



(12)发明专利申请

(10)申请公布号 CN 111662995 A

(43)申请公布日 2020.09.15

(21)申请号 201910168842.0

C12Q 1/04(2006.01)

(22)申请日 2019.03.06

C12N 15/11(2006.01)

C12R 1/685(2006.01)

(71)申请人 勐海茶业有限责任公司

地址 666200 云南省西双版纳傣族自治州
勐海县勐海镇茶厂路9号

申请人 云南大益微生物技术有限公司

(72)发明人 徐丹洋 施佳辉 徐平 唐蜀昆
高林瑞 田飞 高慧英 职晓阳
丁章贵

(74)专利代理机构 北京天昊联合知识产权代理
有限公司 11112

代理人 王静 丁业平

(51)Int.Cl.

C12Q 1/6895(2018.01)

C12Q 1/686(2018.01)

权利要求书1页 说明书10页
序列表6页 附图7页

(54)发明名称

DNA条形码、引物、试剂盒、方法 and 应用

(57)摘要

本发明提供了DNA条形码、引物、试剂盒、方法 and 应用。所述DNA条形码来源于黑曲霉TMCC 70012菌株的基因组,并且包含选自如SEQ ID No.4所示的DNA序列中的至少500bp的序列,并且该DNA条形码的长度为500bp-4000bp。本发明的DNA条形码序列能够实现黑曲霉种内菌种的快速鉴别与区分。由此,本发明建立了普洱茶工业发酵生产菌株黑曲霉TMCC 70012的标准基因序列和样品鉴别方法,相比于传统的形态学鉴定方法,显著提高了鉴定效率。

1. 一种用于鉴别黑曲霉菌株的DNA条形码,该DNA条形码来源于黑曲霉TMCC 70012菌株的基因组,并且包含选自如SEQ ID No.4所示的DNA序列中的至少500bp的序列,并且该DNA条形码的长度为500bp-4000bp,优选为500bp-900bp。

2. 根据权利要求1所述的DNA条形码,其核苷酸序列如SEQ ID No.1、SEQ ID No.4或SEQ ID No.7所示。

3. 一种用于扩增权利要求1或2所述的DNA条形码的引物对。

4. 根据权利要求3所述的引物对,其正向引物的核苷酸序列与黑曲霉TMCC 70012菌株基因组中的这样的序列相同:该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.4所示的核苷酸序列的第1位到如SEQ ID No.4所示的核苷酸序列的第3510位的区域中的序列,并且该正向引物的长度为15-30bp;其反向引物与所述TMCC 70012菌株基因组中的这样的序列反向互补:该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.4所示的核苷酸序列的第471位到如SEQ ID No.4所示的核苷酸序列的最后一位的区域中的序列,并且该反向引物的长度为15-30bp。

5. 根据权利要求4所述的引物对,所述正向引物和反向引物的核苷酸序列分别如下所示:

正向引物:5'-ACGCCTCAGTAGAAGATAGT-3' ;

反向引物:5'-CTTGCCATGGTGGTGGGATA-3' 。

6. 一种用于鉴别黑曲霉菌株的试剂盒,包含根据权利要求3-5中任一项所述的引物对。

7. 一种用于鉴别黑曲霉菌株的方法,包括以下步骤:

a) 提供待测菌株的基因组DNA;

b) 以步骤a)所述的基因组DNA为模板,利用根据权利要求3-5中任一项所述的引物对进行PCR扩增,得到PCR产物;

c) 通过琼脂糖凝胶电泳检测PCR产物,如果没有目标条带,则判定待测菌株不是黑曲霉TMCC 70012菌株,如果有目标条带,则进行步骤d);

d) 对所得PCR产物进行测序,得到待测核苷酸序列;将所述待测序列与权利要求1或2所述的DNA条形码的核苷酸序列进行同源性比对,若同源性在97%以上,则判定待测菌株是黑曲霉TMCC 70012菌株。

8. 根据权利要求1或2所述的DNA条形码在鉴别黑曲霉菌株中的应用。

9. 根据权利要求3-5中任一项所述的引物对在鉴别黑曲霉菌株中的应用。

10. 根据权利要求6所述的试剂盒在鉴别黑曲霉菌株中的应用。

DNA条形码、引物、试剂盒、方法和应用

技术领域

[0001] 本发明属于物种和菌株鉴定领域,具体涉及一种用于鉴别普洱茶发酵生产菌株的DNA条形码引物组合物、DNA条形码、试剂盒、方法及应用。具体而言,所述普洱茶发酵生产菌株为黑曲霉(*Aspergillus niger*) TMCC 70012。

背景技术

[0002] 普洱茶是一种具有云南地理标识范围内生产的后发酵茶,其采用大叶种晒青毛茶作为原料,经过一系列工艺制作而成。传统的普洱茶制作过程为:采摘的新鲜茶叶经揉捻、晒青后制成原料晒青毛茶,之后经除杂、潮水、渥堆、晾干、筛分、压制成型、包装出厂。在普洱茶的生产中,渥堆发酵过程是普洱茶品质形成的主要因素,在此过程中,茶叶中诸如茶多酚、咖啡因以及一些多糖物质等内含成分发生了巨大变化,成就了普洱茶的特殊风味、口感、品质以及各种保健功效。

[0003] 在传统的普洱茶生产中,湿热的环境激活茶叶本身所含的酶,使得茶叶中所含的一部分内含成分转化成微生物能利用的物质;普洱茶发酵过程中微生物也大量生长,产生丰富的胞内酶和胞外酶,催化茶叶中内含成分发生一系列转变,由此形成普洱茶独特的品质。不同的产地,微生物种类及其群落结构的差异,使得普洱茶风味及品质也各具特色。

[0004] 普洱茶除了其独特的风味和文化外,其诸如减肥、降血糖和血脂、预防和改善心血管疾病、抗衰老、抗癌、消炎、助消化以及养胃等保健功效也备受人们关注,并越来越受消费者欢迎。逐渐增加的市场需求拉动了普洱茶产业的发展,促进了云南地区经济增长。

[0005] 目前多数生产商的普洱茶生产仍然是半自然人工渥堆的经验式发酵,虽然渥堆过程中以包括黑曲霉(*Aspergillus niger*)在内的优势共性微生物为主的群落相对稳定,但产品的稳定性还存在较大的提升空间。要进一步获得消费者的青睐和市场的认可、突破外贸壁垒,提升普洱茶企业的市场竞争力,必须实现普洱茶生产的人工可控、清洁化以及高效化,并不断开发新产品、延伸产业链。要做到这些,就必须革新技术,发明一系列安全、清洁、高效、人工可控的自动化普洱茶新工艺,才能确保普洱茶产业的健康发展,给国家和人民带来长远利益。

[0006] 人工接种和普洱茶的纯种发酵是普洱茶可控发酵的新发展方向。为保护普洱茶发酵工艺和优质微生物种质资源,开发对普洱茶发酵用菌种快速、精准鉴别方法势在必行。

[0007] DNA条形码技术可快速、简便地鉴别、区分与其相似菌种,可为普洱茶人工可控纯种发酵工艺开发和资源保护提供理论依据和技术手段,促进普洱茶产业健康发展。

发明内容

[0008] 为克服形态学在鉴定普洱茶发酵生产菌上存在的缺陷,本发明提供了用于鉴别普洱茶发酵菌种黑曲霉(*Aspergillus niger*) TMCC 70012的DNA条形码、引物、试剂盒、方法和应用,从而可以准确地将黑曲霉TMCC 70012菌株从易混淆的物种或复合种中鉴别出来,实现进行快速鉴别和区分,可为普洱茶新发酵工艺提供快速鉴别、评估,防止发酵过程中其他

杂菌的干扰,以及为普洱茶人工可控发酵工艺及菌种滥用提供举证方法和依据。

[0009] 为了实现上述目的,本发明采用如下技术方案:

[0010] (1) 一种用于鉴别黑曲霉菌株的DNA条形码,该DNA条形码来源于黑曲霉TMCC 70012菌株的基因组,并且包含选自如SEQ ID No.4所示的DNA序列中的至少500bp的序列,并且该DNA条形码的长度为500bp-4000bp,优选为500bp-900bp。

[0011] (2) 根据(1)所述的DNA条形码,其核苷酸序列如SEQ ID No.1、SEQ ID No.4或SEQ ID No.7所示。

[0012] (3) 一种用于扩增(1)或(2)所述的DNA条形码的引物对。

[0013] (4) 根据(3)所述的引物对,其正向引物的核苷酸序列与黑曲霉TMCC 70012菌株基因组中的这样的序列相同;该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.4所示的核苷酸序列的第1位到如SEQ ID No.4所示的核苷酸序列的第3510位的区域中的序列,并且该正向引物的长度为15-30bp;其反向引物与所述TMCC 70012菌株基因组中的这样的序列反向互补;该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.4所示的核苷酸序列的第471位到如SEQ ID No.4所示的核苷酸序列的最后一位的区域中的序列,并且该反向引物的长度为15-30bp。

[0014] (5) 根据(4)所述的引物对,所述正向引物和反向引物的核苷酸序列分别如下所示:

[0015] 正向引物:5'-ACGCCTCAGTAGAAGATAGT-3';

[0016] 反向引物:5'-CTTGCCATGGTGGTGGGATA-3'。

[0017] (6) 一种用于鉴别黑曲霉菌株的试剂盒,包含根据(3)-(5)中任一项所述的引物对。

[0018] (7) 一种用于鉴别黑曲霉菌株的方法,包括以下步骤:

[0019] a) 提供待测菌株的基因组DNA;

[0020] b) 以步骤a)所述的基因组DNA为模板,利用根据(3)-(5)中任一项所述的引物对进行PCR扩增,得到PCR产物;

[0021] c) 通过琼脂糖凝胶电泳检测PCR产物,如果没有目标条带,则判定待测菌株不是黑曲霉TMCC 70012菌株,如果有目标条带,则进行步骤d);

[0022] d) 对所得PCR产物进行测序,得到待测核苷酸序列;将所述待测序列与(1)或(2)所述的DNA条形码的核苷酸序列进行同源性比对,若同源性在97%以上,则判定待测菌株是黑曲霉TMCC 70012菌株。

[0023] (8) 根据(1)或(2)所述的DNA条形码在鉴别黑曲霉菌株中的应用。

[0024] (9) 根据(3)-(5)中任一项所述的引物对在鉴别黑曲霉菌株中的应用。

[0025] (10) 根据(6)所述的试剂盒在鉴别黑曲霉菌株中的应用。

[0026] 本发明与现有技术相比具有以下优点和积极效果:

[0027] 1、本发明采用蛋白质基因组学技术从黑曲霉TMCC 70012菌株的基因组中发现了漏注释的肽段,该肽段坐落于UDP-吡喃半乳糖变位酶GLF A基因编码蛋白GLF A中。本发明进一步通过仔细的研究和比较分析,基于该GLF A基因开发了DNA条形码,该条形码序列能实现黑曲霉种内菌种的快速鉴别与区分。

