标化合物选自乙醇、丁醇和1,3-丙二醇。
4. 一种用于制备具有木糖异构酶活性的酵母细胞的方法,所述方法包括:
 - a) 提供酵母细胞;
 - b) 导入编码多肽的异源核酸分子,所述多肽具有木糖异构酶活性并且其氨基酸序列选自SEQ ID NO:1、3和5;其中制备具有木糖异构酶活性的酵母细胞。
5. 根据权利要求4所述的方法,其中所述异源核酸分子是嵌合基因的一部分。
6. 根据权利要求4所述的方法,其中所述具有木糖异构酶活性的酵母细胞具有完整的木糖利用途径并在包含木糖作为碳源的培养基中生长,其中木糖被利用。
7. 根据权利要求6所述的方法,其中所述酵母细胞包含制备目标化合物的代谢途径,其中所述目标化合物选自乙醇、丁醇和1,3-丙二醇。

酵母细胞中的细菌木糖异构酶活性

[0001] 本专利申请要求提交于2013年3月11日的美国国家申请13/792321的权益,其全部内容以引用方式并入。

技术领域

[0002] 本发明涉及酵母的基因工程领域。更具体地,鉴定一组木糖异构酶,其在被工程化以进行其表达的酵母细胞中是活性的。

背景技术

[0003] 目前酵母是用于发酵制备乙醇的生物体或选择。最常见的是使用酿酒酵母(*Saccharomyces cerevisiae*),在使用过程中用从谷物或醪液获得的己糖作为碳水化合物来源。使用从纤维素类生物质制备的水解产物作为发酵的碳水化合物来源是可取的,因为这是易得的可再生资源,不会与食品供应产生竞争。在葡萄糖之后,纤维素类生物质中第二丰富的糖是木糖,一种戊糖。酿酒酵母(*Saccharomyces cerevisiae*)并非天然能够代谢木糖,但能经过工程化(木糖异构酶活性的表达以将木糖转化为木酮糖,以及额外的途径工程化)以代谢木糖。

[0004] 成功地表达在酵母中有活性的异源的细菌木糖异构酶已受限。某些特异性木糖细菌异构酶序列已被报道为酵母中的木糖利用途径提供木糖异构酶活性。例如US 7,622,284公开了一种表达来自厌气性瘤胃真菌属(*Piromyces* sp.)的木糖异构酶的酵母细胞。US 2012/0184020公开了表达分离自生黄瘤胃球菌(*Ruminococcus flavefaciens*)的木糖异构酶的真核细胞。类似地W02011078262公开了来自栖北散白蚁(*Reticulitermes speratus*)和达尔文澳白蚁(*Mastotermes darwiniensis*)中的每种的多种木糖异构酶,和与这些酶具有高度序列同一性的蛋白,以及它们在真核细胞中的表达。W0212009272公开了包含来自软弱贫养菌(*Abiotrophia defectiva*)的木糖异构酶和与之具有序列同一性的其它(酶)的构建体和真菌细胞。

[0005] 但仍需要另外的工程化的酵母细胞,其表达木糖异构酶活性以成功利用木糖,从而在发酵过程中允许有效利用得自的纤维素类生物质的糖类。

发明内容

[0006] 本发明提供了经工程化以表达提供木糖异构酶活性的多肽的重组酵母细胞。

[0007] 因此,本发明提供了包含异源核酸分子的重组酵母细胞,所述异源核酸分子编码多肽,所述多肽具有木糖异构酶活性并且氨基酸序列与选自SEQ ID NO:1,3,5和7的氨基酸序列相比具有至少约85%的序列同一性。

[0008] 在另一方面本发明提供了制备具有木糖异构酶活性的酵母细胞的方法,其包括以下步骤:

[0009] a) 提供酵母细胞;

[0010] b) 导入编码多肽的异源核酸分子,所述多肽具有木糖异构酶活性并且其氨基酸序

列与选自SEQ ID NO:1,3,5和7的氨基酸序列相比具有大于95%的序列同一性；

[0011] 其中制备具有木糖异构酶活性的酵母细胞。

[0012] 序列描述

[0013] 根据下面的详细描述和附带的序列描述可更充分地理解本发明，下面的详细描述和附带的序列描述形成了本专利申请的一部分。

[0014] 下列序列符合37C.F.R.1.821-1.825（“对含有核酸序列和/或氨基酸序列公开的专利申请的要求-序列规则”）并且符合世界知识产权组织(WIPO)标准ST.25(2009)和EPO和PCT的序列表要求(规则5.2和49.5(a-bis)，以及行政性指示的208节和附录C)。用于核苷酸和氨基酸序列数据的符号和格式遵循在37C.F.R. §1.822中列出的规定。

[0015] 表1:木糖异构酶多肽的SEQ ID NO,以及经密码子优化以在酿酒酵母中表达的编码区

[0016]	菌株	SEQ ID NO:	SEQ ID NO:
[0017]		氨基酸	核苷酸密码子选择
	毛螺科菌 (<i>Lachnospiraceae bacterium</i>) ICM7	1	2
	毛螺科菌 (<i>Lachnospiraceae bacterium</i>)	3	4
	口腔菌群类 107 菌株 F0167		
	毛螺科菌 (<i>Lachnospiraceae bacterium</i>)	5	6
	口腔菌群类 082 菌株 F0431		
	砂优杆菌 (<i>Eubacterium saburreum</i>) DSM 3986	7	8
	<i>Ruminococcus chamanellensis</i> 18P13	9	10
	生黄瘤胃球菌 (<i>Ruminococcus flavefaciens</i>) FD-1	11	*nd
	软弱贫养菌 (<i>Abiotrophus defectiva</i>)	12	*nd
	<i>Leptotrichia goodfellowii</i> F0264	13	14
	塞巴鲁德菌属 (<i>Sebaldella termitidis</i>) ATCC (美国典型培养物保藏中心) 33386	15	16

[0018] *nd=未设计

[0019] SEQ ID NO:17是载体pHR81的核苷酸序列,其包含ILVp-xy1A(Hm1)-ILV5t嵌合基因。

[0020] SEQ ID NO:18是P5整合载体的核苷酸序列。

[0021] SEQ ID NO:19是URA3缺失痕的核苷酸序列。

[0022] SEQ ID NO:20是ura3 Δ 后删除区的上游核苷酸序列。

[0023] SEQ ID NO:21是ura3 Δ 后删除区的下游核苷酸序列。

[0024] SEQ ID NO:22是his3 Δ 后删除区的上游核苷酸序列。

[0025] SEQ ID NO:23是his3 Δ 后删除区的下游核苷酸序列。

[0026] SEQ ID NO:24是pJT254的核苷酸序列。

具体实施方式

[0027] 可使用下列定义阐释权利要求和说明书:

[0028] 如本文所用,术语“包含”、“包括”、“具有”或“含有”,或者其任何其它变型旨在包括非排它的包括。例如,包含元素列表的组合物、混合物、工艺、方法、制品或设备不必仅限于那些元素,但可以包括其他未明确列出的元素,或此类组合物、混合物、工艺、方法、制品

或设备固有的元素。此外,除非明确指明相反,“或”是指包含性的“或”而非排他性的“或”。例如,条件A或B满足下列任一项:A为真实的(或存在的)且B为虚假的(或不存在的),A为虚假的(或不存在的)且B为真实的(或存在的),以及A和B均为真实的(或存在的)。

[0029] 此外,涉及元素或组分例子(即出现)的数目在本发明元素或组分前的不定冠词“一个”或“一种”旨在为非限制性的。因此,应将“一个”或“一种”理解为包括一个或至少一个,并且元素或组分的词语单数形式也包括复数指代,除非有数字明显表示单数。

[0030] 如本文所用,术语“发明”或“本发明”是非限制性术语,并且不旨在意指本发明的任何单独实施例,而是涵盖如本说明书和权利要求所述的所有可能的实施例。

[0031] 如本文所用,用术语“约”修饰本发明的成分或反应物的数量时是指数值量的变化,它们可能发生在例如,典型的测量和用于制备浓缩液或实际使用溶液的液体处理程序中;这些程序中的偶然误差中;制造、来源、或用于制备组合物或实施方法的成分的纯度的差异中;等。术语“约”还包括由于对于起因于特定起始混合物的组合物的不同平衡条件而不同的量。无论是否由术语“约”来修饰,权利要求包括量的等同量。在一个实施例中,术语“约”指在报告数值10%范围内,优选地在报告数值5%范围内。

[0032] 术语“木糖异构酶”是指催化D-木糖和D-木酮糖的互变的酶。木糖异构酶(XI)属于被分类为EC 5.3.1.5的酶的组。

[0033] 术语“木糖利用途径”是指包含基因的代谢途径,所述基因编码足以将木糖转化为目标化学制品的酶。在目标化学制品为乙醇的情况下,此类途径通常包含编码以下酶的基因:木酮糖激酶(XKS1)、转醛醇酶(TAL1)、转酮醇酶1(TKL1)、D-核酮糖-5-磷酸3-表异构酶(RPE1)、和核糖5-磷酸酮醇异构酶(RKI1)。该途径的元件对宿主细胞可以是天然的或异源的。

[0034] 术语“基因”是指表达特定蛋白或功能RNA分子的核酸片段,其在编码序列之前和之后可任选地包括调控序列5'非编码序列和3'非编码序列。“天然基因”或“野生型基因”指具有其自身调控序列的天然存在的基因。“嵌合基因”指不是天然基因的任何基因,包含在天然情况下不是一起存在的调控序列和编码序列。因此,嵌合基因可包含来源于不同来源的调控序列和编码序列,或来源于相同来源、但以不同于天然存在的方式排列的调控序列和编码序列。“内源性基因”是指在生物体基因组中其天然位点的天然基因。“外来的”基因是指在宿主生物体中通常不存在、但通过基因转移导入宿主生物体的基因。外来基因可包括插入非天然生物体的天然基因或嵌合基因。

[0035] 术语“启动子”或“启动控制区”是指能够控制编码序列或功能RNA的表达的DNA序列。一般来讲,编码序列位于启动子序列的3'端。启动子可整体来源于天然基因,或由来源于天然存在的不同启动子的不同元件组成,或甚至包含合成的DNA片段。本领域的技术人员应当理解不同的启动子可指导基因在不同组织或细胞类型中的表达,或在不同发育阶段的表达,或响应于不同环境条件的表达。通常将在大多数细胞类型中、在大多数情况下引起基因表达的启动子称为“组成型启动子”。

[0036] 如本文所用,术语“表达”指衍生自基因的编码(mRNA)或功能性RNA的转录和稳定积聚。表达也可指将mRNA翻译成多肽。“过表达”是指在转基因生物体中基因产物制备超过了在正常生物体或非转化过的生物体中的制备水平。

[0037] 如本文所用,术语“转化”指将核酸片段转移至宿主生物体内,导致在基因上稳定

遗传。被转移的核酸可以质粒的形式保持在宿主细胞中,或某些被转移的核酸可整合进宿主细胞的基因组。包含转化的核酸片段的宿主生物被称为“转基因的”或“重组的”或“转化的”生物体。

[0038] 如本文所用,术语“质粒”和“载体”是指通常携带有不属于细胞中心代谢的部分的基因的染色体外元件,并且通常是环状双链DNA分子的形式。此类元件可为来源于任何来源的自主复制的序列、基因组整合序列、噬菌体或核苷酸序列、线性或环状的单链或双链DNA或RNA,其中多个核苷酸序列连接或重组至细胞中能够导入启动子片段和选定基因产物的DNA序列以及合适的3'非翻译序列的独特结构。

[0039] 术语“可操作地连接”指单个核酸片段上的核酸序列的关联,使得其中一个核酸序列的功能受到另一个核酸序列的影响。例如,当启动子能够影响编码序列的表达(即,该编码序列受到该启动子的转录控制)时,则该启动子与该编码序列可操作地连接。编码序列可以按有义或反义的取向可操作地连接到调控序列。

[0040] 术语“选择性标记”是指一种标识因子,通常为抗生素或耐化学品种性基因,其能被基于标记基因的效应(即对抗生素的抗性)被选择,其中所述效应被用于追踪感兴趣的核酸的遗传和/或识别遗传了感兴趣的核酸的细胞或生物体。

[0041] 如本文所用,术语“密码子简并性”指允许核苷酸序列在不影响所编码的多肽的氨基酸序列的情况下发生变化的遗传密码的性质。技术人员非常了解具体宿主细胞在使用核苷酸密码子以确定给定氨基酸时所表现出的“密码子偏倚性”。因此,当合成基因用以改善在宿主细胞中的表达时,希望对基因进行设计,使得其密码子使用频率接近该宿主细胞优选的密码子使用频率。

[0042] 术语“密码子优化的”在其涉及用于转化不同宿主的核酸分子的基因或编码区时是指在不改变由DNA编码的多肽的情况下,改变核酸分子的基因或编码区中的密码子以反映宿主生物体通常的密码子使用。

[0043] 术语“碳底物”或“可发酵碳底物”指能够被微生物代谢的碳源。碳底物的一种类型是“可发酵糖类”,它是指能够在发酵过程中被微生物用作碳源的寡糖和单糖。

[0044] 术语“木质纤维素的”是指一种组合物,其同时包含木质素和纤维素。木质纤维素材料还可包含半纤维素。

[0045] 术语“纤维素的”是指一种组合物,其包含纤维素和附加组分,所述附加组分可包括半纤维素和木质素。

[0046] 术语“糖化”指由多糖产生可发酵糖。

[0047] 术语“经预处理的生物质”是指生物质已经受热、物理和/或化学预处理以增加生物质中多糖对糖化酶的可用性。

[0048] “生物质”是指任何纤维素的或木质纤维素材料并且包括含有纤维素的材料,以及任选地还包含半纤维素、木质素、淀粉、低聚糖和/或单糖。生物质还可包含附加组分,诸如蛋白和/或脂类。生物质可来源于单一来源,或生物质可包括来源于一种以上来源的混合物;例如,生物质可包括玉米芯和玉米秸秆的混合物,或小草和叶片的混合物。生物质包括但不限于:生物能作物、农业残余物、市政固体垃圾、工业固体垃圾、来自造纸业的淤渣、庭院垃圾、木材和林业垃圾。生物质的示例包括但不限于玉米棒、作物残余物诸如玉米壳、玉米秸秆、玉米谷物纤维、草、甜菜浆、小麦秸秆、麦糠、燕麦秸秆、大麦秸秆、大麦壳、干草、稻

秆、稻壳、柳枝稷、细叶芒、大米草、草芦、废纸、甘蔗渣、高粱渣、高粱秸秆、大豆秸秆、从谷物研磨中获得的组分、树、枝、根、叶、木片、锯末、废棕榈、灌木和灌丛、蔬菜、水果、花和动物粪肥。

[0049] “生物质水解产物”指来源于生物质糖化的产物。生物质也可在糖化前进行预处理或预加工。

[0050] 术语“异源”指非天然存在于受关注的位置。例如，异源性基因是指并非在宿主生物中天然存在，但通过基因转移被导入宿主生物体的基因。例如，存在于嵌合基因中的异源性核酸分子是并非与嵌合基因的其它片段天然相关联的核酸分子，诸如具有彼此并非天然相关联的编码区和启动子片段的核酸分子。

[0051] 如本文所用，“分离的核酸分子”是RNA或DNA的聚合物，它是单链-或双链的，任选地包含合成的、非-天然的或改变的核苷酸碱基。DNA聚合物形式的分离的核酸分子可由cDNA、基因组DNA或合成DNA的一个或多个区段构成。

[0052] 术语“目标化合物”或“目标化学制品”是指由微生物经由内源性或重组生物合成途径制成的化合物，所述途径能够代谢可发酵碳源以制备目标化合物。

[0053] 术语“百分比同一性”，如在本领域中已知的，是通过比较序列测定的两个或更多个多肽序列或两个或更多个多核苷酸序列之间的关系。在本领域中，“同一性”还表示多肽或多核苷酸序列之间序列关联的程度，根据具体情况，它由这些序列的序列串之间的匹配程度确定。“同一性”和“相似性”能通过已知方法容易地计算出来，包括但不限于描述于下列的那些：1.) Computational Molecular Biology (Lesk, A.M. 编辑) Oxford University: NY (1988) ; 2.) Biocomputing: Informatics and Genome Projects (Smith, D.W. 编辑) Academic: NY (1993) ; 3.) Computer Analysis of Sequence Data, 部分I (Griffin, A.M. 和 Griffin, H.G. 编辑) Humania: NY (1994) ; 4.) Sequence Analysis in Molecular Biology (von Heijne, G. 编辑) Academic (1987) ; 和 5.) Sequence Analysis Primer (Gribskov, M. 和 Devereux, J. 编辑) Stockton: NY (1991) 。

[0054] 设计确定同一性的优选方法来给出待测试序列之间的最佳匹配。确定同一性和相似性的方法在可公开获得的计算机程序中被编成了代码。序列比对和百分比同一性计算可以用LASERGENE生物信息学计算软件包 (DNASTAR Inc., Madison, WI) 中的MegAlign程序进行。

[0055] 序列的多重比对使用“Clustal比对方法”进行，该方法涵盖若干个不同的算法，包括对应于称为Clustal V比对方法的“Clustal V比对方法”(在Higgins和Sharp, CABIOS 5: 151-153 (1989) ; Higgins, D.G. 等人, Comput. Appl. Biosci., 8: 189-191 (1992) 中有所描述) 并可见于LASERGENE生物信息学计算软件包 (DNASTAR Inc.) 的MegAlign v8.0程序中。对于多重比对，默认值对应于GAP PENALTY=10和GAP LENGTH PENALTY=10。用Clustal方法进行蛋白质序列的百分比同一性成对比对和计算的默认参数为KTUPLE=1、GAP PENALTY=3、WINDOW=5、以及DIAGONALS SAVED=5。对于核酸，这些参数为KTUPLE=2，空位罚分=5，窗=4，以及DIAGONALS SAVED=4。用Clustal V程序比对序列后，可通过查看同一程序中的“序列距离”表来获得“百分比同一性”。

[0056] 此外，也可以使用“Clustal W比对方法”，该方法对应于称为Clustal W的对比方法 (在 Higgins 和 Sharp, CABIOS 5: 151-153 (1989) ; Higgins, D.G. 等人，

Comput. Appl. Biosci. 8:189-191 (1992); Thompson, J.D. 等人, Nucleic Acid Research, 22 (22): 4673-4680, 1994 中有所描述) 并可见于 LASERGENE 生物信息学计算软件包 (DNASTAR Inc.) 的 MegAlign v8.0 程序中。用于多重比对的默认参数 (规定为蛋白质/核酸 (空位罚分 = 10/15, 空位长度罚分 = 0.2/6.66, 延迟分散序列 (%) = 30/30, DNA 转变重量 = 0.5, 蛋白质重量基质 = Gonnet 系列, DNA 重量基质 = IUB)。用 Clustal W 程序比对序列之后, 可通过查看同一程序中的“序列距离”表来获得“百分比同一性”。除非另外指明, 本文所涉及的序列同一性常常应该被认为已经根据上文所示的参数来确定。

