(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 111808934 B
(45)授权公告日 2021.01.26

(21)申请号 202010705819.3

审查员 董珂

(22)申请日 2020.07.21

(65)同一申请的已公布的文献号
申请公布号 CN 111808934 A

(43)申请公布日 2020.10.23

(73)专利权人 广州吉赛生物科技股份有限公司
地址 510000 广东省广州市高新技术产业开发区科学城掬泉路3号广州国际企业孵化器D区D706房

(72)发明人 刘明 李自强 黄宁宁

(74)专利代理机构 广州科沃园专利代理有限公司 44416
代理人 张帅

(51)Int.Cl.
C12Q 1/6869(2018.01)

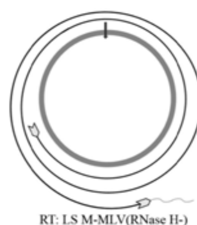
权利要求书2页 说明书14页
序列表16页 附图8页

(54)发明名称

一种环形RNA全长鉴定方法及其试剂盒

(57)摘要

本发明涉及分子生物学领域，涉及一种环形RNA全长鉴定方法及其试剂盒，包括以下步骤：S1.Splice junction序列分析和环形RNA全长序列的对比预测；S2.设计CD Primers和Sjod Primers；S3.LS逆转录反应：取0.1～1μg抽提的总RNA加入N6 Random Primer，RNase-Free H_2O混合得反应液I；然后向反应液I中加入5x LS RT Buffer，将LS Enzyme Mix加入反应液I中，得到cDNA；S4.PCR扩增和PCR产物分析；S5.Sanger测序和交叉分析对比序列。



CN 111808934 B

1.一种环形RNA全长鉴定方法,其特征在于,包括以下步骤:

S1.Splice junction序列分析和环形RNA全长序列的对比预测:首先在UCSC genome blast中对比,确认Splice junction序列对应的AG-GT剪切位点,然后在UCSC genome browser中对比,获得对应的基因组DNA序列,并在Ensembl中对比,获得预测的转录环形RNA全长序列;

S2.设计CD Primers和Sjod Primers:使用软件设计引物在已知序列中搜索得到40bp的引物可选区,从数据库中获得环形RNA的序列和Splice junction序列,选用21~40bp作为正向引物,将1~20bp的序列反向互补后作为反向引物得CD Primers;将Splice junction位置的上游20bp和下游10bp共30bp序列输入引物设计软件中,搜索计算以使引物符合GC含量、Tm值常规的参数要求,从而得到Sjod的上游引物,所述Sjod的上游引物包含上述Splice junction位置的不少于上游13bp和不多于下游7bp的序列,接着根据上游引物的参数和位置使用软件自动设计得到一条Sjod下游引物,获得Sjod Primers;

S3.LS逆转录反应:抽提样品中的总RNA,取0.1～1μg抽提的总RNA加入N6 Random Primer,RNase-Free H_2O混合,在温度为60～70℃反应5～10min,迅速冰浴2～3min,得反应液I;然后向反应液I中加入5x LS RT Buffer,将LS Enzyme Mix加入反应液I中,在温度为20～30℃条件下反应5～10min,升温至50～60℃条件下反应60～65min,再升温至80～90℃条件下反应5～10min得到cDNA;所述LS Enzyme Mix包括逆转录酶LS M-MLV和RNase Inhibitor,所述逆转录酶LS M-MLV基因序列如SEQ ID NO.1;所述逆转录酶LS M-MLV删除了RNase H结构域;所述逆转录酶LS M-MLV突变体为△N23/D108R/T306L/V433K/△aa524-551;

S4.PCR扩增和PCR产物分析:使用步骤S2得到的CD Primers、Sjod Primers和步骤S3得到的cDNA进行扩增,得PCR产物,将PCR产物进行琼脂糖凝胶电泳分析,得直接扩增条带和滚环产物条带;

S5.Sanger测序和交叉分析对比序列:将步骤S4取得的直接扩增条带回收后直接进行Sanger测序,滚环产物条带回收后连接至pMD19T载体中进行Sanger测序;交叉对比CD Primers或Sjod Primers的直接扩增条带序列与滚环产物条带序列,所述直接扩增条带序列确认和验证Splice junction位置的序列,所述滚环产物条带序列中两个Splice junction位点之间为环形RNA的真实全长序列。

2.根据权利要求1所述的环形RNA全长鉴定方法,其特征在于,所述步骤S3中逆转录中反应液I的组成包括:

实验材料	体积
总RNA	0.1~1μg
N6 Random Primer	0.5μL
RNase-Free H_2O	Up to 14 μL

所述步骤S3中LS逆转录中cDNA的组成包括:

实验材料	体积
反应液I	14μL
LS Enzyme Mix	2μL
5× LS RT Buffer	4μL。

3.根据权利要求1所述的环形RNA全长鉴定方法，其特征在于，所述步骤S3中取0.1～1μg抽提的总RNA加入N6 Random Primer，RNase-Free H_2O混合，在温度为65℃反应5min，迅速冰浴2min，得反应液I。

4.根据权利要求1所述的环形RNA全长鉴定方法，其特征在于，所述步骤S3中将LS Enzyme Mix加入反应液I中，在温度为25℃条件下反应5min，升温至55℃条件下反应60min，再升温至85℃条件下反应5min得到cDNA。

一种环形RNA全长鉴定方法及其试剂盒

技术领域

[0001] 本发明涉及分子生物学领域,尤其是涉及一种环形RNA全长鉴定方法及其试剂盒。

背景技术

[0002] 环形RNA(circular RNAs,环形RNAs)是一类具有闭合环状结构的RNA分子,在多物种中广泛存在,比线性RNA更稳定。

[0003] 目前研究多使用circBase等数据库及自有样品的高通量测序获得环形RNA信息,并筛选得到待研究的环形RNA后进行PCR检测、cDNA/gDNA分别扩增验证和RNase R消化等实验验证和确认其真实存在。但这些验证都是基于环形RNA的环化位点接头(Splice junction)序列进行的,缺乏对环化位点内部真实全长序列的鉴定验证,因此极易导致检测的真实准确性失真,特别是对于序列内部有多个外显子且存在可变剪切的情况。例如(Yibing Yang et al..Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis.J Natl Cancer Inst.2018 Mar 1;110(3))中通过PCR扩增和Sanger测序确定FBXW7的exon3和exon4经反向剪切形成620nt的环形RNA,而不是circBase数据库里收录hsa_circ_0001451的1227nt,可见不对环形RNA真实全长序列进行鉴定时极易导致基因识别错误、检测和验证失真。

[0004] 使用circBase等数据库及自有样品的高通量测序获得环形RNA信息中参考序列的注释也局限于技术本身和分析方法。1)高通量测序的reads长度在150nt左右,不能测到环形RNA的真实全长序列。2)环形RNA Junction reads的筛选鉴定和注释是mapping到参考基因组与GenBank数据集中分析的,仅通过已知线性基因的对比分析确定环形RNA全长的参考序列,且以对比匹配最高的转录本作为环形RNA的全长序列,无法区别鉴定外显子的可变剪切和线性基因与环形RNA可能不同的外显子组合。3)来自两条或更多pre-mRNA的剪切会因反式剪切(trans-splicing)形成tsRNA(trans-spliced RNA),这类tsRNA可能和环形RNA有相同的Splice Junction序列,导致环形RNA的筛选识别出现假阳性。

[0005] 因此急需开发一种环形RNA全长鉴定方法,对circBase已收录或新发现的环形RNA都进行全长鉴定以确定其真实的全长序列。

发明内容

[0006] 针对现有技术存在的不足,本发明的目的第一在于提供一种环形RNA全长鉴定方法,解决了环形RNA鉴定验证和研究中容易出现的识别错误、检测和验证失真等问题。

[0007] 为实现上述目的,本发明提供了如下技术方案:

[0008] 一种环形RNA全长鉴定方法,包括以下步骤:

[0009] S1.Splice junction序列分析和环形RNA全长序列的对比预测:首先在UCSC genome blast中对比,确认Splice junction序列对应的AG-GT剪切位点,然后在UCSC genome browser中对比,获得对应的基因组DNA序列,并在Ensembl中对比,获得预测的转录环形RNA全长序列;

[0010] S2.设计CD Primers和Sjod Primers:使用软件设计引物在已知序列中搜索得到40bp的引物可选区,从数据库中获得环形RNA的序列和Splice junction序列,选用21～40bp作为正向引物,将1～20bp的序列反向互补后作为反向引物得CD Primers;将Splice junction位置的上游20bp和下游10bp共30bp序列输入引物设计软件中,搜索计算以使引物符合GC含量、Tm值常规的参数要求,并使包含不少于上游13bp和不多于下游7bp的序列,从而得到Sjod的上游引物,接着根据上游引物的参数和位置使用软件自动设计得到一条Sjod下游引物,获得Sjod Primers;

[0011] S3.LS逆转录反应:抽提各类组织、细胞、血液、体液、外泌体、口腔拭子、肺泡灌洗液样品的总RNA,取0.1～1μg抽提的总RNA加入N6 Random Primer,RNase-Free H_2O混合,在温度为60～70℃反应5～10min,迅速冰浴2～3min,得反应液I;然后向反应液I中加入5x LS RT Buffer,将LS Enzyme Mix加入反应液I中,在温度为20～30℃条件下反应5～10min,升温至50～60℃条件下反应60～65min,再升温至80～90℃条件下反应5～10min得到cDNA;

[0012] S4.PCR扩增和PCR产物分析:使用步骤S2得到的CD Primers、Sjod Primers和步骤S3得到的cDNA进行扩增,得PCR产物,将PCR产物进行琼脂糖凝胶电泳分析,得直接扩增条带和滚环产物条带;

[0013] S5.Sanger测序和交叉分析对比序列。

[0014] 优选地,所述LS Enzyme Mix包括逆转录酶LS M-MLV(RNase H-)和RNase Inhibitor,所述逆转录酶LS M-MLV(RNase H-)突变体为△N23/D108R/T306L/V433K/△aa524-551,所述逆转录酶LS M-MLV(RNase H-)基因序列如SEQ ID NO.1。

[0015] 优选地,所述步骤S3中逆转录中反应液I的组成包括:

[0016]

实验材料	体积
总RNA	Up to 13.5 μL (0.1~1μg)
N6 Random Primer	0.5μL
RNase-Free H_2O	Up to 14μL

[0017] 所述步骤S3中LS逆转录中cDNA的组成包括:

[0018]

实验材料	体积
反应液I	14μL
LS Enzyme Mix	2μL
5× LS RT Buffer	4μL

[0019] 优选地,所述步骤S3中取0.1～1μg抽提的总RNA加入N6 Random Primer,RNase-Free H_2O混合,在温度为65℃反应5min,迅速冰浴2min,得反应液I。

[0020] 优选地,所述步骤S3中将LS Enzyme Mix加入反应液I中,在温度为25℃条件下反应5min,升温至55℃条件下反应60min,再升温至85℃条件下反应5min得到cDNA。

[0021] 优选地,所述PCR产物采用琼脂糖凝胶浓度为1.5～2%进行电泳分析。

[0022] 优选地,所述步骤S5中,将步骤S4取得的直接扩增条带回收后直接进行Sanger测序,滚环产物条带回收后连接至pMD19T载体中进行Sanger测序。

[0023] 优选地,所述步骤S5中,交叉对比CD Primers或Sjod Primers的直接扩增条带序

列与滚环产物条带序列,所述直接扩增条带序列确认和验证Splice junction位置的序列,所述滚环产物条带序列中两个Splice junction位点之间为环形RNA的真实全长序列。

[0024] 本发明的目的第二在于提供一种试剂盒,包括逆转录试剂和PCR扩增试剂,所述逆转录试剂包含LS Enzyme Mix、5×LS RT Buffer、N6 Random Primer和RNase-Free H_2O;所述PCR扩增试剂包含2×HS Taq Mix和RNase-Free H_2O。

[0025] 优选地,所述LS Enzyme Mix包含200U/μL LS M-MLV(RNase H-)、40U/μL RNase Inhibitor、RNase-Free H_2O和50%甘油;所述5×LS RT Buffer包含150mM pH8.3为Tris-HCl、375mM KCl、15mM $MgCl_2$、10mM dNTP和20mM DTT;所述2×HS Taq Mix包含2.5U/μL HS Taq Polymerase、10mM dNTP、125mM pH8.3为Tris-HCl、15mM $MgCl_2$、250mM KCl、500mM Glycine betaine、0.15%Tween-20、5%DMSO和10%甘油。

[0026] 本发明利用逆转录酶M-MLV的天然模板置换活性,通过点突变设计实现高度热稳定性和超强结合力,并删除RNase H结构域,避免逆转录过程中模板链的消化。改造优化的逆转录酶LS M-MLV(RNase H-)可以在55～60℃进行LS逆转录处理,能更好地打开环形RNA的二级结构,获得更高的得率和产量。天然模板置换活性和超强结合力的特性配合优化的反应体系与条件,并降低随机引物浓度,使对环形RNA模板逆转录时的链结合效率和逆转录效率大大增加,得到大量滚环逆转录的产物,即一条cDNA链中有多个Splice junction相连接,为PCR扩增得到环形RNA真实全长序列提供可能。

[0027] 本发明综合使用CD Primers和Sjod Primers两类反向引物,使用CD Primers扩增可能的环形RNA序列组合,同时参考Splice junction设计Sjod Primers以扩增特异的待鉴定环形RNA,两类引物的结果通过交叉对比和分析,能高效快速确定环形RNA的真实全长序列。其中CD Primers的PCR产物中包含一个Splice junction,Sjod Primers的直接扩增条带序列中包含一个Splice junction,Sjod Primers的滚环产物条带序列中包含两个Splice junction。CD Primers的PCR产物或Sjod Primers的直接扩增条带可以验证Splice junction位置的序列,Sjod Primers的滚环产物条带序列中两个Splice junction位点之间即为环形RNA的真实全长序列。

[0028] 由图2知,①为CD Primers的PCR产物示意线,②为Sjod Primers的直接扩增条带示意线,③为Sjod Primers的滚环产物条带示意线。

[0029] 与现有技术相比,本发明具有以下有益效果:

[0030] 本发明的环形RNA全长鉴定方法能对组织、细胞、血液、体液、外泌体、口腔拭子、肺泡灌洗液等各类样品抽提的总RNA直接进行鉴定,无需额外的RNase R消化等处理,简便易行,实验设计和分析方法标准规范,配套的全长鉴定试剂盒体系稳定,适用范围广,能使研究者更简单高效地获得环形RNA的真实全长序列。本发明的环形RNA全长鉴定方法能高效特异地区别鉴定环形RNA外显子的可变剪切以及线性基因与环形RNA不同的外显子组合,并能精确到单碱基的差异,最准确真实地获得环形RNA全长序列,保证环形RNA的识别、检测和验证精准可靠。

附图说明

[0031] 图1为本发明一种环形RNA全长鉴定方法的LS逆转录流程示意图;

[0032] 图2为本发明CD Primers或Sjod Primers和PCR产物的示意图;

[0033] 图3为本发明环形RNAhsa_circ_0007928的Splice junction序列对应的AG-GT剪切位点；

[0034] 图4为本发明环形RNAhsa_circ_0007928的PCR产物电泳检测图；

[0035] 图5为226bp条带Sanger测序峰图；

[0036] 图6为KFR1条带Sanger测序峰图；

[0037] 图7为F2KR条带Sanger测序峰图；

[0038] 图8为本发明环形RNAhsa_circ_0007928的测序结果交叉对比示意图；

[0039] 图9为本发明环形RNArno_novel5139的Splice junction序列对应的AG-GT剪切位点；

[0040] 图10为本发明环形RNArno_novel5139在Ensembl中对比结果示意图；

[0041] 图11为本发明环形RNArno_novel5139的PCR产物电泳检测图；

[0042] 图12为F2R2条带Sanger测序峰图；

[0043] 图13为F1R1条带Sanger测序峰图；

[0044] 图14为本发明环形RNArno_novel5139的测序结果交叉对比示意图；

[0045] 图15为本发明环形RNAmm9_circ_007585的Splice junction序列对应的AG-GT剪切位点；

[0046] 图16为本发明环形RNA mm9_circ_007585在Ensembl中对比结果示意图；

[0047] 图17为本发明环形RNAmm9_circ_007585的PCR产物电泳检测图；

[0048] 图18为本发明环形RNA F1R1条带Sanger测序峰图；

[0049] 图19为F2KR条带Sanger测序峰图；

[0050] 图20为本发明环形RNAmm9_circ_007585的测序结果交叉对比示意图；

具体实施方式

[0051] 以下结合附图及实施例,对本发明作进一步详细说明。

[0052] 如本文所用,除非特别注明,其他实验方法都可采用本领域中的常规方法,例如参考《分子克隆实验指南(第四版)》(科学出版社,2017)。

[0053] 所述LS Enzyme Mix包括逆转录酶LS M-MLV(RNase H-)和RNase Inhibitor,所述逆转录酶LS M-MLV(RNase H-)突变体为△N23/D108R/T306L/V433K/△aa524-551,所述逆转录酶LS M-MLV(RNase H-)基因序列如SEQ ID NO.1;