[0028] 2、本发明进一步发现,相比于其他基因,GLF A基因的序列(例如SEQ ID No.1)具

有通用、易扩增、易比对的特点,其在黑曲霉种内不同菌株之间差异明显。

[0029] 3、本发明建立了普洱茶工业发酵生产菌株黑曲霉TMCC 70012的标准基因序列和样品鉴别方法,相比于传统的形态学鉴定方法,显著提高了鉴定效率。该方法对于样品的完整程度要求低,鉴别指标能够量化,这对于及时判定普洱茶发酵工艺及其种质资源提供了有效的依据,弥补了该类普洱发酵茶生产菌株基于DNA条形码技术的菌株鉴定的空白。此外,进一步加入了形态学混淆种并运用基于聚类分析的系统发育树法设立鉴定规则进行了鉴定,其鉴定的可靠性和准确性也大大优于常规的分子鉴定方法。

附图说明

[0030] 图1示出新鉴定肽段VPSFQDILQGR的质谱谱图。

[0031] 图2示出了合成肽段VPSFQDILQGR质谱图与原鉴定肽段质谱图对比图;所述原鉴定肽段为本发明经过质谱分析并鉴定得到的肽段;图的上方为原鉴定肽段的质谱图,下方为合成肽段的质谱图。

[0032] 图3示出了SEQ ID NO.1的核苷酸序列,灰色部分为内含子。

[0033] 图4示出了SEQ ID NO.1及其所编码的蛋白质的氨基酸序列的对应关系,灰色部分为新鉴定的肽段VPSFQDILQGR。

[0034] 图5A示出TMCC 70012的细胞全蛋白SDS-PAGE分离图谱,加粗显示的条带为GLF A蛋白所在位置;图5B示出GLF A蛋白分子量验证图。图5B中横坐标表示分子量MW的对数,纵坐标表示蛋白质分子量标记的各分子量蛋白在SDS-PAGE电泳分离中的迁移量相对于SDS-PAGE总迁移量的比值。

[0035] 图6示出条形码序列SEQ ID NO.1通过NCBI-BLASTN得到的与之具有同源性的序列比对图,在图6下方的灰色线段显示出同源性序列。

[0036] 图7示出了SEQ ID NO.4的核苷酸序列,浅灰色部分为SEQ ID NO.1,深灰色部分为SEQ ID NO.1中的内含子,黑体字为引物序列,下划线部分为用该引物扩增的产物序列,即SEQ ID NO.7。

[0037] 图8示出条形码序列SEQ ID NO.4通过NCBI-BLASTN得到的与之具有同源性的序列比对图,在图8下方的灰色线段显示出同源性序列。

[0038] 图9示出利用对本发明的黑曲霉TMCC 70012GLF A基因所设计的引物进行PCR扩增所得产物的琼脂糖凝胶电泳结果。

具体实施方式

[0039] 以下通过具体实施方式的描述并参照附图对本发明作进一步说明,但这并非是对本发明的限制,本领域技术人员根据本发明的基本思想,可以做出各种修改或改进,但是只要不脱离本发明的基本思想,均在本发明的范围之内。

[0040] 本发明利用系统的蛋白质基因组学技术从黑曲霉TMCC 70012菌株中发现了一个难以被传统基因预测软件发现的物种特异基因编码蛋白。由于该基因编码蛋白与UDP-吡喃半乳糖变位酶具有高度同源性,因此将其命名为GLF A。本发明发现编码该蛋白的基因GLF A的开放阅读框(SEQ ID No.1)可以将黑曲霉TMCC 70012菌株从易混淆的物种中鉴别出来,因此可以用于开发鉴别普洱茶工业发酵生产菌株黑曲霉TMCC 70012的DNA条形码。与现有

技术相比,本发明通过上述方法获得的DNA条形码的特异性更高。

[0041] 考虑到DNA条形码应当具有合适的长度和足够的菌株间特异性,本发明进一步通过仔细的研究和比较分析,基于上述特有的DNA序列获得了可准确、有效地鉴别黑曲霉TMCC 70012菌株的DNA条形码。

[0042] 由此,本发明的一个方面提供了一种用于鉴别食腺嘌呤节孢酵母菌株的DNA条形码,该DNA条形码来源于黑曲霉TMCC 70012菌株的基因组,并且包含选自如SEQ ID No.4所示的DNA序列中的至少500bp的序列,并且该DNA条形码的长度为500bp-4000bp,优选为500bp-900bp。

[0043] 本发明通过系统的蛋白质基因组学研究发现了一个难以被传统基因预测软件发现的蛋白编码序列,支持该基因编码序列产物存在的肽段及相关蛋白质组学质谱数据准确可靠;基于该蛋白编码序列在基因组中的位置,确定了编码该蛋白可能的基因序列(SEQ ID No.1)和蛋白序列(SEQ ID No.3)。SEQ ID No.1的长度为1964bp,根据比对分析,SEQ ID No.1为黑曲霉TMCC 70012菌株基因组所特有,因此该序列本身即可作为鉴别黑曲霉TMCC 70012菌株的DNA条形码。另外,通过理论分析和实验验证,本发明证明包含该序列、长度更长的DNA条码的特异性更加得到保证。当DNA条形码的长度过长时(例如大于4000bp),对于扩增操作而言较不理想。由此,基于所述基因序列,本发明在黑曲霉TMCC 70012菌株基因组中寻找到了更长、特异性更好的DNA条形码序列。另一方面,通常,DNA条形码的长度不小于500bp时,也能够满足易扩增、易比对的操作要求。因此,本发明通过理论分析和实验验证,证明选自如SEQ ID No.4所示的DNA序列中的至少500bp的序列也能够实现黑曲霉内菌株的快速准确鉴别与区分。

[0044] 本发明的条形码序列能实现黑曲霉种内菌株的快速准确鉴别与区分。

[0045] 特别地,本发明发现在黑曲霉TMCC 70012菌株基因组中,SEQ ID No.4中所包含的如SEQ ID No.7所示的DNA序列在不同菌种间以及在黑曲霉种内不同菌株之间具有特别高的特异性。因此,在本发明优选的实施方案中,所述DNA条形码包含选自如SEQ ID No.7所示的DNA序列中的至少500bp的序列。

[0046] 优选地,所述DNA条形码序列如SEQ ID No.1、SEQ ID No.4或SEQ ID No.7所示。

[0047] 在本文中,术语“漏注释”是指物种完成基因组测序以后,采用基因预测软件(例如GeneMark、Augustus、Glimmer等)未能将某基因或蛋白预测出来,这些基因或蛋白一般表达量不高、在特定的条件下才表达,所以在研究中难以被发现。

[0048] 术语“DNA条形码技术(DNA barcoding)”是指用基因组内一段标准的、短的DNA片段来鉴定物种的一项分子鉴定新技术,其可以快速、准确的进行物种鉴定。

[0049] 术语“六框翻译”是蛋白质组学和基因组学中的已知术语,简述其原理为,DNA编码蛋白质时,采用三联体密码子编码蛋白质,给定一条DNA序列,有3种编码可能性,加上其互补链上的3种编码可能,共有6种编码可能性(+1,+2,+3,-3,-2,-1)。

[0050] 本发明的另一个方面提供了一种用于扩增本发明所述的DNA条形码的引物对。

[0051] 本领域技术人员可以理解,根据本发明提供的用于鉴别食腺嘌呤节孢酵母菌株的DNA条形码序列,可以容易地设计相应的引物对,从而扩增出所需DNA条形码。

[0052] 优选地,所述引物对的正向引物的核苷酸序列与黑曲霉TMCC70012菌株基因组中的这样的序列相同:该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.4所示的核苷

酸序列的第1位到如SEQ ID No.4所示的核苷酸序列的第3510位的区域中的序列,并且该正向引物的长度一般为15-30bp;其反向引物与所述TMCC 70012菌株基因组中的这样的序列反向互补:该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.4所示的核苷酸序列的第471位到如SEQ ID No.4所示的核苷酸序列的最后一位的区域中的序列,并且该反向引物的长度一般也为15-30bp。所述正反引物扩增出的产物即为选自SEQ ID No.4所示的核苷酸序列中的至少500bp的序列。

[0053] 在优选的实施方案中,所述引物对的正向引物的核苷酸序列与黑曲霉TMCC 70012菌株基因组中的这样的序列相同:该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.7所示的核苷酸序列的第1位到如SEQ ID No.7所示的核苷酸序列的第369位的区域中的序列,并且该正向引物的长度一般为15-30bp;其反向引物与所述TMCC 70012菌株基因组中的这样的序列反向互补:该序列为从所述TMCC 70012菌株基因组中如SEQ ID No.7所示的核苷酸序列的第471位到如SEQ ID No.7所示的核苷酸序列的最后一位的区域中的序列,并且该反向引物的长度一般也为15-30bp。所述正反引物扩增出的产物即为选自SEQ ID No.7所示的核苷酸序列中的至少500bp的序列。

[0054] 在更优选的实施方案中,所述正向引物和反向引物的核苷酸序列分别如下所示:

[0055] GLF A-F:5'-ACGCCTCAGTAGAAGATAGT-3'(SEQ ID No.5);

[0056] GLF A-R:5'-GTGGATGGGACGTAAAGCAT-3'(SEQ ID No.6)。

[0057] 本发明的引物能实现对所述DNA条形码序列的特异性扩增。

[0058] 本发明还提供了一种用于鉴别黑曲霉菌株的试剂盒,包含根据本发明所述的引物对。

[0059] 在另一实施方案中,所述试剂盒还包含根据本发明所述的DNA条形码。所述DNA条形码可存在于记录介质上。该记录介质例如为光盘。所述试剂盒还可以包含用于实验操作的任何工具和试剂。

[0060] 本发明的又一方面提供了一种用于鉴别黑曲霉菌株的方法,包括以下步骤:

[0061] a) 提供待测菌株的基因组DNA;

[0062] b) 以步骤a)所述的基因组DNA为模板,利用根据本发明所述的引物对进行PCR扩增,得到PCR产物;

[0063] c) 电泳(例如琼脂糖凝胶电泳)检测PCR产物,如果没有目标条带,则判定待测菌株不是黑曲霉菌株TMCC 70012,如果有目标条带,则进行步骤d);

[0064] d) 对所得PCR产物进行测序,得到待测核苷酸序列;将所述待测核苷酸序列与本发明所述的DNA条形码的核苷酸序列进行同源性比对,若同源性在97%以上(优选98%以上,更优选99%以上,更优选100%),则判定待测菌株是黑曲霉TMCC 70012菌株。

[0065] 在本文中所使用的术语“同一性”和“同源性”具有相同的含义,可以相互交换使用。

[0066] 在本发明的具体实施方案中,所述PCR扩增的程序为:1) 94℃预变性5分钟;2) 94℃变性30秒,56℃退火30秒,72℃延伸1分钟,其中该程序2)进行30个循环;3) 72℃延伸10分钟。

[0067] 在本发明的又一实施方案中,所述方法还可以包括将测序结果得到的待测核苷酸序列与本发明的DNA条形码进行聚类分析(例如系统发育树),如果待测序列与DNA条形码聚