[0057] 术语“序列分析软件”是指可用于分析核苷酸或氨基酸序列的任何计算机算法或软件程序。“序列分析软件”可商购获得或独立开发。典型的序列分析软件包括但不限于: 1.) GCG 程序包 (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI); 2.) BLASTP、BLASTN、BLASTX (Altschul 等人, J. Mol. Biol., 215: 403-410 (1990)); 3.) DNASTAR (DNASTAR, Inc. Madison, WI); 4.) Sequencher (Gene Codes Corporation, Ann Arbor, MI); 以及 5.) 整合了 Smith-Waterman 算法的 FASTA 程序 (W.R. Pearson, Comput. Methods Genome Res., [Proc. Int. Symp.] (1994), 开会日期 1992, 111-20. 编辑: Suhai, Sandor. Plenum: New York, NY)。在本专利申请的上下文中应当理解, 使用序列分析软件进行分析时, 除非另外指明, 否则分析结果将基于所参考程序的“默认值”。在此所用的“默认值”是指在首次初始化软件时软件最初加载的任何值或参数集。

[0058] 本文所用的标准的重组 DNA 和分子克隆技术为本领域所熟知, 并且描述于 Sambrook, J. 和 Russell, D., Molecular Cloning: A Laboratory Manual, 第三版, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2001); 以及 Silhavy, T.J.、Bennan, M.L. 和 Enquist, L.W. 等人, Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1984); 以及 Ausubel, F.M. 等人, Short Protocols in Molecular Biology, 第 5 版, Current Protocols, John Wiley 和 Sons, Inc., N.Y., 2002。此处所用的另外的方法参见 Methods in Enzymology, 第 194 卷, Guide to Yeast Genetics and Molecular and Cell Biology (部分 A, 2004, Christine Guthrie 和 Gerald R. Fink (编辑), Elsevier Academic Press, San Diego, CA)。

[0059] 本发明涉及具有木糖异构酶活性的工程化酵母菌株。将酵母工程化以利用木糖 (其为得自纤维素类生物质的第二最主要糖) 的挑战是在酵母细胞中产生足够的木糖异构酶活性。木糖异构酶催化木糖向木酮糖的转化, 其为木糖利用途径中的第一步。申请人已经发现特定木糖异构酶多肽的表达在酵母细胞中提供了木糖异构酶活性, 而其它木糖异构酶多肽的表达没有提供活性。表达木糖异构酶活性的酵母细胞提供了宿主细胞用于表达完整木糖利用途径, 从而工程化酵母细胞, 其能使用来源于木质纤维素生物质作为碳源的木糖制备目标化合物, 诸如乙醇、丁醇或 1,3-丙二醇。

[0060] 酵母宿主细胞

[0061] 本发明的酵母细胞为包含功能细菌木糖异构酶以及能够产生目标化合物的那些。优选的目标化合物为具有商业价值的那些, 包括但不限于乙醇、丁醇或 1,3-丙二醇。

[0062] 能够产生目标化学制品或能够被工程化以产生目标化学制品的任何酵母细胞可在本文中用作宿主细胞。此类酵母的示例包括但不限于克鲁维酵母属 (*Kluyveromyces*)、假丝酵母属 (*Candida*)、毕赤酵母属 (*Pichia*)、汉逊酵母属 (*Hansenula*)、裂殖酵母属

(*Schizosaccharomyces*)、克勒克酵母属(*Kloeckera*)、*Schwammiomyces*、耶氏酵母属(*Yarrowia*)和酵母属(*Saccharomyces*)的酵母。

[0063] 可根据本领域所熟知的方法将包含活性细菌木糖异构酶的本发明的酵母细胞工程化。例如,可用包含五碳代谢途径(其包含本发明的细菌的木糖异构酶)的多个基因转化具有从六碳糖制备乙醇的天然能力的酵母细胞。此类细胞能够有氧或厌氧发酵制备乙醇。

[0064] 在其它实施例中,可将酵母细胞工程化以表达合成丁醇或1,3-丙二醇的途径。丁醇合成途径的工程化(包括异丁醇、1-丁醇和2-丁醇)已经公开在例如US 8,206,970、US 20070292927、US 20090155870、US 7,851,188和US 20080182308中,所述专利文献以引用方式并入本文。1,3-丙二醇途径的工程化已经公开在例如US 6,514,733、US 5,686,276、US 7,005,291、US 6,013,494和US 7,629,151中,所述专利文献以引用方式并入本文。

[0065] 就将木糖用作碳源而言,将酵母细胞工程化以表达完整的木糖利用途径。工程化酵母(诸如酿酒酵母(*S.cerevisiae*))用于从木糖制备乙醇,其描述于Matsushika等人(*Appl. Microbiol. Biotechnol.* (2009) 84:37-53)以及Kuyper等人(*FEMS Yeast Res.* (2005) 5:399-409)中。在一个实施例中,除如本文所公开的工程化酵母细胞以具有木糖异构酶活性之外,增加了细胞中其它途径酶的活性以提供在作为唯一碳源的木糖上生长的能力。通常增加五种五碳糖途径酶的活性水平:木酮糖激酶(XKS1)、转醛醇酶(TAL1)、转酮醇酶1(TKL1)、D-核酮糖-5-磷酸3-表异构酶(RPE1)、和核糖5-磷酸酮醇异构酶(RKI1)。可使用本领域的技术人员已知的用于增加基因表达的任何方法。例如,如本文所述,在实例1中,这些活性可通过使用高活性启动子表达每个蛋白质的宿主编码区来增加。构建用于表达的嵌合基因并整合进酵母基因组。另选地,可在酵母细胞中表达这些酶的异源编码区以获得增加的酶活性。就工程化能够代谢木糖的酵母的附加方法而言,参见例如,US7622284B2、US8058040B2、US 7,943,366B2、W02011153516A2、W02011149353A1、W02011079388A1、US20100112658A1、US20100028975A1、US20090061502A1、US20070155000A1、W02006115455A1、US20060216804A1和US8129171B2。

[0066] 在一个实施例中,所述酵母细胞具有如下所述的木糖异构酶活性和附加的遗传工程化以提供如上所述的完整木糖利用途径。这些细胞能够在包含木糖作为唯一碳源的培养基中生长。更典型地,这些细胞在包含木糖以及其它糖诸如葡萄糖和阿拉伯糖的培养基中生长。这允许有效使用存在于水解培养基中的糖,所述水解培养基通过预处理和糖化由纤维素类生物质制备。

[0067] 木糖异构酶

[0068] 酵母细胞中的木糖异构酶的表达一直存在问题;具体地讲,已发现当在酵母细胞中表达时,许多细菌木糖异构酶具有很少或不具有活性。在所述重组酵母细胞中,通过表达异源性核酸分子提供了木糖异构酶活性,所述异源性核酸分子编码多肽,所述多肽的氨基酸序列与SEQ ID NO:1,3,5,或7的氨基酸序列具有至少约85%的序列同一性。用来自生黄瘤胃球菌(*Ruminococcus flavefaciens*)FD-1的木糖异构酶序列(SEQ ID NO:11)和来自*Ruminococcus chamanellensis* 18P13的木糖异构酶序列(SEQ ID NO:9),通过BLAST检索GenBank数据库(National Center for Biotechnology Information (NCBI);Benson等人*Nucleic Acids Research*,2011 Jan;39(数据库专刊):D32-7)对这些序列进行鉴定。SEQ ID NO:11与US 2012/0184020中的生黄瘤胃球菌(*Ruminococcus flavefaciens*)木糖异构

酶SEQ ID NO:31相同。

[0069] SEQ ID NO:1,3,5和7是分别来自毛螺科菌 (*Lachnospiraceae bacterium*) ICM7 (本文中被称为Hm1)、毛螺科菌 (*Lachnospiraceae bacterium*) 口腔菌群类107菌株F0167 (本文中被称为Hm2)、毛螺科菌 (*Lachnospiraceae bacterium*) 口腔菌群类082菌株F0431 (本文中被称为Hm3) 以及砂优杆菌 (*Eubacterium saburreum*) DSM 3986 (本文中被称为Hm4) 的细菌木糖异构酶的氨基酸序列。如表2中所示,这四个序列与生黄瘤胃球菌 (*Ruminococcus flavefaciens*) FD-1 (SEQ ID NO:11) 和 *Ruminococcus chamanellensis* 18P13 (SEQ ID NO:9) 序列的同一性在60.9%和62.6%之间。如表2中所示,这四个序列与来自软弱贫养菌 (*Abiotrophis defectiva*) ATCC 49176的一种假定蛋白 (SEQ ID NO:12; Accession#ZP 04453767) (其与WO 2102/009272的SEQ ID NO:2相同并在其中被鉴定为软弱贫养菌 (*Abiotrophis defectiva*) 木糖异构酶) 的同一性在71.7%和73.2%之间。

[0070] 本文 (实例3) 发现,编码Hm1, Hm2, Hm3, 和Hm4的核酸分子在酿酒酵母 (*S.cerevisiae*) 中的表达允许包含木糖利用途径但缺乏木糖异构酶活性的酿酒酵母 (*S.cerevisiae*) 菌株在包含作为唯一糖类的木糖的培养基中生长。由酵母细胞利用木糖并制备乙醇。因此,Hm1, Hm2, Hm3, 和Hm4中的每个的表达提供了木糖异构酶活性以使酵母细胞中木糖利用途径完整。在Hm1, Hm2, Hm3, 和Hm4其间,如表2中所示,序列同一性在92.2%至95.7%的范围内。

[0071] 在该酵母细胞中可表达具有木糖异构酶活性并且与SEQ ID NO:1,3,5和7的任一个具有至少约85%同一性的任何多肽。在各种实施例中,所述多肽可具有与SEQ ID NO:1、3、5和7中的任一个约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或高达100%的氨基酸序列同一性。

[0072] 当用本发明的木糖异构酶转化时,酿酒酵母展示当在包含木糖的培养基中生长时,增加生长、木糖利用率和乙醇收率。具有与SEQ ID NO:1,3,5和7多达95%的同一性的木糖异构酶蛋白的效果并不相同,表明在酵母宿主中有活性的酶的能力可能不取决于序列。具体地,来自 *Leptotrichia goodfellowii* F0264 (本文中被称为0ra1-2; SEQ ID NO:13) 和塞巴鲁德菌属 (*Sebaldella termitidis*) ATCC 33386 (本文中被称为Term-1; SEQ ID NO:15) 的序列具有与生黄瘤胃球菌 (*Ruminococcus flavefaciens*) FD-1 (SEQ ID NO:11) 和 *Ruminococcus chamanellensis* 18P13 (SEQ ID NO:9) 木糖异构酶相似的序列同一性Hm1、Hm2、Hm3、和Hm4 (参见表2),但未提供如本文实例3中测试的酵母细胞中的木糖异构酶活性。

[0073] 表2:木糖异构酶氨基酸序列同一性的比较

	Hm1	Hm2	Hm3	Hm4	Oral2	Term1	R. f. XI	R. c. XI
Hm1								
Hm2	95.5							
Hm3	93.6	92.3						
Hm4	92.9	93.2	95.7					
<i>Leptotrichia goodfellowii</i> F0264 (Oral2)	57.4	57.2	56.1	57.0				
[0074] 塞巴鲁德菌属 (<i>Sebaldella termitidis</i>) ATCC 33386 (Term1)	55.8	55.1	54.9	55.8	85.2			
来自生黄瘤胃球菌 (<i>R. flavefaciens</i>) 的 XI	62.2	61.5	60.9	60.9	61.7	59.9		
来自 <i>R. chamanellensis</i> 的 XI	62.6	61.9	61.8	62.2	60.7	61.2	77.4	
来自软弱贫养菌 (<i>A. defectiva</i>) 的 XI	73.2	72.5	71.7	72.1	57.4	54.9	61.9	61.0

[0075] 在酵母细胞中提供木糖异构酶活性的所述氨基酸序列不是酵母细胞原生的,因此其编码核酸序列对于酵母细胞是异源性的。就表达而言,可使用对于期望的酵母细胞优化的密码子来设计编码本发明多肽的核酸分子,如本领域技术人员所熟知的。例如,为了在酿酒酵母 (*Saccharomyces cerevisiae*) 中表达Hm1、Hm3、Hm5、或Hm7,使用针对在酿酒酵母 (*S.cerevisiae*) 表达中的密码子优化设计被命名为xy1A (Hm1) (SEQ ID NO:2)、xy1A (Hm2) (SEQ ID NO:4)、xy1A (Hm3) (SEQ ID NO:6) 和xy1A (Hm4) (SEQ ID NO:8) 的核酸分子。

[0076] 酵母中基因表达的方法是本领域已知的(参见例如Methods in Enzymology, 第194卷,Guide to Yeast Genetics and Molecular and Cell Biology,部分A,2004, Christine Guthrie和Gerald R.Fink (编辑),Elsevier Academic Press, San Diego,CA)。基因在酵母中的表达通常需要可操作地连接至所关注的编码区的启动子,和转录终止子。可使用多个酵母启动子来构建编码期望的蛋白质的基因的表达盒,包括但不限于组成型启动子FBA1、GPD1、ADH1、GPM、TPI1、TDH3、PGK1、ILV5p,以及诱导型启动子GAL1、GAL10、和CUP1。适宜的转录终止子包括但不限于FBAt、GPDt、GPMt、ERG10t、GAL1t、CYC1、ADH1t、TAL1t、TKL1t、ILV5t和ADHt。

[0077] 可将合适的启动子、转录终止子和编码区克隆进大肠杆菌 (*E.coli*) -酵母穿梭载体,并转化进酵母细胞。这些载体允许在大肠杆菌 (*E.coli*) 和酵母菌株两者中增殖。

[0078] 载体通常包含选择性标记和在目标宿主中允许自主复制或染色体整合的序列。在酵母中通常使用的质粒是穿梭载体pRS423、pRS424、pRS425和pRS426(美国典型培养物保藏中心, Rockville, MD),它们包含大肠杆菌复制起点(例如, pMB1)、酵母2μ复制起点,以及用于营养选择的标记。这四种载体的选择标记物是His3(载体pRS423)、Trp1(载体pRS424)、Leu2(载体pRS425)和Ura3(载体pRS426)。可用的另外的载体包括pHR81(ATCC#87541)和pRS313(ATCC#77142)。可通过在大肠杆菌 (*E.coli*) 中标准分子克隆技术或酵母中缺口修复重组方法进行编码期望蛋白的嵌合基因的表达载体的构建。

[0079] 缺口修复克隆方法利用了酵母中的高效同源重组。典型地,酵母载体DNA被消化(例如,在其多克隆位点),以在其序列中产生“缺口”。然后将“带缺口的”载体和具有顺序重

叠末端(彼此重叠并且与带缺口的载体末端重叠,以期望的插入序列顺序排列)的插入DNA共转化酵母细胞,将所述酵母细胞涂布在包含合适的化合物混合物(允许质粒上的营养选择标记互补)的培养基上。能使用从所选择的细胞制备的质粒DNA通过PCR分型确认正确插入序列组合的存在。然后将从酵母分离的质粒DNA转化进大肠杆菌(E.coli)菌株,例如TOP10,随后通过微量制备和限制分型以进一步验证所述质粒构建体。最后,所述构建体能通过序列分析进行验证。

[0080] 与缺口修复技术类似,整合进酵母基因组也利用了酵母中的同源重组系统。通常,包含编码区加控制元件(启动子和终止子)以及营养缺陷标记的盒是用高保真DNA聚合酶PCR扩增的,使用的引物与盒杂交并且包含与期望插入的基因组区域的5'和3'区同源的40-70个碱基对的序列。然后将PCR产物转化到酵母细胞中,其被铺板于包含适当的化合物混合物的培养基上,所述混合物允许对所整合的营养缺陷型标记物的选择。可通过菌落PCR或通过对染色体DNA直接测序验证转化体。

[0081] 本发明提供按照上述教导内容制备酵母细胞的方法,所述酵母细胞具有木糖异构酶活性。在一个实施例中,将编码多肽的异源核酸分子导入酵母菌株,所述多肽具有木糖异构酶活性并且其氨基酸序列与SEQ ID N0:1,3,5,或7的氨基酸序列的任一个具有至少85%的序列同一性。在各种实施例中,所述多肽的氨基酸序列具有SEQ ID N0:1、3、5和7中的任一个的至少约85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或高达100%。

[0082] 在一个实施例中,被导入的核酸分子是嵌合基因的一部分,如上所述,该嵌合基因被导入酵母细胞用于表达。

[0083] 在一个实施例中,将所述核酸分子导入具有其它遗传修饰的酵母细胞,如上所述对于酵母宿主细胞,一旦引入木糖异构酶活性,将提供完整的木糖利用途径。木糖异构酶活性的引入和另外的遗传修饰可以任何顺序进行,和/或与两种或更多种引入/修饰同时进行。这些细胞能够在包含木糖作为唯一碳源的培养基中生长。更典型地,这些细胞在包含木糖以及其它糖诸如葡萄糖和阿拉伯糖的培养基中生长。这允许有效使用存在于水解培养基中的糖,所述水解培养基通过预处理和糖化由纤维素类生物质制备。

[0084] 在另一个实施例中,将所述核酸分子导入具有制备目标化学制品的代谢途径的酵母细胞。木糖异构酶活性和代谢途径的引入可以任何顺序进行,和/或与两种或更多种遗传修饰同时进行。目标化合物的示例包括乙醇、丁醇和1,3-丙二醇。上文描述了包含用于制备目标化学制品的代谢途径的酵母细胞。

[0085] 实例

[0086] 本发明将在以下实例中进一步限定。应该理解,这些实例尽管说明了本发明的优选实例,但仅是以例证的方式给出的。通过上述论述和这些实例,本领域的技术人员可确定本发明的必要特征,并且在不脱离本发明的实质和范围内的前提下,可对本发明进行各种变化和修改以适应多种用途和条件。

[0087] 基本方法

[0088] 缩写词的含义如下:“kb”是指千碱基,“bp”是指碱基对,“nt”是指核苷酸,“hr”是指小时,“min”是指分钟,“sec”是指秒钟,“d”是指天数,“L”是指升,“ml”或“mL”是指毫升,“μL”是指微升,“μg”是指微克,“ng”是指纳克,“mg”是指毫克,“mM”是指毫摩,“μM”是指微