[0054] 其中N6 Random Primer为合成的寡核苷酸5'-N_6-3',浓度50μM。

[0055] 实施例1

[0056] 在本实施例中,环形RNA参考circBase数据库中hsa_circ_0007928序列和Splice junction序列。

[0057] 从circBase数据库中下载hsa_circ_0007928的参考序列389nt如SEQ ID NO.2;

[0058] 从circBase数据库中获得hsa_circ_0007928的Splice junction序列100nt如下：5'-CTGAAGGCATGGAGAAATTTTGTGAAGACATTGGTGTTGAACCAGAAAAC-3'-5'-ATTTTCAGCTGAACTCTCATCTCTCAACACTGGCAAATATTCATAAGATC-3'。

[0059] 一种环形RNA全长鉴定方法,包括以下步骤：

[0060] S1.Splice junction序列分析和环形RNA全长序列的对比预测：首先在UCSC

genome blast中对比，确认Splice junction序列对应的AG-GT剪切位点(参考图3)，然后在UCSC genome browser中对比，获得对应的基因组DNA序列，并在Ensembl中对比，获得预测的转录环形RNA全长序列；

[0061] S2.设计CD Primers和Sjod Primers：利用Clone Mgr Suite软件对参考的389nt序列分别设计CD Primers和Sjod Primers，其中一些引物可以共用，将序列发送至生工生物进行合成，PAGE纯化。

[0062] CD Primers：

[0063] C7928-TZ-F1:TGCGGTCTTGCAGTTCTTCAGA

[0064] C7928-TZ-R1:GACTTCCTGTTTGGTGATCGTC

[0065] 扩增大小389bp，或389+n*389bp

[0066] C7928-TZ-F2:CCTGCCTCTGGAGATGATTTA

[0067] C7928-TZ-R2:GTCTTCTCTTTTTCCTTGGTGGC

[0068] 扩增大小390bp，或390+n*389bp

[0069] Sjod Primers：

[0070] C7928-TZ-KF:GTTGAACCAGAAAACATTTTC

[0071] C7928-TZ-R1:GACTTCCTGTTTGGTGATCGTC

[0072] 扩增大小132bp，滚环产物扩增大小521bp

[0073] C7928-TZ-F2:CCTGCCTCTGGAGATGATTTA

[0074] C7928-TZ-KR:GAGTTCAGCTGAAAATGTTTTC

[0075] 扩增大小226bp，滚环产物扩增大小615bp

[0076] S3.RNA抽提和LS逆转录反应：使用Trizol Reagent抽提人食管癌细胞系ECA109的总RNA，然后进行LS逆转录处理。逆转录反应体系和程序如下：

[0077] ①按表1配置反应液，在温度为65℃反应5min，迅速冰浴2min。

[0078] 表1 LS逆转录步骤①

[0079]

试剂	体积
ECA109 RNA	2.5 μL (1 μg)
N6 Random Primer	0.5μL
RNase-Free H_2O	11μL

[0080] ②按表2配置反应液，反应条件25℃反应5min，升温至55℃反应60min，再升温至85℃反应5min。

[0081] 表2 LS逆转录步骤②

[0082]

试剂	体积
① 的反应液 I	14 μL
LS Enzyme Mix	2 μL
5× LS RT Buffer	4 μL

[0083] S4.PCR扩增和PCR产物分析:以ECA109cDNA为模板,使用四对引物分别进行PCR扩增,反应体系如表3所示。反应程序温度为95℃反应5min,40循环(95℃30s,58℃30s,72℃80s),温度为72℃反应5min,得PCR产物。

[0084] 表3 PCR扩增反应体系

[0085]

试剂	体积
2× HS Taq Mix	25 μL
ECA109 cDNA	1 μL
Primer-F (10 μM)	2 μL
Primer-R (10 μM)	2 μL
RNase-Free H_2O	20 μL

[0086] 根据预测环形RNA长度和预期扩增的PCR产物大小,选用合适的琼脂糖凝胶浓度,预测环形RNA长度低于750nt时,使用2%的浓度,预测cirRNA长度大于750nt时,使用1.5%的浓度,对PCR产物进行琼脂糖凝胶电泳分析,根据条带大小分析辨别预期的直接扩增条带和滚环产物条带,结果见图4。

[0087] 结果显示F1/R1和F2/R2的扩增产物条带有较多杂带,KF/R1和F2/KR的产物条带大小符合预期,其中132bp和226bp为直接扩增条带,KFR1和F2KR为滚环产物条带。分别标记为KFR1、F2KR和226bp并回收。

[0088] S5.Sanger测序和交叉分析对比序列。

[0089] 回收直接扩增条带226bp并直接进行Sanger测序,结果验证hsa_circ_0007928的Splice junction序列与预期一致,测序峰图如5。

[0090] 回收滚环产物条带KFR1,连接至pMD19T载体中进行Sanger测序,得到测序序列如下,测序峰图如6,显示521bp条带内有C7928-TZ-KF和C7928-TZ-R1引物序列,并有两个环化剪切位点,两个剪切位点之间的序列即为hsa_circ_0007928的真实全长序列。下划线标记为目的基因的全长序列。

[0091] 如:

[0092] >GZ19011701145(KFR1)M13-48_J_G11

[0093] TTGGCACGCCTGCCGTTCGACGATT**GTTGAACCAGAAAACATTTTC**AGCTGAACTCT
CATCTCTCAACACTGGCAAATATTCATAAGATCTACCACACCCTTAATAAGCTGAACCTA
ACAGAAGACATTGGCCAAGACGATCACCAAACAGGAAGTCTGCGGTCTTGCAGTTCT
TCAGACTGCTTTAATAAAGTGATGCCACCAAGGAAAAAGAGAAGACCTGCCTCTGGA
GATGATTTATCTGCCAAGAAAAGTAGACATGATAGCATGTATAGAAAATATGATTCGACT
AGAATAAAGACTGAAGAAGAAGCCTTTTCAAGTAAAAGGTGCTTGGAATGGTTCTATG

[0094] AATATGCAGGAACTGATGATGTTGTAGGCCCTGAAGGCATGGAGAAATTTTGTGAAGA
CATTGGTGTTGAACCAGAAAACCTCTCATCTCTCAACACTGGCAAATATTCATAAGATC
TACCACACCCTTAATAAGCTGAACCTAACAGAAGACATTGGCCAA**GACGATCACCAA**
ACAGGAAGTCAATCTCTGGAAGATCCGCGCGTACCGAGTTCTAATTCACTGGCCGTCG
TTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGC
ACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCC
CAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCA
TCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGC
ATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGT
CTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTC
AGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGC

[0095] 回收滚环产物条带F2KR，连接至pMD19T载体中进行Sanger测序，得到测序序列如下，测序峰图如7，显示615bp条带内有C7928-TZ-F2和C7928-TZ-KR引物序列，并有两个环化剪切位点，两个剪切位点之间的序列即为hsa_circ_0007928的真实全长序列。加粗标记为引物序列，下划线标记为目的基因的全长序列。

[0096] 如：>GZ19011701148(F2KR)M13-48_J_B12

[0097] AGGGCAGGAACATGATTACGCCAGTTTGCACGCCTGCCGTTCGACGATT**CCTGCCTCT**
GGAGATGATTTATCTGCCAAGAAAAGTAGACATGATAGCATGTATAGAAAATATGATTC
GACTAGAATAAAGACTGAAGAAGAAGCCTTTTCAAGTAAAAGGTGCTTGGAATGGTTC
TATGAATATGCAGGAACTGATGATGTTGTAGGCCCTGAAGGCATGGAGAAATTTTGTGA
AGACATTGGTGTTGAACCAGAAAACATTTTCAGCTGAACTCTCATCTCTCAACACTGG
CAAATATTCATAAGATCTACCACACCCTTAATAAGCTGAACCTAACAGAAGACATTGGC
CAAGACGATCACCAAACAGGAAGTCTGCGGTCTTGCAGTTCTTCAGACTGCTTTAATA
AAGTGATGCCACCAAGGAAAAAGAGAAGACCTGCCTCTGGAGATGATTTATCTGCCAA
GAAAAGTAGACATGATAGCATGTATAGAAAATATGATTCGACTAGAATAAAGACTGAAG
AAGAAGCCTTTTCAAGTAAAAGGTGCTTGGAATGGTTCTATGAATATGCAGGAACTGAT
GATGTTGTAGGCCCTGAAGGCATGGAGAAATTTTGTGAAGACATTGGTGTTGAACCA**G**
AAAACATTTTCAGCTGAACTCAATCTCTGGAAGATCCGCGCGTACCGAGTTCTAATTC
ACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAAT
CGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCG
ATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTT
CTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGC
TCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCT
GACGGGC

[0098] 参见图8，交叉对比226bp条带Sanger测序得到的Splice junction序列和KFR1与F2KR条带Sanger测序得到的全长序列，分析得到hsa_circ_0007928的真实全长序列389nt如SEQ ID NO.3。

[0099] 实施例2

[0100] 高通量测序得到rno_novel5139的Splice juntion序列；

[0101] S1.Splice junction序列分析和环形RNA全长序列的对比预测：

[0102] 已知rno_novel5139的Splice junction序列如下。

[0103] >rno_novel5139_junction_seq

[0104] ACTGGTGGCCATCCTGTTGTGAAGATCTTGAACCAGATGATTGTGCAT

[0105] 在UCSC genome blast中对比，确认rno_novel5139的Splice junction序列对应的AG-GT剪切位点如图9。

[0106] 在UCSC genome browser中对比，获得对应的基因组DNA序列如SEQ ID NO.4。

[0107] 参考图10，在Ensembl中对比，显示rno_novel5139与Nfatc3基因序列重叠，预测rno_novel5139包含两个外显子，其参考全长序列1301nt如SEQ ID NO.5。

[0108] S2.设计CD Primers和Sjod Primers：利用Clone Mgr Suite软件对参考的1301nt序列分别设计CD Primers和Sjod Primers，其中一些引物可以共用，将序列发送至生工生物进行合成，PAGE纯化。

[0109] CD Primers：

[0110] rno5139-YZ-F2:AACAGTATGGACCTGGACATT

[0111] rno5139-YZ-R1:TGTGGTAAGCCAAGTGATGAA

[0112] 扩增大小810bp，或810+n*1301bp

[0113] rno5139-YZ-F2:AACAGTATGGACCTGGACATT

[0114] rno5139-YZ-R2:ATACTGTTGATGCCAGGACTC

[0115] 扩增大小1309bp，或1309+n*1301bp

[0116] Sjod Primers：

[0117] rno5139-YZ-F1:CATCCTGTTGTGAAGATCTTG

[0118] rno5139-YZ-R1:TGTGGTAAGCCAAGTGATGAA

[0119] 扩增大小103bp，滚环产物扩增大小1404bp

[0120] S3.RNA抽提和LS逆转录反应：使用Trizol Reagent抽提大鼠肝脏组织的总RNA，然后进行LS逆转录处理。逆转录反应体系和程序如下：

[0121] 按表4配置反应液，反应条件65℃5min，迅速冰浴2min。

[0122] 表4 LS逆转录步骤①

[0123]

试剂	体积
rno RNA	3.1 μL (1 μg)
N6 Random Primer	0.5μL
RNase-Free H_2O	10.4μL

[0124] 按表5配置反应液，反应条件25℃5min，55℃60min，85℃5min。

[0125] 表5LS逆转录步骤②

[0126]

试剂	体积
② 的反应液	14μL
LS Enzyme Mix	2μL
5× LS RT Buffer	4μL

[0127] S4.PCR扩增和PCR产物分析：

[0128] 以rno cDNA为模板，使用三对引物分别进行PCR扩增，反应体系如表6所示。反应程序95℃5min，40循环(95℃30s，58℃30s，72℃180s)，72℃5min。

[0129] 表6PCR扩增反应体系

[0130]

试剂	体积
2× HS Taq Mix	25 μL
rno cDNA	1 μL
Primer-F (10 μM)	2 μL
Primer-R (10 μM)	2 μL
RNase-Free H_2O	20 μL

[0131] 根据预测环形RNA长度和预期扩增的PCR产物大小，选用合适的琼脂糖凝胶浓度，预测环形RNA长度低于750nt时，使用2%的浓度，预测cirRNA长度大于750nt时，使用1.5%的浓度，对PCR产物进行琼脂糖凝胶电泳分析，根据条带大小分析辨别预期的直接扩增条带和滚环产物条带。

[0132] 电泳检测结果如图11，结果显示F2/R1扩增产物条带大小错误，可能是非特异扩增。F2/R2和F1/R1的产物条带大小符合预期，其中103bp为直接扩增条带，F2R2和F1R1为滚环产物条带。分别标记为F2R2和F1R1并回收。

[0133] 如图12，回收滚环产物条带F2R2并直接进行Sanger测序，结果验证rno_novel5139的Splice junction序列与预期一致。回收滚环产物条带F1R1，连接至pMD19T载体中进行Sanger测序，得到测序序列如下，测序峰图如图13，显示1404bp条带内有rno5139-YZ-F1和rno5139-YZ-R1引物序列，并有两个环化剪切位点，两个剪切位点之间的序列即为rno_novel5139的真实全长序列。加粗标记为引物序列，下划线标记为目的基因的全长序列。

[0134] >GZ18061302232(5139-F1R1)M13+_-

[0135] TGGAAAATTGGACATGGATTACGCCAGTTTGCACGCCTGCCGTTCGACGATT**CATCCT**
GTTGTGAAGATCTTGAACCAGATGATTGTGCATCCATTTACATCTTTAATGTAGACCCG
CCTCCATCTACTTTAAATTCATCACTTGGCTTACCACATCATGGACTGCTGCAGTCTCAC
TCTTCTGTTTTGTCACCATCATTTCAGCTCCAAGGTTTCAAAAATTATGAAGGAACTGA
TGATATTTCTGAATCTAAATATAGCTCATTAAGTGGTCCTAAACCCTTTGAATGCCCAAG
TATTCAAATTACATCCATCTCTCCTAACTGTCATCAAGAAACAGATGCTCATGAAGATGA
CCTACATGTAAATGACCCAGAAAGGGAATATTTGGAAAGGCCTTCTAGAGATCATCTCT
ATCTCCCACTTGAACCGTCCTACCGGGAATCTTCCCTTAGTCCTAGTCCTGCCAGCAGC
GTTTCTTCTAGGAGTTGGTTCTCAGATGCATCTTCTTGTGAGTCTCTCTCACACATTTAT
GACGATGTGGACTCAGAGTTGAATGAAGCTGCTGCACGATTTACTCTTGGCTCACCTCT
GACTTCTCCAGGTGGCTCTCCAGGAGGTTGCCCTGGAGAAGAGTCCTGGCATCAACA
GTATGGACCTGGACATTCCTTGTCACCTAGGCAATCTCCTTGCCACTCTCCTAGATCCA
GTATCACTGATGAGAATTGGCTGAGCCCCAGACCAGCCTCAGGACCCTCATCCAGGCC
CACTTCTCCTTGTGGTAAACGACGGCACTCCAGTGCTGAAGTATGTTATGCTGGCTCTC
TTTCACCCCATCACTCACCTGTTCCGTCCCCTGGTCACTCGCCTAGAGGGAGTGTAACT
GAAGATACCTGGCTCACTGCTCCTGTCCACACTGGATCAGGCCTCAGCCCTGCACCGT
TTCCATTTCAGTACTGTGTAGAGACTGACATCCCTTTGAAAACAAGGAAGACTTCTGA
CGATCAAGCTGCCATACTACCAGGAAAATTAGAGGTCTGTTCAGATGATCAAGGGAGC
TTATCCCCATCCCGGGAGACATCAGTAGATGATGGCCTTGGATCTCAGTATCCTTTAAAG
AAAGATTCATCTGGTGACCAATTTCTTTCAGTTCCTTCACCTTTTACCTGGAGCAAACC
AAAGCCTGGCCACACTCCTATATTTCGCACATCTTCATTACCTCCATTAGACTGGCCTTT
ACCAACTCACTTTGGACAATGTGAATTGAAAATAGAAGTGCAACCTAAAACTCACCAT
AGAGCCCATTATGAAACTGAAGGTAGCCGAGGGGCAGTGAAAGCCTCTACTGGTGGC
CATCCTGTTGTGAAGATCTTGAACCAGATGATTGTGCATCCATTTACATCTTTAATGTAG
ACCCGCCCCCATCTACTTTAAA**TTCATCACTTGGCTTACCACA**AATCTCTGGAAGATC
CGCGCGTACCGAGTTCTAATCACTGCCGTTT