类在一起,则确定待测菌株为黑曲霉菌株TMCC 70012。

[0068] 在本发明一个具体的实施方案中,利用本发明的引物对提取自待鉴别的菌株的基因组DNA进行PCR扩增,然后进行琼脂糖凝胶电泳检测。基于检测是否存在PCR产物来鉴别菌株:如果待鉴定的菌株没有扩增出相应的目标条带,则说明该菌株不是TMCC 70012;如果扩增出相应的目标条带,则证明该菌株可能是TMCC 70012。为了进一步进行鉴别,对PCR产物进行测序,将DNA测序结果与DNA条形码序列进行同源性比对,获得序列间的相似性(即同源性),如果序列同源性小于97%,则判定待测菌株不是黑曲霉菌株TMCC 70012。如果序列同源性大于或等于97%,则判定待测菌株是黑曲霉菌株TMCC 70012。

[0069] 如果进行聚类分析,例如系统发育树,则将所述DNA条形码和各待鉴定菌株的DNA测序结果(即待测序列)一起应用MEGA 6或PAUP软件构建NJ系统发育树。如果待鉴定菌株的待测序列与黑曲霉TMCC 70012的DNA条形码聚类在一起,则鉴定为黑曲霉TMCC 70012。

[0070] 在此所使用的术语“聚类”是指经系统发育树分析后,处于同一个分支,且进化距离相同。

[0071] 本发明还提供了本发明所述的DNA条形码在鉴别黑曲霉菌株中的应用。

[0072] 本发明还提供了本发明所述的引物对在鉴别黑曲霉菌株中的应用。

[0073] 本发明还提供了本发明所述的试剂盒在鉴别黑曲霉菌株中的应用。

[0074] 实施例

[0075] 下面结合具体的实施例对本发明做进一步说明。实施例中所用方法在未具体说明的情况下,均采用常规方法和已知工具。

[0076] 实施例1:GLF A基因和DNA条形码的获得

[0077] 1、利用高覆盖蛋白质组技术,采用pFind和pAnno软件对普洱茶工业发酵菌黑曲霉TMCC 70012(*Aspergillus niger* TMCC 70012)进行了蛋白质组的深度覆盖研究,并对其基因组进行了注释编码基因的验证。为了发现新的蛋白编码区,利用系统蛋白质基因组学中的六框翻译(Six Frame Translation)策略,获得了黑曲霉TMCC 70012基因组数据的六框翻译数据库,穷举了基因组的6种编码可能(+1,+2,+3,-1,-2,-3),并且将该核酸序列称为“六框翻译核酸序列”,蛋白序列称为“六框翻译蛋白序列”。通常而言,六框翻译核酸序列是从一个终止子到下一个终止子的序列,在本文中,也将其称为“蛋白编码框”。利用这个数据库,采用pFind和pAnno软件对TMCC 70012菌株总细胞蛋白的高覆盖蛋白质组质谱数据进行了新肽段和新蛋白质的鉴定。

[0078] 经鉴定,发现了一条现有技术中黑曲霉TMCC 70012注释基因中未发现的肽段VPSFQDILQGR,质谱谱图如图1所示。

[0079] 手工检查质谱谱图结果显示,检测到了肽段VPSFQDILQGR二级质谱谱图(MS2)的几乎所有y离子序列,匹配好,信号强,结果可信。

[0080] 为进一步确证这个鉴定结果,按照新鉴定肽段VPSFQDILQGR的氨基酸序列化学合成了该肽段,并对合成肽段产生的高能量碰撞MS2进行了核实,一级母离子和二级子离子均符合理论值,表明合成的肽段序列正确,参见图2。

[0081] 在此基础上,人工检查了根据大规模蛋白质组数据鉴定到的新肽段序列合成肽段的MS₂和大规模鉴定新肽段谱图,两者几乎完全一致,以子离子相似性获得的cosin值高达0.99,由此证明从普洱茶工业发酵菌黑曲霉TMCC 70012鉴定到的新肽段正确无误。

[0082] 2、根据新肽段所在的位置,以前一个终止密码子和后一个终止密码子包括的区域为界,得到六框翻译核酸序列(蛋白编码框)SEQ ID NO.2。

[0083] 3、为了进一步确定编码该蛋白的基因的编码起始位点和终止位点,将上述蛋白编码框向上游和下游分别扩展1000bp,采用GeneMark.hmm进行基因预测,参考物种选用粟酒裂殖酵母(*Schizosaccharomyces pombe*)。在该区域预测到了该蛋白编码基因的存在,其读码框与上述蛋白读码框一致,在ORF序列中找到了基因的起始子,但ORF序列的终止子实际位于内含子部分,所以根据基因预测,在后面的序列中找到了真正的终止子,这样根据预测结果,确定了这个基因可能的编码起始和终止位点,其完整的从起始子到终止子的基因序列为SEQ ID NO.1,如图3所示。

[0084] SEQ ID NO.1及其所编码的蛋白质的氨基酸序列的对应关系如图4所示。肽段VPSFQDILQGR坐落在该蛋白的N端附近。

[0085] 该基因(SEQ ID NO.1)的核苷酸序列包括终止子在内共1964bp,除去内含子区域,共编码547个氨基酸,理论分子量为61.0kDa,理论编码的氨基酸序列如SEQ ID NO.3所示。

[0086] 4、为了进一步确定所鉴定的序列的正确性,需要确定所鉴定的VPSFQDILQGR的ORF所在区域编码蛋白的分子量,因此测定了其实验分子量(图5)。在YPD培养基中培养TMCC 70012菌株并提取其中的总蛋白,进行SDS-PAGE分离,电泳胶经考马斯亮蓝染色,根据蛋白质组样品制备时的分子量特征验证分子量,在SDS-PAGE中,蛋白迁移距离和染料前沿的迁移距离的比值与蛋白分子量对数呈一定关系,见图5。整个泳道按分子量和蛋白质丰度共切成30个胶条(fraction),并进行胶内消化和质谱分析,其中GLF A蛋白在第13个胶条的质谱数据中被鉴定到了,与预测的分子量大小匹配。理论分子量为61.0kDa,理论分子量与该蛋白所属的胶条所在SDS-PAGE中的位置是一致的。结果显示鉴定到的GLF A的实验分子量大小在59.2-61.8kDa之间,与理论分子量相符,证实了所鉴定的蛋白的正确性。

[0087] 5、以该基因理论编码产物的氨基酸序列(SEQ ID NO.3)进行NCBI-BLASTP分析,检索到了序列相似性较高的蛋白。Blastp结果表明我们检测到的TMCC-70012菌株中漏注释GLF A基因产物有结构域HemY超家族,说明该序列构成的蛋白在生物大分子中具有特异结构和独立功能。上述blastp比对结果参见表1。

[0088] 表1与TMCC 70012 GLF A蛋白序列具有较高同源性的序列

[0089]

	描述	最大得分	总分	覆盖度	E 值	同源性	登录号
1	假定蛋白 ASPUDRAFT_120157 [塔宾曲霉 (<i>Aspergillus tubingensis</i>) CBS 134.48]	1098	1098	100%	0.0	97%	OJI85387.1
2	假定蛋白 ASPFODRAFT_127725 [琉球曲霉 (<i>Aspergillus luchuensis</i>) CBS 106.47]	1097	1097	100%	0.0	97%	OJZ89409.1
3	假定蛋白 ASPBRDRAFT_38309 [巴西曲霉 (<i>Aspergillus brasiliensis</i>) CBS 101740]	1093	1093	100%	0.0	96%	OJJ75961.1
4	UDP-半乳糖变位酶 [黑曲霉 (<i>Aspergillus niger</i>) CBS 513.88]	1091	1091	100%	0.0	96%	XP_001399999.2
5	UDP-半乳糖变位酶 [河内曲霉 (<i>Aspergillus kawachii</i>) IFO 4308]	1088	1088	95%	0.0	99%	GAA84661.1
6	UDP-半乳糖变位酶 [黑曲霉 (<i>Aspergillus niger</i>)]	1087	1087	95%	0.0	99%	GAQ36662.1
7	FAD/NAD(P)-含结合结构域蛋白 [赭色针状曲霉 (<i>Aspergillus ochraceoroseus</i>) IBT 24754]	1057	1057	95%	0.0	97%	PLB26011.1
8	UDP-半乳糖变位酶 [卢氏曲霉 (<i>Aspergillus luchuensis</i>)]	1054	1054	95%	0.0	97%	GAT23622.1
9	假定蛋白 ALT_1536 [图鲁斯曲霉 (<i>Aspergillus lentulus</i>)]	1050	1050	100%	0.0	91%	PKX88520.1
10	FAD/NAD(P)-含结合结构域蛋白 [新烟曲霉 (<i>Aspergillus novofumigatus</i>) IBT 16806]	1050	1050	100%	0.0	91%	PKX88520.1

[0090] 表2与TMCC 70012 GLF A基因序列具有较高同源性的序列

[0091]

	描述	最大得分	总分	覆盖度	E 值	同源性	登录号
1	黑曲霉 (<i>Aspergillus niger</i>) 重叠群 (contig) An02c0270, 基因组重叠群	3083	3083	100%	0.0	95%	AM270022.1
2	TPA: 构巢曲霉 (<i>Aspergillus nidulans</i>) FGSC A4 染色体 VI	1421	1540	80%	0.0	84%	BN001306.1
3	嗜松篮状菌 (<i>Talaromyces pinophilus</i>) 菌株 1-95 染色体 1, 完整序列	1230	1230	74%	0.0	82%	CP017344.1
4	选择序列 XM_001399962.2 黑曲霉 (<i>Aspergillus niger</i>) CBS 513.88 UDP-半乳糖变位酶, mRNA	1022	2728	81%	0.0	98%	XM_001399962.2
5	选择序列 FJ746724.1 烟曲霉 (<i>Aspergillus fumigatus</i>) 分离 D141 UDP-半乳糖变位酶(glfA)基因, 完整 cds	867	1564	88%	0.0	81%	FJ746724.1
6	米曲霉 (<i>Aspergillus oryzae</i>) RIB40 DNA, SC012	797	1687	83%	0.0	84%	AP007161.1
7	选择序列 XM_022725047.1 带状青霉菌 (<i>Penicillium zonata</i>) CBS 506.65 假定蛋白 (ASPZODRAFT_144858), 部分 mRNA	756	1352	71%	0.0	83%	XM_022725047.1
8	选择序列 XM_001263193.1 费舍尔曲霉 (<i>Aspergillus fischeri</i>) NRRL 181 保守假定蛋白部分 mRNA	745	1656	69%	0.0	89%	XM_001263193.1
9	选择序列 AM920433.1 产黄青霉 (<i>Penicillium chrysogenum</i>) Wisconsin 54-1255 完整基因组, 重叠群 Pc00c18	734	1414	71%	0.0	83%	AM920433.1
10	蛹虫草菌 (<i>Cordyceps militaris</i>) 菌株 ATCC 34164 染色体 III, 完整序列	730	730	71%	0.0	76%	CP023326.1