摩,“nm”是指纳摩,“ μmol ”是指微摩尔,“pmol”是指皮摩尔,“XI”是指木糖异构酶,“nt”是指核甘酸。

[0089] 本文使用的标准重组DNA和分子克隆技术是本领域所熟知的,并且描述于:Sambrook, J., Fritsch, E.F. 和 Maniatis, T., Molecular Cloning: A Laboratory Manual, 第2版, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989) (下文中称为“Maniatis”); 以及 Silhavy, T.J.、Bennan, M.L. 和 Enquist, L.W., Experiments with Gene Fusions, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); 以及 Ausubel 等人, Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ 出版 (1987), 以及 Methods in Yeast Genetics, 2005, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY。

[0090] HPLC分析

[0091] 在固定的时间间隔采集细胞培养样品并分析EtOH和木糖,所述分析使用Waters HPLC系统(Alliance system, Waters Corp., Milford, MA)或Agilent 1100 Series LC进行; 条件=0.6mL/min的0.01N H₂SO₄,注射体积=10 μL ,自动取样机温度=10°C,柱温=65°C,运行时间=25分钟,通过折射率进行检测(保持在40°C)。HPLC柱购自BioRad (Aminex HPX-87H, BioRad Inc., Hercules, CA)。通过折射率检测定量分析物并与已知的标准品进行比较。

[0092] 实例1

[0093] 酿酒酵母中天然戊糖途径的增量调节

[0094] 除了活性木糖异构酶的表达之外,稳健戊糖途径对于在酿酒酵母中在限氧条件下有效使用木糖和乙醇制备而言也是必要的。戊糖途径由五种酶组成。在酿酒酵母(*S. cerevisiae*)中,这些蛋白是木酮糖激酶(XKS1)、转醛醇酶(TAL1)、转酮醇酶1(TKL1)、D-核酮糖-5-磷酸3-表异构酶(RPE1)、和核糖5-磷酸酮醇异构酶(RKI1)。为了增加这些蛋白质的表达,将来自酿酒酵母基因组的其编码区克隆以在不同启动子下表达并整合到酿酒酵母染色体中。对于整合选择编码醛糖还原酶的GRE3基因座。为构建此类菌株,第一步为在GRE3中构建被称为P5整合载体的整合载体。

[0095] SEQ ID NO:18给出了GRE3中P5整合载体的序列,后面的数字是指在该载体序列中的核苷酸位置。给定nt数之间的间隙包括包含限制位点的序列区。TAL1编码区(15210至16217)用TPI1启动子(14615至15197)表达并使用TAL1t终止子。RPE1(13893至14609)编码区用FBA1启动子(13290至13879)表达,并且使用在TPI1启动子上游末端的终止子。RKI1编码区(nt 11907至12680)用TDH3启动子(11229至11900)表达并使用GPDt(先前被称为TDH3t)终止子。TKL1编码区(nt 8830至10872)用PGK1启动子(nt 8018至8817)表达并使用TKL1t终止子。XKS1编码区(nt 7297至5495至)用I1v5启动子(nt 8009至7310)表达并使用ADH终止子。在该整合载体中,URA3标记物(nt 332至1135)侧面与loxP位点(nt 42至75和nt 1513至1546)相接以循环利用标记物。所述载体包含GRE3基因座(nt 1549至2089以及nt 4566至5137)的整合臂。在整合前,该GRE3中的P5整合载体可通过用KasI酶消化来线性化。

[0096] 本研究选用的酵母菌株是BP1548,其为来源于原养型二倍体菌株CBS8272(Centraalbureau voor Schimmelcultures (CBS) 荷兰真菌生物多样性中心)的单倍体菌株。该菌株属于于酿酒酵母菌株的CEN.PK谱系。BP1548包含MAT α 交配型,缺失URA3和HIS3基

因。

[0097] 为制备BP1548,首先使CBS 8272形成孢子,并且使用标准程序将四分染色体分裂成四个单倍体菌株(Amberg等人,Methods in Yeast Genetics,2005)。选择MAT α 单倍体中的一个,PNY0899,用于进一步修饰。使用侧面与loxP位点、引物结合位点相接的KanMX缺失盒以及待删除的URA3区之外的同源序列,通过同源重组来删除URA3编码序列(ATG至终止密码子)和URA3编码序列的序列上游的130bp。在使用cre重组酶移除KanMX标记后,剩下包含侧翼为引物结合位点的loxP位点的95bp序列作为基因组中的URA3删除痕迹(SEQ ID NO:19)。该序列位于基因组中URA3上游序列(SEQ ID NO:20)和URA3下游序列(SEQ ID NO:21)之间。使用无痕方法通过同源重组移除所述HIS3编码序列(ATG最多至终止密码子)。所述缺失将HIS3编码序列的原初上游的基因组序列(SEQ ID NO:22)和下游的基因组序列(SEQ ID NO:23)连接起来。使用来自Zymo Research (Irvine, CA) 的Frozen-EZ酵母转化II试剂盒,将包含GRE3中P5整合载体的所有五个戊糖途径基因的所述KasI整合片段转化进BP1548菌株。在缺乏尿嘧啶的合成缺陷型培养基上选择转化体。为了回收URA3标记,将CRE重组酶载体pJT254 (SEQ ID NO:24) 转化进这些整合的菌株。该载体来源于pRS413,cre编码区(核苷酸2562至3593)在GAL1启动子(核苷酸2119至2561)的控制下。选择在SD(-尿嘧啶)培养基上不再生长的菌株。采用在YPD培养基上进一步传代以固化质粒pJT257。所得到的菌株被命名为C52-79。

[0098] 实例2

[0099] 细菌木糖异构酶的选择和表达

[0100] 为了鉴定在酵母中表达时可具有活性的候选细菌木糖异构酶,在对GenBank数据库进行BLAST检索(National Center for Biotechnology Information (NCBI); Benson等人Nucleic Acids Research, 2011Jan; 39 (数据库专刊): D32-7)中,我们使用了来自生黄瘤胃球菌(*Ruminococcus flavefaciens*)FD-1的木糖异构酶的氨基酸序列(SEQ ID NO:11)和来自*Ruminococcus chamanellensis* 18P13的木糖异构酶的氨基酸序列(SEQ ID NO:9)。基于序列同一性,从这个检索中选取了六种细菌木糖异构酶用于测试。这些是来自毛螺科菌(*Lachnospiraceae bacterium*) ICM7 (SEQ ID NO:1)、毛螺科菌(*Lachnospiraceae bacterium*) 口腔菌群类107菌株F0167 (SEQ ID NO:3)、毛螺科菌(*Lachnospiraceae bacterium*) 口腔菌群类082菌株F0431 (SEQ ID NO:5)、砂优杆菌(*Eubacterium saburreum*) DSM 3986 (SEQ ID NO:7)、*Leptotrichia goodfellowii* F0264 (SEQ ID NO:13) 和塞巴鲁德菌属(*Sebaldella termitidis*) ATCC 33386 (SEQ ID NO:15) 的推定的木糖异构酶。使用针对在酿酒酵母(*S.cerevisiae*)中表达的密码子优化合成编码这些蛋白的DNA序列,分别命名为xylA (Hm1) (SEQ ID NO:2)、xylA (Hm2) (SEQ ID NO:4)、xylA (Hm3) (SEQ ID NO:6)、xylA (Hm4) (SEQ ID NO:8)、xylA (Ora1-2) (SEQ ID NO:14) 和xylA (Term1) (SEQ ID NO:X16)。此外,合成了针对*Ruminococcus chamanellensis* 18P13木糖异构酶的密码子优化的编码区并命名为xylA-10 (SEQ ID NO:10)。

[0101] 使用酿酒酵母(*S.cerevisiae*)乙酰羟酸还原异构酶基因的1184个核苷酸的启动子(ILV5p)和酿酒酵母(*S.cerevisiae*)乙酰羟酸还原异构酶基因的635个核苷酸的终止子(ILV5t)表达所述合成的xylA编码区。嵌合基因位于pHR81基穿梭载体中的NotI和XhoI位点之间,其中编码区位于PmeI和SfiI位点之间。pHR81载体(ATCC#87541)包含pMB1起点和氨苄

青霉素抗性 (ampR) 标记分别允许在大肠杆菌 (E.coli) 中质粒增殖和选择。此外, pHR81 具有 2 微米复制起点、URA3 选择标记和用于在酵母中增殖和选择的 LEU 2-d, 当在缺乏亮氨酸的培养基中生长时所述 (元件) 赋予在酿酒酵母 (S.cerevisiae) 中的高拷贝数。包含 ILVp-xy1A (Hm1) -ILV5t 嵌合基因的 pHR81 载体的序列是 SEQ ID NO:17。包含其它编码区的载体是相同的, 不同之处在于在 ILV5p 和 ILV5t 之间、在 PmeI 和 SfiI 位点之间每个独立编码区的替换。xy1A (Hm1) 载体被命名为 pHR81 ilv5p xy1A (Hm1), 其它载体有相同名称, 除了替换了特定的 xy1A 编码区命名。将这些构建体转化到 C52-79 菌株 (实例 1) 中, 并且在包含缺少尿嘧啶的合成葡萄糖培养基的板上选择转化体: 6.7g/L 不具有氨基酸的酵母氮基质 (Amresco, Solon, OH), 0.77g/L 负尿嘧啶除去补充剂 (Clontech Laboratories, Mountain View, CA), 20g/L 葡萄糖。然后测试转化体的生长和乙醇制备。

[0102] 实例 3

[0103] 包含不同细菌木糖异构酶的酿酒酵母 (S.cerevisiae) 的生长和乙醇制备

[0104] 酿酒酵母菌株 C52-79 (实例 1) 缺乏使用木糖作为能量和碳源的能力, 因为其缺乏木糖异构酶活性。在 YPX 培养基 (10g/1 酵母提取物, 20g/1 蛋白胨, 和 40g/1 木糖) 中测试了表达 xy1A (Hm1)、xy1A (Hm2)、xy1A (Hm3)、xy1A (Hm4)、xy1A (Ora1-2)、xy1A (Term1) 和 xy1A-10 嵌合基因的酵母菌株。为了进行这个测试, 将菌株接种至 50ml 组织培养管中的 10ml YPX 培养基中, 起始 OD₆₀₀ 为 0.5。将盖子紧密闭合并将试管放置在设定为 225 rpm 速度的 30°C 旋转摇动器中。在不同时间间隔 (24 小时, 4 小时和 72 小时) 取样并通过如基本方法中所述的 HPLC 分析测定木糖和乙醇浓度, 以及记录 OD₆₀₀。对于每个菌株进行三种单独培养物的生长和分析。将每组三次平行测定的结果取平均值。在相同时间点测定带有 xy1A (Hm1)、xy1A (Ora1-2)、xy1A (Term-1) 和 xy1A-10 的菌株。在相同时间点测定带有 xy1A (Hm2) 和 xy1A (Hm3) 的菌株。单独测定带有 xy1A (Hm4) 的菌株。在表 3 中给出了所有的结果。

[0105] 表 3: 表达各种木糖异构酶的酵母菌株的生长、木糖消耗和乙醇制备

[0106]

菌株中的载体	OD ₆₀₀		消耗的木糖 (g/L)		制备的乙醇 (g/L)	
	24 小时后					
	平均	标准 偏差	平均	标准 偏差	平均	标准 偏差
pHR81 ilv5p xylA (Hm1)	11.85	0.40	33.86	0.70	13.07	0.29
pHR81 ilv5p xylA (Hm2)	11.60	0.36	32.08	2.34	12.56	0.82
pHR81 ilv5p xylA (Hm3)	10.36	0.20	24.25	1.33	9.51	0.60
pHR81 ilv5p xylA (Hm4)	6.54	0.30	7.65	0.72	2.57	0.30
pHR81 ilv5p xylA (Oral-2)	2.88	0.09	0.52	0.08	0.00	0.00
pHR81 ilv5p xylA (Term-1)	2.35	0.56	0.47	0.07	0.00	0.00
pHR81 ilv5p xylA (xylA-10)	3.00	0.41	0.40	0.13	0.00	0.00
44 小时后						
pHR81 ilv5p xylA (Hm1)	12.79	0.51	40.00	0.00	15.82	0.11
pHR81 ilv5p xylA (Hm2)	13.07	0.21	39.92	0.00	15.17	0.33
pHR81 ilv5p xylA (Hm3)	12.48	0.26	39.92	0.00	16.20	0.18
pHR81 ilv5p xylA (Hm4)	11.26	0.84	31.50	2.67	11.65	1.20
pHR81 ilv5p xylA (Oral-2)	2.88	0.09	0.60	0.15	0.00	0.00
pHR81 ilv5p xylA (Term-1)	2.78	0.29	0.04	0.09	0.00	0.00
pHR81 ilv5p xylA (xylA10)	3.22	0.50	0.64	0.12	0.00	0.00
72 小时后						
pHR81 ilv5p xylA (Oral-2)	2.70	0.35	0.71	0.24	0.00	0.00
pHR81 ilv5p xylA (Term-1)	2.61	0.17	1.20	0.06	0.00	0.00
pHR81 ilv5p xylA (xylA-10)	3.02	0.13	1.00	0.10	0.00	0.00

[0107] 如表3中所示,在24小时点测量时,包含表达Hm1、Hm2、Hm3和Hm4的嵌合基因的酵母菌株消耗木糖的同时产生乙醇。在温育44小时后,基本上所有的木糖都被消耗了,并且由表达Hm1、Hm2和Hm3的菌株制备超过15g/L的乙醇。对于表达Hm4的菌株,在44小时后消耗了大多数的木糖,制备了约11g/L的乙醇。这些结果表明在酿酒酵母(*S.cerevisiae*)中Hm1、Hm2、Hm3和Hm4表达为有活性的木糖异构酶。然而,甚至在72小时后表达其它的Oral-2、Term-1和XylA10的菌株几乎没有消耗木糖并且没有产生乙醇。

序列表

<110> Zhongqiang Chen
Kelly, Kristen
Ye, RIck Weizhang

<120> 酵母细胞中的细菌木糖异构酶活性

<130> CL5984

<160> 24

<170> PatentIn版本3.5

<210> 1

<211> 442

<212> PRT

<213> 毛螺科菌ICM7

<400> 1

Met Lys Glu Phe Phe Pro Ser Ile Ser Pro Ile Lys Phe Glu Gly Ser
1 5 10 15

Glu Ser Lys Asn Pro Leu Ser Phe Lys Tyr Tyr Asp Ala Lys Arg Val
20 25 30

Ile Met Gly Lys Thr Met Glu Glu His Leu Ser Phe Ala Met Ala Trp
35 40 45

Trp His Asn Leu Cys Ala Ser Gly Val Asp Met Phe Gly Gln Gly Thr
50 55 60

[0001] Ala Asp Lys Gly Phe Gly Glu Asn Leu Gly Thr Met Glu His Ala Lys
65 70 75 80

Ala Lys Val Asp Ala Gly Ile Glu Phe Met Gln Lys Leu Gly Ile Lys
85 90 95

Tyr Tyr Cys Phe His Asp Thr Asp Ile Val Pro Glu Asp Gln Glu Asp
100 105 110

Ile Asn Val Thr Asn Ala Arg Leu Asp Glu Ile Thr Asp Tyr Ile Leu
115 120 125

Glu Lys Thr Lys Gly Thr Asp Ile Lys Cys Leu Trp Ala Thr Cys Asn
130 135 140

Met Phe Ser Asn Pro Arg Phe Met Asn Gly Ala Gly Ser Ser Asn Ser
145 150 155 160

Ala Asp Val Phe Cys Phe Ala Ala Gln Ala Lys Lys Gly Leu Glu
165 170 175

Asn Ala Val Lys Leu Gly Ala Lys Gly Phe Val Phe Trp Gly Gly Arg
180 185 190

Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Met Lys Leu Glu Glu
195 200 205

Asn Ile Ala Thr Leu Phe Thr Met Cys Arg Asp Tyr Gly Arg Ser Ile
210 215 220

Gly Phe Lys Gly Asp Phe Tyr Ile Glu Pro Lys Pro Lys Glu Pro Met
 225 230 235 240

Lys His Gln Tyr Asp Phe Asp Ala Ala Thr Ala Ile Gly Phe Leu Arg
 245 250 255

Lys Tyr Gly Leu Asp Lys Asp Phe Lys Met Asn Ile Glu Ala Asn His
 260 265 270

Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Val Ser Ala
 275 280 285

Ile Asn Gly Met Leu Gly Ser Val Asp Ala Asn Gln Gly Asp Thr Leu
 290 295 300

Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Val Tyr Asp Thr Thr
 305 310 315 320

Leu Ala Met Tyr Glu Ile Leu Lys Ala Gly Gly Leu Ser Gly Gly Leu
 325 330 335

Asn Phe Asp Ser Lys Asn Arg Arg Pro Ser Asn Thr Ala Glu Asp Met
 340 345 350

Phe Tyr Gly Phe Ile Ala Gly Met Asp Thr Phe Ala Leu Gly Leu Ile
 355 360 365

[0002] Lys Ala Ala Gln Ile Ile Glu Asp Gly Arg Ile Asp Glu Phe Val Lys
 370 375 380

Glu Arg Tyr Ser Ser Tyr Asn Ser Gly Ile Gly Glu Lys Ile Arg Asn
 385 390 395 400

Arg Ser Val Thr Leu Val Glu Cys Ala Glu Tyr Ala Leu Lys Met Lys
 405 410 415

Lys Pro Glu Leu Pro Glu Ser Gly Arg Gln Glu Tyr Leu Glu Thr Val
 420 425 430

Val Asn Asn Ile Phe Phe Asn Ser Lys Leu
 435 440

<210> 2
 <211> 1326
 <212> DNA
 <213> 人工序列

<220>
 <223> 经优化以在酿酒酵母中表达的Hml的编码区

<400> 2
 atgaaggagt tttcccatc catctctcca atcaaggatcg aagggttccga atccaagaac 60
 ccattgtctt tcaagtacta cgacgctaag agagttatca tggtaaaac catggaagaa 120
 cacttgtctt tcgctatggc ttgggtggcac aacttgtgtc cttccgtgt tgacatgttc 180
 ggtcaaggta ctgctgacaa gggtttgggt gaaaacttgg gtactatgga acacgctaag 240
 gctaagggtt acgctggat cgagttcatg caaaagggtt ggatacaacttgc tcaactgtttc 300

cacgacacccg atatcggtcc	agaagaccaa gaagatatac	acgtcaccaa cgtagattg	360
gacgaaatca ctgattacat	cttggaaaag accaagggtt	ctgacatcaa gtgttgg	420
gctacttgc	aatgttctc taacccaa	gttgcacaa gttcaactct	480
gtgtacgtt	tctgttgc	tgctgtcaaa gctaagaagg	540
ttgggtgcta	agggttgc	gttggaaa cgctgttaag	600
actgacatga	agttggaga	agaaaacatc gctacccgt	660
ggttagatcta	tcggttcaa	gggtgactc tacatgaac	720
aagcacaat	acgacttgc	caaagccaa ggaaccaatg	780
gacaaggatt	tcaagatgaa	catcgaagct aaccacgta	840
caacacgaat	ttagatttc	ccttggctgg tacacacttc	900
ggtgacactt	tggtgggtt	ggacacgat caattccaa ctaacgttta	960
ttggctatgt	acgaaatctt	cgacacttgc ttgggttgc	1020
aagaacagaa	gaccatccaa	gacatgttct acgggttcat	1080
gacacttgc	cttgggtt	cgctgtatgc gatcgttgc	1140
gaatttgc	aggaaagata	ctttccatc aactctggta	1200
agatccgtt	cttgggtc	tcggtaaaa gatcgttgc	1260
ccagaatctg	gttagacaaga	atactggaa accgtcgta	1320
aagttg		acaacatctt cttaactct	
			1326