[0136] S5.Sanger测序和交叉分析对比序列

[0137] 如图14,交叉对比F2R2条带Sanger测序得到的Splice junction序列和F1R1带Sanger测序得到的全长序列,分析得到rno_novel5139的真实全长序列1301nt如SEQ ID NO.6。

[0138] 实施例3

[0139] 从circBase数据库中获取mm9_circ_007585的参考序列和Splice junction序列;

[0140] S1.Splice junction序列分析和环形RNA全长序列的对比预测:从circBase数据库中下载mm9_circ_007585的参考序列449bp如下,如SEQ ID NO.7;

[0141] 从circBase数据库中获得mm9_circ_007585的Splice junction序列100bp如下:

[0142] 5’-TGGACAGCCTGGCTTGGCTACTCATTCTTCTGTTAACAGGGTCCTGGCAG-3’-5’-GAGGCTGCTGGAGTCTTCCCTCATTTCATTATCCCGTTATGATGGAGCAG-3’

[0143] 在UCSC genome blast中对比,确认mm9_circ_007585的Splice junction序列对应的AG-GT剪切位点如图15。

[0144] 在UCSC genome browser中对比,获得对应的基因组DNA序列如下,如SEQ ID NO.8。

[0145] 如图16,在Ensembl中对比,显示mm9_circ_007585与Ambra1基因序列重叠,预测

mm9_circ_007585最多包含四个外显子，按最长序列预测其参考全长序列449nt如下，如SEQ ID NO.9。

[0146] S2.设计CD Primers和Sjod Primers:利用Clone Mgr Suite软件对参考序列分别设计CD Primers和Sjod Primers，其中一些引物可以共用，将序列发送至生工生物进行合成，PAGE纯化。

[0147] CD Primers:

[0148] mc7585-YZ-F1:AATCGGCTTCGTTCTTCCACCTCC

[0149] mc7585-YZ-R1:ATTCTGTTGGTAGCGCATGGAGC

[0150] 扩增大小452bp，或452+n*449bp

[0151] Sjod Primers:

[0152] mc7585-YZ-F2:CTGCACCTTCACTTGGACGATT

[0153] mc7585-YZ-KR:AAGACTCCAGCAGCCTCCTGC

[0154] 扩增大小149bp，滚环产物扩增大小598bp。

[0155] S3.RNA抽提和LS逆转录反应:使用Trizol Reagent抽提小鼠NC1469细胞系的总RNA，然后进行LS逆转录处理。逆转录反应体系和程序如下:

[0156] 按表7配置反应液，反应条件65℃5min，迅速冰浴2min。

[0157] 表7 LS逆转录步骤①

[0158]

试剂	体积
NC1469 RNA	1.3 μL (1 μg)
N6 Random Primer	0.5μL
RNase-Free H_2O	12.2μL

[0159] 按表8配置反应液，反应条件25℃5min，55℃60min，85℃5min。

[0160] 表8 LS逆转录步骤②

[0161]

试剂	体积
③ 的反应液	14μL
LS Enzyme Mix	2μL
5× LS RT Buffer	4μL

[0162] S4.PCR扩增和PCR产物分析:

[0163] 以NC1469 cDNA为模板，使用两对对引物分别进行PCR扩增，反应体系如表9所示。反应程序95℃5min，40循环(95℃30s，58℃30s，72℃90s)，72℃5min。

[0164] 表9 PCR扩增反应体系

[0165]

试剂	体积
2× HS Taq Mix	25 μL
NC1469 cDNA	1 μL
Primer-F (10 μM)	2 μL
Primer-R (10 μM)	2 μL
RNase-Free H_2O	20 μL

[0166] 根据预测环形RNA长度和预期扩增的PCR产物大小，选用合适的琼脂糖凝胶浓度，预测环形RNA长度低于750nt时，使用2%的浓度，预测cirRNA长度大于750nt时，使用1.5%的浓度，对PCR产物进行琼脂糖凝胶电泳分析，根据条带大小分析辨别预期的直接扩增条带和滚环产物条带。

[0167] 电泳检测结果如图17，结果显示F1/R1和F2/KR的产物条带大小符合预期，其中149bp为直接扩增条带，F2KR为滚环产物条带。分别标记为F1R1和F2KR并回收。

[0168] S5.Sanger测序和交叉分析对比序列

[0169] 如图18，回收F1R1条带并直接进行Sanger测序，结果验证mm9_circ_007585的Splice junction序列与预期一致。

[0170] 回收滚环产物条带F2KR，连接至pMD19T载体中进行Sanger测序，得到测序序列如下，测序峰图如图19，显示598bp条带内有mc7585-YZ-F2和mc7585-YZ-KR引物序列，并有两个环化剪切位点，两个剪切位点之间的序列即为mm9_circ_007585的真实全长序列。加粗标记为引物序列，下划线标记为目的基因的全长序列。

[0171] >GZ17122700658(7585-F2KR)M13-47_J_C01

[0172]
GGGGCATGCGTAGTGATTAGACTCGGTACGCGCGGATCTTCCAGAGATT**CTGCAC**
CTTCACTTGGACGATTTGTTCCAAGGCGTTTTTTGCTGCCTGAGTACTTGCCTTATGC
TGGGATTTTTCATGAACGTGGACAGCCTGGCTTGGCTACTCATTCTTCTGTTAACAGGG
TCCTGGCAGGAGGCTGCTGGAGTCTTCCCTCATTTCATTATCCCGTTATGATGGAGCAG
GATCCAGAGAGCACCCAATTTACCCAGACCCAGCGAGATTATCTCCTGCTGCATACTAC
GCCCAGAGGATGATCCAGTATCTCTCAAGGAGAGACAGTATTCGCCAGCGCTCCATGC
GCTACCAACAGAATCGGCTTCGTTCTTCCACCTCCTCTTCTTCCTCAGACAACCAGGGT
CCATCAGTAGAGGGAACCGACTTGGAATTTGAGGACTTTGAGGACAATGGTGACAGAT
CTAGGCACCGAGCTCCCCGAAATGCCCGAATGTCTGCACCTTCACTTGGACGATTTGTT
CCAAGGCGTTTTTTGCTGCCTGAGTACTTGCCTTATGCTGGGATTTTTCATGAACGTGG
ACAGCCTGGCTTGGCTACTCATTCTTCTGTTAACAGGGTCCTG**GCAGGAGGCTGCTG**
GAGTCTTAATCGTCGAACGGCAGGCGTGCAAACTTGGCGTAATCATGGTCATAGCTGT
TTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATA
AAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTC
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAA
CGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACT
CGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAT
ACGGTTATCC

[0173] 如图20，交叉对比F1R1条带Sanger测序得到的Splice junction序列和F2KR条带Sanger测序得到的全长序列，分析得到mm9_circ_007585的真实全长序列449nt如SEQ ID NO.10。

[0174] 实施例4

[0175] 一种鉴定环形RNA全长序列的试剂盒，所述试剂盒包括逆转录试剂和PCR扩增试剂和使用说明，所述逆转录试剂包含LS Enzyme Mix、5×LS RT Buffer、N6 Random Primer和RNase-Free H_2O；所述PCR扩增试剂包含2×HS Taq Mix和RNase-Free H_2O。

[0176] 所述LS Enzyme Mix包含200U/μL LS M-MLV(RNase H-)、40U/μL RNase Inhibitor、RNase-Free H_2O和50％甘油；所述5×LS RT Buffer包含150mM pH8.3为Tris-HCl、375mM KCl、15mM $MgCl_2$、10mM dNTP和20mM DTT；所述2×HS Taq Mix包含2.5U/μL HS Taq Polymerase、10mM dNTP、125mM pH8.3为Tris-HCl、15mM $MgCl_2$、250mM KCl、500mM Glycine betaine、0.15％Tween-20、5％DMSO和10％甘油。

[0177] 使用说明给出了所述试剂盒的使用方法和鉴定环形RNA真实全长序列的标准化实验设计和分析方法，为全长鉴定中的逆转录、序列分析和全长序列的对比预测、引物设计、PCR扩增与电泳检测、sanger测序及序列分析，提供指导、指引或教导。

[0178] 本具体实施方式的实施例均为本发明的较佳实施例，并非依此限制本发明的保护

范围，故：凡依本发明的结构、形状、原理所做的等效变化，均应涵盖于本发明的保护范围之内。

[0001] 序列表

[0002] <110> 广州吉赛生物科技股份有限公司

[0003] <120> 一种环形RNA全长鉴定方法及其试剂盒

[0004] <130> 2020.7.17

[0005] <160> 10

[0006] <170> SIPOSequenceListing 1.0

[0007] <210> 1

[0008] <211> 623

[0009] <212> PRT

[0010] <213> SEQ ID NO.1

[0011] <400> 1

[0012] Met Thr Trp Leu Ser Asp Phe Pro Gln Ala Trp Ala Glu Thr Gly Gly

[0013] 1 5 10 15

[0014] Met Gly Leu Ala Val Arg Gln Ala Pro Leu Ile Ile Pro Leu Lys Ala

[0015] 20 25 30

[0016] Thr Ser Thr Pro Val Ser Ile Lys Gln Tyr Pro Met Ser Gln Glu Ala

[0017] 35 40 45

[0018] Arg Leu Gly Ile Lys Pro His Ile Gln Arg Leu Leu Asp Gln Gly Ile

[0019] 50 55 60

[0020] Leu Val Pro Cys Gln Ser Pro Trp Asn Thr Pro Leu Leu Pro Val Lys

[0021] 65 70 75 80

[0022] Lys Pro Gly Thr Asn Arg Tyr Arg Pro Val Gln Asp Leu Arg Glu Val

[0023] 85 90 95

[0024] Asn Lys Arg Val Glu Asp Ile His Pro Thr Val Pro Asn Pro Tyr Asn

[0025] 100 105 110

[0026] Leu Leu Ser Gly Leu Pro Pro Ser His Gln Trp Tyr Thr Val Leu Asp

[0027] 115 120 125

[0028] Leu Lys Asp Ala Phe Phe Cys Leu Arg Leu His Pro Thr Ser Gln Pro

[0029] 130 135 140

[0030] Leu Phe Ala Phe Glu Trp Arg Asp Pro Glu Met Gly Ile Ser Gly Gln

[0031] 145 150 155 160

[0032] Leu Thr Trp Thr Arg Leu Pro Gln Gly Phe Lys Asn Ser Pro Thr Leu

[0033] 165 170 175

[0034] Phe Asp Glu Ala Leu His Arg Asp Leu Ala Asp Phe Arg Ile Gln His

[0035] 180 185 190

[0036] Pro Asp Leu Ile Leu Leu Gln Tyr Val Asp Asp Leu Leu Leu Ala Ala

[0037] 195 200 205

[0038] Thr Ser Glu Leu Asp Cys Gln Gln Gly Thr Arg Ala Leu Leu Gln Thr

[0039] 210 215 220

[0040] Leu Gly Asn Leu Gly Tyr Arg Ala Ser Ala Lys Lys Ala Gln Ile Cys

[0041] 225 230 235 240

[0042] Gln Lys Gln Val Lys Tyr Leu Gly Tyr Leu Leu Lys Glu Gly Gln Arg
[0043] 245 250 255
[0044] Trp Leu Thr Glu Ala Arg Lys Glu Thr Val Met Gly Gln Pro Thr Pro
[0045] 260 265 270
[0046] Lys Thr Pro Arg Gln Leu Arg Glu Phe Leu Gly Leu Ala Gly Phe Cys
[0047] 275 280 285
[0048] Arg Leu Trp Ile Pro Gly Phe Ala Glu Met Ala Ala Pro Leu Tyr Pro
[0049] 290 295 300
[0050] Leu Thr Lys Thr Gly Thr Leu Phe Asn Trp Gly Pro Asp Gln Gln Lys
[0051] 305 310 315 320
[0052] Ala Tyr Gln Glu Ile Lys Gln Ala Leu Leu Thr Ala Pro Ala Leu Gly
[0053] 325 330 335
[0054] Leu Pro Asp Leu Thr Lys Pro Phe Glu Leu Phe Val Asp Glu Lys Gln
[0055] 340 345 350
[0056] Gly Tyr Ala Lys Gly Val Leu Thr Gln Lys Leu Gly Pro Trp Arg Arg
[0057] 355 360 365
[0058] Pro Val Ala Tyr Leu Ser Lys Lys Leu Asp Pro Val Ala Ala Gly Trp
[0059] 370 375 380
[0060] Pro Pro Cys Leu Arg Met Val Ala Ala Ile Ala Val Leu Thr Lys Asp
[0061] 385 390 395 400
[0062] Ala Gly Lys Leu Thr Met Gly Gln Pro Leu Lys Ile Leu Ala Pro His
[0063] 405 410 415
[0064] Ala Val Glu Ala Leu Val Lys Gln Pro Pro Asp Arg Trp Leu Ser Asn
[0065] 420 425 430
[0066] Ala Arg Met Thr His Tyr Gln Ala Leu Leu Leu Asp Thr Asp Arg Val
[0067] 435 440 445
[0068] Gln Phe Gly Pro Val Val Ala Leu Asn Pro Ala Thr Leu Leu Pro Leu
[0069] 450 455 460
[0070] Pro Glu Glu Gly Leu Gln His Asn Cys Leu Asp Ile Leu Ala Glu Ala
[0071] 465 470 475 480
[0072] His Gly Thr Arg Pro Asp Leu Thr Asp Gln Pro Leu Pro Asp Ala Asp
[0073] 485 490 495
[0074] His Thr Trp Tyr Thr Ala Leu Pro Ala Gly Thr Ser Ala Gln Arg Ala
[0075] 500 505 510
[0076] Glu Leu Ile Ala Leu Thr Gln Ala Leu Lys Met Ala Glu Gly Lys Lys
[0077] 515 520 525
[0078] Leu Asn Val Tyr Thr Asp Ser Arg Tyr Ala Phe Ala Thr Ala His Ile
[0079] 530 535 540
[0080] His Gly Glu Ile Tyr Arg Arg Arg Gly Leu Leu Thr Ser Glu Gly Lys
[0081] 545 550 555 560
[0082] Glu Ile Lys Asn Lys Asp Glu Ile Leu Ala Leu Leu Lys Ala Leu Phe
[0083] 565 570 575

[0084] Leu Pro Lys Arg Leu Ser Ile Ile His Cys Pro Gly His Gln Lys Gly
[0085] 580 585 590
[0086] His Ser Ala Glu Ala Arg Gly Asn Arg Met Ala Asp Gln Ala Ala Arg
[0087] 595 600 605
[0088] Lys Ala Ala Ile Thr Glu Thr Pro Asp Thr Ser Thr Leu Leu Ile
[0089] 610 615 620
[0090] <210> 2
[0091] <211> 389
[0092] <212> DNA
[0093] <213> SEQ ID NO.2
[0094] <400> 2
[0095] attttcagct gaactctcat ctctcaacac tggcaaatat tcataagatc taccacaccc 60
[0096] ttaataagct gaacctaaca gaagacattg gccaagacga tcaccaaaca ggaagtctgc 120
[0097] ggtcttgcag ttcttcagac tgctttaata aagtgatgcc accaaggaaa aagagaagac 180
[0098] ctgcctctgg agatgattta tctgccaaga aaagtagaca tgatagcatg tatagaaaat 240
[0099] atgattcgac tagaataaag actgaagaag aagccttttc aagtaaaagg tgcttggaat 300
[0100] ggttctatga atatgcagga actgatgatg ttgtaggccc tgaaggcatg gagaaatttt 360
[0101] gtgaagacat tggtgttgaa ccagaaaac 389
[0102] <210> 3
[0103] <211> 389
[0104] <212> DNA
[0105] <213> SEQ ID NO.3
[0106] <400> 3
[0107] attttcagct gaactctcat ctctcaacac tggcaaatat tcataagatc taccacaccc 60
[0108] ttaataagct gaacctaaca gaagacattg gccaagacga tcaccaaaca ggaagtctgc 120
[0109] ggtcttgcag ttcttcagac tgctttaata aagtgatgcc accaaggaaa aagagaagac 180
[0110] ctgcctctgg agatgattta tctgccaaga aaagtagaca tgatagcatg tatagaaaat 240
[0111] atgattcgac tagaataaag actgaagaag aagccttttc aagtaaaagg tgcttggaat 300
[0112] ggttctatga atatgcagga actgatgatg ttgtaggccc tgaaggcatg gagaaatttt 360
[0113] gtgaagacat tggtgttgaa ccagaaaac 389
[0114] <210> 4
[0115] <211> 4637
[0116] <212> DNA
[0117] <213> SEQ ID NO.4
[0118] <400> 4
[0119] atcttgaacc agatgattgt gcatccattt acatctttaa tgtagacccg cctccatcta 60
[0120] ctttaaattc atcacttggc ttaccacatc atggactgct gcagtctcac tcttctgttt 120
[0121] tgtcaccatc atttcagctc caaggtttca aaaattatga aggaactgat gatatttctg 180
[0122] aatctaaata tagctcatta agtggtccta aaccctttga atgcccaagt attcaaatta 240
[0123] catccatctc tcctaactgt catcaagaaa cagatgctca tgaagatgac ctacatgtaa 300
[0124] atgacccaga aagggaatat ttggaaaggc cttctagaga tcatctctat ctcccacttg 360
[0125] aaccgtccta ccgggaatct tcccttagtc ctagtcctgc cagcagcgtt tcttctagga 420