[0092] 6、将所鉴定的GLF A基因的序列(即SEQ ID NO.1)进行NCBI-BLASTN分析,结果见

图6。BLASTN结果总结在表2中。

[0093] 表2的结果表明GLF A基因的序列(即SEQ ID NO.1)与黑曲霉中的另一菌株有一条同源性较高的序列,但与黑曲霉以外的其他物种之间的差异性较大,可以作为区分TMCC-70012菌株和其他菌种的DNA条形码。另外,如果设定高的同源性比对筛选标准,例如同源性大于等于99%,则该SEQ ID NO.1也可作为区分TMCC-70012菌株和其他黑曲霉菌株的DNA条形码。

[0094] 7、进一步考虑了GLF A基因序列(即SEQ ID NO.1)前后转录间隔区的序列,得到SEQ ID NO.4,如图7所示。对SEQ ID NO.4进行同源性比较,如图8所示。可从图8得知,SEQ ID NO.4在NCBI数据库中特异性较高,在SEQ ID NO.4的2250bp之后的片段几乎没有同源性匹配。说明SEQ ID NO.4可有效地区分TMCC-70012和近源菌株,因此可以作为DNA条形码。

[0095] 实施例2.利用DNA条形码鉴别菌株

[0096] 依据待测样品的扩增结果和与黑曲霉TMCC 70012菌株GLF A基因序列的扩增结果、序列同源性,来判定待测样品是否为普洱茶工业应用菌株黑曲霉TMCC 70012。

[0097] (1)基于SEQ ID NO.4,采用NCBI引物设计工具设计PCR引物,得到正反引物序列分别为:

[0098] GLF A-F:5'-ACGCCTCAGTAGAAGATAGT-3' ;

[0099] GLF A-R:5'-GTGGATGGGACGTAAAGCAT-3'。

[0100] 引物的位置见图7。所扩增的序列为SEQ ID NO.7 (838bp)。

[0101] (2) 菌株来源

[0102] 表3选用的相关菌株的信息

[0103]

物种	菌株	生境	分离地
黑曲霉	CBS 513.88	不详	不详
黑曲霉	CBS 120.49	不详	不详
黑曲霉	CBS 554.65 ^T	单宁没食子酸发酵	美国
黑曲霉	CBS 101697	外指型咖啡豆	肯尼亚
黑曲霉	CBS 116681	杏仁	荷兰
黑曲霉	CBS 117784	海枣 (Date)	摩洛哥
黑曲霉	CBS 119557	葡萄	意大利
黑曲霉	CBS 118725	风信子球茎	荷兰

[0104] (3) 分别提取菌株DNA:采用OMEGA E.Z.N.A.TM的Yeast DNA试剂盒提取基因组DNA,并用灭菌的去离子水将样品的DNA浓度稀释到0.5μg/μL。

[0105] (4) 扩增DNA片段,进行聚合酶链式 (PCR) 反应,所用引物序列分别为:

[0106] GLF A-F:5'-ACGCCTCAGTAGAAGATAGT-3' ;

[0107] GLF A-R:5'-GTGGATGGGACGTAAAGCAT-3'。

[0108] PCR反应体系为50μL,PCR试剂为Thermo ScientificTM的Taq DNA Polymerase(重组):ddH₂O 37.7μL,MgCl₂ 5μL,dNTPs 4μL,正向引物1μL,反向引物1μL,Taq DNA聚合酶0.3μL、DNA模板1μL,不含染料。扩增程序为:94℃预变性5分钟;接下来94℃变性30秒,56℃退火30秒,72℃延伸1分钟,一共进行30个循环;最后72℃延伸10分钟。

[0109] (5) 扩增产物检测:用1.0%的琼脂糖凝胶、1×TBE电泳液电泳检测,使用DNA分子

量marker判断PCR片段大小。若待测菌株没有预期838bp大小的扩增条带,说明该菌株不是黑曲霉TMCC70012;若出现明显清晰的条带,且无杂带,送生物测序公司进行DNA片段测序。

[0110] (6) 该引物仅能在黑曲霉TMCC 70012中实现扩增,在同物种的黑曲霉CBS 513.88、CBS 554.65T、CBS 120.49、CBS 101697、CBS 116681、CBS 119557、CBS 117984、CBS 118725中则不能实现扩增,结果如图8所示。该PCR引物的理论扩增序列为838bp,实际扩增结果和预期相符。为了进一步验证扩增的DNA的序列,进行了测序和同源序列比较。该结果证明SEQ ID NO.7可以非常特异且有效地区分黑曲霉TMCC 70012菌株,从而作为黑曲霉TMCC 70012菌株的DNA条形码。

[0111] (7) 对于有条带的序列的测序结果,首先用软件Chromas检查测序后得到的序列峰图质量,确定峰图质量达到数据分析的要求之后,用DNASTAR软件包中的SeqMan进行正反向序列的拼接。将测序结果进行手工校对、序列拼接,如果待测菌株DNA片段与黑曲霉TMCC 70012标准DNA条形码的同源性在97%以上,即可判断所述的待测菌株很可能为黑曲霉TMCC 70012菌株。

[0112] (8) 还可以进一步进行聚类分析,例如系统发育树,这可以进一步佐证经上述步骤的判断结果。

SEQUENCE LISTING

<110> 勐海茶业有限责任公司

<120> DNA条形码、引物、试剂盒、方法和应用

<130> FI-190367-59:52/C

<160> 7

<170> PatentIn version 3.5

<210> 1

<211> 1964

<212> DNA

<213> 黑曲霉 (*Aspergillus niger*)

<400> 1

```

atgctcagcc tcgcccgcag gactttgaac cgtgtcecca gctttcaaga tattctacaa 60
ggcaggatga cccaccccga tatgtaagga gacctttccc cctccctcaa acggcaagct 120
tgccttcact acaacaagaa gctcattgga acaattgctg accgttgtct ttgtgcttga 180
attcagctcc gtcgacgttc tcgtcattgg tgccggccct actggtctcg gtgccgcgaa 240
gcgtcttaac cagattgtac gcattgcgcg cccctatcg tataccttca ctccaataaa 300
ggctaattgg acaattcgcg ggtgtaggat ggcccttctt ggttgatcgt tgacagcaac 360
gagacccttg gtggtcttgc ttccaccgat gtgacccccg aaggcttcgt atgttgattc 420
ccactttccc ctcaatgggc aggcatgca ctggtcaagc aaacagcgac tgactcgtga 480
caatcacagc tcttcgatgt tgggtgtcac gtcatcttct cccactacaa gtacttcgac 540
gactgcatca acgaggctct cccaaggat gacgactggt acaccacca gcgtatctcc 600
tacgttcgct gccagggcca atgggttccc tacccttcc agaacaacat ctccatgctt 660
cccaagaatg agcaggtcg ctgtatcgat ggctgatcg acgctgccct tgaggctcgt 720
gtcgccaaca ccaagcccca gaacttcgat gagtggattg ttcgccagat ggggtgtcgt 780
atcgccgacc tcttcgatgag accctacaac ttcaaggttt gggctgtgcc tacgaccaag 840
gtaagatacg atgtaccaca tggcgagtaa cggcgtgctg attctcatca gatgcaatgt 900
gcctggttgg gtgagcgtgt tgctgcccct aacgtcaagg ccgtgacgac caacgtcatc 960
cttaacaaga ccgctggtaa ctggggtcct aacgtactt tccgtttccc cgcccgtggt 1020
ggtaccggtg gtatctggat cgctgtcgcc gacactatcc ccaaggagaa gactcgcttc 1080
ggtgagaagg gcaaggctgt caaggtcaat gccacaaca agaccgtcac cctgggtgat 1140
ggcaccactg tcggctacaa gaagctcgtc tccaccatgg ctgtcgactt cctcgctgag 1200
cagatcggcg accaggagct cgttggcctc accaagcagc tcttctactc ctccactcac 1260
gtcattggtg tcggtatccg tggtaaccgc cccgagagaa tcggtgacaa gtgctgggta 1320
agttgacctt ggatggaaac catgaatgac ttcaaaacta acatgctcgt tactagctct 1380
acttccccga ggacaactgc cccttctacc gtgccaccat cttctccaac tactcccccc 1440
acaaccagcc cgagggctcc aagaagcttc ccactctgca gcttgccgat ggctctaagc 1500
cccagagcac tgaggctcag gaggtcctt actggtccat catgttgag gtttccgagt 1560
cttcgatgaa gcctgtcaac cacgagactc tcctggctga ttgcatccag ggtctcgtca 1620

```

acaccgagat gctgaagccc accgatgaga ttgtctccac ctaccaccgc cgcttcgacc 1680
 acggctaccc caccctctcc ctggagcgtg aggggtgctct taccagatc ctgcccagac 1740
 tccaggagaa ggacatctgg acccgtggtc gcttcggtag ctggcgctac gaggtcggtg 1800
 accaggacca ctctttcatg ctcggtgttg aggctgttga caacattgtc aacggcgctg 1860
 tcgagctgac ccttaactac cctgacttcg tcaacggtag acagaacacc gagcgtcggc 1920
 tggttgacgg cgcccagggtt ttcgccaaga gccaggcgca gtaa 1964

<210> 2

<211> 294

<212> DNA

<213> 黑曲霉(Aspergillus niger)

<400> 2

tgaaacgat tccaagcgcc cccaccgct tgcttgteac gtttcgcttt cccgaccgac 60
 cgcaggacgg atacacctat accttcccgc ccattcccc ttcttgctcc ctccatcacc 120
 aacctctctt ctctctctct ctctctcccc tctcatctct ctcttataa ctctgctgtc 180
 cctgcagctt cacaaggctc gttttctatg ctacgctcg cccgcaggac tttgaaccgt 240
 gtccccagct ttcaagatat tctacaaggc aggatgaccc accccgatat gtaa 294

<210> 3

<211> 547

<212> PRT

<213> 黑曲霉(Aspergillus niger)