[0003]

<210> 3
<211> 442
<212> PRT
<213> 毛螺科菌口腔菌群107

<400> 3

Met	Lys	Glu	Phe	Phe	Pro	Gly	Ile	Ser	Pro	Val	Lys	Phe	Glu	Gly	Ser
1					5				10				15		

Glu	Ser	Lys	Asn	Pro	Leu	Ser	Phe	Lys	Tyr	Tyr	Asp	Ala	Lys	Arg	Val
				20				25				30			

Ile	Met	Gly	Lys	Thr	Met	Glu	Glu	His	Leu	Ser	Phe	Ala	Met	Ala	Trp
					35			40				45			

Trp	His	Asn	Leu	Cys	Ala	Ser	Gly	Val	Asp	Met	Phe	Gly	Gln	Gly	Thr
					50			55			60				

Ala	Asp	Lys	Gly	Phe	Gly	Glu	Ser	Ser	Gly	Thr	Met	Gly	His	Ala	Lys
					65			70			75		80		

Ala	Lys	Val	Asp	Ala	Gly	Ile	Glu	Phe	Met	Lys	Lys	Leu	Gly	Ile	Lys
					85			90				95			

Tyr	Tyr	Cys	Phe	His	Asp	Thr	Asp	Ile	Val	Pro	Glu	Asp	Gln	Glu	Asp
					100			105				110			

Ile	Asn	Val	Thr	Asn	Ala	Arg	Leu	Asp	Glu	Ile	Thr	Asp	Tyr	Ile	Leu
					115			120				125			

Glu Lys Thr Lys Gly Ser Asp Ile Lys Cys Leu Trp Thr Thr Cys Asn
 130 135 140

Met Phe Gly Asn Pro Arg Phe Met Asn Gly Ala Gly Ser Ser Asn Ser
 145 150 155 160

Ala Asp Val Phe Cys Phe Ala Ala Ala Gln Ala Lys Lys Gly Leu Glu
 165 170 175

Asn Ala Val Lys Leu Gly Ala Lys Gly Phe Val Phe Trp Gly Gly Arg
 180 185 190

Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Met Lys Leu Glu Glu Glu
 195 200 205

Asn Ile Ala Thr Leu Phe Thr Met Cys Arg Asp Tyr Gly Arg Ser Ile
 210 215 220

Gly Phe Lys Gly Asp Phe Tyr Ile Glu Pro Lys Pro Lys Glu Pro Met
 225 230 235 240

Lys His Gln Tyr Asp Phe Asp Ala Ala Thr Ala Ile Gly Phe Leu Arg
 245 250 255

Lys Tyr Gly Leu Asp Lys Asp Phe Lys Leu Asn Ile Glu Ala Asn His
 260 265 270

Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Val Ser Ala
 [0004] 275 280 285

Ile Asn Gly Met Leu Gly Ser Val Asp Ala Asn Gln Gly Asp Thr Leu
 290 295 300

Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Ile Tyr Asp Thr Thr
 305 310 315 320

Phe Ala Met Tyr Glu Ile Leu Lys Ala Gly Gly Leu Ser Gly Gly Leu
 325 330 335

Asn Phe Asp Ser Lys Asn Arg Arg Pro Ser Asn Thr Ala Glu Asp Met
 340 345 350

Phe Tyr Gly Phe Ile Ala Gly Met Asp Thr Phe Ala Leu Gly Leu Ile
 355 360 365

Lys Ala Ala Gln Ile Ile Glu Asp Gly Arg Ile Asp Glu Phe Ile Lys
 370 375 380

Glu Arg Tyr Ser Ser Tyr Ser Thr Gly Ile Gly Glu Lys Ile Arg Asn
 385 390 395 400

Lys Ser Val Thr Leu Glu Glu Cys Ala Glu Tyr Ala Ala Lys Leu Lys
 405 410 415

Lys Pro Glu Leu Pro Glu Ser Gly Arg Gln Glu Tyr Leu Glu Thr Val
 420 425 430

Val Asn Asn Ile Leu Phe Asn Ser Lys Leu
435 440

<210> 4
<211> 1326
<212> DNA
<213> 人工序列

<220>
<223> 经优化以在酿酒酵母中表达的L bact ot 107 XI的编码区
<400> 4
atgaaggagt tttcccagg tatctctcca gtcaaggatcg aaggttctga atccaagaac 60
ccattgtctt tcaagtacta cgatgctaag agagttatca tggtaaaaac catgaaagaa 120
cacttgtctt tcgctatggc ttggggcac aacttgttg ctccgggtgt tgacatgttc 180
ggtcaaggta ctgctgacaa gggtttcggt gaatcttccg gtactatggg tcacgctaag 240
gctaagggtt acgctggtat cgagttcatg aagaagttgg gtatcaagta ctactgttc 300
caacgacaccg atatcggtcc agaagaccaa gaagatatac acgtactaa cgctagattg 360
gacgaaatca ccgattacat ctggaaaag actaagggtt ctgacatcaa gtgttgg 420
accacttgc acatgttcgg taaccaaga ttcatgaac gtgttgg 480
gctgacgttt tctgttgcg tgctgctcaa gctaagaagg gttggaaaa cgctgttaag 540
ttgggtgcta agggttcgt ctctgggtt gtagagaag gttacgaaac ctgttgaac 600
actgacatga agttggaga agaaaacatc gctacctgtt tcactatgtg tagagactac 660
ggtagatcta tcggttcaa gggtagttt tacatgaaac caaaggccaa ggaaccaatg 720
aageaccaat acgacttgc tgctgctacc gctatcggtt tcttgagaaa gtacggttt 780
gacaaggatt tcaaggtaa catcgaaatc aaccacgcta cttggctgg tcacacttc 840
caacacgaat tgagagttt tgctatcaac ggtatgttg gtccgttga cgctaaacca 900
ggtgacactt tggtgggtt ggacaccgat caattccaa ctaacatcta cgacaccact 960
ttcgctatgt acgaaatctt gaaggctggt gttttgtctg gtggttgaa ctgcactct 1020
aagaacagaa gaccatccaa caccgctgaa gacatgtt acggttcat cgctggat 1080
gacacttgc ctgggttt gatcaaggct gctcaaatca tcgaagacgg tagaatcgat 1140
gagttcatca agggaaagata ctcttcctac tctaccgtt tcggtaaaaa gatcagaac 1200
aagtccgttta ctggaaaga atgtgctgaa tacgctgta agttgaagaa gccagaattt 1260
ccagaatctg gtagacaaga atacttggaa accgtcgta acaacatctt gttcaactct 1320
aagttt 1326

<210> 5
<211> 439
<212> PRT
<213> 毛螺科菌口腔菌群082
<400> 5

Met Lys Glu Phe Phe Pro Gly Ile Ser Pro Val Lys Phe Glu Gly Lys
1 5 10 15

Asp Ser Lys Asn Pro Leu Ser Phe Lys Tyr Tyr Asp Ala Lys Arg Val
20 25 30

Ile Met Gly Lys Thr Met Glu Glu His Leu Ser Phe Ala Met Ala Trp
35 40 45

Trp His Asn Leu Cys Ala Cys Gly Val Asp Met Phe Gly Gln Gly Thr
 50 55 60

Ile Asp Lys Ser Phe Gly Ala Leu Pro Gly Thr Met Glu His Ala Lys
 65 70 75 80

Ala Lys Val Asp Ala Gly Ile Glu Phe Met Gln Lys Leu Gly Ile Lys
 85 90 95

Tyr Tyr Cys Phe His Asp Thr Asp Ile Val Pro Glu Asp Gln Glu Asp
 100 105 110

Ile Asn Val Thr Asn Ala Arg Leu Asp Glu Ile Thr Asp Tyr Ile Leu
 115 120 125

Glu Lys Thr Lys Gly Thr Asp Ile Lys Cys Leu Trp Thr Thr Cys Asn
 130 135 140

Met Phe Ser Asn Pro Arg Phe Met Asn Gly Ala Gly Ser Ser Asn Ser
 145 150 155 160

Ala Asp Val Phe Cys Phe Ala Ala Gln Ala Lys Lys Gly Leu Glu
 165 170 175

Asn Ala Val Lys Leu Gly Ala Lys Gly Phe Val Phe Trp Gly Gly Arg
 180 185 190

[0006] Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Met Lys Leu Glu Glu
 195 200 205

Asn Ile Ala Thr Leu Phe Thr Met Cys Arg Asp Tyr Gly Arg Ser Ile
 210 215 220

Gly Phe Met Gly Asp Phe Tyr Ile Glu Pro Lys Pro Lys Glu Pro Met
 225 230 235 240

Lys His Gln Tyr Asp Phe Asp Ala Ala Thr Ala Ile Gly Phe Leu Arg
 245 250 255

Lys Tyr Gly Leu Glu Lys Asp Phe Lys Met Asn Ile Glu Ala Asn His
 260 265 270

Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Val Cys Ala
 275 280 285

Val Asn Gly Met Ile Gly Ser Val Asp Ala Asn Gln Gly Asp Thr Leu
 290 295 300

Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Val Tyr Asp Thr Thr
 305 310 315 320

Leu Ala Met Tyr Glu Ile Leu Lys Ala Gly Gly Leu Arg Gly Gly Leu
 325 330 335

Asn Phe Asp Ser Lys Asn Arg Arg Pro Ser Asn Thr Ala Asp Asp Met
 340 345 350

Phe Tyr Gly Phe Ile Ala Gly Met Asp Ala Phe Ala Leu Gly Leu Ile
 355 360 365

Lys Ala Ala Glu Ile Ile Glu Asp Gly Arg Ile Asp Glu Phe Val Lys
 370 375 380

Glu Arg Tyr Ser Ser Tyr Asn Ser Gly Ile Gly Glu Lys Ile Arg Asn
 385 390 395 400

Arg Ala Val Thr Leu Val Glu Cys Ala Glu Tyr Ala Ala Lys Leu Lys
 405 410 415

Lys Pro Glu Leu Pro Asp Ser Gly Lys Gln Glu Tyr Leu Glu Ser Val
 420 425 430

Val Asn Asn Ile Leu Phe Gly
 435

<210> 6

<211> 1317

<212> DNA

<213> 人工序列

<220>

<223> 经优化以在酿酒酵母中表达的L bact ot 082 XI的编码区

<400> 6

atgaaggagt tcttcccagg tatctccccca gtcaaggatcg aaggcaagga ctccaagaac 60

ccattgtctt tcaagtacta cgatgtcaag agagttatca tgggtaaaac catggaagaa 120

[0007]

cacttgtctt tcgctatggc ttgggtggcac aacttgtgtc cttgtgggtg tgacatgttc 180

ggtcaaggta ctatcgataa gtccttcggt gcttgcacag gtactatgga acacgctaag 240

gctaagggtt acgctggat cgagttcatg caaaagggtt ggatcaagta ctactgttc 300

cacgacactg atatcggtcc agaagaccaa gaagatatca acgtcaccaa cgctagattt 360

gacgaaatca ctgattacat cttggaaaag accaagggtt ctgacatcaa gttttgtgg 420

accacttgta acatgttctc taacccaaaga ttcatgaacg gtgcgtggc ttctaaactct 480

gtcgacgttt tctgtttcgc tgctgtcaaa gctaagaagg gtttggaaa cgctgttaag 540

ttgggtgcta agggttcgt cttctgggtt ggttagagaag gttacgaaac cttgttgaac 600

actgacatga agttggaaaga agaaaacatc gtcacattgt tcactatgt tagagactac 660

gttagatcta tcgggttcat ggggtgactt tacatcgaa caaagccaaa ggaaccaatg 720

aagcaccaat acgacttcga tgctgttacc gtcacgggtt tcttgagaaa gtacggttt 780

gaaaaggact tcaagatgaa catcgaaatc aaccacgcta cttggctgg tcacacttcc 840

caacacgaat ttagatgttgc ggtatgttgc gttctgttgc cgctaaaccaa 900

ggtagacact tgggtgggtt ggacacccgtt caattccaa ctaacgtcta cgacaccact 960

ttggctatgt acgaaatctt gaaggctgtt gggttggagag gtgggttgc cttcgttct 1020

aagaacacaa gaccatccaa cactgtgttgc gatatgttgc acgggttcat cgctgggtat 1080

gacgtttcg ctttgggtttt gatcaaggct gtcgttgc tcaaggacgg tagaatcgat 1140

gaatttttgc tggaaagata ctcttgcac aactctggta tcgggtggaaa gatcggaaac 1200

agagctgttgc ttttgggttgc atgtgttgc aactctggta tcgggtggaaa gatcggaaac 1260

ccagactccg gcaagcaaga atactggaa tcggcgtca acaacatctt gttcggt 1317

<210> 7
<211> 457
<212> PRT
<213> 砂优杆菌

<400> 7

Met Lys Thr Lys Asn Asn Ile Ile Cys Thr Ile Ala Leu Lys Gly Asp
1 5 10 15

Ile Phe Met Lys Glu Phe Phe Pro Gly Ile Ser Pro Val Lys Phe Glu
20 25 30

Gly Arg Asp Ser Lys Asn Pro Leu Ser Phe Lys Tyr Tyr Asp Ala Lys
35 40 45

Arg Val Ile Met Gly Lys Thr Met Glu Glu His Leu Ser Phe Ala Met
50 55 60

Ala Trp Trp His Asn Leu Cys Ala Cys Gly Val Asp Met Phe Gly Gln
65 70 75 80

Gly Thr Val Asp Lys Ser Phe Gly Glu Ser Ser Gly Thr Met Glu His
85 90 95

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

[0008] Ile Lys Tyr Tyr Cys Phe His Asp Thr Asp Ile Val Pro Glu Asp Gln
115 120 125

Glu Asp Ile Asn Val Thr Asn Ala Arg Leu Asp Glu Ile Thr Asp Tyr
130 135 140

Ile Leu Glu Lys Thr Lys Asp Thr Asp Ile Lys Cys Leu Trp Thr Thr
145 150 155 160

Cys Asn Met Phe Ser Asn Pro Arg Phe Met Asn Gly Ala Gly Ser Ser
165 170 175

Asn Ser Ala Asp Val Phe Cys Phe Ala Ala Ala Gln Ala Lys Lys Gly
180 185 190

Leu Glu Asn Ala Val Lys Leu Gly Ala Lys Gly Phe Val Phe Trp Gly
195 200 205

Gly Arg Glu Gly Tyr Glu Thr Leu Leu Asn Thr Asp Met Lys Leu Glu
210 215 220

Glu Glu Asn Ile Ala Thr Leu Phe Thr Met Cys Arg Asp Tyr Gly Arg
225 230 235 240

Ser Ile Gly Phe Met Gly Asp Phe Tyr Ile Glu Pro Lys Pro Lys Glu
245 250 255

Pro Met Lys His Gln Tyr Asp Phe Asp Ala Ala Thr Ala Ile Gly Phe

260

265

270

Leu Arg Lys Tyr Gly Leu Asp Lys Asp Phe Lys Leu Asn Ile Glu Ala
 275 280 285

Asn His Ala Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Arg Val
 290 295 300

Cys Ala Val Asn Gly Met Met Gly Ser Val Asp Ala Asn Gln Gly Asp
 305 310 315 320

Thr Leu Leu Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Val Tyr Asp
 325 330 335

Thr Thr Leu Ala Met Tyr Glu Ile Leu Lys Ala Gly Gly Leu Arg Gly
 340 345 350

Gly Leu Asn Phe Asp Ser Lys Asn Arg Arg Pro Ser Asn Thr Ala Asp
 355 360 365

Asp Met Phe Tyr Gly Phe Ile Ala Gly Met Asp Thr Phe Ala Leu Gly
 370 375 380

Leu Ile Lys Ala Ala Glu Ile Ile Glu Asp Gly Arg Ile Asp Asp Phe
 385 390 395 400

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

[0009]