[0126] gttggttctc agatgcatct tcttgtgagt ctctctcaca catttatgac gatgtggact 480
[0127] cagagttgaa tgaagctgct gcacgattta ctcttggctc acctctgact tctccaggtg 540
[0128] gctctccagg aggttgccct ggagaagagt cctggcatca acagtatgga cctggacatt 600
[0129] ccttgtcacc taggcaatct ccttgccact ctcctagatc cagtatcact gatgagaatt 660
[0130] ggctgagccc cagaccagcc tcaggaccct catccaggcc cacttctcct tgtggtaaac 720
[0131] gacggcactc cagtgctgaa gtatgttatg ctggctctct ttcaccccat cactcacctg 780
[0132] ttccgtcccc tggtcactcg cctagaggga gtgtaactga agatacctgg ctcactgctc 840
[0133] ctgtccacac tggatcaggc ctcagccctg caccgtttcc atttcagtac tgtgtagaga 900
[0134] ctgacatccc tttgaaaaca aggaagactt ctgacgatca agctgccata ctaccaggaa 960
[0135] aattagaggt ctgttcagat gatcaaggga gcttatcccc atcccgggag acatcagtag 1020
[0136] atgatggcct tggatctcag tatcctttaa agaaagattc atctggtgac caatttcttt 1080
[0137] cagttccttc accttttacc tggagcaaac caaagcctgg ccacactcct atatttcggt 1140
[0138] gagttggtag aaatggctgc tgatcagttt ttcatgctta taggtcattt gcattgtata 1200
[0139] actacattct gacctttttt cttctccttg cccttctccc tttcatttct tttctctctc 1260
[0140] ccctctccct gccctcatct tccttgtttt tcattttttt aaaaaattaa ttcttctggt 1320
[0141] caaataactt agaggtcctt tttatttttc tttaagtctt taaaagtgtt tttaaattga 1380
[0142] ctaaagttac attttaaaag tatatagatg aagaaagtat gtgatggatt gtagagaaca 1440
[0143] gtaaatgaaa ttgaaggttc cttgcggaat tccttatcta gaaaagatta ctgatgggga 1500
[0144] gtggggaggg gaaatgggga gtggggaggg gaaattgggg agtggggagg ggaaatccca 1560
[0145] tatagaatga gaggaggagg ggttaggggg atgttggcct ggaaaccggg aaagggaata 1620
[0146] acaattgaaa tgtaaataag aaatacccaa tttaataaag atggagaaaa aactgtaaat 1680
[0147] gtcattcaaa aggtattttt atataaatgt acagttaggc agtaatacct tgaaaatcca 1740
[0148] tcatttcttt ctaatttcat ttagtaaagt aaaaaatact aaaaaaaaaa aaaagattac 1800
[0149] tgtacaggtg ggttattata ttatgaatga attaactgaa tttacatatc tgaaggactg 1860
[0150] tgtgggagtg tgtaagtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg 1920
[0151] tgtaggccag aagacagcca tgggtattat ttctcaggag tcatacatct tgtcatcttg 1980
[0152] tttttgaaac aagttttctt gttggcctag agctcaccaa gctgctggct ttagaaaaaa 2040
[0153] aaatacgtgt gtgaatgtgc tcatgtgtga catgccctgt gtgttgtatt cagagaacgg 2100
[0154] ctttctggag gtgtagctgt tctttccttt tacctttatg tagattctgg ggattgagct 2160
[0155] caggtcctaa ggcttgtata tactgacttt tcctttgtcc ttttaatggc aagcattttc 2220
[0156] tcccacaagg tatttacttt tttttttttt ttttggttct tttttttttt tttcggagct 2280
[0157] ggggaccgaa cccagggcct tgcgcttcct aggcaagcgc tctgccactg agccaaatcc 2340
[0158] ccaaccccag gtatttactt ttaaatattt atttaattga gatggatct tgctacattg 2400
[0159] cctgtattga cccgtgctga cctaaaactg tatgatccta ctattccagt ctcccaaagt 2460
[0160] agctgggact acagatccat gctattgtgc ttggccttct taactttttt ttcccctcca 2520
[0161] aaatctttca gcagtcttag tttctttctt tgagattagt ttctatgtag ctcagactga 2580
[0162] ccttaagttt acaacataac tggcaatggc cttgaagttc cgatcttcct acatccaccg 2640
[0163] ccaaaattct gggataacag gtgtgtgtca ctacacctag ttttatgtag tggtagggtt 2700
[0164] caaactcagg ctttctcaca tgctgctagg caagaactct acttatcagc tgagcttttt 2760
[0165] gtcctttact atttgatctg tggatatgta aataatttta cagttctatg ttttctctgt 2820
[0166] gtgtgtgcac atgtgcgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt atgtgtgtgt 2880
[0167] gtgtgtaggt ccatgtggaa gccagaggtt gacattggat atcttcctaa ttcatttgtt 2940

[0168] tctccacttg atttttgttt atttgtttgg ttggtttggc ttttcaagtc aaggtttctc 3000
[0169] tatgtagccc tggctgtcct ggaacttgct ctgtagacca ggctggcctc taagtcagag 3060
[0170] atctgcctgc ctctgccttc tgagttctag gatcaaagtt gtataccaca ctgcccagct 3120
[0171] tccccttggt tttaaaaaag gatctctctt tgaacctggt tggctggaaa gcctcctatc 3180
[0172] tccctcccca gtgctggaat tacaggcttg agccacagtg ctggtttttta tgtgtgtttg 3240
[0173] ggggatttga accaaggccc tcattataga atggcaagta cccttaatac tgaacaacct 3300
[0174] tccaagcgct tgtaatattt tcaatattaa taggcctaaa ttaacgttta ggtcttcaca 3360
[0175] gtttaagttc ttactaatca agggaaatat taacctttta aaaacatact aaaatgcttc 3420
[0176] atgtctgttt caagcagtca tgcactgaaa ctaaactttt catcagtcac aggactgctt 3480
[0177] gggattagca gctcctcaga gagaagagtg gaaggaagtg gaggagacca taggcataat 3540
[0178] agattgattt ccataggtca gattagagca gagaaagatt catactgaaa gaaatataaa 3600
[0179] aatttagatt ttaaaggtac ttagtcttgg gactggatag attactcagg tttttttttt 3660
[0180] tgtttttgt tttttttctt tttcttttct cttttctttt tttttttgg agctggggat 3720
[0181] tgaacccagg gccttgcgct tgctaggcaa gcgctctacc actgagctaa atccccaacc 3780
[0182] ccgattactc agtttttaag agtacctgct gttcctctga agacccaggc tcaattccta 3840
[0183] gcatccacat agcagctata accatctgta actctaactc tgagggatac aatgccctct 3900
[0184] tctggcctct gcagctattg ggcctgcaca tggtacacag acacacgtgc aggcaaaaca 3960
[0185] ttcatacaca taaaatgaaa atgaataaaa ttacttattc atagctacat atgtgtgtat 4020
[0186] gtgtatgtat gtatatatat atacacatac atatatacat atatgtgtat atatatacac 4080
[0187] acatagaaag atgagtatct ttgctttaag aaattgaact actagtggga ctagggcgac 4140
[0188] aaacaacttt aatccctctt gggaggcaga ggcaagtgga tttctgtgag tttgaggaca 4200
[0189] gactggccta cagattaagt tccaggacag ccaaagctac atagagaaac cctgtctcaa 4260
[0190] aaaacccaaa ccaaccaaac aaagaaaaaa aagaatagcc cattaaatac ctcccatgat 4320
[0191] tggttctgta acttactaag tgaagtaagt tgattttgga atattttttt aagccactgg 4380
[0192] aaaaaacatt gaacctggca aagagatgct ttgagttttt gaatctgctt tgcaatgtta 4440
[0193] atatgatgaa attgttcgtg tgtttttttt atagcacatc ttcattacct ccattagact 4500
[0194] ggcctttacc aactcacttt ggacaatgtg aattgaaaat agaagtgcaa cctaaaactc 4560
[0195] accatagagc ccattatgaa actgaaggta gccgaggggc agtgaaagcc tctactggtg 4620
[0196] gccatcctgt tgtgaag 4637
[0197] <210> 5
[0198] <211> 1301
[0199] <212> DNA
[0200] <213> SEQ ID NO.5
[0201] <400> 5
[0202] atcttgaacc agatgattgt gcatccattt acatctttaa tgtagacccg cctccatcta 60
[0203] ctttaaattc atcacttggc ttaccacatc atggactgct gcagtctcac tcttctgttt 120
[0204] tgtcaccatc atttcagctc caaggtttca aaaattatga aggaactgat gatatttctg 180
[0205] aatctaaata tagctcatta agtggtccta aacccttga atgcccaagt attcaaatta 240
[0206] catccatctc tcctaactgt catcaagaaa cagatgctca tgaagatgac ctacatgtaa 300
[0207] atgacccaga aagggaatat ttggaaaggc cttctagaga tcatctctat ctcccacttg 360
[0208] aaccgtccta ccgggaatct tcccttagtc ctagtcctgc cagcagcgtt tcttctagga 420
[0209] gttggttctc agatgcatct tcttgtgagt ctctctcaca catttatgac gatgtggact 480

[0210] cagagttgaa tgaagctgct gcacgattta ctcttggctc acctctgact tctccaggtg 540
[0211] gctctccagg aggttgccct ggagaagagt cctggcatca acagtatgga cctggacatt 600
[0212] ccttgtcacc taggcaatct ccttgccact ctcctagatc cagtatcact gatgagaatt 660
[0213] ggctgagccc cagaccagcc tcaggaccct catccaggcc cacttctcct tgtggtaaac 720
[0214] gacggcactc cagtgctgaa gtatgttatg ctggctctct ttcaccccat cactcacctg 780
[0215] ttccgtcccc tggtcactcg cctagaggga gtgtaactga agatacctgg ctcactgctc 840
[0216] ctgtccacac tggatcaggc ctcagccctg caccgtttcc atttcagtac tgtgtagaga 900
[0217] ctgacatccc tttgaaaaca aggaagactt ctgacgatca agctgccata ctaccaggaa 960
[0218] aattagaggt ctgttcagat gatcaaggga gcttatcccc atcccgggag acatcagtag 1020
[0219] atgatggcct tggatctcag tatcctttaa agaaagattc atctggtgac caatttcttt 1080
[0220] cagttccttc accttttacc tggagcaaac caaagcctgg ccacactcct atatttcgca 1140
[0221] catcttcatt acctccatta gactggcctt taccaactca ctttggacaa tgtgaattga 1200
[0222] aaatagaagt gcaacctaaa actcaccata gagcccatta tgaaactgaa ggtagccgag 1260
[0223] gggcagtgaa agcctctact ggtggccatc ctgttgtgaa g 1301
[0224] <210> 6
[0225] <211> 1301
[0226] <212> DNA
[0227] <213> SEQ ID NO.6
[0228] <400> 6
[0229] atcttgaacc agatgattgt gcatccattt acatctttaa tgtagacccg cctccatcta 60
[0230] ctttaaattc atcacttggc ttaccacatc atggactgct gcagtctcac tcttctgttt 120
[0231] tgtcaccatc atttcagctc caaggtttca aaaattatga aggaactgat gatatttctg 180
[0232] aatctaaata tagctcatta agtggtccta aaccctttga atgcccaagt attcaaatta 240
[0233] catccatctc tcctaactgt catcaagaaa cagatgctca tgaagatgac ctacatgtaa 300
[0234] atgacccaga aagggaatat ttggaaaggc cttctagaga tcatctctat ctcccacttg 360
[0235] aaccgtccta ccgggaatct tcccttagtc ctagtcctgc cagcagcgtt tcttctagga 420
[0236] gttggttctc agatgcatct tcttgtgagt ctctctcaca catttatgac gatgtggact 480
[0237] cagagttgaa tgaagctgct gcacgattta ctcttggctc acctctgact tctccaggtg 540
[0238] gctctccagg aggttgccct ggagaagagt cctggcatca acagtatgga cctggacatt 600
[0239] ccttgtcacc taggcaatct ccttgccact ctcctagatc cagtatcact gatgagaatt 660
[0240] ggctgagccc cagaccagcc tcaggaccct catccaggcc cacttctcct tgtggtaaac 720
[0241] gacggcactc cagtgctgaa gtatgttatg ctggctctct ttcaccccat cactcacctg 780
[0242] ttccgtcccc tggtcactcg cctagaggga gtgtaactga agatacctgg ctcactgctc 840
[0243] ctgtccacac tggatcaggc ctcagccctg caccgtttcc atttcagtac tgtgtagaga 900
[0244] ctgacatccc tttgaaaaca aggaagactt ctgacgatca agctgccata ctaccaggaa 960
[0245] aattagaggt ctgttcagat gatcaaggga gcttatcccc atcccgggag acatcagtag 1020
[0246] atgatggcct tggatctcag tatcctttaa agaaagattc atctggtgac caatttcttt 1080
[0247] cagttccttc accttttacc tggagcaaac caaagcctgg ccacactcct atatttcgca 1140
[0248] catcttcatt acctccatta gactggcctt taccaactca ctttggacaa tgtgaattga 1200
[0249] aaatagaagt gcaacctaaa actcaccata gagcccatta tgaaactgaa ggtagccgag 1260
[0250] gggcagtgaa agcctctact ggtggccatc ctgttgtgaa g 1301
[0251] <210> 7