<400> 3

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

Asn	Asn	Ile	Ser	Met	Leu	Pro	Lys	Asn	Glu	Gln	Val	Arg	Cys	Ile	Asp
145					150				155						160
Gly	Leu	Ile	Asp	Ala	Ala	Leu	Glu	Ala	Arg	Val	Ala	Asn	Thr	Lys	Pro
				165					170					175	
Gln	Asn	Phe	Asp	Glu	Trp	Ile	Val	Arg	Gln	Met	Gly	Val	Gly	Ile	Ala
				180					185					190	
Asp	Leu	Phe	Met	Arg	Pro	Tyr	Asn	Phe	Lys	Val	Trp	Ala	Val	Pro	Thr
				195					200					205	
Thr	Lys	Met	Gln	Cys	Ala	Trp	Leu	Gly	Glu	Arg	Val	Ala	Ala	Pro	Asn
						215						220			
Val	Lys	Ala	Val	Thr	Thr	Asn	Val	Ile	Leu	Asn	Lys	Thr	Ala	Gly	Asn
225						230					235				240
Trp	Gly	Pro	Asn	Ala	Thr	Phe	Arg	Phe	Pro	Ala	Arg	Gly	Gly	Thr	Gly
						245					250				255
Gly	Ile	Trp	Ile	Ala	Val	Ala	Asp	Thr	Ile	Pro	Lys	Glu	Lys	Thr	Arg
						260								270	
Phe	Gly	Glu	Lys	Gly	Lys	Val	Val	Lys	Val	Asn	Ala	Asn	Asn	Lys	Thr
														285	
Val	Thr	Leu	Gly	Asp	Gly	Thr	Thr	Val	Gly	Tyr	Lys	Lys	Leu	Val	Ser
														300	
Thr	Met	Ala	Val	Asp	Phe	Leu	Ala	Glu	Gln	Ile	Gly	Asp	Gln	Glu	Leu
305						310									320
Val	Gly	Leu	Thr	Lys	Gln	Leu	Phe	Tyr	Ser	Ser	Thr	His	Val	Ile	Gly
						325									335
Val	Gly	Ile	Arg	Gly	Thr	Arg	Pro	Glu	Arg	Ile	Gly	Asp	Lys	Cys	Trp
						340								350	
Leu	Tyr	Phe	Pro	Glu	Asp	Asn	Cys	Pro	Phe	Tyr	Arg	Ala	Thr	Ile	Phe
														365	
Ser	Asn	Tyr	Ser	Pro	His	Asn	Gln	Pro	Glu	Gly	Ser	Lys	Lys	Leu	Pro
														380	
Thr	Leu	Gln	Leu	Ala	Asp	Gly	Ser	Lys	Pro	Gln	Ser	Thr	Glu	Ala	Gln
385						390									400
Glu	Gly	Pro	Tyr	Trp	Ser	Ile	Met	Leu	Glu	Val	Ser	Glu	Ser	Ser	Met
						405									415
Lys	Pro	Val	Asn	His	Glu	Thr	Leu	Leu	Ala	Asp	Cys	Ile	Gln	Gly	Leu
														430	
Val	Asn	Thr	Glu	Met	Leu	Lys	Pro	Thr	Asp	Glu	Ile	Val	Ser	Thr	Tyr
														445	
His	Arg	Arg	Phe	Asp	His	Gly	Tyr	Pro	Thr	Pro	Ser	Leu	Glu	Arg	Glu

450	455	460
Gly Ala Leu Thr Gln Ile Leu Pro Arg Leu Gln Glu Lys Asp Ile Trp		
465	470	475
Thr Arg Gly Arg Phe Gly Ser Trp Arg Tyr Glu Val Gly Asn Gln Asp		
	485	490
His Ser Phe Met Leu Gly Val Glu Ala Val Asp Asn Ile Val Asn Gly		
	500	505
Ala Val Glu Leu Thr Leu Asn Tyr Pro Asp Phe Val Asn Gly Arg Gln		
	515	520
Asn Thr Glu Arg Arg Leu Val Asp Gly Ala Gln Val Phe Ala Lys Ser		
	530	540
Gln Ala Gln		

545

<210> 4

<211> 3979

<212> DNA

<213> 黑曲霉(*Aspergillus niger*)

<400> 4

```

atgctcagcc tcgcccgcag gactttgaac cgtgtcccca gctttcaaga tattctacaa 60
ggcaggatga cccaccccga tatgtaagga gacctttccc cctccctcaa acggcaagct 120
tgcccttact acaacaagaa gctcattgga acaattgctg accgttgtct ttgtgcttga 180
attcagctcc gtcgacgttc tcgtcattgg tgccggccct actggtctcg gtgccgcgaa 240
gcgtcttaac cagattgtac gcattgcgcg cccctatcg tataccttca ctgaataaa 300
ggctaattgg acaattcgcg ggtgtaggat ggcccttctt ggttgatcgt tgacagcaac 360
gagacccctg gtggtcttgc ttccaccgat gtgacccccg aaggcttcgt atgttgattc 420
ccactttccc ctcaatgggc aggcattgca ctggtcaagc aaacagcgac tgactcgtga 480
caatcacagc tcttcgatgt tggtaggtcac gtcatcttct cccactacaa gtacttcgac 540
gactgcatca acgaggctct cccaaggat gacgactggt acaccacca gcgtatctcc 600
tacgttcgct gccagggcca atgggttccc tacccttcc agaacaacat ctccatgctt 660
cccaagaatg agcaggctcg ctgtatcgat ggctgatcg acgtgccct tgaggctcgt 720
gtcgccaaca ccaagcccca gaacttcgat gagtggattg ttcgccagat ggggtgctggt 780
atcgccgacc tcttcatgag accctacaac ttcaaggttt gggctgtgcc tacgaccaag 840
gtaagatacg atgtaccaca tggcgagtaa cggcgtgctg attctcatca gatgcaatgt 900
gcctggttgg gtgagcgtgt tgetgcccct aacgtcaagg cctgacgac caacgtcatc 960
cttaacaaga ccgctggtaa ctggggtcct aacgtactt tccgtttccc cgcccgtggt 1020
ggtagcgggt gtatctggat cgtgtcgcc gacactatcc ccaaggagaa gactcgttc 1080
ggtagagaagg gcaaggtcgt caaggtcaat gccacaaca agaccgtcac cctgggtgat 1140
ggcaccactg tcggctacaa gaagctcgtc tccaccatgg ctgtcgactt cctcgctgag 1200
cagatcggcg accaggagct cgttggcctc accaagcagc tcttctactc ctccactcac 1260

```

```

gtcattgggtg tcggtatccg tggtagccgc cccgagagaa tcggtgacaa gtgctgggta 1320
agttgacctt ggatggaaac catgaatgac ttcaaaacta acatgctcgt tactagctct 1380
acttccccga ggacaactgc cccttctacc gtgccacat cttctccaac tactcccccc 1440
acaaccagcc cgagggctcc aagaagcttc ccactctgca gcttgcgat ggctctaagc 1500
cccagagcac tgaggctcag gagggtcctt actgggtccat catgttgag gtttccgagt 1560
cttcgatgaa gcctgtcaac cacgagactc tcctggctga ttgcatccag ggtctcgtca 1620
acaccgagat gctgaagccc accgatgaga ttgtctccac ctaccaccgc cgcttcgacc 1680
acggctaccc cccccctcc ctggagcgtg aggggtcctt taccagatc ctgccagac 1740
tccaggagaa ggacatctgg acccgtggtc gcttcggtag ctggcgctac gaggtcggta 1800
accaggacca ctctttcatg ctcggtgttg aggtgttga caacattgtc aacggcgctg 1860
tcgagctgac ccttaactac cctgacttcg tcaacgtag acagaacacc gagcgtcggc 1920
tggttgacgg cgcccagggtt ttcgccaaga gccaggcgca gtaaagggtt gaaatgttaa 1980
tagaattaca cgggttttaa ttttttctt gaatgattcc ggggggcaaa agatttgttt 2040
agagatctga ttacgatcaa cgggagccat atttcgtcat ttttttactt cttgtcctat 2100
ctatttaact tattttacca attttctttt tgctcttttg ctcatgaaa gcgatgtatg 2160
ggagatgttc ttggcatttt tgtgatcgt atcatcatcc actttaccag ccgtcttttg 2220
ctccatctac tagacatgaa tagaagttgc ttatcaaagc cgtgtctttc tgttggggac 2280
ttaatttgct gatttacgta tgccggttac atgtctcgcc gcgaccccggt ggagatcgaa 2340
tccagcttca tgacagatga cacatgtatc accaccgaa gtgaggcagt gtgtgcagca 2400
catcagccca tgcaggcggt ttctaaattc gtcacttct gtcccgatca agtatagcaa 2460
tgtgatatat agcagcatcg caacaactga gtatggagta tggagtatgc agaacggagt 2520
agtatggaat acatcaaacc cacacatatg gagtcatcta ctactacttc gtacttatgt 2580
acaccaatca tctagagaat aaatactact gcagtgggcc aatcggtatt ccgtgtaaga 2640
catactaccg tacgtagtga aacatgggat ctaccgttag ctattaatca atatctgaca 2700
tactatctat ctatcctacc gtgtgataac tccatattct cctcagtact aaccacatc 2760
cccccatcc acaaccacc caccgcccgt ctcataaggc tgagtgagtg agctcgagct 2820
cgagctcgga gggtaggat gaagaagaga gtgcactgac tatgtcgga attttcttta 2880
ctcattgttc tactactgta tatctgagta tgctagtaac tactagtagt aatcatgcaa 2940
gtatcccata gttgagcatg tgaattttat atcagtagta gtatcacagc agttactata 3000
tatgatcatg ctgcaggcat acttacatac atatattcac atgtacatac tatggggacc 3060
atcatgtgaa tactacttac atagatacta ctaagtacgt agtacttaac ttagcattcg 3120
cctccaatac tagagaccac cagcctcag tagaagatag taatttaccg tgataatatt 3180
attaaccacc aaccttacta gtacttactt actatcctcc ccccggaaca tccattacc 3240
gcaaccaacc atgccaagaa ggggaaccac cccacacacc taaagcggtg tagcgactg 3300
gggcaggacg cgtaagtaga actaaacca agaaaggaaa aaaaaaaaaa gaatgaaat 3360
tgggaggggag agtcgtgcag aaaaaatagt atccactatt tatccagcat ccacaggcag 3420
aggaaaatta attcccaatg taaaaataga ttttgattat tagtagtagt ctagcccagg 3480
taggacagggt ggtcagaact atcttactct attatttttg cccctcatgg atggccgatg 3540
acgattaaat ggagtgtgat ggcgatgagg gtggtttgtt tcatggatag tagatactct 3600

```

gtagtactct aggattatgc tagtattgag gtgatactga ggtaataatg gtgatgtcta 3660
 atttggttatt tgcagtaatg atagtatctt tcagggaaga gggataccta catacacatg 3720
 gttacataac gcaagttggc agctttggat tccatccgct tcagccttat tgtcggtcgg 3780
 ttttgctttc ttatttgagc tccccagtct acaccgtcac tcctctacat tacactacac 3840
 tacaccatca ttactgaatt ttctgcagct tgcgtgtcga ctgggctcat tgggatggat 3900
 gacgaatagc ctcacgagat ctcacccttc attgggtccg gacaccaact ccgcccccca 3960
 tgctttacgt cccatccac 3979

<210> 5

<211> 20

<212> DNA

<213> 人工序列(artificial sequence)

<400> 5

acgcctcagt agaagatagt 20

<210> 6

<211> 20

<212> DNA

<213> 人工序列(artificial sequence)

<400> 6

gtggatggga cgtaaagcat 20

<210> 7

<211> 838

<212> DNA

<213> 黑曲霉(*Aspergillus niger*)