Arg Asn Arg Lys Val Thr Leu Ile Glu Cys Ala Glu Tyr Ala Ala Lys
 420 425 430

Leu Lys Lys Pro Glu Leu Pro Glu Ser Gly Arg Gln Glu Tyr Leu Glu
 435 440 445

Ser Val Val Asn Asn Ile Leu Phe Gly
 450 455

<210> 8

<211> 1371

<212> DNA

<213> 人工序列

<220>

<223> 经优化以在酿酒酵母中表达的E. bact XI的编码区

<400> 8

atagaagacca agaacaacat catctgtact atcgcttga agggtgacat cttcatgaag 60
 gagttcttcc caggtaatctc tccagttaaat ttcgagggtt gagaactctaa gaaccattt 120
 tccttcaagt actacgacgc taagagaggat atcatgggtt aaaccatggg agaacacttg 180
 tcttgcgtt tggcttgggtt gcacaacttg tggcttgggtt gtgttgacat gttcggttca 240
 ggtactgtcg ataagtccctt cggtaatctt tccggtaacta tggaacacgc tagagctt 300
 gttgacgtt gttactgtt catgaagaatgg tgggttacta agtactactt tttccacgtt 360
 actgatatacg ttccagaaga ccaagaatgtt atcaacgtt ccaacgtt tagggacgaa 420
 atcaactgtt acatcttggaa aaagaccaag gacactgttata tcaagtgttt gtggaccact 480
 tgtaacatgt tctttaaccc aagattcatg aacgggtgtt gttttccaa ctccgttac 540

gtttctgtt tcgctgctgc tcaagctaag aagggtttgg aaaacgctgt taagttgggt	600
gctaagggtt tcgtcttcg ggggtttaga gaagggttacg aaacctgtt gaacactgac	660
atgaagttgg aagaagaaaa catcgctacc ttgttcacta tgttagaga ctacggtaga	720
tctatcggtt tcatgggtga cttctacate gaaccaaage caaaggaacc aatgaagcac	780
caatacgact tcgatgctgc taccgtatc ggttcttga gaaagtacgg tttggacaag	840
gatttcaagt tgaacatcga agctaaccac gctaccttgg ctggcacac ttccaacac	900
gaatttggag tttgtgtgtt caacggatgt atgggttctg ttgacgttta ccaagggttac	960
actttgttgg gttgggacac cgatcaattt ccaactaactc tctacgacac cactttggct	1020
atgtacgaaa tcttgaaggc tgggtgtttt agagggttgg tgaacttcaacttcaac	1080
agaagaccat ccaacaccgc tgacgatatg ttctacggtt tcatcgctgg tatggacact	1140
ttcgctttgg gtttggatcaa ggctgttcaatcatcgaa acggtagaaat cgacgatttgc	1200
gttaaggaaa gatacgcttc ttacaactcc ggtatcggtt aaaagatcag aaacagaaag	1260
gtcaccttga tcaatgtgc tgaatacgct gctaagggttga agaagccaga attgccagaa	1320
tccggtagac aagaataactt ggaatccgtc gtcaacaaca tcttggatcggtt	1371
<210> 9	
<211> 441	
<212> PRT	
<213> Ruminococcus champanellensis	
<400> 9	
[0010] Met Ser Glu Phe Phe Thr Gly Ile Ser Lys Ile Pro Phe Glu Gly Lys	
1 5 10 15	
Ala Ser Asn Asn Pro Met Ala Phe Lys Tyr Tyr Asn Pro Asp Glu Val	
20 25 30	
Val Gly Gly Lys Thr Met Arg Glu Gln Leu Lys Phe Ala Leu Ser Trp	
35 40 45	
Trp His Thr Met Gly Gly Asp Gly Thr Asp Met Phe Gly Val Gly Thr	
50 55 60	
Thr Asn Lys Lys Phe Gly Gly Thr Asp Pro Met Asp Ile Ala Lys Arg	
65 70 75 80	
Lys Val Asn Ala Ala Phe Glu Leu Met Asp Lys Leu Ser Ile Asp Tyr	
85 90 95	
Phe Cys Phe His Asp Arg Asp Leu Ala Pro Glu Ala Asp Asn Leu Lys	
100 105 110	
Glu Thr Asn Gln Arg Leu Asp Glu Ile Thr Glu Tyr Ile Ala Gln Met	
115 120 125	
Met Gln Leu Asn Pro Asp Lys Lys Val Leu Trp Gly Thr Ala Asn Cys	
130 135 140	
Phe Gly Asn Pro Arg Tyr Met His Gly Ala Gly Thr Ala Pro Asn Ala	
145 150 155 160	

Asp Val Phe Ala Phe Ala Ala Ala Gln Ile Lys Lys Ala Ile Glu Ile
 165 170 175

Thr Val Lys Leu Gly Gly Lys Gly Tyr Val Phe Trp Gly Gly Arg Glu
 180 185 190

Gly Tyr Glu Thr Leu Leu Asn Thr Asn Met Gly Leu Glu Leu Asp Asn
 195 200 205

Met Ala Arg Leu Leu His Met Ala Val Asp Tyr Ala Arg Ser Ile Gly
 210 215 220

Phe Thr Gly Asp Phe Tyr Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys
 225 230 235 240

His Gln Tyr Asp Phe Asp Thr Ala Thr Val Ile Gly Phe Leu Arg Lys
 245 250 255

Tyr Asn Leu Asp Lys Asp Phe Lys Met Asn Ile Glu Ala Asn His Ala
 260 265 270

Thr Leu Ala Gln His Thr Phe Gln His Glu Leu Arg Val Ala Arg Glu
 275 280 285

Asn Gly Phe Phe Gly Ser Ile Asp Ala Asn Gln Gly Asp Thr Leu Leu
 290 295 300

[0011] Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Thr Tyr Asp Ala Ala Leu
 305 310 315 320

Cys Met Tyr Glu Val Leu Lys Ala Gly Gly Phe Thr Asn Gly Gly Leu
 325 330 335

Asn Phe Asp Ser Lys Ala Arg Arg Gly Ser Phe Glu Met Glu Asp Ile
 340 345 350

Phe His Ser Tyr Ile Ala Gly Met Asp Thr Phe Ala Leu Gly Leu Lys
 355 360 365

Ile Ala Gln Lys Met Ile Asp Asp Gly Arg Ile Asp Gln Phe Val Ala
 370 375 380

Asp Arg Tyr Ala Ser Trp Asn Thr Gly Ile Gly Ala Asp Ile Ile Ser
 385 390 395 400

Gly Lys Ala Thr Met Ala Asp Leu Glu Ala Tyr Ala Leu Ser Lys Gly
 405 410 415

Asp Val Thr Ala Ser Leu Lys Ser Gly Arg Gln Glu Leu Leu Glu Ser
 420 425 430

Ile Leu Asn Asn Ile Met Phe Asn Leu
 435 440

<210> 10
 <211> 1323
 <212> DNA

<213> 人工序列

<220>

<223> 经优化以在酿酒酵母中表达的R champ XI的编码区

<400> 10

atgtccgagt	tcttcactgg	tatctctaag	atccccattcg	aaggcaaggc	ttctaacaac	60
ccaatggcct	tcaagtacta	caaccacagac	gaagttgtcg	gtggtaaaac	catgagagaa	120
caattgaagt	tcgctttgtc	ttgggtggcac	accatgggtg	gtgacggtac	tgatatgttc	180
ggtgtggta	ctactaaca	gaagttcggt	ggtaactgacc	caatggatat	cgctaagaga	240
aaggtaacg	ctgcttcga	attgatggac	aagttgtcca	tcgattactt	ctgttccac	300
gacagagatt	tggcttcaga	agctgacaac	ttgaaggaaa	ccaaccaaag	attggatgaa	360
atcactgaat	acatcgctca	aatgatgca	ttgaaccac	acaagaaggt	tttgtgggt	420
actgctaact	gtttcggtaa	cccaagatac	atgcacggtg	ctggtaactgc	tccaaacgct	480
gacgttttcg	cttgcgtgc	tgctcaa	aagaaggcta	tcgaaatcac	cgttaagttg	540
ggtgtttaag	gttacgtctt	ctgggggtgt	agagaaggtt	acgaaacctt	gttgaacact	600
aacatgggtt	tggaattgga	caacatggct	agattgtgc	acatggctgt	tgactacgct	660
agatctatcg	gtttcacccg	tgacttctac	atcgaacca	agccaaagga	accaactaag	720
caccaatacg	acttcgatac	cgctactgtc	atcggttct	tgagaaagta	caacttggac	780
aaggatttca	agatgaacat	cgaagctaac	cacgctac	tggctcaaca	cacttccaa	840
cacgaattga	gagttgtcgt	agaaaacggt	ttttcggtt	ctatcgacgc	taaccaaggt	900
gacaccttgc	tgggttggga	cactgatcaa	ttcccaacca	acacttacga	cgctgtttg	960
tgtatgtacg	aagtcttgc	ggctgggtgt	ttcaccaacg	gtggtttgc	cttcgactct	1020
aaggctagaa	gagggttcc	cgaaatggaa	gacatcttc	actcttacat	cgctggat	1080
gacactttcg	cttgggttt	gaagatgc	caaaagatga	tcgacgttg	tagaatcgac	1140
caattcggt	ctgatagata	cgcttctgg	aacaccgta	tcgggtctga	catcatctcc	1200
ggtaaagcta	ccatggctga	cttggaaagct	tacgcttgc	ctaagggtga	cgttactgct	1260
tccttgaagt	ccggtagaca	agaattgttgc	aatctatct	tgaacaacat	catgttcaac	1320
ttg						1323

<210> 11

<211> 439

<212> PRT

<213> 生黄瘤胃球菌

<400> 11

Met	Glu	Phe	Phe	Lys	Asn	Ile	Ser	Lys	Ile	Pro	Tyr	Glu	Gly	Lys	Asp
1				5				10				15			

Ser	Thr	Asn	Pro	Leu	Ala	Phe	Lys	Tyr	Tyr	Asn	Pro	Asp	Glu	Val	Ile
				20				25				30			

Asp	Gly	Lys	Lys	Met	Arg	Asp	Ile	Met	Lys	Phe	Ala	Leu	Ser	Trp	Trp
				35		40			45						

His	Thr	Met	Gly	Gly	Asp	Gly	Thr	Asp	Met	Phe	Gly	Cys	Gly	Thr	Ala
					50			55		60					

Asp	Lys	Thr	Trp	Gly	Glu	Asn	Asp	Pro	Ala	Ala	Arg	Ala	Lys	Ala	Lys
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

65	70	75	80
Val Asp Ala Ala Phe Glu Ile Met Gln Lys Leu Ser Ile Asp Tyr Phe			
85	90	95	
Cys Phe His Asp Arg Asp Leu Ser Pro Glu Tyr Gly Ser Leu Lys Asp			
100	105	110	
Thr Asn Ala Gln Leu Asp Ile Val Thr Asp Tyr Ile Lys Ala Lys Gln			
115	120	125	
Ala Glu Thr Gly Leu Lys Cys Leu Trp Gly Thr Ala Lys Cys Phe Asp			
130	135	140	
His Pro Arg Phe Met His Gly Ala Gly Thr Ser Pro Ser Ala Asp Val			
145	150	155	160
Phe Ala Phe Ser Ala Ala Gln Ile Lys Lys Ala Leu Glu Ser Thr Val			
165	170	175	
Lys Leu Gly Gly Thr Gly Tyr Val Phe Trp Gly Gly Arg Glu Gly Tyr			
180	185	190	
Glu Thr Leu Leu Asn Thr Asn Met Gly Leu Glu Leu Asp Asn Met Ala			
195	200	205	
Arg Leu Met Lys Met Ala Val Glu Tyr Gly Arg Ser Ile Gly Phe Lys			
210	215	220	
[0013]			
Gly Asp Phe Tyr Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys His Gln			
225	230	235	240
Tyr Asp Phe Asp Thr Ala Thr Val Leu Gly Phe Leu Arg Lys Tyr Gly			
245	250	255	
Leu Asp Lys Asp Phe Lys Met Asn Ile Glu Ala Asn His Ala Thr Leu			
260	265	270	
Ala Gln His Thr Phe Gln His Glu Leu Cys Val Ala Arg Thr Asn Gly			
275	280	285	
Ala Phe Gly Ser Ile Asp Ala Asn Gln Gly Asp Pro Leu Leu Gly Trp			
290	295	300	
Asp Thr Asp Gln Phe Pro Thr Asn Ile Tyr Asp Thr Thr Met Cys Met			
305	310	315	320
Tyr Glu Val Ile Lys Ala Gly Gly Phe Thr Asn Gly Gly Leu Asn Phe			
325	330	335	
Asp Ala Lys Ala Arg Arg Gly Ser Phe Thr Pro Glu Asp Ile Phe Tyr			
340	345	350	
Ser Tyr Ile Ala Gly Met Asp Ala Phe Ala Leu Gly Tyr Lys Ala Ala			
355	360	365	
Ser Lys Leu Ile Ala Asp Gly Arg Ile Asp Ser Phe Ile Ser Asp Arg			

370

375

380

Tyr Ala Ser Trp Ser Glu Gly Ile Gly Leu Asp Ile Ile Ser Gly Lys
 385 390 395 400

Ala Asp Met Ala Ala Leu Glu Lys Tyr Ala Leu Glu Lys Gly Glu Val
 405 410 415

Thr Asp Ser Ile Ser Ser Gly Arg Gln Glu Leu Leu Glu Ser Ile Val
 420 425 430

Asn Asn Val Ile Phe Asn Leu
 435

<210> 12
 <211> 440
 <212> PRT
 <213> 软弱贫养菌

<400> 12

Met Ser Glu Leu Phe Gln Asn Ile Pro Lys Ile Lys Tyr Glu Gly Ala
 1 5 10 15

Asn Ser Lys Asn Pro Leu Ala Phe His Tyr Tyr Asp Ala Glu Lys Ile
 20 25 30

Val Leu Gly Lys Thr Met Lys Glu His Leu Pro Phe Ala Met Ala Trp
 35 40 45

[0014]

Trp His Asn Leu Cys Ala Ala Gly Thr Asp Met Phe Gly Arg Asp Thr
 50 55 60

Ala Asp Lys Ser Phe Gly Leu Glu Lys Gly Ser Met Glu His Ala Lys
 65 70 75 80

Ala Lys Val Asp Ala Gly Phe Glu Phe Met Glu Lys Leu Gly Ile Lys
 85 90 95

Tyr Phe Cys Phe His Asp Val Asp Leu Val Pro Glu Ala Cys Asp Ile
 100 105 110

Lys Glu Thr Asn Ser Arg Leu Asp Glu Ile Ser Asp Tyr Ile Leu Glu
 115 120 125

Lys Met Lys Gly Thr Asp Ile Lys Cys Leu Trp Gly Thr Ala Asn Met
 130 135 140

Phe Ser Asn Pro Arg Phe Val Asn Gly Ala Gly Ser Thr Asn Ser Ala
 145 150 155 160

Asp Val Tyr Cys Phe Ala Ala Ala Gln Ile Lys Lys Ala Leu Asp Ile
 165 170 175

Thr Val Lys Leu Gly Gly Arg Gly Tyr Val Phe Trp Gly Gly Arg Glu
 180 185 190

Gly Tyr Glu Thr Leu Leu Asn Thr Asp Val Lys Phe Glu Gln Glu Asn
 195 200 205

Ile Ala Asn Leu Met Lys Met Ala Val Glu Tyr Gly Arg Ser Ile Gly
 210 215 220

Phe Lys Gly Asp Phe Tyr Ile Glu Pro Lys Pro Lys Glu Pro Met Lys
 225 230 235 240

His Gln Tyr Asp Phe Asp Ala Ala Thr Ala Ile Gly Phe Leu Arg Gln
 245 250 255

Tyr Gly Leu Asp Lys Asp Phe Lys Leu Asn Ile Glu Ala Asn His Ala
 260 265 270

Thr Leu Ala Gly His Ser Phe Gln His Glu Leu Arg Ile Ser Ser Ile
 275 280 285

Asn Gly Met Leu Gly Ser Val Asp Ala Asn Gln Gly Asp Met Leu Leu
 290 295 300

Gly Trp Asp Thr Asp Glu Phe Pro Phe Asp Val Tyr Asp Thr Thr Met
 305 310 315 320

Cys Met Tyr Glu Val Leu Lys Asn Gly Gly Leu Thr Gly Gly Phe Asn
 325 330 335

Phe Asp Ala Lys Asn Arg Arg Pro Ser Tyr Thr Tyr Glu Asp Met Phe
 340 345 350

[0015] Tyr Gly Phe Ile Leu Gly Met Asp Ser Phe Ala Leu Gly Leu Ile Lys
 355 360 365

Ala Ala Lys Leu Ile Glu Glu Gly Thr Leu Asp Asn Phe Ile Lys Glu
 370 375 380

Arg Tyr Lys Ser Phe Glu Ser Glu Ile Gly Lys Lys Ile Arg Ser Lys
 385 390 395 400

Ser Ala Ser Leu Gln Glu Leu Ala Ala Tyr Ala Glu Glu Met Gly Ala
 405 410 415

Pro Ala Met Pro Gly Ser Gly Arg Gln Glu Tyr Leu Gln Ala Ala Leu
 420 425 430

Asn Gln Asn Leu Phe Gly Glu Val
 435 440

<210> 13
 <211> 440
 <212> PRT
 <213> Leptotrichia goodfellowii

<400> 13

Met Lys Glu Phe Phe Pro Glu Ile Lys Glu Ile Lys Tyr Glu Gly Ala
 1 5 10 15

Glu Ser Lys Asn Asp Leu Ala Phe Lys Tyr Tyr Asn Lys Asp Glu Val
 20 25 30

Leu Gly Gly Lys Thr Met Lys Glu His Leu Arg Phe Ala Met Ser Tyr
 35 40 45

Trp His Thr Leu Lys Ala Gln Gly Val Asp Met Phe Gly Gly Glu Thr
 50 55 60

Met Asp Arg Glu Trp Asn Lys Tyr Glu Asn Val Leu Glu Arg Ala Lys
 65 70 75 80

Ala Arg Ala Asn Ala Gly Phe Glu Phe Met Gln Lys Leu Gly Leu Glu
 85 90 95

Tyr Phe Cys Phe His Asp Arg Asp Ile Ile Asp Glu Ser Met Met Leu
 100 105 110

Ala Asp Ser Asn Lys Leu Leu Asp Glu Ile Val Asp His Ile Glu Glu
 115 120 125

Leu Met Lys Lys Thr Gly Arg Lys Leu Leu Trp Gly Thr Thr Asn Ala
 130 135 140

Phe Ser His Pro Arg Phe Val His Gly Ala Ser Thr Ser Pro Asn Ala
 145 150 155 160

Asp Val Phe Ala Tyr Ala Ala Gln Val Lys Lys Ala Met Asp Ile
 165 170 175

[0016] Thr Asn Arg Leu Gly Gly Glu Asn Tyr Val Leu Trp Gly Gly Arg Glu
 180 185 190

Gly Tyr Glu Thr Leu Leu Asn Thr Asn Ser Glu Leu Glu Tyr Asp Asn
 195 200 205

Phe Ala Arg Phe Leu Lys Met Val Val Asp Tyr Lys Glu Lys Ile Gly
 210 215 220

Phe Lys Gly Gln Leu Leu Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys
 225 230 235 240

His Gln Tyr Asp Phe Asp Thr Ala Thr Val Leu Ala Phe Leu Arg Lys
 245 250 255

Tyr Asn Leu Asp Lys Tyr Tyr Lys Val Asn Ile Glu Ala Asn His Ala
 260 265 270

Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Asn Leu Ala Arg Ile
 275 280 285

Asn Gly Val Leu Gly Ser Ile Asp Ala Asn Gln Gly Asp Met Leu Leu
 290 295 300

Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Ile Tyr Asp Thr Thr Leu
 305 310 315 320