[0252] <211> 449

[0253] <212> DNA

[0254] <213> SEQ ID NO.7

[0255] <400> 7

[0256] gaggctgctg gagtcttccc tcatttcatt atcccgttat gatggagcag gatccagaga 60

[0257] gcacccaatt tacccagacc cagcgagatt atctcctgct gcatactacg cccagaggat 120

[0258] gatccagtat ctctcaagga gagacagtat tcgccagcgc tccatgcgct accaacagaa 180

[0259] tcggcttcgt tcttccacct cctcttcttc ctcagacaac cagggtccat cagtagaggg 240

[0260] aaccgacttg gaatttgagg actttgagga caatggtgac agatctaggc accgagctcc 300

[0261] ccgaaatgcc cgaatgtctg caccttcact tggacgattt gttccaaggc gtttttttgct 360

[0262] gcctgagtac ttgccttatg ctgggatttt tcatgaacgt ggacagcctg gcttggctac 420

[0263] tcattcttct gttaacaggg tcctggcag 449

[0264] <210> 8

[0265] <211> 20329

[0266] <212> DNA

[0267] <213> SEQ ID NO.8

[0268] <400> 8

[0269] gaggctgctg gagtcttccc tcatttcatt atcccgttat gatggagcag gatccagaga 60

[0270] gcacccaatt tacccagacc cagcgaggta aggccaaaat gagaaatggc actaggaggg 120

[0271] ggggaggggc ggcccttggg gttgccaagt tgtgggaacc cccacctgtg cctctctggt 180

[0272] cctggtgctc tgttgctctt tagtcacagc attgaggcag gtgtttcagg tgattgggga 240

[0273] ttgttgcagg taagtgaatg gcatctctgg tgacagaaaa ttataattgc gctattgtcc 300

[0274] tttgggactc tgtctgtcag tggggcacta gggtacctgt ggagggcata ggtagtcctt 360

[0275] cctactactc aagagtctgt tggctgtcac atggacaacc aaaagaagag ccaaagtttt 420

[0276] gagcatttta tttgtaggga agatagttat attaaacaaa tttacaatat caaaagggta 480

[0277] aaaaaaaatc aagtcaatgt ttagacactt tcccaaagaa gagatgaagg aagcataatg 540

[0278] gaatgaaatt gtcattttac tgaagtgctt gctttattct gttttaggag acatttgaaa 600

[0279] ggttattgaa gaccctgaat gtgtttgtgt gtgttatgca tatgtgtgtt tgtggtatgg 660

[0280] gcacaggcag aggccacatg aagactttgg agtctttctc tgttcttcac attatcattt 720

[0281] ttttctcttc ctcctactcc accatcatca tcatcaccat gacatcatcg ttcatttttg 780

[0282] agatggcatt ttttactgaa ctagaaactt gccattgtat ctagacatac ctgtgatctg 840

[0283] tctttctgta tcttgtattg ctggggtcac agatacctgt gtcatgccca gtttttacat 900

[0284] gggtgctgga gagttgcaca caggcccttt tgcttttaca gcaaactccc ttatccaccc 960

[0285] agtcaaatat ttttctaagc ctgaagtttt taaaagttag gggaaagatg tttttcattt 1020

[0286] gagcctaagg gaccatccta aagtccctag tgctgatggt ggcgctgata acaaaaccaa 1080

[0287] agctggaaca ttgagttgta ttacaacatt tatcagtaaa gagcagactc aggtggttgt 1140

[0288] catagctgga tttgatgact tattagagat ctgaaaaatg ctaacttgac tattatttaa 1200

[0289] atagtatcag aaaaagctat aaggaaaagg atagtttaca gtggtgggct aggcaagtca 1260

[0290] gtagtatagg ccagataaaa ctaaatgagc tttcccgtct atctggaaac atctttaaag 1320

[0291] accaatttga ctctagagat agtaaatgat tttgttgttg ccacagttta gttgcagtag 1380

[0292] tgcccattct taccaagata gtagtggaac aaaaacttgc catggcctct tactgtcttt 1440

[0293] tccatatccc agaagtagca aatccaagga tctagacagt agagagggat ggggacatta 1500

[0294] aaaatagata ggcaattcag atggtgattg aagcaaggtc agctatatta gtctctaatt 1560
[0295] atgttggaaa acatacaagt ccttaaaaag ggaagggaaa gtttcacatt ttcgggaacc 1620
[0296] ataaaagact cgacacaact tctttccata aatagtttct tttcatgctg tgacattgca 1680
[0297] cacctttaac tccagcagtt gggaagcaga ggtaggtaga tctctgtgaa tttggggcca 1740
[0298] gcctgctcta cctagtgact ttcagatatt ggtaggaatc tcattctcag aaattttaga 1800
[0299] ctgggaaact ttaacgattt gtttcggaga cagggccaga tgtacagcct gggctacctc 1860
[0300] taggtcactg tgtagccaat ggaggctggt ttcaaacttt tgatctttct gtctctgtct 1920
[0301] cagaatgtag agattacaga tactgccctt tcatgcagtc tgttattaat tcctactctg 1980
[0302] atttctgtcc tagtgaaagg cttagatacc taagatctca ctcggaacac ttcctggttg 2040
[0303] ttggatggtc ctgtggctgc tcggtggagg gtatttgtac catcttctca cattaactca 2100
[0304] tttctttctc ctcttagttg tgttctacat tatacggata gggaaagctt ttgtgagaca 2160
[0305] ttttctaaaa catgcagaaa ccttccttcc cacctggata cataagaggg caaaactctg 2220
[0306] tcagaactct atcttccttt atttgtttat gtgtactact ggtgatttag tttcatgtcc 2280
[0307] tgtagtgtaa ttaaacgact aggggctatg tgatgctgtg cttacatctc tatacaaaag 2340
[0308] tgtagtcctt attcaatatt acatcaacag cgctatttct acagccatta tcagtaacag 2400
[0309] tccatatttt tctgttcttt ctggccatcc ataagctgag aaatgggtgt gcttgttgtc 2460
[0310] acttacgttc tcttccacat cagttctctt caattctttc ctgttacata tttttttctcc 2520
[0311] agtctaattt ctttgtgttt aatccatgtt caatttctgc tggaatattt tagtatattt 2580
[0312] ggtaacagta agcacaacat gaaggcactg acttctgtgc aatgtctggt ggttagtggt 2640
[0313] tgaatgcttg gcaagctagc ctatgaatag tgaaaatacc atgagttgag aaacttttgg 2700
[0314] aaaacagttg gtgttgtctg attagtttga tctaggcttc tgcatgcccc tgcccaccta 2760
[0315] ctgtacaaag gtcatttgga gtcattgtgc tcagtgtagc atattatttt aacatgtaaa 2820
[0316] cttttgagtt ccagtgatta ctagctagct ccttgagaat tctatttgat aagtgaccca 2880
[0317] aaggacaatt ttaatgacag ccttaggtcc cagtgcagtt gactacaaga tctcttttag 2940
[0318] cctaggaagg tagaacttct gagaacaaag tatcaaaggt actggaagtg gaactgcaag 3000
[0319] atgtttgcag ttttgggggt gcctctgtta ctaatataag ttttggggtg cctctgttac 3060
[0320] taatataagt tttgggggtg cctctgttac taatataagt tgcccaactg acctgagctg 3120
[0321] aggctcttca gcttcttgag tagctgagat taaagtcaca caggtgtgcc caagaataaa 3180
[0322] tacttttatt tcatttagat acagctattc taagtcagtt tttaagtatt ctcaataaaa 3240
[0323] agaagacaaa acacaccttc ccttttagat atctaaacat tataaaaatg taaaatctcc 3300
[0324] caggtttgta acactaggtt tatgacttct ttttaaatgc agaaaaaaaa gtgtgtgtgt 3360
[0325] gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtat gaacgacagt 3420
[0326] atatatttca ctaacccttt ctctatttta catttttctt ttggaagagt tgccttatga 3480
[0327] gtagataagt ataattgtta catgtacata catcttaaaa tgttgagaga gggtctcaag 3540
[0328] ctaacttcaa acttgtgtgg ttgagactga ccttgaactc ctgatccatc tgctctaggt 3600
[0329] ttgtagtgct gagagtatag gcttatgtca ctgtgcccag ccaattcttt ttaatggctt 3660
[0330] cctggtgttt ctctgtatgg aaataattta cttaaatctg ttggtgggtt tctgggttgt 3720
[0331] ttcttggatt ttccccactg atgacagtgt tgcagatgaa ctttctcatg taatttatat 3780
[0332] aacattcatt tatattcttg tgtatttctg tgtgatataa tttactagtg gacttgcatc 3840
[0333] attgtgagtt ctttaaagtc ttacagttgg tttagcaatc tcttaacact atcacataag 3900
[0334] acatttttct gttgtcctca gtttatagtc tctaattaca gaacagtgtt gtaaagatgg 3960
[0335] aggaaaacct taagaaaggg aaggagtatg tagaattgtg tgatattaga aattgtcagg 4020

[0336] ggcctggaga gatggctcag tggttaagag caccgactgc tcttccaaag gtcctgagtt 4080
[0337] caaatcccag caaccacatg gtggctcaca accttcagtc atgagatctg acaccctctt 4140
[0338] ctggtgcatc tgaagacaac tacagtgtac ttagatataa taataaataa atctttaaaa 4200
[0339] aaaaaaaaag aaattgtcag agagctaaga catgtcgtcc ttggtttcaa ctggagtcta 4260
[0340] tagactgatg atggtttatt ggaaccctag cagaggagat gtattctcca ttgctgtaat 4320
[0341] tctacaaatt gaaaattctt ttttctatta ggttttctat tctcccaact acaaaatgtc 4380
[0342] tctcactccc tatcctggat agttttccat gtacaaatac ctatgattaa tttgtgaggg 4440
[0343] agggaagata caggtttcaa agagacaata ggagttaggg gaaagctgga gctgtgtgtg 4500
[0344] tgttgaagaa agaagtcctg gtaaagagga tgcctctggg tacctttgtt aatcatctaa 4560
[0345] gtgttctacc gccttctggc ttgtttttc atctataaaa catccagact tgattactaa 4620
[0346] ggtcttttc atttcttaaa ctttgtgtat ggacactgag caaggaagaa ctatttaagg 4680
[0347] acacccttaa atgctaagcc aggttccata agagcctgag aaggaagaaa gacccagtga 4740
[0348] taatcctctt gggaaggggc cactcttgat cagaagatga ctggttacat ggcctatttg 4800
[0349] cctattctca gagcaggagc actggggtgg tgtggcacac attattctgc agcacacacc 4860
[0350] tcctggcgag ttcaatgctc tatagttttt gtatatccag agctgtatca gcagagagat 4920
[0351] gccatgggca agttggcttt ttctatcttt cacagtattc tccacattag tgctagtttg 4980
[0352] tgtttgtcta tctctttctc tctcagatta tctcctgctg catactacgc ccagaggatg 5040
[0353] atccagtatc tctcaaggag agacagtatt cgccagcgct ccatgcgcta ccaacagaat 5100
[0354] cggcttcgtt cttccacctc ctcttcttcc tcagacaacc agggtccatc agtagaggga 5160
[0355] accgacttgg aatttgagga ctttgagtaa gtatattttc agtttgcttc cttctttctc 5220
[0356] ccagactaag agtgcttctg agttggcctc tgtttctgta gcagtggtga gcattcttgc 5280
[0357] ctctaacccc taagttatta tttatgtgca gcaaagtggg atgaggttca gagggtttta 5340
[0358] ggtacctaga gcactcctat ttctgggagt tacattgccg attcgccaga gcctttggga 5400
[0359] agagacaacc cattgggcca caggccggag tggaaagagg atggctgctg ctctcctgac 5460
[0360] cacacaccag acaaaggaag caggaggctt ggtgcttatt ctcttaacta gagtagaggt 5520
[0361] gacaatttaa gtttcatcgt aagtatagat ttccttgaat aaactgttcc ttgagatctg 5580
[0362] actctcttgc ctcttcctgc tcctagcccc acctgtcaga ttaaggtggc ttggggagtg 5640
[0363] actccttgat gccactgtca gctggggccg atgctgcatc tcctcttcct ccactccctt 5700
[0364] ctctttttc tttgtttcct tatgttggtg agtggggtaa agacaaggcc taagagcagg 5760
[0365] gccagcccca agttccagag tccatgcttt ccagccctac tgctgagttc tgcctctcag 5820
[0366] atattctatt cctgaccttc tagccattta ttgtgcctcc agtttgaaat tgagttttcc 5880
[0367] tagcttgtat caaattagac atttggcttt tgttttgatt ttgtattttg acatttttta 5940
[0368] acatgcaaaa gaaaaaaaga aaaagaagca tttccctccc tcttccccaa ctccctcttg 6000
[0369] gcagctgcat gaatggttcc tcacaagcta acttgcaata gagctataga gtatatttta 6060
[0370] tctttttgcc agaaagtagt ctgaaaaaag attttaatgc tcaacaaatg caggtttagc 6120
[0371] atgaagcaga cagagcatga gtgtgctggt cctgtcagta atccctgaga ggagcccagc 6180
[0372] acctctgacc cacagccttt tttctgtctg tcctcctggg taagatggct gcatgtacaa 6240
[0373] acagtgaaag gagccagata gatgtatgtt attctggagc tgtagtagtt tttctagcac 6300
[0374] aagggtcctt ttagggaagg agtttaaaca tacatgccaa tggggttggg atgtaattca 6360
[0375] cttgcttagc atgtgaggtc ctgtgtttga ttcttcagca ctgcaaatat ataaacacac 6420
[0376] agacacacac agagtgggga gtaatgcggg gggtgggggg agagaaataa acaaacaaac 6480
[0377] aaacaaacac ttaaacactg aggctacggt agtagaaaaa cctgggtgac tttagttcag 6540

[0378] ttattctgca gtgcgctagg actctgaact ctggactctg gcttgccttg cctcttcttt 6600
[0379] ctatgcatgc ttagaagttg aaactgcctc tcaaccagtg gtgtgatgtg gggatcatgc 6660
[0380] ctctcagctg taccctgcct cgtgacttgt ccttcacaat ggagttctgg gttagtttaa 6720
[0381] aatgttctgt tctcatttta ccttcccttt cttctgctgc cacctcctcg ctctcccctt 6780
[0382] ccttcccccct cgattttcct ctgtcagact gccttctgtg tgtgttcccc caacccacct 6840
[0383] gtgctctcat catcttcctg ggcttcagca tgacctggtt ttacattgcc tttttttacct 6900
[0384] ttccttcatc agcttaccag gtttttcctc ggtgtcgtta ttagccatgg cctaatctcc 6960
[0385] ctaatgccgc acctatatgt ccttgtccag atgtttgcct tttggaagat ggttttggta 7020
[0386] caaactgtaa tcttgctttt gatagcagct atgactcagt taccttaata tgagtagtac 7080
[0387] aggtcttcta gagaaagagg ttgatactct cttgccacct gtttcttctt tttctatctt 7140
[0388] ctctgtcctc ctcctggtcc tatacgctga gtactggttg taattgctcc ttttcgcctc 7200
[0389] tagtgctcag gggagaggca gaaactaggg attgtagcag cctgttctaa tttatttgcc 7260
[0390] ataatcctat tccttagctt tcataactag aaaggatgtt atttaggcct gccctccctc 7320
[0391] cctcccttct ttccttcctt cttgtttttt gagacagggt ttttctgtgt agtcctggct 7380
[0392] gtcctggaac ttgctctgta gaccaggctg gcctcaaact cagagatcca tctgcctgtt 7440
[0393] tctgccatct gagtgctgag attaaaggcc tgtgccacta ccacccagtt tgggatatta 7500
[0394] gtttttttaaa taatgtttta acttatgtgt atggatgttt tgcctgcatg tatatctgtg 7560
[0395] accatcgtgg ttcctggtgc ccagagagat ctgactagag tcaggtctaa tagaactgta 7620
[0396] gttacagaca ttgggagctc ccatcttcca tgtgggtggg tgcttagaac tgagcctagg 7680
[0397] tcctctgcaa gagcaacaat tgcttttaat tgctgagcca tatctccaac cctcagaaaa 7740
[0398] gagggatatg tgggggggtg gggtaggggg atgttataca gtcattttat agttatttta 7800
[0399] aaattatgtt attttaatat acagaagcag gtaaggtttt tgtgatctgg aatctgccta 7860
[0400] aatggacacc attaagagac actgttctat tgttttttgt tgttgttgtt gttgtttgtt 7920
[0401] tgtttgtttg tttgtttgtt tttctgagca cattacaatc ttacctgtaa ggcctcatat 7980
[0402] aatcctttcc actctccctt tttctaagtc atagattcat tttactgttc ttgagcagtt 8040
[0403] ttcatccaga tgcataaagt ccccaagaaa tcagatcatg tcttctttgg ttttctttca 8100
[0404] tcatcaaaaa ttatcagtta ctgaagagtt catatgcata attgttaact acttatattt 8160
[0405] atagtatata taagcatata tacagtatgt atagttatag ccaaagtatc cacctcatat 8220
[0406] atgtatatgt agcaattgct aatactttgc atacagatgt agattatttg catatctttt 8280
[0407] tatgtataaa taatagattt aaaatttttg ttgaatcctt tttaaaaata tattatttat 8340
[0408] ttattattat ttatatgagt acactgtagc tgtctgcaga cacaccagaa gagggcatca 8400
[0409] tatcctatta cagatggttg tgagccacca aatagttgct gggaattgaa ctcaggacct 8460
[0410] ctggaagagc agtcagtgat cttaaccact gagccatctc tccagccctg ttgaatccct 8520
[0411] ttaatggccc aacaagaggg ctaataaatc attctcatct gttgttcttt gactgaggat 8580
[0412] atcaccatat cagagatgta tcagattaga gaacaggaat gagtatattt attgtctaaa 8640
[0413] gatctaaaac cttttaacag tgatatacaa ttatatttgg gtcctttcac atacttgttg 8700
[0414] tggaataggt ctttccaatt tatcttctgg gaggatgccc ttggaaattg gtcaggttaa 8760
[0415] attttctatt tttttgtagt ctggattgat gactatcttc taagcacgca tcttctcctg 8820
[0416] tcattcatct ctatctgttg aagattgaaa tggcaggatc ccagtgttta tcaaaatgta 8880
[0417] caaataggat gtccagaaac ttgctgtcag caggtctttt ttgaagtcct cagctagctg 8940
[0418] tctttgcata aagttcacaa gcagggtaaa ggtgtgtctg cctattgaca ttactgtctg 9000
[0419] cctgtgttaa agcttagact gcttgcttga atgtagttgt catataagca taacaaaaat 9060