<400> 7

acgcctcagt agaagatagt aatttaccgt gataatatta ttaaccacca accttactag 60
 tacttactta ctatcctccc cccggaacat ccattaccgg caaccaacca tgccaagaag 120
 gggaaccacc ccacacacct aaagcggat agcgcactgg ggcaggacgc gtaagtagaa 180
 ctaaacccaa gaaaggaaaa aaaaaaaaaag aatgaaaatt gggagggaga gtcgtgcaga 240
 aaaatatgta tccactatct atccagcatc cacaggcaga ggaaaattaa ttcccaatgt 300
 aaaaatagat tttgattatt agtagtagtc tagcccagggt aggacagggtg gtcagaacta 360
 tcttactcta ttatttttgc cctcatgga tggccgatga cgattaaatg gagtgtgatg 420
 gcgatgaggg tggtttgttt catggatagt agatactctg tagtactcta ggattatgct 480
 agtattgagg tgatactgag gtaataatgg tgatgtctaa tttgttattt gcagtaatga 540
 tagtatcttt cagggaagag ggatacctac atacacatgg ttacataacg caagttggca 600
 gctttggatt ccatccgctt cagccttatt gtcggtcggg tttgctttct tatttgagct 660
 cccagtccta caccgtcact cctctacatt aactacact acaccatcat tactgaattt 720
 tctgcagctt gcgtgtcgac tgggtcatt gggatggatg acgaatagcc tcacgagatc 780
 tcacccttca ttgggtccgg acaccaactc cgcccccat gctttacgtc ccatccac 838

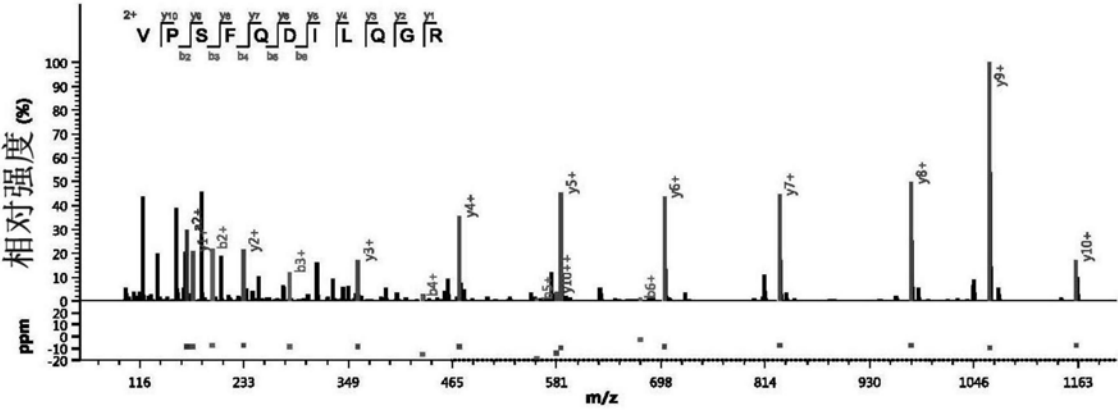


图1

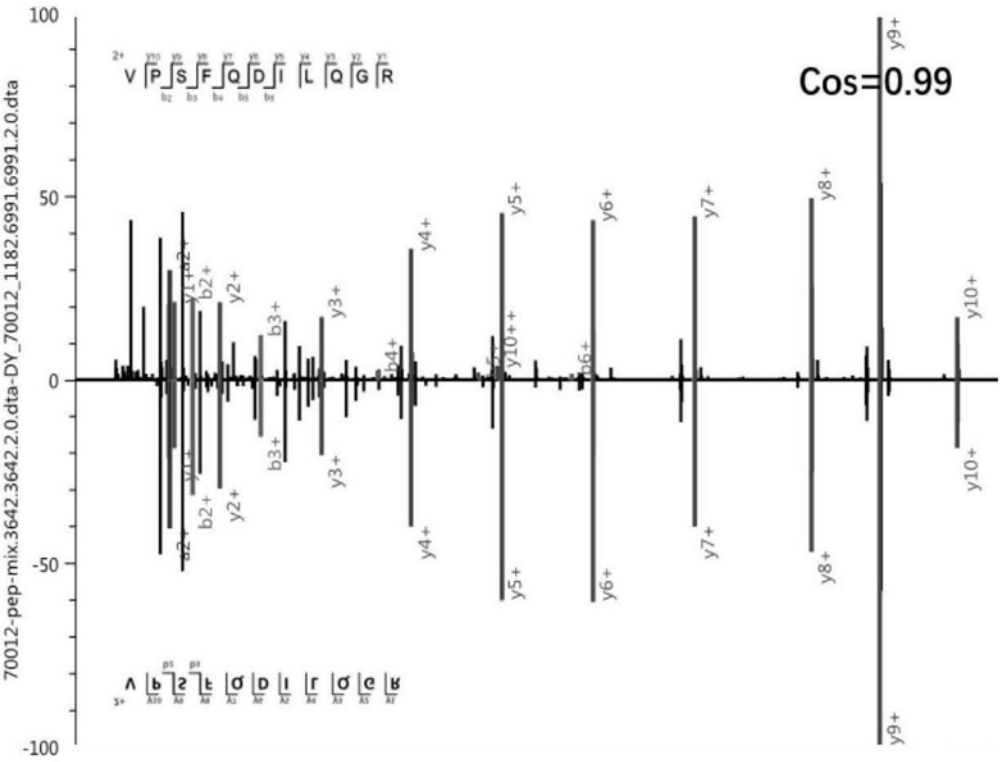


图2

ATGCTCAGCCTCGCCCGCAGGACTTTGAACCGTGTCCCCAGCTTTCAAGATATTCTACAAGGCAGGATGACCCACCCGATATGTA
AGGAGACCTTTCCCCCTCCCTCAAACGGCAAGCTTGCCTTCACTACAACAAGAGCTCATTGGAACAATTGCTGACCGTTGTCTTT
GTGCTTGAATTCAGCTCCGTCGACGTTCTCGTCATTGGTGCCGGCCCTACTGGTCTCGGTGCCGCGAAGCGTCTTAACCAGATTGT
ACGCATTGCGCGCCCCCTATCGTATACCTTCACTCGAATAAAGGCTAATTGGACAATTCGCGGGTGTAGGATGGCCCTTCTTGTT
GATCGTTGACAGCAACGAGACCCCTGGTGGTCTTGCTTCCACCGATGTGACCCCGAAGGCTTCGTATGTTGATTCCCACTTTCCC
CTCAATGGGCAGGCATTGCACTGGTCAAGCAAACAGCGACTGACTCGTGACAATCACAGCTCTTCGATGTTGGTGGTCACGTCATC
TTCTCCCACTACAAGTACTTCGACGACTGCATCAACGAGGCTCTCCCAAGGATGACGACTGGTACACCCACCAGCGTATCTCCTA
CGTTCGCTGCCAGGGCCAATGGGTTCCCTACCCCTTCCAGAACAACATCTCCATGCTTCCCAAGAATGAGCAGGTCCGCTGTATCG
ATGGCCTGATCGACGCTGCCCTTGAGGCTCGTGTGCCAACACCAAGCCCCAGAACTTCGATGAGTGGATTGTTGCCAGATGGGT
GTCGGTATCGCCGACCTCTTCATGAGACCCTACAACCTCAAGGTTTGGGCTGTGCCTACGACCAAGGTAAGATACGATGTACCACA
TGGCGAGTAACGGCGTGCTGATTCTCATCAGATGCAATGTGCCTGGTTGGGTGAGCGTGTGCTGCCCCCTAACGTCAAGGCCGTGA
CGACCAACGTCATCCTTAACAAGACCGCTGGTAACCTGGGGTCTTAACGCTACTTTCCGTTTCCCCGCCCGTGGTGGTACCGGTGGT
ATCTGGATCGCTGTGCGCGACACTATCCCAAGGAGAAGACTCGCTTCGGTGAGAAGGGCAAGGTCGTCAAGGTCAATGCCAACAA
CAAGACCGTCACCTGGGTGATGGCACCCTGTCGGCTACAAGAAGCTCGTCTCCACCATGGCTGTCGACTTCTCGCTGAGCAGA
TCGGCGACCAGGAGCTCGTTGGCCTCACCAGCAGCTCTTCTACTCCTCCACTCAGTCATTGGTGTGCGTATCCGTGGTACCCGC
CCCGAGAGAATCGGTGACAAGTGCTGGGTAAGTTGACCTTGGATGGAAACCATGAATGACTTCAAACTAACATGCTCGTTACTAG
CTCTACTTCCCCGAGGACAACCTGCCCCCTTCTACCGTGCCACCATCTTCTCCAACTACTCCCCCAACAACAGCCCGAGGGCTCCAA
GAAGCTTCCCACTCTGCAGCTTGCGGATGGCTCTAAGCCCCAGAGCACTGAGGCTCAGGAGGGTCCTTACTGGTCCATCATGTTGG
AGGTTTCCGAGTCTTCGATGAAGCCTGTCAACCACGAGACTCTCCTGGCTGATTGCATCCAGGGTCTCGTCAACACCGAGATGCTG
AAGCCACCGATGAGATTGTCTCCACCTACCACCGCGCTTCGACCACGGCTACCCACCCCTCCCTGGAGCGTGAGGGTGCTCT
TACCCAGATCCTGCCAGACTCCAGGAGAAGGACATCTGGACCCGTGGTTCGCTTCGGTAGCTGGCGCTACGAGGTGGTAACCAGG
ACCACTCTTTCATGCTCGGTGTTGAGGCTGTTGACAACATTGTCAACGGCGCTGTGAGCTGACCCCTAACTACCTGACTTCGTC
AACGGTAGACAGAACACCGAGCGTCGGCTGGTTGACGGCGCCAGGTTTTCGCCAAGAGCCAGGCGCAGTAA

图3

1 ATGCTCAGCTCGCCGCGAGGACTTTGAACCGTGTCCCGAGCTTTCAAGATATTCTACAAGGCAGGATGACCCACCCCGATATCTCATTG
1 M L S L A R R T L N R **V P S F Q D I L Q G R** M T H P D I S L

91 GAACAATTGCTGACCGTTGTCTTTGTGCTGAATTCAGCTCCGTCGACGTTCTCGTCATTGGTGCCGCCCTACTGGTCTCGGTGCCGCG
31 E Q L L T V V F V L E F S S V D V L V I G A G P T G L G A A

181 AAGCGTCTTAACCAGATTGATGGCCCTTCTTGGTTGATCGTTGACAGCAACGAGACCCCTGGTGGTCTTGCTTCCACCGATGTGACCCCC
61 K R L N Q I D G P S W L I V D S N E T P G G L A S T D V T P

271 GAAGGCTTCTCTTCGATGTTGGTGGTCACGTCATCTTCTCCACTACAAGTACTTCGACGACTGCATCAACGAGGCTCTCCCAAGGAT
91 E G F L F D V G G H V I F S H Y K Y F D D C I N E A L P K D

361 GACGACTGGTACACCCACCGATATCTCTACGTTGCGTGCAGGGCCAATGGGTTCCTACCCCTTCCAGAACAACATCTCCATGCTT
121 D D W Y T H Q R I S Y V R C Q G Q W V P Y P F Q N N I S M L

451 CCCAAGAATGAGCAGGTCCGCTGTATCGATGGCTGATCGACGCTGCCCTTGAGGCTCGTGTGCCAACACCAAGCCCCAGAACTTCGAT
151 P K N E Q V R C I D G L I D A A L E A R V A N T K P Q N F D