Ala Met Tyr Glu Val Val Lys Asn Lys Gly Leu Gly Ser Gly Gly Leu
 325 330 335

Asn Phe Asp Ala Lys Val Arg Arg Gly Ser Phe Glu Asp Lys Asp Leu
340 345 350

Phe Leu Ala Tyr Ile Ala Gly Met Asp Thr Phe Ala Lys Gly Leu Lys
355 360 365

Ile Ala Tyr Arg Leu Tyr Glu Asp Lys Val Phe Glu Asp Phe Ile Asp
370 375 380

Lys Arg Tyr Glu Ser Tyr Lys Thr Gly Ile Gly Lys Asp Ile Ile Asp
385 390 395 400

Gly Lys Val Gly Phe Glu Glu Leu Ser Lys Tyr Ala Glu Thr Leu Thr
405 410 415

Glu Val Lys Asn Asn Ser Gly Arg Gln Glu Met Leu Glu Ser Lys Leu
420 425 430

Asn Gln Tyr Ile Phe Glu Val Lys
435 440

<210> 14
<211> 1320
<212> DNA
<213> 人工序列

<220>
<223> 经优化以在酿酒酵母中表达的Lep goodf XI的编码区
<400> 14
[0017] atgaaggagt tttccaga aatcaaggaa atcaagtacg aagggtctga atctaagaac 60
gatttggctt tcaagtacta caacaaggac gaagttttgg gtggtaaaac catgaaggaa 120
cacttggat tcgttatgtc ttactggcac accttgaagg ctcaagggtgt tgacatgttc 180
ggtgtgaaa ctatggatag agaatggAAC aagtacgaaa acgtcttgg aagagctaag 240
gcttagagcta acgctgggtt cgagttcatg caaaagtgg gtttggata cttctgtttc 300
cacgacagag atatcatcga cgaatctatg atgttggctt attccaaacaa gttgtggac 360
gaaatcggtt atcacatcga agaattgtatg aagaagactg gtagaaagtt gttgtgggt 420
actactaacg ctttctctca cccaaagatc gtccacgggt cttctacctc cccaaacgct 480
gacgtttcg cttacgctgc tgctcaagtc aagaaggctt tggacatcac taacagattt 540
ggtgtgaaa actacgtttt gtgggggtgtt agagaagggtt acgaaacctt gttgaacact 600
aactccgaat tggataacga caacttcgtt agatttttga agatgggtt cgattacaag 660
gaaaatcg gtttcaaggg tcaattttgtt atcgaaccaa agccaaaggg accaaccaag 720
caccataacg acttcgatac cgtactgtt ttggctttt tgagaaagta caacttggac 780
aagtactaca aggtcaacat cgaagctaac cacgttaccc tggctggta cacttccaa 840
cacgaattga acttggcttag aatcaacgtt gtcttgggtt ctatcgacgc taaccaagg 900
gacatgttgc tgggttggaa caccgttcaa ttcccaacta acatctacga caccacttt 960
gttatgtacg aagttgttcaaa gaacaagggtt ttgggttctg gtgggttggaa cttcgacgt 1020
aaggtagaa gaggttcctt cgaagacaag gatttttttct tggttttacat cgctggat 1080
gacacccctcg ctaagggtttt gaagatcgtt tacagattgtt acgaaagacaa ggtcttcgaa 1140
gacttcatcg ataagagata cgaatcttac aagactggta tggtaaaga catcatcgat 1200

ggtaaaagttg gttcgaaga attgtccaag tacgctgaaa ccttgactga agtcaagaac 1260
 aactccggta gacaagaaat gttgaatct aagttgaacc aatacatctt cgaagtcaag 1320

<210> 15
 <211> 440
 <212> PRT
 <213> 塞巴鲁德菌属

<400> 15

Met Lys Glu Tyr Phe Pro Glu Ile Lys Glu Ile Lys Tyr Glu Gly Pro
 1 5 10 15

Glu Ser Lys Asn Val Met Ala Phe Lys Tyr Tyr Asn Lys Asp Glu Val
 20 25 30

Ile Gly Gly Lys Pro Met Arg Glu His Leu Lys Phe Ala Met Ser Tyr
 35 40 45

Trp His Thr Leu Lys Ala Gln Gly Leu Asp Met Phe Gly Gly Asp Thr
 50 55 60

Met Asp Arg Ala Trp Asn Arg Tyr Asp Asp Ala Leu Glu Gln Ala Lys
 65 70 75 80

Ala Arg Ala Asp Ala Gly Phe Glu Phe Met Gln Lys Ile Gly Met Asp
 85 90 95

[0018] Tyr Phe Cys Phe His Asp Arg Asp Ile Ile Asn Glu Ala Met Thr Leu
 100 105 110

Lys Glu Thr Asn Arg Leu Leu Asp Glu Ile Val Asp His Leu Glu Gly
 115 120 125

Leu Met Lys Thr Gly Ile Lys Leu Leu Trp Gly Thr Thr Asn Ala
 130 135 140

Phe Ser His Pro Arg Phe Leu His Gly Gly Ala Thr Ala Pro Asn Ala
 145 150 155 160

Asp Val Phe Ala Tyr Ala Ala Gln Val Lys Lys Ala Met Glu Ile
 165 170 175

Thr Lys Arg Leu Gly Gly Glu Asn Tyr Val Leu Trp Gly Gly Arg Glu
 180 185 190

Gly Tyr Glu Thr Leu Leu Asn Thr Lys Ser Asp Leu Glu Tyr Asp Asn
 195 200 205

Phe Ala Arg Phe Leu Gln Met Val Val Asp Tyr Lys Glu Lys Ile Gly
 210 215 220

Phe Glu Gly Gln Leu Leu Ile Glu Pro Lys Pro Lys Glu Pro Thr Lys
 225 230 235 240

His Gln Tyr Asp Phe Asp Thr Ala Thr Val Leu Gly Phe Leu Arg Lys
 245 250 255

Tyr Asn Leu Asp Lys His Tyr Lys Met Asn Ile Glu Ala Asn His Ala
 260 265 270

Thr Leu Ala Gly His Thr Phe Gln His Glu Leu Asn Leu Ala Arg Ile
 275 280 285

Asn Asn Val Met Gly Ser Ile Asp Ala Asn Gln Gly Asp Met Leu Leu
 290 295 300

Gly Trp Asp Thr Asp Gln Phe Pro Thr Asn Ile Tyr Asp Ala Val Leu
 305 310 315 320

Ala Met Tyr Glu Val Ile Lys Asn Asn Gly Leu Gly Lys Gly Leu
 325 330 335

Asn Phe Asp Ala Lys Val Arg Arg Gly Ser Phe Glu Asp Lys Asp Leu
 340 345 350

Phe Leu Ala Tyr Ile Ala Gly Met Asp Thr Phe Ala Lys Gly Leu Thr
 355 360 365

Ile Ala Tyr Arg Leu Tyr Glu Asp Lys Val Phe Glu Asp Phe Gln Asp
 370 375 380

Lys Arg Tyr Glu Ser Tyr Lys Thr Gly Ile Gly Lys Asp Ile Val Glu
 385 390 395 400

[0019] Gly Lys Val Gly Phe Glu Glu Leu Ala Glu Tyr Val Glu Asn Leu Ala
 405 410 415

Glu Ile Lys Asn Thr Ser Gly Arg Gln Glu Met Leu Glu Ser Ile Leu
 420 425 430

Asn Ser Tyr Ile Leu Glu Ala Lys
 435 440

<210> 16
 <211> 1320
 <212> DNA
 <213> 人工序列

<220>
 <223> 经优化以在酿酒酵母中表达的Seb term XI的编码区

<400> 16		
atgaaggaaatctccaga aatcaaggaa atcaagtacg aagggtccaga atccaagaac	60	
gttatggctt tcaagtacta caacaaggac gaagttatcg gtggtaaacc aatgagagaa	120	
cacttgaagt tcgctatgtc ttactggcac accttgaagg ctcaagggtt ggacatgttc	180	
ggtgttgaca ctatggatag agcttggAAC agatacgcg atgcttgga acaagctaag	240	
gcttagagctg acgctggttt cgagttcatg caaaagatcg gtatggatta ctctgtttc	300	
cacgacagag atatcatcaa cgaagctatg accttgaagg aaactaacag attgttgac	360	
gaaatcggtt atcacttggaa aggtttgatg aagaagaccg gtatcaagtt gtgtgggt	420	
actactaacg ctttctctca cccaaagattc ttgcacgggt gtgcgtaccgc tccaaacgct	480	
gacgtttcg cttacgctgc tgctcaagtc aagaaggcta tggaaatcac taagagattt	540	

ggtgtggaaa	actacgtt	gtggggtggt	agagaagg	tttggaaacact	600	
aagtccgact	tggaaatacga	taacttcgct	agattcttc	aatgggttgc	660	
aaaaagatcg	gttcaagg	tcaattgtt	atcgaaccaa	agccaaagg	720	
caccaatacg	acttcgatac	cgtctactgtt	ttgggttct	tgagaaagta	780	
aagcactaca	agatgaacat	cgaagetaac	cacgctacct	tggctgtca	840	
cacgaattga	acttggctag	aatcaacaac	gtcatgggtt	ctatcgacgc	900	
gacatgttgc	tgggtggaa	caccgatcaa	ttcccaacta	acatctacga	960	
gctatgtacg	gaatcatcaa	gaacaacgg	ttgggtaaag	gtgggttgc	1020	
aaggtcagaa	gagggtcctt	cgaagacaag	gatttgcctt	tggcttacat	1080	
gacacccctcg	ctaagggttt	gactatcgct	tacagattgt	acgaagacaa	1140	
gacttccaag	ataagagata	cgaatcttac	aagaccgta	tcggtaaaga	1200	
ggtaaagttt	gttcaaga	attggctgaa	tacgtcgaaa	acttggctga	1260	
acttccggta	gacaagaaat	gttggaaatct	atcttgaact	cctacatctt	1320	
<210>	17					
<211>	9910					
<212>	DNA					
<213>	人工序列					
<220>						
<223>	用于Hm1表达的构建载体					
<400>	17					
aggccagagg	aaaataat	caagtgcgtt	aaacttttc	tcttggaaatt	tttgcacat	60
caagteatag	tcaattgaaat	tgacccaatt	tcacattaa	gattttttt	tttcatcgc	120
acatacatct	gtacactagg	aaggccgtt	tttctgaagc	agettcaat	atataatattt	180
tttacatatt	tattatgatt	caatgaacaa	tctaattaaa	tcgaaaacaa	gaaccgaaac	240
gccaataat	aatttattta	gatggtgaca	agtgtataag	tcctcatcgg	gacagctacg	300
atttcttctt	cggtttggc	tgagctactg	gttgcgtga	cgcagcggca	ttagcgcggc	360
gttatgagct	accctcgtgg	cctgaaagat	ggcgggaaata	aagcggact	aaaaattact	420
gactgagcca	tattgaggc	aatttgtaa	ctcgtcaagt	cacgtttgg	ggacggcccc	480
tttccaacga	atcgatata	ctaacatgcg	cgegcttct	atatacacat	atacatatata	540
atataatat	atatgtgtc	gtgtatgtt	acacctgtat	ttaatttcct	tactcgcggg	600
tttttctttt	ttctcaattt	ttggcttctt	ctttctcgag	cggaccggat	cctccgcgg	660
gccggcagat	ctatttaat	ggcgccgca	cgtcagggtt	cactttcgg	ggaaatgtgc	720
gccaaccccc	tatttggta	tttttctaaa	tacattcaaa	tatgtatcgg	ctcatgagac	780
aataaccctg	ataaatgctt	caataatatt	aaaaaaggaa	gagttatgat	attcaacatt	840
tcctgtcgc	ccttatttccc	tttttgcgg	cattttgcct	tcctgttttt	gctcaccccg	900
aaacgctgg	gaaagtaaaa	gatgtgaag	atcagttggg	tgcacgagtg	ggttacatcg	960
aactggatct	caacagcgg	aagatcctt	agagtttcg	ccccgaagaa	cgttttccaa	1020
tgtatgagc	ttttaaagtt	ctgtatgtc	gcccgggtatt	atcccgtatt	gacgcggggc	1080
aagageaact	cggtcggcgc	atacactatt	ctcagaatga	cttgggttgc	tactcaccag	1140
tcacagaaaa	gcatcttaac	gatggcatga	cagtaagaga	attatgcagt	gctgccataa	1200

[0021]	ccatgagtga taacactgcg gccaacttac ttctgacaac gatcgaggaa ccgaaggagc	1260
	taaccgcctt ttgcacaac atggggatc atgtaactcg cttgategt tggaaaccgg	1320
	agctgaatga agccatacca aacgacgagc gtgacaccac gatgcctgta gcaatggcaa	1380
	caacggtcgcaaaactttaa actggcgaac tacttactct agcttccggcaaaactttaa	1440
	tagactggat ggaggcggat aaagttgcag gaccacttct ggcgtcgcccttccggctg	1500
	gctggtttat tgctgataaa tctggagccg gtgagcgtgg gtctcgccgtt atcattgcag	1560
	cactggggcc agatggtaag ccctccgta tcgtgttat ctacacgacg gggagtcagg	1620
	caactatgga tgaacgaaat agacagatcg ctgagatagg tgcctcactg attaagcatt	1680
	ggtaactgtc agaccaagtt tactcatata tactttatgt tgatttaaaa cttcattttt	1740
	aatttaaaag gatcttaggtg aagatcctt ttgataatct catgacccaa atcccttaac	1800
	gtgagtttc gttccactga ggcgtcagacc ccgtagaaaa gatcaaagga tcttctttag	1860
	atcctttttt tctgcgcgta atctgctgtc tgcaaaacaaa aaaaccaccc ctaccagcgg	1920
	tggttgtttt gccggatcaa gagctaccaa ctcttttcc gaaggttaact ggcttcagca	1980
	gagcgcagat accaaatact gttttcttag tgcgtccgtt gttaccatgt gctgtcgcc	2040
	gtggcgatataa gtcgtgtt accgggttgg actcaagacg atagttacgg gataaggcgc	2100
	agcggtcggg ctgaacgggg ggttcgtgca cacagccag cttggagcga acgacccata	2160
	ccgaaactgag atacccatcg cgtgagctat gagaaagcgc caccgttccc gaagggagaa	2220
	aggcggacag gtatccgta agcggcagg tcggaacagg agagcgcacg agggagcttc	2280
	cagggggaaa cgcctggat ctttatagtc ctgtcggtt tcgcccaccc tgacttgagc	2340
	gtcgatttttt gtgatgtcg tcagggggc ggagccatgt gaaaaacgcg agcaacgcgg	2400
	ccttttacg gttccctggcc ttttgcgtc cttttgcata catgttctt cctcggttat	2460
	ccctgatttc tgcgtccgtt cgtattaccg cttttgatgt agctgataacc gtcgtccgca	2520
	gccgaacgac cgagcgcagc gagtcgtga gcgaggaagc ggaagagcgc ccaatacgca	2580
	aaccgcctt ccccgccgt tgcccgattt attaatgcag ctggcacgac aggttcccg	2640
	actggaaagc gggcagttag cgcaacgcaa ttaatgttag ttagctact cattaggcac	2700
	cccaggctt acactttatg ctccggctc gtatgttgc tgaaattgtg agcggataac	2760
	aatttcacac agggaaacagc tatgaccatg attacgccaat gcttttctt tccaattttt	2820
	ttttttcgat cattataaaa atcattacga ccgagattcc cggtaataa ctgatataat	2880
	taaattgttgcgt ctctaatggat tgcgttttagt atacatgttgc tttttttttt tccaattttt	2940
	tagtttgcgt ggcgcacatct tctcaaatat gcttccagc ctgttttctt gtaacgttca	3000
	ccctctaccc tgcgtccgtt tcccttgcata aatagtcctt ttccaaacaa aataatgtca	3060
	gtatgtcgat agaccacatc atccacgggtt ctatactgtt gacccatgc gtcgtccgtt	3120
	tcatctaaac ccacacccggg tgcataatc aaccaatgtt aacccatc tttccaccc	3180
	atgtctttt gagaataaaa gccgataaca aaatcttgcgtt cgctttcgc aatgtcaaca	3240
	gtacccttag tatatttcgcgtt agtagatagg gagccctgcgtt atgacaaatc tgctaaatc	3300
	aaaaggccctc taggttcctt tgttacttct tctgcgtccgtt gttcaaaacc gctaaacaaata	3360
	cctggccca ccacacccggg tgcataatc aaccaatgtt aacccatc tttccaccc	3420
	3480	

cccgagagt actgcaattt gactgttata ccaatgtcg caaattttct gtcttcgaag 3540
agtaaaaaat tgtacttggc ggataatgcc tttagggcgt taactgtgccc ctccatggaa 3600
aaatcagtca agatatccac atgtgtttt agtaaacaaa tttgggacc taatgttca 3660
actaactcca gtaattcctt ggtggtaacga acatccaatg aagcacacaa gtttgggttc 3720
tttcgtgca tgatattaaa tagcttggca gcaacaggac taggatgagt agcagcacgt 3780
tccttatatg tagcttcga catgatttat ctgcgttcc tgcagggttt tggtctgtgc 3840
agttgggtta agaatactgg gcaatttcat gtttcttcaa cactacatat gcgtatata 3900
accaatctaa gtctgtgctc ctcccttgcgt tttccttctt gttcgagat taccgaatca 3960
aaaaaaatttca aaggaaacccg aaatcaaaaaaa aaagaataaa aaaaaaaatga tgaattgaaa 4020
agcttgcattt cctgcaggc gactctagta tactccgtct actgtacat acacttccgc 4080
tcaggtcctt gtccttaac gaggccttac cactctttt ttactctatt gatccagctc 4140
agcaaaaggca gtgtgatcta agattctatc ttgcgtatgt agtaaaacta gctagaccga 4200
gaaagagact agaaatgca aaggcacttc tacaatggct gccatcatta ttatccgatg 4260
tgacgctgca ttttttttt ttttttttt ttttttttt ttttttttt ttttttttt 4320
ttgtacaaat atcataaaaaaa aagagaatct tttaagcaaa ggatttctt aacttcttc 4380
gctgacagcat caccgacttc ggtggtaactg ttggaaaccac ctaaatcacc agttctgata 4440
cctgcatttca aaacctttt aactgcattt tcaatggctt taccttcttca aggcaagttc 4500
aatgacaatt tcaacatcat tgcacgac aagatagtgg cgatagggtt gaccttattt 4560
tttggcaaat ctggagcgga accatggcat gttcgtaca aaccaaatttgc ggtgttctt 4620
tctggcaag aggccaaggga cgcatggc aacaaacccca aggagcctgg gataacggag 4680
gcttcatgg agatgatatac accaaacatg ttgctggta ttataatacc atttaggtgg 4740
gttgggttct taacttaggt catggggca gaatcaatca attgatgtt aacttcaat 4800
gtagggaatt cgttcttgat ggttcttcc acagtttttcc tccataatct tgaagaggcc 4860
aaaacattag ctttatccaa ggaccaaaaat ggcattgggtt gtcattgtt tagggccatg 4920
aaagcggcca ttcttgat tcttgact tctggacgg tggatgttcc actatccaa 4980
gctgacaccat caccatgtc ttcccttctt ttaccaaaatg aaatacccttcc cactaatttct 5040
ctaaacaacaa cgaagtcaatg accttagca aatttgggtt tgattggaga taagtctaa 5100
agagagtcgg atgcaaaatg acatggctt aagttggcgt acaatttgaag ttcttacgg 5160
attttagta aaccttggc aaggcttaaca ctaccggat cccatggat accacccaca 5220
gcacaccaaa acacggcattt acgccttctt gaggcttccaa ggccttccatc tggaaatgg 5280
acacctgttagt catcgatagc agcaccacca attaaatgtt ttgcatttttca gaaacttgcata 5340
tttggaaatgg catcagaaat agcttaaga accttataatgg ctccgggtt gatttttgcata 5400
ccaaacgttggt cacctggcaaa aacgacgatc ttcttagggg cagacatttac aatggatata 5460
ccttggaaata tatataaaaaaa aaaaaaaaaaaa aaaaaatgcag cttctcaatg 5520
atattcgat aacgcttgcgtt gagatacage ctaatatccg acaaaactgtt ttacagattt 5580
acgatcgatc ttgttacca tcattgttattt ttaacatcc gaaacctgggaa gttttccctg 5640
aaacagatag tatatttgcata cctgtataat aatataatgtt ctgcatttttca gaaacttgcata 5700
atgtatgtt atcggttccctt ggagaaacta ttgcattat tgcatttttca gaaacttgcata 5760