[0420] cctattttaa tttataactt tcccattaga taaaggacag gtagcttaca aaaaaagctc 9120
[0421] gagaaaagtg cttttttttcc agtagttgct aagacttgca agagaataat gtgtgtgtgt 9180
[0422] gtgtgtgtgt gtgtttatga taattctagc tagtctttat ttttaattaa ttaaattaaa 9240
[0423] taataattta ataccaggtt ttagtgacaa acatttacat gctgagctat ctcactaacc 9300
[0424] cttgattctc gctagtcttg aacctttctt tcttttatct gttattctaa ttatgattct 9360
[0425] agaatctttc tccactcatt tccctgccct tttaacttct ttttgtagct gctttgcctt 9420
[0426] ctgaggcaaa ggatgaataa cctgatccga tacttgctac tgccagaact agagatgttt 9480
[0427] cagattagat aagccacaaa taacttcctc cacacataca gccactggta attatactca 9540
[0428] gcttcttaaa gatcgttgct agcatagaag agaataaacc ctttctttat tcatggtttt 9600
[0429] gacactgggc tgtatttgga aaatctcagt gagaacctta aagttttcag tttaagatag 9660
[0430] agtctgcagg ttagttagaa ttgtttcccc ttcctctcct ctacccttcc tccttttctt 9720
[0431] cctttcttcc cttctttctt tttttggagg ggttgggtgg gaacagtagt aggaattgaa 9780
[0432] cctagggcct cttgcctgct aggcaggtta cctgatcaca gttgggccct ttgttcttat 9840
[0433] ttctagacag tctcactatt gtagcctagg ctaacttgta agtctttatt ttcatgctgc 9900
[0434] agcctcccga gtgctgagat tataggtgtg tgtaccacca tgctcaattt tttaaaaaaa 9960
[0435] ttttatttat tttatttatg agtatactgt agctatcttc agacacacca gaagagggca 10020
[0436] tcagatccca ttacagatgg ttgtgagcta ccatgtggtt gctgggattt gaactcagga 10080
[0437] ccttcagaag agcagtcagt gctcttaact actgagccat ccctccagcc cgctcaatta 10140
[0438] cttttttttg ttgtttttg ttttgttttg tttttcgaga cagagtttct ctgtgtagct 10200
[0439] ctggctgtcc tggaactcac tctgtagacc aggctggcct caaactcaga aatccacctg 10260
[0440] cctctgcctc ccaagtgctg ggattaaagg catgtgccac cactgcccag ctgctcaact 10320
[0441] acttttaaat aaatactttt aaaagattta tttactttat tcttttatac atatgtgtgt 10380
[0442] ttagcctgca tatatgtgtg taccacatgt atgcctagtc agaaggagaa cattggactg 10440
[0443] gagtttcaaa aagtggttag ctgtgagtat tgggtattaa actcaggtct ttcttctaca 10500
[0444] agagcaccaa gtattagtag cctttgagcc aactctctac cccattttgt ttctgagaca 10560
[0445] gggttttgat gtatttccaa ggcagatctc aaactttgga gtttgtgtaa tgtttccacc 10620
[0446] caagcttctc cagtagttga gactataaga ctataagcac atagttgaga ctatgtgcac 10680
[0447] cacctcacct gcccgagctt ctctttctct gctccccttg ctaggcatct tgtggcatat 10740
[0448] atgttcttaa ttttgtgttt ggtatagctg aggagagttc caaaggctat cagctctaac 10800
[0449] atctggctga atatgacagg aaaagttgtt gacctttagt ttacaatata ctcttcttcc 10860
[0450] tctgcccttt tggtataact gagatttagt ccagcctttg gcaaaggaac tatgttaaaa 10920
[0451] accacgtaag tattgcttct catttcaggt gataatcctg aatgagactc ctacatttac 10980
[0452] ttcaaatctt ggctaggaat tccatcaaag aagccacatg tcttcagtag tacttgttct 11040
[0453] tagtttcttt tctgagctag gaagttccag gatttctgta ggagctcttg acttgtcttt 11100
[0454] tattttattt tattttattt tttgacttgt ctttaataac aaatataact cagagccaag 11160
[0455] aagaagggca gcttttgcca tctgaaatcg ttagagagat atggctgtat cattttgaaa 11220
[0456] tggtctttat ttcttccttt cacatataaa ctttagaagt gtaaagatgc actataaagc 11280
[0457] actatgaaga tgttgttctt gctgggaggt gatagtgcac acctttaatc ccagctcttg 11340
[0458] gaagatggag gcaggcagat ctctgtgagt ttgaaaatag actggtctac agagcgagtt 11400
[0459] ccaggcaatc agggctacac atagagaaac cctttttcta aaaacaaaca aacaaacaaa 11460
[0460] caaacaaaca acattcaggt gttgttcttt gtaggctgct aggcaagtgc tcagcaaatt 11520
[0461] gcatctgtag ccaggcatca gtcttaaaaa aatagggtct ttcaatagtt aaacttagga 11580

[0462] gacagcaaaa caaaacaaga aaagcatagt tattttgaca gttggctttt cttgattctt 11640
[0463] gatagatgca actatattag tcactgtaaa gttcaaatct atccaaaaag atgagacttg 11700
[0464] ggagtcatag aagtgcagat cttaagagtt gtttttgcct ttgagatgag ctttatccta 11760
[0465] ttaagatggt gacattaaaa cattttagtt atagaataaa atgatcagat ttttatataa 11820
[0466] gtatatgaaa gggggctgga aagatggttc agaagtccat agcactggct ggtcttttag 11880
[0467] gggaccctgg ctcagttccc agaacccaca tggcagttcc cagctattta taactccagt 11940
[0468] tacaggcaag tgcctctggc ctctgtggtc actgcatgta catagtgtcc atacaaacat 12000
[0469] gcaggccaaa cactcatata gacataaaca ataaaataca aagtaaaaag tacatatgtg 12060
[0470] aagagagact gtgacttaaa gagtttgacc agctctgcag gaaggaagaa cagattttgg 12120
[0471] aaaaatggga gactccatgg cttgagcttg attttgaaga ctgaatagga ctactttttg 12180
[0472] ttttgtttgt ttttaaattg cgtggggttg ttggcatagt ggggaagagt tgcatcccac 12240
[0473] tgcatggatg tattgggttt caccttacct gactttggtt ctgttttgtt catagttgta 12300
[0474] tgaccaactt aggtgattga gtttctgaga attttcttgt ctctgcctat ctccctagag 12360
[0475] gggtacactg ggattctaca ccagtgacca gttatgtcct tatattgtct ccagcagaag 12420
[0476] gtgtagccca gattaaaggt gtgttccgcc acacctttaa tcccagatga aaggcgtagc 12480
[0477] ccagattaaa ggtgtgttcc ttaaactcgg agattcaatc ttctggaatc cattgccatt 12540
[0478] atggctcaag atctccatac caagatccag ataaggatct ccaagcctcc agataagggt 12600
[0479] cactggtgag ccttccaatt ctggattgta gttcattcca aatatagtca agttgacaac 12660
[0480] caggaataac cactacaatc cacccсttgt caacttgaca caaataatat ctcatgttca 12720
[0481] catgaaacag taacaaggtt gtgaatacgc ctaacatgat ataactatcc ctcgtacaat 12780
[0482] cgcaaacgca tttgtcaatt tacaaggggg cattcatatt actttataat cctcgtttct 12840
[0483] gcaactggtt acgtggcctt aattggtatt atctagtttt aacatgggtt ctggggatcc 12900
[0484] aaactcaggt tgttgggctt tcgtagcagt acttacccgc tgagccattt ccccaccccc 12960
[0485] tggactttga aagatgaatc tgtgaggatg gtgtagaaag aatatgctag ataaaggcaa 13020
[0486] taacaaagta aaagctcaga aacagatttt actcaaggtt tattcagaag cttgaataga 13080
[0487] tctctgtgta gagctatagt gtgctggtga ttgagaaaga tcagaaaact gagaactgtg 13140
[0488] aagggcttgc aatgtatggc agaaaatttg gattgtgtgt tgacttggcc tcctgagtag 13200
[0489] gggcattaca agtataggtc actatgcaag cgtacttgat gttttccaag actgagaggc 13260
[0490] ctatttgtgg gtgtctttgt aaagacagcc atgtcatcta cttctccaaa tagaatgctt 13320
[0491] cctcagcctt tattaaatat gtgctatagg aagctaactg aacaaccctg attaaggtga 13380
[0492] tacaatgtga caagtcaact gatgaagaac tagttcagta gtgatggctt agtataagtg 13440
[0493] aagatgtttt aaatatttat tgtgaaagtc cttagtgttt tcatcgtaga taccctttta 13500
[0494] aaaacgaaaa ccagggtctt attttgtagc cttagggatt gccctgtttc tgcctctgga 13560
[0495] ataatgggac taaagatgct caccatcatg cttatctggg tgtcaaattt ttgttgttgt 13620
[0496] ttttcaagcc tggttttttg tatgatgctg aggctgaacc tgaatttctg ggctcatatg 13680
[0497] attggcctga ctgggcctcc tgagtagggg cattacaagt gtaggccact atgcaagcgt 13740
[0498] acttgatgtt ttccaagact gagaggccta tttgtcgatg tctttgtaaa gacagccatg 13800
[0499] tcatctactt ctccaaatag aatacttcct cagtctttga ttactattac tggagaccat 13860
[0500] tttgaagtaa gtatttgtaa gttaacaagt gatccttaga acatcattgc ttgcaattgg 13920
[0501] aatcttttat tattgttgct agctctgttg actgggttct agtcttggtt gagcctctgc 13980
[0502] tgtagaaagc ttggaacatc taactgagaa gtttggattt tgtatgtatt tggggcagcc 14040
[0503] attattcagt ttttttttt tttgtttttt gtttttgtt tttaatttag gcagcaaact 14100

[0504] ctaaaaaaac agtggctggg ggggacctac tttggaaagt gtactgatgt agaagaacaa 14160
[0505] gtttgagaac aagtcctatt cccaaggtcc cagtgtctga tagtaaagga aattgaacaa 14220
[0506] cactggtgct tcttttaacc ttgaggttct tggatctcta ttggtgttaa ctgccttaga 14280
[0507] agatttccct tgatactcag caaagagaaa ataatgagaa ttctcagaaa gtaagagggc 14340
[0508] ctcctgaaga ggaagaggga aaagctatct agtgcttgtc aagcctcaat atgtttccaa 14400
[0509] aatatctact tatgcactac tacttcttta ccaaggatta tgacctgttg cttacaaatg 14460
[0510] tgtaacattg cttgacagct aaagaaatcc tctttctgaa aatataaagt accttatgaa 14520
[0511] tgtttatttt gggttatgga gaataagaaa tgagcaaagg aaagtagaat atatctagat 14580
[0512] tgtgcctggg ctgaaaatgg aaaaatacag acagaaagaa tgaaaaatac agccaggaaa 14640
[0513] tacagggagg aagacattac ctcttagatt tatatgttgt tttatcagaa aatgcacaat 14700
[0514] atgaatataa atttgcataa tgtaggcaaa tgagattagg aggtagaata agggaggaca 14760
[0515] agaatttgta gattataatt atctcatatt acttttcctt ggagaaatct aaggctctgt 14820
[0516] tgttccagga gacgtgcatt gttaggtttc ttacaaagga tgtcaactgg gacaaaaatt 14880
[0517] taagatcttt gtaagtgtcg tctctggaaa tagaggctgt gcctcagttt ctaataggag 14940
[0518] ctagtaagaa aaggacttga ccaagatcac tgtgccattc agagacagaa atgctaagta 15000
[0519] ggtcttctga tatttgttct tagacacact ttctttttaa atatttagca aaatcaacag 15060
[0520] tactgttggg aagacaatga cctttgagc tttttgatgg aatagattca tcctcaaggc 15120
[0521] tcatgtcaga gctctgtttg tctgaggaac aatgacaacc tgactgtctg gtgagggctc 15180
[0522] agacaacagc ctgtagggat atttttattt cctttaaatt atggtctttc tttgaagctc 15240
[0523] ttgatcctgg gcatctagca gattgctgtt gtgtctctct tttgtcctga agatgtgttg 15300
[0524] cctttcattc aacacatgtc ttagaagctt gtttgaagga ttttttttt ctttataaaa 15360
[0525] taatttgaaa caaaatttta gtcgttcccc cacccaaaaa aataccccac tgagagttaa 15420
[0526] tcaaagatgc atagagttca acaactggga aggccattaa tagtacagtc agtttctatc 15480
[0527] ctgctgtctt aaatgcaact tctgaatgtt aaaattgatg aggatttctc tgatgttttc 15540
[0528] ttaaaaagta gtccttcttg gtggaaagat tgtactagca tatttgtgta ctttgttctt 15600
[0529] catttctttc ccataaaagg ttgttggcta tcttatgtta gagcctagtt tttagcactt 15660
[0530] gctactattg ctaatgtcaa ccaactgtag atctgttcat ctggcattat cttcagtccc 15720
[0531] gttttgactt atttcactgc tttctaactc ttgcagcttg ctttccgctg accttgtgtt 15780
[0532] agcagtgaag tgaatggcag gctcctcgtc ttcttaaaat atcttatttt aagtgcatct 15840
[0533] aattcttttg gtcttttctt ggaactcttg tcattatctt tctctcaaat ctcccatgca 15900
[0534] caaaatttgt tcttctatcc ccagttgcct ctaaaccccc ttctttctag gcaataaaat 15960
[0535] cagttatttc tctcatcttt agtttatttg gagctgtgtc aaaattattt tcccgtctag 16020
[0536] tcacccactt tggcttttgg acattctgat cattgctacc ttgacacatt ctttcttttc 16080
[0537] aaggtttatc ctgttagtaa ccatttgtgc tgcttatctc tcatgctgtt ttctcctaga 16140
[0538] cggttttgct ggtttctccc ttaacatctc atttgtgtgt ttctcatatt tcatattttc 16200
[0539] ctaataatcc tggttcctgt gctgtccttt ttctctgttg ctgctggagt cctaacactg 16260
[0540] aattccgagg cttacctgca tggtgttgag ttactactca gctcctgctt acccatcctt 16320
[0541] agccagaatg ctaagattaa gacctgcata gagctttcca ctcacttgta tgctgttcca 16380
[0542] taatgccatc catttgagaa tgtgcctata tctgagctag ctatgcttac tgtctgtctc 16440
[0543] tgtctctctg tgattagtga ttgggtctta ctatgcagcc cagactggtc tcctgagtac 16500
[0544] tgggatcact atatatgtct caccatacct gatttgatct tatttttcct ttatatatca 16560
[0545] tatttcagta gtctaaacat tttaaagcta ttacctgtgg ctgtatttca gcccaccctc 16620