541 GAGTGGATTGTTGCGCAGATGGGTGCGGTATCGCCGACCTCTTCATGAGACCCTACAACCTCAAGGTTTGGGCTGTGCCTACGACCAAG
181 E W I V R Q M G V G I A D L F M R P Y N F K V W A V P T T K

631 ATGCAATGTGCCTGGTTGGGTGAGCGTGTGCTGCCCTAACGTCAAGCCGTGACGACCAACGTCATCCTTAACAAGACCGCTGGTAAC
211 M Q C A W L G E R V A A P N V K A V T T N V I L N K T A G N

721 TGGGGTCTAACGCTACTTTCGTTTCCCGCCCGTGGTGGTACCGGTGGTATCTGGATCGCTGTGCCGACACTATCCCCAAGGAGAAG
241 W G P N A T F R F P A R G G T G G I W I A V A D T I P K E K

811 ACTCGCTTCGGTGAGAAGGGCAAGGTGCTCAAGTCAATGCCAACAACAAGACCGTCACCCCTGGGTGATGGCACCCTGTGCGGTACAAG
271 T R F G E K G K V V K V N A N N K T V T L G D G T T V G Y K

901 AAGCTCGTCTCCACCATGGCTGTGACTTCTCGCTGAGCAGATCGGGACAGGAGCTCGTTGGCCTACCAAGCAGCTCTTCTACTCC
301 K L V S T M A V D F L A E Q I G D Q E L V G L T K Q L F Y S

991 TCCACTCAGTCATTGGTGTGCGGTACCGGTACCGCCCCGAGAGAATCGGTGACAAGTGCTGGCTCTACTTCCCGAGGACAACCTGC
331 S T H V I G V G I R G T R P E R I G D K C W L Y F P E D N C

1081 CCCTTCTACCGTGCACCATCTTCTCCAATACTCCCCCACAACCAGCCCGAGGGCTCCAAGAAGCTTCCACTCTGCAGCTTGC GGAT
361 P F Y R A T I F S N Y S P H N Q P E G S K K L P T L Q L A D

1171 GGCTCTAAGCCCCAGAGCACTGAGGCTCAGGAGGTCCTTACTGGTCCATCATGTTGGAGGTTTCCGAGTCTTCGATGAAGCCTGTCAAC
391 G S K P Q S T E A Q E G P Y W S I M L E V S E S S M K P V N

1261 CACGAGACTCTCTGGCTGATTGCATCCAGGCTCTCGTCAACACCGAGATGCTGAAGCCCACCGATGAGATTGTCTCCACCTACCACCGC
421 H E T L L A D C I Q G L V N T E M L K P T D E I V S T Y H R

1351 CGCTTCGACCACGGTACCCCAACCCCTCCCTGGAGCGTGAGGTGCTCTTACCCAGATCCTGCCAGACTCCAGGAGAAGGACATCTGG
451 R F D H G Y P T P S L E R E G A L T Q I L P R L Q E K D I W

1441 ACCCGTGGTTCGCTTCGGTAGCTGGCGCTACGAGTGGTAACCGAGGACCTCTTTCATGCTCGGTGTTGAGGCTGTTGACAACATTGTC
481 T R G R F G S W R Y E V G N Q D H S F M L G V E A V D N I V

1531 AACGGCGCTGTCGAGCTGACCCCTTAACCTACCTGACTTCGTCAACGGTAGACAGAACACCGAGCGTCGGCTGGTTGACGGCGCCAGGTT
511 N G A V E L T L N Y P D F V N G R Q N T E R R L V D G A Q V