[0023]	tcgcatcccc ggttcattt ctgcgttcc atcttgact tcaatagcat atcttggta	5820
	acgaagcata tgtgctcat ttgttagaac aaaaatgcaa cgcgagagcg ctaattttc	5880
	aaacaaagaa tctgagctgc attttacag aacagaaatg caacgcgaaa gcgtatTTT	5940
	accacaagaa aatctgtgct tcattttgt aaaacaaaaa tgcaacgcga gagcgtata	6000
	tttcaaaaca aagaatctga gctgcattt tacagaacag aaatgcaacg cgagagcgct	6060
	atTTTaccAA caaagaatct atacttctt ttgttctac aaaaatgcat cccgagagcg	6120
	ctatTTTctt aacaaagcat cttagattac ttTTTtctc ctttgtgcgc tctataatgc	6180
	agtctcttga taacttttgc cactgttagt ccgttaaggt tagaagaagg ctactttgg	6240
	gtctatTTTc tctccataa aaaaagcctg actccacttcc cgcgcgttac tgattactag	6300
	cgaagctgcg ggtgcattt ttcaagataa aggcatcccc gattatattc tataccgatg	6360
	tggattgcgc atactttgtg aacagaaagt gatagcgttg atgattctc attggcaga	6420
	aaattatgaa cggttttctc tattttgtct ctatatacta cgtataggaa attttacat	6480
	tttctgtatttgc tttcatttgc actctatgaa tagttcttac tacaatTTT ttgtctaaag	6540
	agtaatacta gagataaaaca taaaaatgt agaggcggag ttttagatgca agttcaaggaa	6600
	gaaaaagggtg gatgggtagg ttatataggg atatagcaca gagatataa gcaaaagagat	6660
	acttttgcgc aatgtttgtg gaagcggat tgcataattt ttagtagctc gttacagtcc	6720
	ggtgcgtttt tggttttttgg aaagtgcgtc ttcagagegc ttgtgtttt caaaagcgt	6780
	ctgaagtcc tatactttctt agagaatagg aacttcggaa taggaacttcc aagcgtttcc	6840
	cggaaacggag cgcttccaa aatgcaacgc gagctgcgc catacagctc actgttcac	6900
	tcgcacccat atctgcgtgt tgccgttata tatataataca tgagaagaac ggcatagtgc	6960
	gtgtttatgc ttatgcgtt acttatatgc gtctattttt gttaggtgaa aggtgtctat	7020
	gtaccccttgc tgatatttgc ccattccatg cgggtatcg tatgttctt tcagcactac	7080
	ccttagctg ttctatatgc tgccacttcc caattggatt agtctcatecc ttcaatgcta	7140
	tcattccctt tgatatttgc tcatatgc agtaccgaga aacttagagga tctccattt	7200
	ccgacatttgc ggcgtatac gtgcataatgt tcatgtatgt atctgttattt aaaaacttt	7260
	tgtattttt ttctatatgc atgtgtatag gtttatacgg atgatTTTtca tattacttca	7320
	ccacccttta ttccaggctg atatcttagc cttgttacta gtcaccgggtg gcccgcac	7380
	ctggtaaaac ctctagtggaa gtagtagatg taatcaatga agcggaaagcc aaaagaccag	7440
	agtagaggcc tatagaagaa actgcgatac cttttgtatgg ggcataaaca acagacatct	7500
	ttttatgtt ttttacttctt gtatatgcgtt aagtagtaag tgataagcgtt atttggctaa	7560
	gaacgttgcgtt agtgaacaag ggacctttt tgccttcaa aaaaggattt aatggagttt	7620
	atcattgaga tttagtttgcgtt gttagattctt gtatcccttta ataactccctt taccggacgg	7680
	gaaggcaca aagacttggaa taatagcaaa cggccatgtt ccaagaccaaa ataataacttag	7740
	agttaacttgc tggctttaaa caggcattttt gttgtgtttt ccaagaccaaa tataaaaaat	7800
	atcgtataatgtt tatttttttttgcgtt caccatTTTtta aggacgttac atcaggacag tagtaccatt	7860
	cctcagagaa gaggtatataca taacaagaaaa atcgcgttgcgtt caccttataat aacttagccc	7920
	gttatttgcgtt taaaaaaaccc ttgcataatTTTtta agaataacttcc agacgtgtata	7980
	aaaatTTTactt ttcataacttctt ttcacgttgcgtt cccctatgtt ttcttccgtt ctaccgttgcgtt	8040

[0024]	aaataaagca tcgagtacgg cagttcgctg tcactgaact aaaacaataa ggctagttcg	8100
	aatgatgaac ttgcttgcgtc tcaaacttct gagttccgc tgatgtgaca ctgtgacaat	8160
	aaattcaaac cggttatagc ggtctccctc ggtaccggtt ctgccacctc caatagagct	8220
	cagtaggagt cagaacctct ggggtggctg tcagtgcact atccgcgtt cgtaagttgt	8280
	gcccgtgcac atttcgcccgttccctca tcttgcagca ggcggaaatt ttcacacgc	8340
	tgttaggacgc aaaaaaaaaa taattaatcg tacaagaatc ttggaaaaaa aattgaaaaa	8400
	ttttgtataa aagggtatgac ctaacttgc tcaatggctt ttacacccag tattttccct	8460
	ttccttgttt gttacaatta tagaagcaag acaaaaacat atagacaacc tattcctagg	8520
	agtttatattt ttttacccta ccagcaatat aagttaaaaa ctgtttaaac agtatgaagg	8580
	agttttccctt atccatctctt ccaatcaagt tcgaagggttc cgaatccaag aaccattgt	8640
	ctttcaagta ctacgacgct aagagagttt tcatggtaa aaccatggaa gaacacttgt	8700
	cttcgctat ggcttgggtgg cacaacttgtt gtgcctccgg tggatcatg ttccgtcaag	8760
	gtactgctga caagggtttc ggtggaaact tgggtactat ggaacacgct aaggctaaagg	8820
	ttgacgctgg tatcgagttc atgcaaaagt tgggtatcaa gtactactgt ttccacgaca	8880
	ccgatatcgatccagaagac caagaagata tcaacgtc acacgttga ttggacgaaa	8940
	tcaactgatcaatccatggaa aagaccaagg gtactgacat caagtgtttt tgggtacttt	9000
	gttacatgtt ctctaaaccatca agattcatga acgggtgtgg ttcttctaaac tctgtgtacg	9060
	ttttctgtttt cgctgtgtcaatcaaga aggggttggaa aacgtgtttt aagttgggtt	9120
	ctaaagggtttt cgtcttcgg ggtggtagag aagggttacga aaccttggtaa aacactgaca	9180
	tgaagttggaa agaagaaaac atcgctaccc ttgttactat gtgttagagac tacggtagat	9240
	ctatcggtttt caagggtgtac ttctacatcg aaccaaaagcc aaaggaacca atgaagcacc	9300
	aatacgactt cgatgctgtaccgttacatcg gtttcttggag aaggtacggt ttggacaagg	9360
	atttcaagat gaacatcgaa gctaaccacg ctaccttggc tggcacact ttccaaacacg	9420
	aatttggatgtt ttcgttgc ttcgttgc aacatcgaa gtaaccacg ctaccttggc tggcacact	9480
	ctttgttggg ttggacacc gatcaattcc caactaacgt ttacgacacc acttggctaa	9540
	tgtacgaaat ttttgcgtt ggtggtttgc tgggtgggtt gaaacttgcac tctaagaaca	9600
	gaagaccatc caacaccgctt gaaatcgatgt ttttgcgtt ttttgcgtt atggacactt	9660
	tgcgtttttt ttttgcgtt ggtggtttgc tgggtgggtt gaaacttgcac tctaagaaca	9720
	tcaaggaaag atacttcc tacaactctg gtatcggttga aaagatcaga aacagatccg	9780
	tttactttgtt cgaatgtgttcaatcgctt ttttgcgtt ggtggtttgc tgggtgggtt	9840
	ctggtagaca agaataacttg gaaaccgtcg tcaacaacat ttttgcgtt ggtggtttgc	9900
	ggggccctgc	9910
	<210> 18	
	<211> 16404	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 构建质粒	
	<400> 18	
	gatccacatcgat cgcattgggg attacgttattt ctaatgttca gtaccgttgc tataatgtat	60

[0026]	ctcggtatac gcctatTTT ataggTTAAT gtcatgataa taatggTTc tttagacgtca	2400
	ggTggcactt ttcggggaaa tggcgcggA accccTattt gtttattttt ctaaatacat	2460
	tcaaatatgt atccgctcat gagacaataa ccctgataaa tgcttcaata atattgaaaa	2520
	aggaagagta tggatttca acatttcgt gtcgcctta ttccctttt tggcattt	2580
	tgccttcctg ttttgcTca cccagaaacg ctggtaaag taaaagatgc tgaagatcag	2640
	ttgggtgcac gagggtgttA catgaaactg gatctcaaca gcggtaaagat ctttagagt	2700
	tttgcggccg aagaacgtt tccatgtg agcactttt aagtctgtc atgtggcgcg	2760
	gtattatccc gtattgacgc cggcaagag caactcggtc gcccataca ctatttcag	2820
	aatgacttgg ttgagttactc accagtcaca gaaaagcata ttacggatgg catagacgta	2880
	agagaattat gcaatgtc cataaccatg agtataaca ctgcggccaa ctatcttg	2940
	acaacgatcg gaggaccgaa ggagctaaacc gcttttgc acaacatggg ggatcatgta	3000
	actcgcttgc acgttggga accggagctg aatgaagcca taccaaacga cgagcggtac	3060
	accacgatgc ctgttagcaat ggcaacaacg ttgcgcAAac tattactgg cgaactactt	3120
	actctagttt cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca	3180
	cttctgcgtt cggcccttcc ggctggctgg tttattgtc ataaatctgg agccgggtgag	3240
	cgtgggtctc ggggtatcat tgcagcactg gggccagatg gtaaggccctc cctatcgta	3300
	gttatctaca cgacggggag tcaaggcaact atggatgaaac gaaatagaca gatcgctgag	3360
	atagggtcct cactgattaa gcatggtaa ctgtcagacc aagtttactc atatatactt	3420
	tagattgatt taaaactca ttttaattt aaaaggatct aggtgaagat ctttttgc	3480
	aatctcatga caaaatccc ttaacgttag tttcggtcc actgagcgctc agacccgta	3540
	gaaaagatca aaggatctt ttgagatctt ttttttgc gcttaatctg ctgcttgcAA	3600
	acaaaaaaaac caccgctacc agcggtggg tggggccgg atcaagagct accaactctt	3660
	tttccgaagg taactggcgtt cagcagagcg cagataccaa atactgtct tctagtgtag	3720
	cccttagttt gccaccactt caagaactct gtgcacccgc ctacatacct cgctctgtca	3780
	atccgttac cagtggctgc tgecagtggc gataagtctgt gtcttaccgg gttggactca	3840
	agacgatagt taccggataa ggcgcagcgg tgggtgtaa cgggggggtt gtcacacag	3900
	cccaagttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa	3960
	agcgccacgc ttcccgagg gagaaggcg gacaggatc eggtaaaggc cagggtcgga	4020
	acaggagacg gcacgaggga gttccaggg ggaaacgcct ggtatctta tagtctgtc	4080
	gggttgcgc acctctgtact tgagcgatc tttttgtat gtcgtcagg gggggggagc	4140
	ctatggaaaa acgccagcaa cggggcctt ttacgggtcc tggccctttt ctggcccttt	4200
	gctcacatgt ttttgcgtc gttatcccct gattctgtgg ataaccgtat taccgccttt	4260
	gagttagtgc ataccgtcg ccgcagcgg acgaccgacg gcaagcgatc agtgagcgag	4320
	gaagcggaag agcgccaaat acgcaaaacgg cctctcccg cgctggcc gattcattaa	4380
	tgcagctggc acgacagggtt tcccgactgg aaagcgggca gtgagcgaa cgaattaaat	4440
	gtgagttacg tcactcatta ggcaccccg gctttacact ttatgttcc ggctgtatg	4500
	ttgtgtggaa ttgtgagcggtt ataacaattt cacacaggaa acagctatga ccatgattag	4560
	gcgcctactt ctagggggcc tatcaagtaa attactctgt gtacactgaa gtatataagg	4620

[0028]	cttttccgg ttttttattc aattgctcta acagagattc ggcttggag gaccagtaga	6960
	cagaccgtg ctgctggcag gaccctgaga cggccataac tttgttcaat ggaaatttag	7020
	cctcgcgata ttgcgagaga accagatcta gaggctctaa ccacatggct acgggacatt	7080
	cgatagtgc ggcgtgtata tagacaccct tctttgtgt ataatgcgga agatccttt	7140
	caaattccac tggttctgaa tggacaattt ttaggtcctg gttaatggcg agacatttca	7200
	gttgggggt cgaaagatca aacccaagat agtatgagtc taaagacatt gtgttgaaa	7260
	cctcttctgt ctgtctctga attactgaa acAACatact agtcgtacgg ttttatttt	7320
	tacttatatt gctggtaggg taaaaaaaata taactcttag gaataggtt tctatatgtt	7380
	tttgcttc ttttataatt gtaacaaaca agggaaaggga aaatactggg tgtaaaagcc	7440
	attgagtca gtttaggtcat ccctttata caaaattttt caattttttt tccaagattc	7500
	ttgtacgatt aattttttt ttttgcgtc ctacagcgtg atgaaaattt ccgcctgctg	7560
	caagatgagc gggAACgggc gaaatgtgca cgcgcacaac ttacgaaacg cggatgagtc	7620
	actgacagcc accgcagagg ttctgactcc tactgagtc tattggaggt ggcagaaccg	7680
	gtaccggagg agaccgctat aaccggttt aattttatgt cacagtgtca catcagcggc	7740
	aactcagaag tttgacagca agcaagttc a tcatcgaac tagccttatt gttttatgtt	7800
	agtgacagcg aactgcgtta ctgcgtcatt tatttctac ggttagcgg aagaacagat	7860
	aggggcagcg tgagaagagt tagaaagtaa attttatca cgtctgaatg atttttattc	7920
	atagaaatt ttgcaagggtt ttttagctca ataacgggt aagttatata aggtgttac	7980
	gcgattttct ttttatgtat acctttctg gcgcgcctt ttttattaac cttatatttt	8040
	attttagatt cctgacttca actcaagacg cacagatatt ataacatctg cataataggc	8100
	atttgcaga attactcgta agtaaggaaa gagtgaggaa ctatgcata cctgcattta	8160
	aagatgccga tttggcgcg aatcctttat tttggctca ccctcataact attatcagg	8220
	ccagaaaaag gaagtgttc cctccttctt gaattgtatg taccctcata aagcacgtgg	8280
	cctcttatcg agaaagaaat taccgtcgct cgtgattgt ttgcaaaaag aacaaaactg	8340
	aaaaaaccga gacacgtcg acttcctgtc ttcctattga ttgcagctc caatttcgtc	8400
	acacaacaag gtcctagcga cggctcacag gttttgtiac aagcaatcga aggttctgg	8460
	atggcggaa agggtttagt accacatgtt atgatgccc ctgtgtatc cagagcaaag	8520
	ttcgttcgat cgtactgtta ctctctctt ttcaaacaga attgtccgaa tgcgtgtaca	8580
	acaacagcct gttctcacac actctttct tctaaccag ggggtggtt agtttagtag	8640
	aacctcgta aacttacatt tacatataata taaacttgc aaaaattggc aatgcaagaa	8700
	atacatattt ggtttttctt aattcgtagt tttcaagtt cttagatgtt ttcttttct	8760
	ctttttaca gatcatcaag gaagtaattt tctactttt acaacaaata taaaacacgt	8820
	acgactagta tgactcaatt cactgacatt gataagtgg cgcgtccac cataagaatt	8880
	ttggctgtgg acaccgtatc caaggccaa tcaggtcacc caggtgcctc attgggtatg	8940
	gcaccagctg cacacgtct atggagtcaa atgcgtatga acccaacaa cccagactgg	9000
	atcaacagag atagatttgtt ctgtctaa ggtcacggc tgcgtttgtt gtattctatg	9060
	ctacatttga ctggttacga tctgtctattt gaagacttga aacagtttag acagttgggt	9120
	tccagaacac caggcatcc tgaatttgag ttgccaggtt ttgaagttac taccggcata	9180