[0546] cccatttctt tagttggtgt ttatttgtat gtgtgtgcgc gcgcatgtgt acacattata 16680
[0547] gcactaaaat tactactttt gcccccccttt taaatgttat ttattttata agcattgatt 16740
[0548] ggtgttttgc ctgaatgcat gtctgtatga ggctattaga tcccatgggg cttgagttac 16800
[0549] agacagttgt aagtttccat gtgggtcctg gaaattgaac actggtcctc tggaagagca 16860
[0550] gccagtgttc ttaaccactg atccatctct ccagccccta tcccctcttt ttttaaaaaa 16920
[0551] atagaaagaa taactgcatt taggttttag gctttaaaat agcaaaaaga ggtttgtgat 16980
[0552] aatatgtatt gtaagagtca aacatttccc cactcatcct tgtcacttgt catccttatt 17040
[0553] tctcaccagg ctttatgctc tctcaaaaag cattgcacct catcagtact gtccactacc 17100
[0554] ctgcccttca tatctgtatt cacagataat agtcccatgt tacccgaaaa ggtggatcta 17160
[0555] tggacttgta ggaacgtgag tcagctccac aggtgatttg taggaagaac agctcagtcc 17220
[0556] agagccatgg gatagaatct accctataat gctgctatct agctttgttt tcagtcttct 17280
[0557] taaggcctgg gggcctctct atgttccttt catgtcacac ttttaaactt taaaatgttt 17340
[0558] tctaaatgcc tttcatgata attgtttaca atcattaata aattcataat aacaaatgct 17400
[0559] atcattactc tagttgtact tttaattggt actttaactt atcataacca aaatcatata 17460
[0560] tttgaaaaat aacaatattt agagctgtct tatttgttta atattcagct tctttgtgcg 17520
[0561] catgttagga gctctgcaat aattatattt atgcagcttc tgatatattt gtgtatatgt 17580
[0562] ggatgtgtgt gtacatgtgt attatgtacc tctgtgtata tgtagatgta gacatctgtt 17640
[0563] gaggaggatg tgtgcacacg ctcctgcgat tggcttgaag atgatgttga gtatcttcct 17700
[0564] tggttgttcc tttacttttt gttaaggcag ggtctcctga tgaacccata gctccctgaa 17760
[0565] tctggttaac ctagctagtt agccagcttg cctttgggac cctctgtctg cctccacagc 17820
[0566] cctagcagta taagctcctc tataccagcc cagccctttt tttttttttt tttttttgg 17880
[0567] taaataataa atgatttatt tattttgtat acaaaatgtt ctgacttgat ggctagatga 17940
[0568] gggcaccaga gctcattgta gatggttatg agccaccatg tggttgctgg gaattaaact 18000
[0569] cggaacccct ggaagaacag ccagtgctct taacctccaa gccatctctc cagcccccaa 18060
[0570] gcccagcttt tatatgactt ctgatgatct gaactccatt cctcatgctt acatggcaaa 18120
[0571] cttgtctttc tccttatcct tccttttata cccagagttt caagtcaaca gaatacaata 18180
[0572] tcttcggaaa acaagagtgt atcttttcat tgggatgggg gacccaaata cttctatcgg 18240
[0573] atcagctttc aatagccttg gtattgattg ctcttaatgt gtgaaaccct taaaagcttt 18300
[0574] tgtttctagc tggtacagct cccagtgggc cacctaaaag ccaaagtgtt taaggttgtt 18360
[0575] tccagggcta gattgaagat tatctcattg atttgaagat aattcaaaga gagaattcag 18420
[0576] atatcccaag tcaggctttc agaggttcat gtacttcttg atttgcttaa gcagctaatt 18480
[0577] aactgctctg tccagccagt gccctttaag ttattgtttc tttgatgtgg tggaaactta 18540
[0578] agccacaaga agttcctttc ctcttttctt gtaggaaaca gtaaaatgca gtgaagagat 18600
[0579] ccaaagctct ttcctgaatc atcaactaca ttctagcctg tgaccttgga tttacctgtc 18660
[0580] tgattcttta gtttccttat gtaaaatcag actaataaca cctcatggaa tagctagaga 18720
[0581] ttaaatagga tgttaacaac actgaactta gcaggcactt aataaatggt gcttaaacat 18780
[0582] ctataataaa aaattttttt caaatgacgt tgtttaggag tgcatttttc atcatggcca 18840
[0583] gttttcaaag aaggaactgt ggactttttc ttatttctgt ctttctgccc tcttgccttt 18900
[0584] ttcatgatac taaggaatga acgtgagacc ttgcatatgc caggcaggca agagctgtca 18960
[0585] cactgagtta catctctagc ttccaacccc catctctctc tgtcagtctt tctgccaagg 19020
[0586] gggtcaaatg aagaaaaagg tattcctggg caagatgtgg caatgcatgc cattagcctc 19080
[0587] agcactcagg atgcagaggc aggtggatct ctgaatttga ggccaccttg gtctacacag 19140

[0588] tgagtttcag gacagccagg gctacataga gggactctgt cttaaaaatt aaaggaaaga 19200
[0589] ataaattctt atgtagcctt tccagattct aagatttcag atactacggg gttcaaagac 19260
[0590] cacattgact cagacctgct tttgggtact gaattaaagt tagtcaatag gtaactgctt 19320
[0591] actttggatt cctttcctga gtgtaggact ttagtttgca gatgcttagt tttggtggca 19380
[0592] ggtgaataat tctgcttcag taactgtctt atatgtagtc tctctgtttt ctccagggac 19440
[0593] aatggtgaca gatctaggca ccgagctccc cgaaatgccc gaatgtctgc accttcactt 19500
[0594] ggacgatttg ttccaaggta agaactgggt tgtaaatatt gacggatcct tgatgaatgc 19560
[0595] ctaaggtttc atttttgtgt gtttggatta ctttggcttt atagagctgg aactgggacc 19620
[0596] cagggccttc cacatactag agaggcacac tgcctctgag ctacaccaac ctctggcctt 19680
[0597] tggattaccc atatatagaa ttctctccta gcctctatat tctggtcctg ccatttaaac 19740
[0598] atttcatggg agattattag ccacagtgcc tagagataga aaaatggcac caagagcata 19800
[0599] ctgttcaaag tctgtcagag tctctctttg tagttctagt tggcctgaac tcatcatgta 19860
[0600] gactgggctg gcctggaact catgtatcta cctgctgtgc ctgtgattgc tggggttaaa 19920
[0601] ggcgtgcatt actcctgact aacattactt tagaagaagg tcagtaagga gtcacacatt 19980
[0602] tataaaagga caaatcctat agattttatg gatttttagt tagtatatcc cttaggctat 20040
[0603] attttgttat aacgtatctt ttgggaccca gtctggtctt gaacttgcaa tccttcctta 20100
[0604] gtcttccatc tttataaacc atgccattta tgcccagcct tacttagcca acgattttga 20160
[0605] aactgtagtt tttcctctgt tgtttggtag tatttagttg gctttttaca tgtgattcct 20220
[0606] tcttacaggc gtttttgct gcctgagtac ttgccttatg ctgggatttt tcatgaacgt 20280
[0607] ggacagcctg gcttggctac tcattcttct gttaacaggg tcctggcag 20329
[0608] <210> 9
[0609] <211> 449
[0610] <212> DNA
[0611] <213> SEQ ID NO.9
[0612] <400> 9
[0613] gaggctgctg gagtcttccc tcatttcatt atcccgttat gatggagcag gatccagaga 60
[0614] gcacccaatt tacccagacc cagcgagatt atctcctgct gcatactacg cccagaggat 120
[0615] gatccagtat ctctcaagga gagacagtat tcgccagcgc tccatgcgct accaacagaa 180
[0616] tcggcttcgt tcttccacct cctcttcttc ctcagacaac cagggtccat cagtagaggg 240
[0617] aaccgacttg gaatttgagg actttgagga caatggtgac agatctaggc accgagctcc 300
[0618] ccgaaatgcc cgaatgtctg caccttcact tggacgattt gttccaaggc gtttttgct 360
[0619] gcctgagtac ttgccttatg ctgggatttt tcatgaacgt ggacagcctg gcttggctac 420
[0620] tcattcttct gttaacaggg tcctggcag 449
[0621] <210> 10
[0622] <211> 449
[0623] <212> DNA
[0624] <213> SEQ ID NO.10
[0625] <400> 10
[0626] gaggctgctg gagtcttccc tcatttcatt atcccgttat gatggagcag gatccagaga 60
[0627] gcacccaatt tacccagacc cagcgagatt atctcctgct gcatactacg cccagaggat 120
[0628] gatccagtat ctctcaagga gagacagtat tcgccagcgc tccatgcgct accaacagaa 180
[0629] tcggcttcgt tcttccacct cctcttcttc ctcagacaac cagggtccat cagtagaggg 240

[0630] aaccgacttg gaatttgagg actttgagga caatggtgac agatctaggc accgagctcc 300

[0631] ccgaaatgcc cgaatgtctg caccttcact tggacgattt gttccaaggc gttttttgct 360

[0632] gcctgagtac ttgccttatg ctgggatttt tcatgaacgt ggacagcctg gcttggctac 420

[0633] tcattcttct gttaacaggg tcctggcag 449

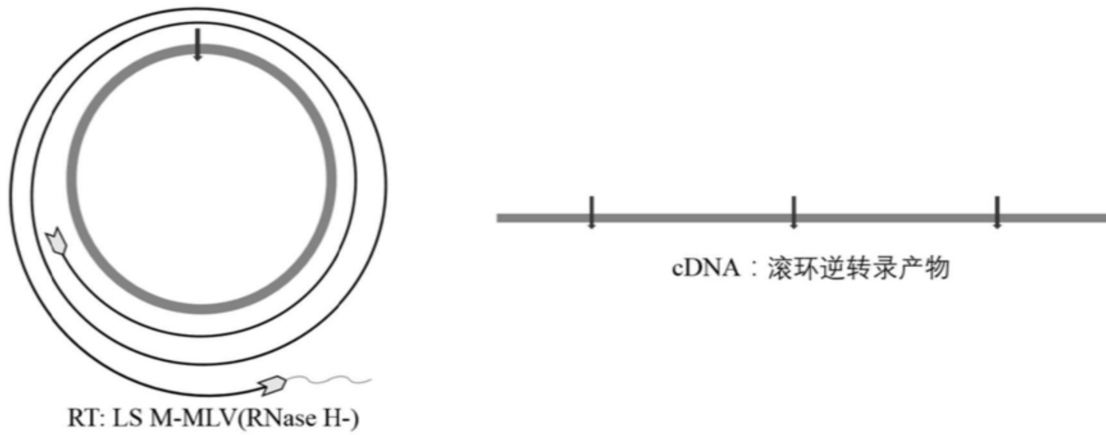


图1

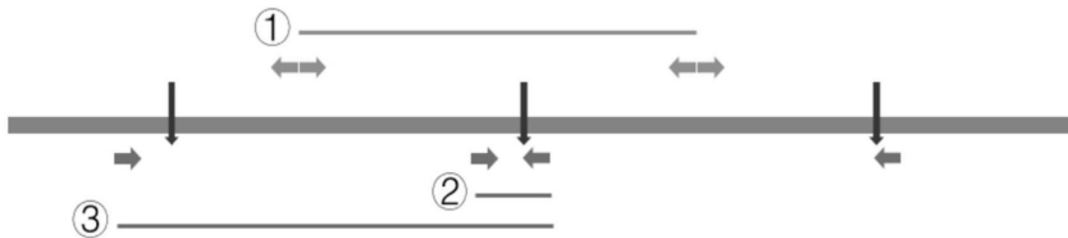


图2

Genomic chr4 :

```
tagtatttaa ttatatcaaa acatttgttt attttctgag tttgtttttt  51863386
tacaatattg gaaataaatg acgctgacat ttttcctttt tctttttcaag 51863436
ATTTTCAGCT GAACTCTCAT CTCTCAACAC TGGCAAATAT TCATAAGATC  51863486
TACCACACCC TTAATAAGCT Ggtaagtcat tcttttaaag acaaaattac  51863536
caactgtttg aagtagtcat tttgtataag tgtatttgat aaaaatttag  51863586

agacttgatt tttttttcca aaaggagtta atggccttta agtcttactt  51886486
gttgacagta taatactact acagtggtaa tagtatggtg tgcatggtca  51886536
gatttaattt acataagact gtatttcaca gGAACTGATG ATGTTGTAGG  51886586
CCCTGAAGGC ATGGAGAAAT TTTGTGAAGA CATTGGTGTT GAACCAGAAA  51886636
AGgtgagtca aacttactga gttgggtgaa tcagttggtt gtttttcata  51886686
cttaaatctt tgttctttag caaataaata aataattaaa aagtagtggt  51886736
```

图3

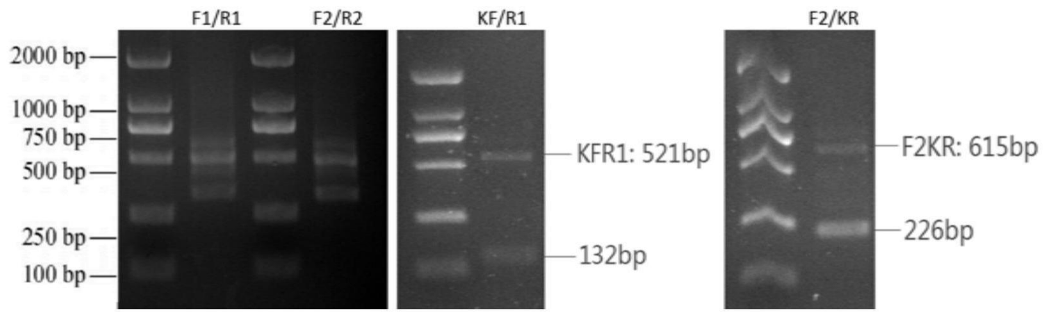


图4

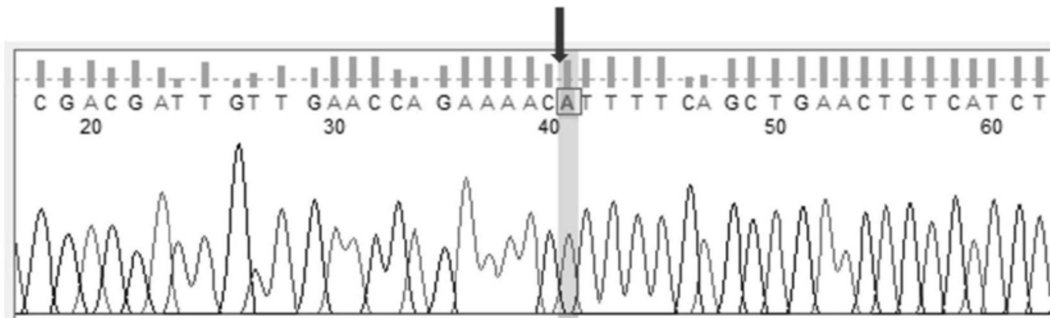


图5

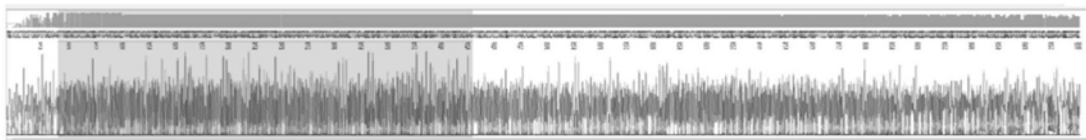


图6

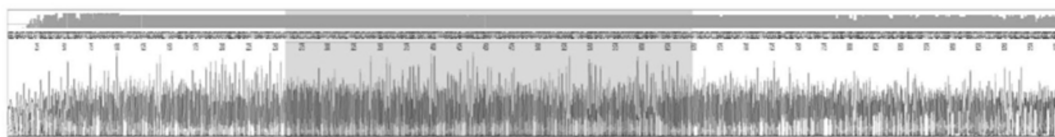


图7

```
   1 --------------------------------------------------------------------------------------------------------------
   1 ttggcacgcctgccgttcgacgattgttgaaccagaaaac----------------------------------------------------------------------
   1 agggcaggaacatgattacgccagtttgcacgcctgccgttcgacgattcctgcctctggagatgatttatctgccaagaaaagtagacatgatagcatgtatagaaaat

   1 --------------------------------------------------------------------------------------------------------------
  41 --------------------------------------------------------------------------------------------------------------
 111 atgattcgactagaataaagactgaagaagaagccttttcaagtaaaaggtgcttggaatggttctatgaatatgcaggaactgatgatgttgtaggccctgaaggcatg

   1 ---------------------------------------attttcagctgaactctcatctctcaacactggcaaatattcataagatctaccacacccttaataagctg
  41 ---------------------------------------attttcagctgaactctcatctctcaacactggcaaatattcataagatctaccacacccttaataagctg
 221 gagaaattttgtgaagacattggtgttgaaccagaaaacattttcagctgaactctcatctctcaacactggcaaatattcataagatctaccacacccttaataagctg

  72 aacctaacagaagacattggccaagacgatcaccaaacaggaagtctgcggtcttgcagttcttcagactgctttaataaagtgatgccaccaaggaaaaagagaagacc
 112 aacctaacagaagacattggccaagacgatcaccaaacaggaagtctgcggtcttgcagttcttcagactgctttaataaagtgatgccaccaaggaaaaagagaagacc
 331 aacctaacagaagacattggccaagacgatcaccaaacaggaagtctgcggtcttgcagttcttcagactgctttaataaagtgatgccaccaaggaaaaagagaagacc

 182 tgcctctggagatgatttatctgccaagaaaagtagacatgatagcatgtatagaaaatatgattcgactagaataaagactgaagaagaagccttttcaagtaaaaggt
 222 tgcctctggagatgatttatctgccaagaaaagtagacatgatagcatgtatagaaaatatgattcgactagaataaagactgaagaagaagccttttcaagtaaaaggt
 441 tgcctctggagatgatttatctgccaagaaaagtagacatgatagcatgtatagaaaatatgattcgactagaataaagactgaagaagaagccttttcaagtaaaaggt

 292 gcttggaatggttctatgaatatgcaggaactgatgatgttgtaggccctgaaggcatggagaaattttgtgaagacattggtgttgaaccagaaaac------------
 332 gcttggaatggttctatgaatatgcaggaactgatgatgttgtaggccctgaaggcatggagaaattttgtgaagacattggtgttgaaccagaaaacattttcagctga
 551 gcttggaatggttctatgaatatgcaggaactgatgatgttgtaggccctgaaggcatggagaaattttgtgaagacattggtgttgaaccagaaaacattttcagctga

 390 --------------------------------------------------------------------------------------------------------------
 442 actctcatctctcaacactggcaaatattcataagatctaccacacccttaataagctgaacctaacagaagacattggccaagacgatcaccaaacaggaagtcaatct
 661 actcaatctctggaagatccgcgcgtaccgagttctaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagc

 390 --------------------------------------------------------------------------------------------------------------
 552 ctggaagatccgcgcgtaccgagttctaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccc
 771 acatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcctgatgcggtattttctcc

 390 --------------------------------------------------------------------------------------------------------------
 662 tttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatggcgaatggcgcctgatgcggtattttctccttacgcatc
 881 ttacgcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcg

 390 --------------------------------------------------------------------------------------------------------------
 772 tgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacgg
 991 ccctgacgggc---------------------------------------------------------------------------------------------------

 390 --------------------------------------------------------------------------------------------------------------
 882 gcttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcct
1002 --------------------------------------------------------------------------------------------------------------

 390 ----------
 992 cgtgatacgc
1002 ----------
```