1621 TTCGCCAAGAGCCAGGCGCAGTAA
541 F A K S Q A Q *

图4

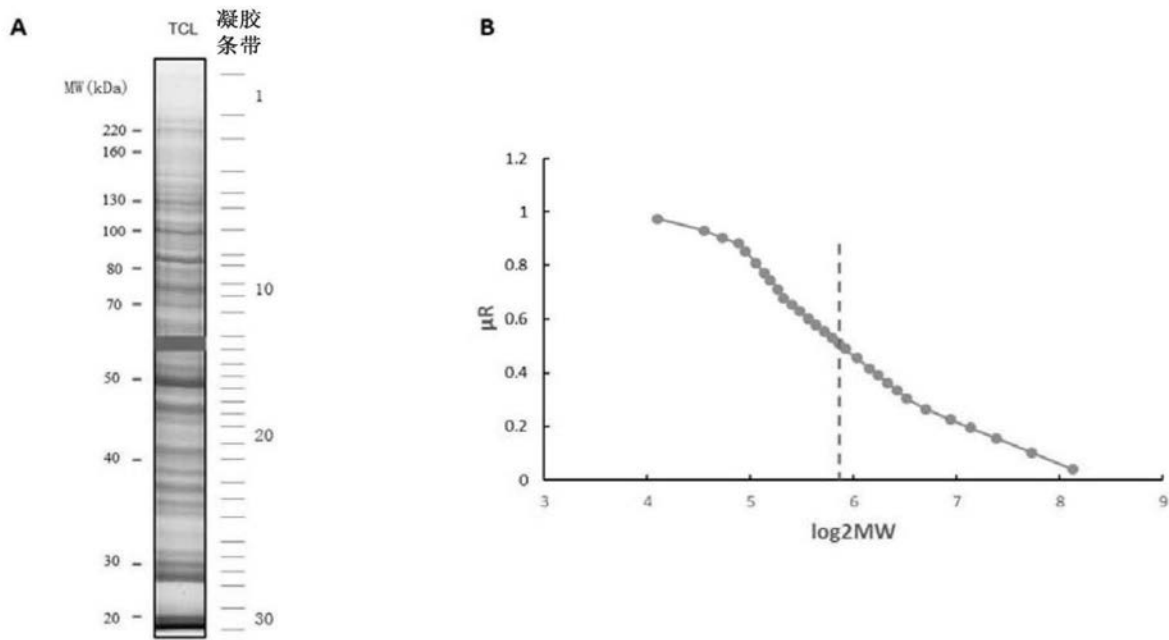


图5

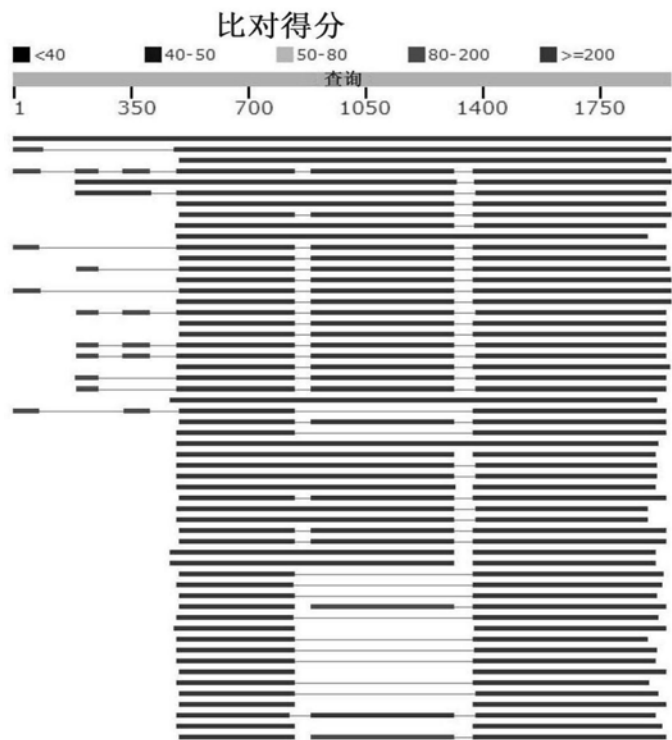


图6

ATGCTCAGCCTCGCCCGCAGGACTTTGAACCGTGTCCCAGCTTTCAAGATATTTACAAAGGCAGGATGACCCACCCGATATGTAAGGAGACCTTTCCCTCCCTCAA
ACGGCAAGCTTGCCCTTACTACAACAAGAAGCTCATTGGAACAATTGCTGACCGTTGCTTTGTGCTTGAATTCAGCTCCGTCGACGTTCTCGTCAATTGGTGCCGGCCCT
ACTGGTCTCGGTGCCGGAAGCGCTTAACAGATTGTACGCATTGCGCGCCCTATCGTATACCTTCACTCGAATAAAGGCTAATTGGACAATTTCGCGGGTGTAGGAT
GGCCCTTCTTGGTTGATCGTTGACAGCAACGAGACCCCTGGTGGTCTTGCTTCCACCGATGTGACCCCGAAGGCTTCGTATGTTGATTCCCACTTTCCCTCAATGGG
AGGCATTGCAGTGGTCAAGCAAACAGCGACTGACTCGTGACAATCACAGCTCTTCGATGTTGGTGGTCAAGTCATCTTCTCCCACTACAAGTACTTCGACGACTGCATCA
ACGAGGCTCTCCCAAGGATGACGACTGGTACCCACCAGCGTATCTCTACGTTGCTGCCAGGGCCAATGGGTTCCCTACCCCTTCCAGAACACATCTCCATGCTT
CCCAAGAATGAGCAGGTCCGCTGTATCGATGGCTGATCGACGCTGCCCTTGAGGCTCGTGTGCCAACACCAAGCCCGAGAAGTTCGATGAGTGGATTGTTCGCCAGAT
GGGTGTCGGTATCGCCGACCTTTCATGAGACCTTACAACCTCAAGGTTTGGGCTGTGCTACGACCAAGGTAAGATACGATGTACCACATGGCGAGTAACGGCGTGTG
ATTCTCATCAGATGCAATGTGCCTGGTGGGTGAGCGTGTGCTGCCCTAACGTCAAGGCCGTGACGACCAACGTCATCCTTAACAAGACCGCTGGTAACGGGGTCT
AACGCTACTTTCCGTTTCCCGCCCGTGGTGGTACCGGTGGTATCTGGATCGCTGTGCCGACACTATCCCAAGGAGAAGACTCGCTTCGGTGAGAAGGCAAGGTCGT
CAAGGTCAATGCCAACAAAGACCGTCAACCTGGGTGATGACCACTGTGCGCTACAAGAAGCTCGTCTCCACCATGGCTGTCGACTTCTCGCTGAGCAGATCGCG
ACCAGGAGCTCGTTGGCTTCAACAGCAGCTTCTACTCTCCACTCACGTCATTGGTGTGGTATCCGTTGTTACCGCCCCGAGAGAATCGGTGACAAGTGTGGGTA
AGTTGACCTTGGATGGAACCATGAATGACTTCAAACTAACATGCTGTTACTAGCTCTACTTCCCGAGGACAAGTCCCTTCTACCGTGCCACCATCTTCTCCAAC
TACTCCCCCACAACAGCCCGAGGGCTCCAAGAAGTTCCTCACTCTGCAGCTTGGGATGGCTCTAAGCCCGAGGACTGAGGCTCAGGAGGGTCTTACTGGTCCAT
CATGTGGAGGTTTCCGAGTCTTCGATGAAGCCTGTCAACACGAGACTCTCTGGCTGATTGCATCCAGGGTCTCGTCAACACCGAGATGTGAAGCCACCGATGAGA
TTGTCTCCACTACCACCGCCGTTCCGACCAAGGCTACCCACCCCTCCCTGGAGCGTGAGGGTGTCTTACCAGATCTGCCAGACTCCAGGAGAAGGACATCTGG
ACCCGTGGTCTGCTTCGGTAGCTGGCGCTACGAGGTGGTAACAGGACCACTCTTTCATGCTCGGTGTTGAGGCTGTTGACAACATTGTCAACGGCGCTGTCGAGCTGAC
CCTTAACACCTGACTTCGTCAACGGTAGACAGAACACCGAGCGTGGCTGGTGTGACGGCGCCAGGTTTTCGCCAAGGCCAGGCGAGTAAAGGGTTGAAATGTAA
TAGAATTACAGGGTTTAAATTTTCTTGAATGATTCGGGGGGGAAAAGATTGTTTAGAGATCTGATTACGATCAACGGGAGCCATATTCGTCATTTTTTACTT
CTTGCTCTATCTATTTAACTTATTTTACCAATTTCTTTTGTCTTTGGCTCATTGAAAGCGATGTATGGGAGATGTTCTTGGCATTTTTGTGATGCGTATCATATCC
ACTTTACAGCGCTCTTTTGTCTTACTACTAGACATGAATAGAAGTTGCTTATCAACCGCTGTCTTTCTGTTGGGACTTAATTTGTGATTTACGTATGCCGGTTAC
ATGCTCGCGGACCCCGTGGAGATCGAATCCAGCTTCATGACAGATGACACATGTATCACCCACCGAAGTGAGGCAGTGTGTGAGCAGATCAGCCCATGCAGGGCGT
TTCTAAATTGCTCACTTCTGTCCGATCAAGTATAGCAATGTATATAGCAGCATCGAACAACAGTATGAGTATGGAGTATGAGATGAGAACCGAGTATGGAAT
ACATCAAAACCACACATATGGAGTCACTACTACTTCTGATCTATGTACACCAATCATCTAGAGAATAAATACTACTGCAGTGGTCAATCGGTATTCGGTGAAGA
CATACTACCGTACGTAGTGAACATGGGATCTACCGTTAGCTATTAATCAATATCTGACATACTATCTATCTATCTCTACCGTGTGATACTCCATATTCTCTCAGTACT
AACCACATCCCCCATCCACAACCCACCCCGCGTCTATAAGGCTGAGTGTGAGCTCGAGCTCGAGCTCGGAGGGTATGGATGAAGAAGAGAGTGCAGTGC
TATGTGGAATTTTCTTTACTATTGTTCTACTACTGTATATCTGAGTATGTAGTAAGTACTAGTAGTAATCATGCAAGTATCCCATAGTTGAGCATGTGAATTTTAT
ATCAGTAGTAGTATCACGAGTACTATATATGATCATGCTGCAGGCATCTTACATACATATATTACATGTACATACTATGGGACCATCATGTGAATACTACTTAC
ATAGATACTACTAAGTACGTAGTACTTAACTTAGCATTGCGCTCCAATAGTAGAGACCACCACGCCCTCAGTAGAAGATAGTAATTTACCGTGATAATATTATTAACACC
AACCTTACTAGTACTTACTTACTATCTCCCCCGGAACATCCATTACCCGCAACCAACCATGCCAAGAAGGGGAACACCCACACACCTAAAGCGGTATAGCGCACTG
GGGAGGACGCGTAAGTAGAACTAAACCAAGAAAGGAAAAAAAAAAGAATGAAAATTGGGAGGGAGAGTCGTGCAGAAAAATATGTATCCACTATTTATCCAGCAT
CCACAGGAGAGGAAATTAATTCCTAATGTAAAAATAGATTTTGATTATTAGTAGTAGCTAGCCAGGTAGGACAGGTGGTGCAGAACTATCTTACTCTATTATTTTG
CCCCTCATGGATGGCGATGACGATTAATGGAGTGTGATGGCGATGAGGGTGGTTGTTTATGGATAGTAGATACTCTGTAGTACTTAGGATTATGCTAGTATTGAG
GTGATACTGAGGTAATAATGGTGTGCTAATTTGTTATTGTCAGTAATGATAGTATCTTTCAGGGAAGAGGATACCTACATACATGTTACATAACGCAAGTTGGC
AGCTTTGGATTCCATCCGCTTACGCTTATTGTCGGTGGTTTTGTCTTCTTATTGAGCTCCCGAGTCTACACCGTCACTCTCTACATTACACTACACTACACCATCA
TTACTGAATTTCTGAGCTTGGCTGTGACTGGGCTCATTGGGATGGATGACGAATAGCCTACGAGATCTCACCTTCATTGGGTCCGGACCAACTCCGCCCCCA
TGCTTTACGTCCCATCCAC

图7

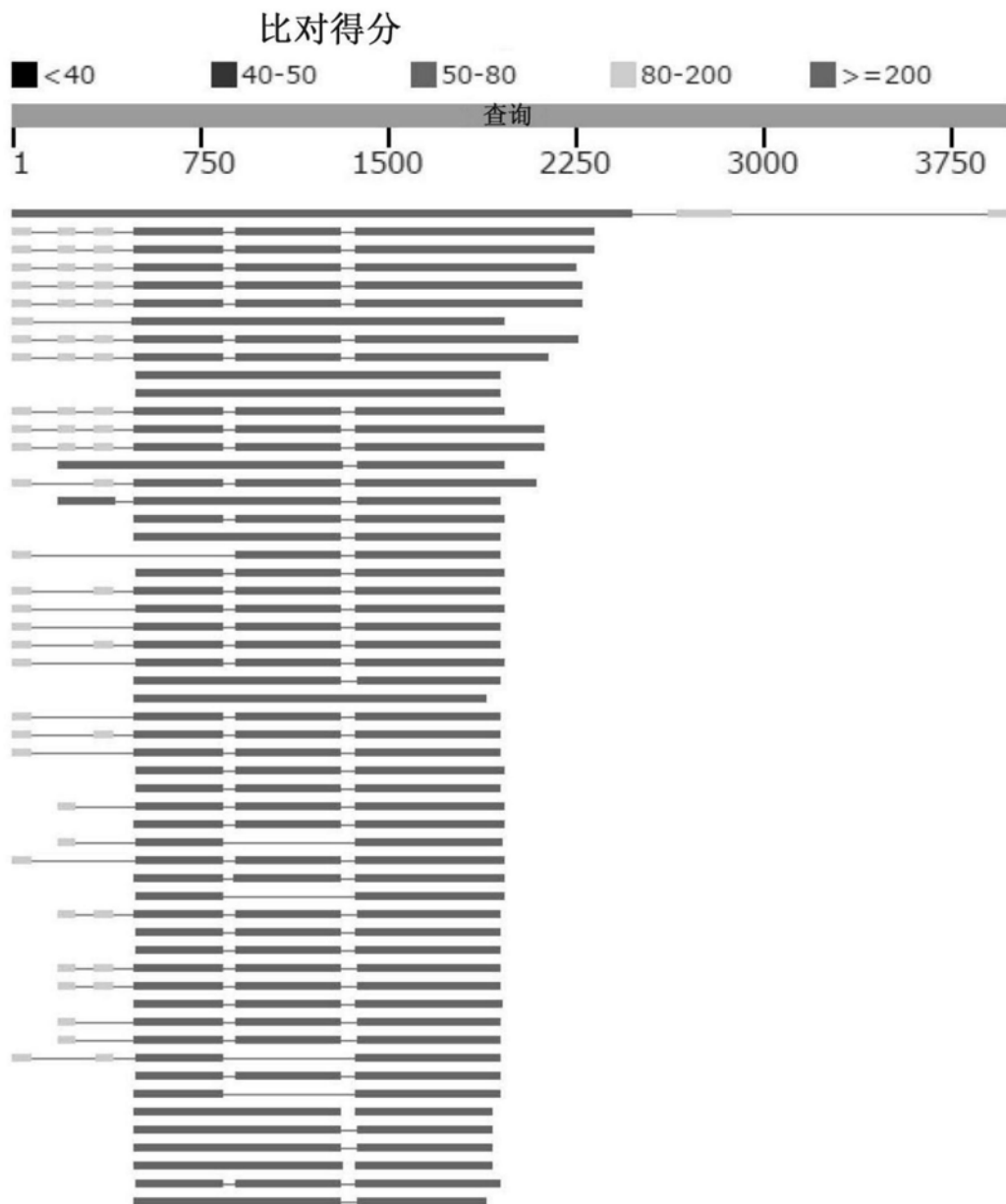


图8

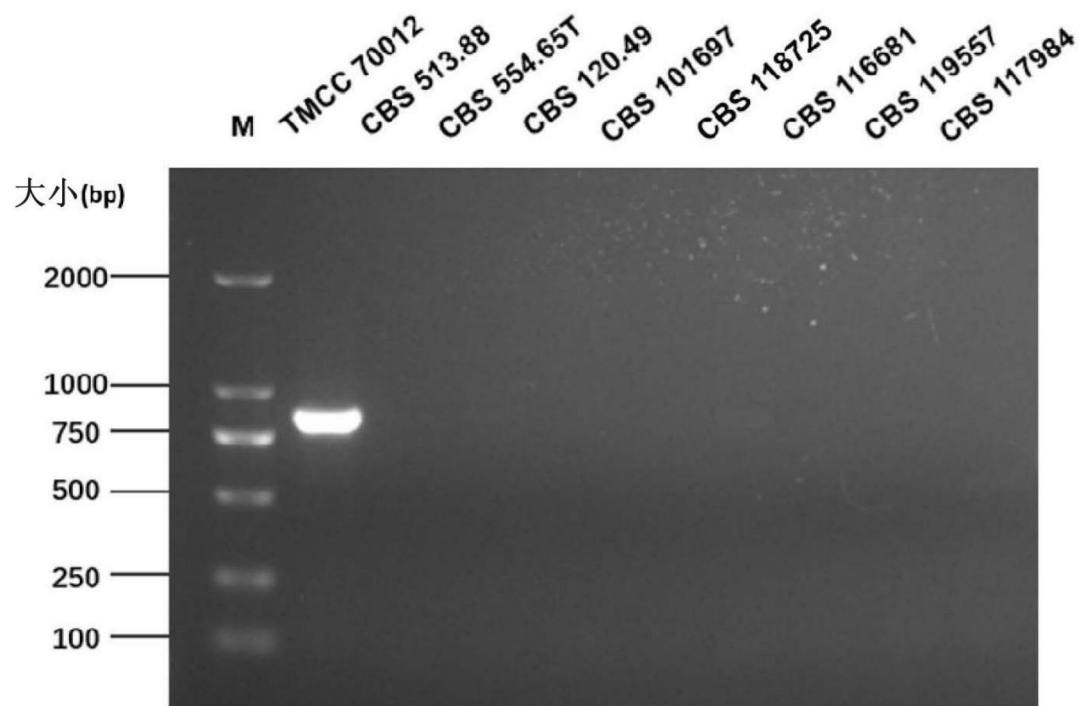


图9