tttaggtcaag	gtatctccaa	cgttgttgg	atggccatgg	ctcaagctaa	cctggctgg	9240
acttacaaca	agccgggctt	taccctgtct	gacaactaca	cctatgtttt	cttgggttgac	9300
ggttgttgc	aagaaggtat	ttcttcagaa	gettcteect	tggtggtca	tttgaatttg	9360
ggtaacttga	ttgccatcta	cgtgacaac	aagatcacta	tcgtggtgc	taccagtatc	9420
tcattcgatg	aagatgtgc	taagagatac	gaagcctacg	gttgggaagt	tttgtacgt	9480
gaaaatggta	acgaagatct	agccggtatt	gccaggctta	ttgtcaaggc	taagttatcc	9540
aaggacaaac	caacttgc	caaaatgacc	acaaccattg	gttacggttc	cttgcattgc	9600
ggctctact	ctgtgcacgg	tgccccattt	aaagcagatg	atgttaaaca	actaaagagc	9660
aaatcggtt	tcaaccaga	caagtcctt	gttgttccac	aagaagttt	cgaccactac	9720
caaaagacaa	ttttaaagcc	agggtcgaa	gccaacaaca	agtggacaa	gttggcagc	9780
gaataccaaa	agaaattccc	agaatttagt	gtgttttttt	ctagaagatt	gagcggccaa	9840
ctacccgcaa	attgggaaatc	taagttgcca	acttacaccg	ccaaggactc	tgccgtggcc	9900
actagaaaaat	tatcagaacac	tgttcttgag	gatgtttaca	atcaatttgc	agagttgatt	9960
ggtgttctg	ccgatttaac	acettctaac	ttgaccagat	ggaaggaaagc	ccttgcattc	10020
caacccctt	cttccggttc	aggtaactac	tctggtagat	acattaggta	cggtattttaga	10080
gaacacgcta	tgggtccat	aatgaacggt	atttcagtt	tgggtccaa	ctacaacacca	10140
tacggtggtt	ctttttgaa	tttcgtttct	tatgctgtt	gttgcgttag	attgtccgt	10200
tttgttggcc	accctgttat	ttgggttgc	acacatgact	ctatcggtt	cggtgaagat	10260
ggtccaacac	atcaacctat	tgaaacttta	gcacacttca	gatccctacc	aaacattcaa	10320
gtttggagac	cagctgtatgg	taacgaagtt	tctggccct	acaagaactc	tttagaaatcc	10380
aagcataactc	caagtatcat	tgcattgtcc	agacaaaact	tgcacacatt	ggaaggtagc	10440
tctattgaaa	gctgttctaa	gggtggttac	gtactacaag	atgttgc	cccagatatt	10500
attttagtgg	ctactggttc	cgaagtgtt	tttagtgc	aaagtc	ttttggcc	10560
gcaaaagaaca	tcaaggctcg	tgttgc	ctaccagatt	tcttca	tgacaaacaa	10620
cccctagaat	acagactatc	agtc	tttacatc	gacaacgtt	caatcatgc	10680
ttggctacca	catgttgggg	caaatacgt	catcaatct	tgggtattga	cagatttgg	10740
gcctccggta	aggcaccaga	agttcaag	tttccgtt	tcacccaga	agggtttgt	10800
gaaagagctc	aaaagaccat	tgcattctat	aagggtgaca	agcttattc	tccttgc	10860
aaagcttct	aaattctgtat	cgtatcat	cagatttgc	atgtattt	ttgtgaaaaa	10920
atgaaataaa	actttataca	actttaatac	aactttttt	ataaacgtt	aagcaaaaaa	10980
atagttcaaa	acttttaaca	atattccaa	cactcgttcc	ttttcc	tatattatag	11040
gtgtacgtat	tatgaaaaaa	tttcaatgtat	tacttttct	ttttttcc	ttgttaccagc	11100
acatggccga	gtttgaatgt	taaacccttc	gagagaatca	caccattca	gtataaagcc	11160
aataaagaat	ataactctta	aaaggctaat	tgaaccctg	tgattttgc	ccgggtttaa	11220
ggcgcgcct	ttatcattat	caatactgc	atttcaaga	atacgtaat	attaatagt	11280
agtgttttc	ctaaactttat	ttgtcaaaaa	aatttcgtt	ttaatttgc	tgttacccgt	11340
acatgccccaa	aatagggggc	gggttacaca	gaatataaa	cacgttgc	gttgcgttgc	11400
acagtttatt	cctggcattcc	actaaatata	atggagcccg	cttttaagc	ttggcatccag	11460

[0031]	tctgcacttg atttattata aaaagacaaa gacataatac ttctctatca atttcagttt ttgttcttcc ttgcgttatt cttctgttct tcttttctt ttgtcatata taaccataac caagtaatac atattcaaac ttaagactcg agatggtcaa accaattata gctcccagta tccttgccttc tgacttcgccc aacttgggtt gccaatgtca taaggtcata aacgcggccg cagattggtt acatatcgat gtcataggacg gccattttgt tccaaacatt actctggcc aaccattgt tacctcccta cgtcggtctg tgccacgccc tggcgatgtc agcaacacag aaaagaagcc cactgcgttc ttgcattgtc acatgtatgt tgaaaatctt gaaaatgg tcgacgattt tgctaatgt ggtgctgacc aatttacgtt ccactacgag gccacacaaag acccttgcata tttagttaag ttgattaagt ctaagggcat caaagctgca tgcgecatca aacctggtaa ttctgttgc gttttatgtt aactagctcc tcatttggat atggcttctt ttatgactgt ggaacctggg tttggaggcc aaaaattcat ggaagacatg atgcacaaag tggaaacttt gagagccaag ttccccatt tgaatatcca agtcgtatgtt ggtttggca aggagaccat cccgaaagcc gccaaagccg gtgcacacgt tattgtcgtt ggtaccatgt tttcactgc agctgacccg cacgtatgtt ttccttcata gaaagaagaa gtctcgaagg aattgcgttc tagagattt ctagattaga cgtctgtttt aagattacgg atatttaact tacttagaat aatgcattt tttgagttt taataatctt acgttagtgt gagegggatt taaactgtga ggacctaatac acattcagac acttctgcgg tatcacccta cttattccct tcgagattt atcttaggaac ccatcagggtt ggtggagat taccgttctt aagactttt agctccctt attgtatgtt caccgtggaca cccctttctt ggcacatccagt ttttatactt cagttggcatg tgagatttcc cggaaattataa taaagcaatc acacaatttc ctggatacc acctcggtt aactgacag gtgggttgc acgtatgtca atgcaagga gcctatatac cttggctcg gctgtgtaa cagggatat aagggcage ataatttagg agtttagtga acttgcacaa ttactatattt tcccttcata cgtaaatattt ttttttttta attctaaatc aatcttttca aattttttgtt ttgtattttt ttcttgcata aatctataac tacaaaaaac acatacataa actaaaaacgt acgacttagta tgctcgaacc agctcaaaag aaacaaaagg ttgctaaaca ctctctagaa caattgaaag cctccggcac tgctgtgtt ggcacactg gtgatttgcg ctctatttgc aagtttcaac ctcagaactc cacaactaac ccatcattga tcttggctgc tgccaaagcaa ccaacttacg ccaagttgat cgatgttgcg gtggataacg gtaagaagca tggtaaagacc accgaagaac aagtcgaaaa tgctgtggac agatttttag tcgaatttgcg taaggagatc ttaaagattt ttcaggccg acgttccacc gaagttgtat ctagatttgc ttttgcactt caagctacca ttgaaaaggc tagacatatac attaaatttgt ttgacaaaga aggtgttcc aaggaaagag tccttattaa aattgttcc acttggaaag gtattcaacg tgccaaagaa ttgaaagaaaa aggacggtat ccactgtat ttgacttcat tatttcctt cgttcaagca gttgcctgtg ccgaggccca agttactttt atttcccat ttgttggtag aattctagac tggtacaaat ccagcactgg taaagattac aagggtgaag ccgacccagg tggatattcc gtcacaaaaa tctacaacta ctacaagaag tacggttaca agactattgt tatgggtgtc tcttcagaa gcaactgacga aatcaaaaaac ttggctgg ttgactatct aacaatttctt ccagtttat tggacaagtt gatgaacagt actgaaccc	13800 13860 13920 13980 14040 14100 14160 14220 14280 14340 14400 14460 14520 14580 14640 14700 14760 14820 14880 14940 15000 15060 15120 15180 15240 15300 15360 15420 15480 15540 15600 15660 15720 15780 15840 15900 15960 16020
--------	--	--

teccaagagt tttggaccct gtctccgcta agaaggaagc cggcgacaag atttcttaca	16080
tcagcgacga acttaaattc agattcgact tgaatgaaga cgctatggcc actgaaaaat	16140
tgtccgaagg tatcagaaaa ttctctgccc atattgtac tctattcgac ttgattgaaa	16200
agaaaagttac cgcttaagga agtatctcg aaatattaat ttaggccatg tcttatgca	16260
cgtttcttt gatacttacg ggtacatgtt cacaagtata tctatataataaattatg	16320
aaaatcccct atttatataat atgactttaa cgagacagaa cagttttta ttttttatcc	16380
tatttcatgtt atgatacagt ttgc	16404
<210> 19	
<211> 95	
<212> DNA	
<213> 人工序列	
<220>	
<223> 在使用cre重组酶除去KanMX标记物之后, 由侧面与引物结合位点相接的loxP位点组成的95bp序列作为基因组中的URA3缺失序列特异性扩增区域而保留	
<400> 19	
gcattgcggta ttacgttattc taatgttcag ataacttcgt atagcataca ttatacgaag	60
ttatccagtg atgataacaac gagtttagcca aggtt	95
<210> 20	
<211> 100	
<212> DNA	
<213> 酿酒酵母	
<400> 20	
gtccataaag ctttcaatt catttttttt tttttgttc tttttttga ttccggtttc	60
tttggaaattt ttttgattcg gtaatctccg agcagaagga	100
<210> 21	
<211> 100	
<212> DNA	
<213> 酿酒酵母	
<400> 21	
aaaactgtat tataagtaaa tgcgtatata ctaaactcac aaatttagagc ttcaatttaa	60
ttatatacgat tattaccgg gaatctcggt cgtaatgatt	100
<210> 22	
<211> 100	
<212> DNA	
<213> 酿酒酵母	
<400> 22	
attggcatta tcacataatg aattatacat tatataaagt aatgtgattt ctgcgaagaa	60
tatactaaaa aatgaggcagg caagataaac gaaggcaaag	100
<210> 23	
<211> 100	
<212> DNA	
<213> 酿酒酵母	
<400> 23	
tagtgacacc gattatataa agctgcagca tacgatataat atacatgtgt atatatgtat	60
acctatgaat gtcagtaagt atgtatacga acagttatgtat	100
<210> 24	
<211> 6728	

<212>	DNA						
<213>	人工序列						
<220>							
<223>	构建载体						
<400>	24						
acatatttga	atgtatttag	aaaaataaaac	aaatagggt	tccgcgcaca	tttccccaa	60	
aagtgcacc	tgggtcctt	tcatcacgtg	ctataaaaat	aattataatt	taaattttt	120	
aatataaata	tataaattaa	aaatagaag	taaaaaaaga	aattaaagaa	aaaatagttt	180	
ttgtttccg	aagatgtaaa	agactctagg	gggatcgcca	acaaatacta	cctttatct	240	
tgccttcct	gctctcaggt	attaatgcgg	aattgtttca	tcttgctgt	gtagaagacc	300	
acacacgaaa	atcctgtgat	tttacatttt	acttacgtt	aatcgaatgt	atatctattt	360	
aatctgctt	tcttgctaa	taaatatata	tgtaaagtac	gcttttgtt	gaaattttt	420	
aaacctttgt	ttatttttt	ttcttcattt	cgtaactctt	ctacctctt	tatttacttt	480	
ctaaaatcca	aatacaaacc	ataaaaataa	ataaacacag	agtaaattcc	caaattattc	540	
catcattaaa	agatacgagg	cgcgtgtaa	ttacaggca	gcgcgcgtc	ctaagaaacc	600	
attattatca	tgacattaac	ctataaaaat	aggcgatca	cgaggccctt	tcgtctcg	660	
cggttcgg	atgacggta	aaaccttga	cacatgcage	tccggagac	ggtcacagct	720	
tgtctgtaa	cggtgcgg	gaggcagaca	gccgcgtcagg	gcgcgtcage	cggtgttggc	780	
gggtgtcgg	gctggctaa	ctatgcggca	tcaagcaga	ttgtactgag	agtgcaccat	840	
aaattccgt	ttaaagagct	tggtgagcgc	taggagtcac	tgccaggtat	cgtttgaaca	900	
[0033]	cgccattagt	caggaaagtc	ataacacagt	ccittccgc	aatttcttt	ttcttattact	960
cttggcctcc	tctagtagac	tctatatttt	tttatgcctc	ggtaatgatt	ttcattttt	1020	
ttttccct	agcggatgac	tctttttt	tcttagcgat	ttgcatttac	acataatgaa	1080	
ttatacatta	tataaagtaa	tgtgatttc	tcaagaata	tactaaaaaa	tgagcaggca	1140	
agataaacga	aggcaaagat	gacagagcag	aaagccctag	taaagcgtat	tacaaatgaa	1200	
accaagattc	agattgcgt	ctctttaaag	gggtgtcccc	tagegataga	gcactcgatc	1260	
ttcccagaaa	aagaggcaga	agcagtagca	gaacaggca	cacaatcgca	agtgattaac	1320	
gtccacacag	gtatagggtt	tctggaccat	atgatacatg	ctctggccaa	gcatccggc	1380	
tggtcgctaa	tctttgagtg	cattggtgac	ttcacatag	acgaccatca	caccactgaa	1440	
gactgcgg	ttgctctcg	tcaagctttt	aaagaggccc	tactggcg	tggagtaaaa	1500	
aggtttggat	caggattgc	gcctttggat	gaggcactt	ccagagcggt	ggtagatctt	1560	
tcaacaggc	cgtacgcgt	tgtcgaactt	ggtttgcaaa	gggagaaaat	aggagatctc	1620	
tcttgcgaga	tgatccgca	ttttcttga	agtttgcag	aggtagcag	aattaccctc	1680	
cacgttgatt	gtctgcgagg	caagaatgt	catcacgt	gtgagagtgc	gttcaaggct	1740	
cttgcgggtt	ccataagaga	agccacctcg	cccaatgtt	ccaaacgtat	tccctccacc	1800	
aaagggttcc	ttatgtatgt	acaccgatta	tttaaagctg	cagcatacga	tatataataca	1860	
tgtgtatata	tgtatatact	tgaatgtcag	taagtatgt	tacgaacagt	atgatactga	1920	
agatgacaag	gtaatgcac	attctatacg	tgtcattctg	aacgaggcgc	gctttccctt	1980	
tttctttttt	ctttttctt	ttttttctt	tgaactcgac	ggatctatgc	ggtgtgaaat	2040	
accgcacaga	tgcgtaaagga	aaaaataccg	catcaggaaa	ttgtaaacgt	taatattttg	2100	

taggtcagg tgcgttcgtt ggtatagcat gaggtcgctc ttattgacca cacccttacc	4440
ggcatggca gcaaatgcct gcaaatcgct ccccaatttca cccaaattgtt gatatgcata	4500
ctccagcaat gagttgtatga atctcggtgt gtatttatg tcctcagagg acaacacccgt	4560
tgggtttctt gageggccgc caccgggtg gagctccagc ttttggccc ttttagtgagg	4620
gttaatttgcg cgcttggcgt aatcatggtc atagttgttt cctgtgtgaa attgttatacc	4680
gctcacaatt ccacacaaca taggagccgg aagcataaag tgtaaagcct ggggtgccta	4740
atgagtgagg taactcacat taattgcgtt gcgtcaactg cccgtttcc agtccggaaa	4800
cctgtcggtc cagctgcatt aatgaatcg ccaacgcgcg gggagaggcg gtttgcgtat	4860
tgggcgtct tccgcgtt cgcgtcaactg ctcgcgtgc tcggcgttgc ggctgcggcg	4920
agcggatatac gtcactcaa aggccgtaat acggttatcc acagaatcag gggataaacgc	4980
aggaaagaac atgtgagcaa aaggccagea aaaggccagg aaccgtaaaa aggcgcgtt	5040
gctggcgtt ttccataggc tccgcgttcc tgacgagcat cacaataatc gacgtcaag	5100
tcagagggtgg cgaaaccgcg caggactata aagataccag gcgttcccc ctggaaagetc	5160
cctcggtgc ctcctgttc cgaccctgccc gttaccggc tacctgtccg ctttctccc	5220
tccggaaagc gtggcgttt ctcatagtc acgtgttagg tatctcagtt cgggttaggt	5280
cgttcgttcc aagctggcgt gtgtgcacga acccccccgtt cagccgcacc gtcgcgttcc	5340
atccggtaac tategttttgc agtccaaacc gtaagacac gacttacgc cactggcage	5400
agccactggta aacaggatttta gcagagcgag gtatgttagc ggtgtacag agttttgaa	5460
gtggcgttcc aactacggcgt acactagaag gacagtattt ggtatctgcg ctctgtgaa	5520
gccagttacc ttccggaaaaa gagttggtag ctcttgcatttcc ggcacaaacaa ccaccgcgtt	5580
tagcggtgtt tttttgtttt gcaaggcagea gattacgcgc agaaaaaaaag gatctcaaga	5640
agatcccttttgc atctttcttca cggggcttgc cgcgtactgg aacgaaaact cacgttaagg	5700
gatttggcgtt atgagatttcaaaaaggat cttcacctag atccctttaa attaaaaatg	5760
aagttttaaa tcaatctaaa gtatataatgta gtaaacttgg tctgacagtt accaatgttt	5820
aatcagtgttgc gcacccatctt cagcgatctg tctatccgt tcatccatag ttgcctgtact	5880
ccccgtcgtagtataacta cgcatacggcgg gggcttacca tctggccccc gtgtgcata	5940
gataccgcga gacccacgtt caccggctcc agattttatca gcaataaacc agccagccgg	6000
aaggccgcggcgcg cgcagaagtg gtcctgcacac ttatccgc tccatccagt ctatatttg	6060
ttgcggggaa gctagagtaa gtagttcgcc agttatagt ttgcgcacac ttgttgcatt	6120
tgctacaggcgtt atcggtgttgc caccgtcgatc gtttggatgt gtttcattca gtcgggttcc	6180
ccaaacgtatca aggccgttta catgtatcccc catgtgtgc aaaaaaggcg ttagctcatt	6240
cggccctccg atcggtgtca gaagtaagtt ggcgcagtg ttatcactca tggttatggc	6300
agcactgtcatca aattcttta ctgtcatgcc atccgtatggaa tgctttctg tgactgggtgt	6360
gtactcaacc aagtccatttgc gagaatagtg tatggggcga ccgagttgttcc ttggccggc	6420
gtcaataacgg gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa	6480
acgttccatcg gggcgaaaac tctcaaggat cttaccgcgtt ttgagatcca gttcgatgtt	6540
acccactgttgc gcacccaaact gatcttcgcgtt atctttacttccaccgcgtt ttctgggtgt	6600
agcaaaaaca ggaaggccaaa atgccgcacaa aaaggaaata agggcgacac ggaaatgttgc	6660
aatactcata ctcttcctt ttcataatatttttta ttgaaggcatt tatcagggtt attgtctcat	6720
gagccggat	6728