图8

Genomic chr19 :

```
ctgtttcatc taataattta taccottgaa ttttctaatt tcttctttta  38057506
ttgacttttt tggttgtggg tttttttttt tctttttttt tttttcatag  38057556
ATCTTGAACC AGATGATTGT GCATccattt acatctttaa tgtagacccg  38057606
cctccatcta ctttaaattc atcacttggc ttaccacatc atggactgct  38057656
tcactttgga caatgtgaat tgaaaataga agtgcaacct aaaactcacc  38062119
atagagccca ttatgaaact gaaggtagcc gaggggcagt gaaagcctct  38062169
ACTGGTGGCC ATCCTGTTGT GAAGgtatga gacttgggag cttttcccta  38062219
taaatatata catgttaatg attgaataag tatattgatt tagctaaaag  38062269
```

图9

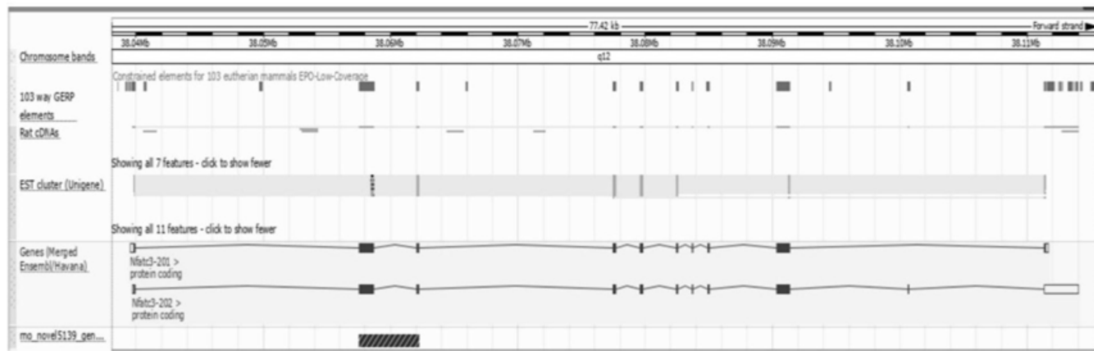


图10

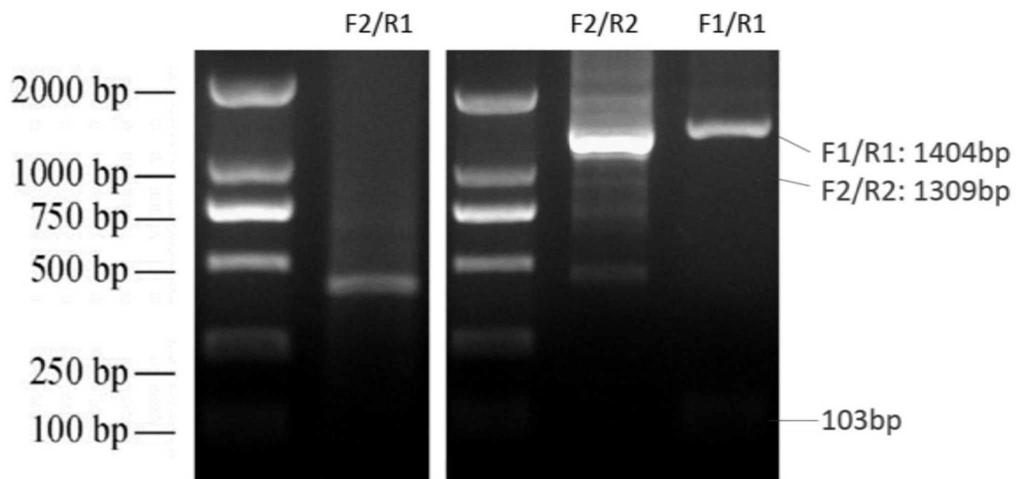


图11

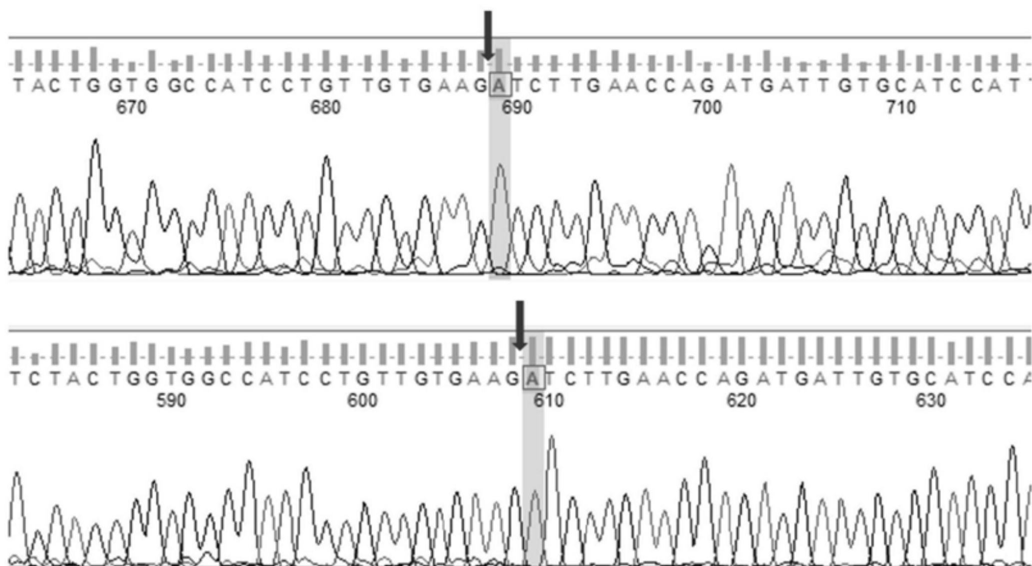


图12

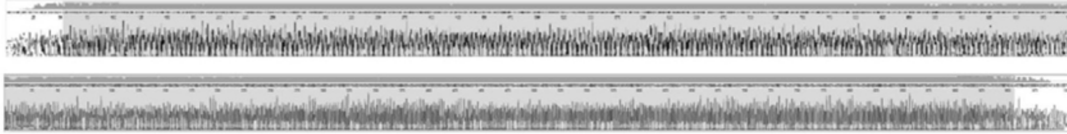


图13

```
   1 tggaaaattggacatggattacgccagtttgcacgcctgccgttcgacgattcatcctgttgtgaagatcttgaaccagatgattgtgcatccatttacatctttaatgtagacccgcct
   1 ------------------------------------------------------------------------------------------------------------------------
   1 -------------------------------------------------------------------atcttgaaccagatgattgtgcatccatttacatctttaatgtagacccgcct

 121 ccatctactttaaattcatcacttggcttaccacatcatggactgctgcagtctcactcttctgttttgtcaccatcatttcagctccaaggtttcaaaaattatgaaggaactgatgat
   1 ------------------------------------------------------------------------------------------------------------------------
  54 ccatctactttaaattcatcacttggcttaccacatcatggactgctgcagtctcactcttctgttttgtcaccatcatttcagctccaaggtttcaaaaattatgaaggaactgatgat

 241 atttctgaatctaaatatagctcattaagtggtcctaaaccctttgaatgcccaagtattcaaattacatccatctctcctaactgtcatcaagaaacagatgctcatgaagatgaccta
   1 ------------------------------------------------------------------------------------------------------------------------
 174 atttctgaatctaaatatagctcattaagtggtcctaaaccctttgaatgcccaagtattcaaattacatccatctctcctaactgtcatcaagaaacagatgctcatgaagatgaccta

 361 catgtaaatgacccagaaagggaatatttggaaaggccttctagagatcatctctatctcccacttgaaccgtcctaccgggaatcttcccttagtcctagtcctgccagcagcgtttct
   1 ------------------------------------------------------------------------------------------------------------------------
 294 catgtaaatgacccagaaagggaatatttggaaaggccttctagagatcatctctatctcccacttgaaccgtcctaccgggaatcttcccttagtcctagtcctgccagcagcgtttct

 481 tctaggagttggttctcagatgcatcttcttgtgagtctctctcacacatttatgacgatgtggactcagagttgaatgaagctgctgcacgatttactcttggctcacctctgacttct
   1 ---------------------gcatcttcttgtgagtctctctcacacatttatgacgatgtggactcagagttgaatgaagctgctgcacgatttactcttggctcacctctgacttct
 414 tctaggagttggttctcagatgcatcttcttgtgagtctctctcacacatttatgacgatgtggactcagagttgaatgaagctgctgcacgatttactcttggctcacctctgacttct

 601 ccaggtggctctccaggaggttgccctggagaagagtcctggcatcaacagtatggacctggacattccttgtcacctaggcaatctccttgccactctcctagatccagtatcactgat
 100 ccaggtggctctccaggaggttgccctggagaagagtcctggcatcaacagtatggacctggacattccttgtcacctaggcaatctccttgccactctcctagatccagtatcactgat
 534 ccaggtggctctccaggaggttgccctggagaagagtcctggcatcaacagtatggacctggacattccttgtcacctaggcaatctccttgccactctcctagatccagtatcactgat

 721 gagaattggctgagccccagaccagcctcaggaccctcatccaggcccacttctccttgtggtaaacgacggcactccagtgctgaagtatgttatgctggctctctttcaccccatcac
 220 gagaattggctgagccccagaccagcctcaggaccctcatccaggcccacttctccttgtggtaaacgacggcactccagtgctgaagtatgttatgctggctctctttcaccccatcac
 654 gagaattggctgagccccagaccagcctcaggaccctcatccaggcccacttctccttgtggtaaacgacggcactccagtgctgaagtatgttatgctggctctctttcaccccatcac

 841 tcacctgttccgtcccctggtcactcgcctagagggagtgtaactgaagatacctggctcactgctcctgtccacactggatcaggcctcagccctgcaccgtttccatttcagtactgt
 340 tcacctgttccgtcccctggtcactcgcctagagggagtgtaactgaagatacctggctcactgctcctgtccacactggatcaggcctcagccctgcaccgtttccatttcagtactgt
 774 tcacctgttccgtcccctggtcactcgcctagagggagtgtaactgaagatacctggctcactgctcctgtccacactggatcaggcctcagccctgcaccgtttccatttcagtactgt

 961 gtagagactgacatccctttgaaaacaaggaagacttctga-------------------------------------------------------------------------------
 460 gtagagactgacatccctttgaaaacaaggaagacttctgacgatcaagctgccatactaccaggaaaattagaggtctgttcagatgatcaagggagcttatccccatcccgggagaca
 894 gtagagactgacatccctttgaaaacaaggaagacttctgacgatcaagctgccatactaccaggaaaattagaggtctgttcagatgatcaagggagcttatccccatcccgggagaca

1002 ------------------------------------------------------------------------------------------------------------------------
 580 tcagtagatgatggccttggatctcagtatcctttaaagaaagattcatctggtgaccaatttctttcagttccttcaccttttacctggagcaaaccaaagcctggccacactcctata
1014 tcagtagatgatggccttggatctcagtatcctttaaagaaagattcatctggtgaccaatttctttcagttccttcaccttttacctggagcaaaccaaagcctggccacactcctata

1002 ------------------------------------------------------------------------------------------------------------------------
 700 tttcgcacatcttcattacctccattagactggcctttaccaactcactttggacaatgtgaattgaaaatagaagtgcaacctaaaactcaccatagagcccattatgaaactgaaggt
1134 tttcgcacatcttcattacctccattagactggcctttaccaactcactttggacaatgtgaattgaaaatagaagtgcaacctaaaactcaccatagagcccattatgaaactgaaggt

1002 ------------------------------------------------------------------------------------------------------------------------
 820 agccgaggggcagtgaaagcctctactggtggccatcctgttgtgaagatcttgaaccagatgattgtgcatccatttacatctttaatgtagacccgcccccatctactttaaattcat
1254 agccgaggggcagtgaaagcctctactggtggccatcctgttgtgaag------------------------------------------------------------------------

1002 --------------------------------------------------------------
 940 cacttggcttaccacaaatctctggaagatccgcgcgtaccgagttctaatcactgccgttt
1302 --------------------------------------------------------------
```

图14

Genomic chr3 :

```
gcgaaggggg atcttattta cctctccccc atttctctta tgactgagct  80713687
ctgagttgga aattgacagg gattattgcc atctttttaa tctctgctag  80713737
GAGGCTGCTG GAGTCTTCCC TCATTTCATT ATCCCGTTAT GATGGAGCAG  80713787
GATCCAGAGA GCACCCAATT TACCCAGACC CAGCGAGgta aggccaaaat  80713837
tacttagcca acaattttga aactctgtag tagctccttt gttgtttggt  80732787
aatctttagt tgactttcac atgtgattcc ttcttacagG CGTTTTTTGC  80732837
TGCCTGAGTA CTTGCCTTAT GCTGGGATTT TTCATGAgCG TGGACAGCCT  80732887
GGCTTGGCTA CTCATTCTTC TGTTAACAGG GTCCTGGCAG gtaaggaagc  80732937
```

图15

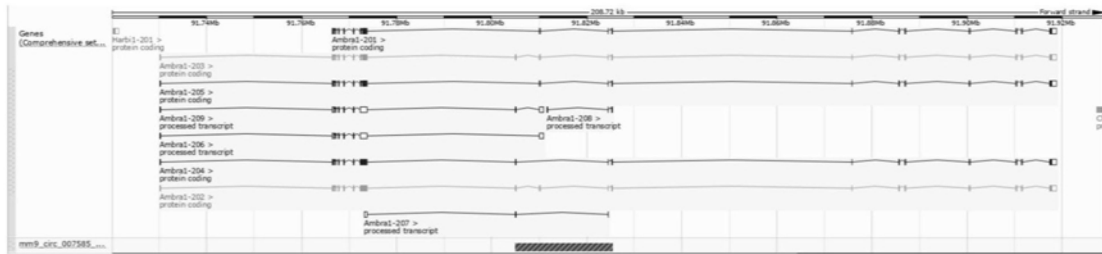


图16

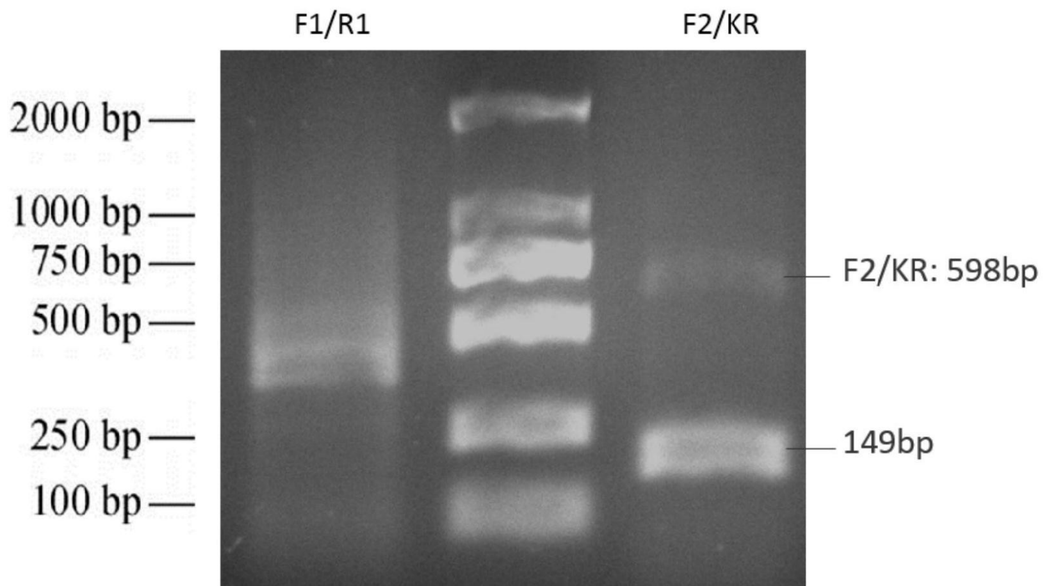


图17

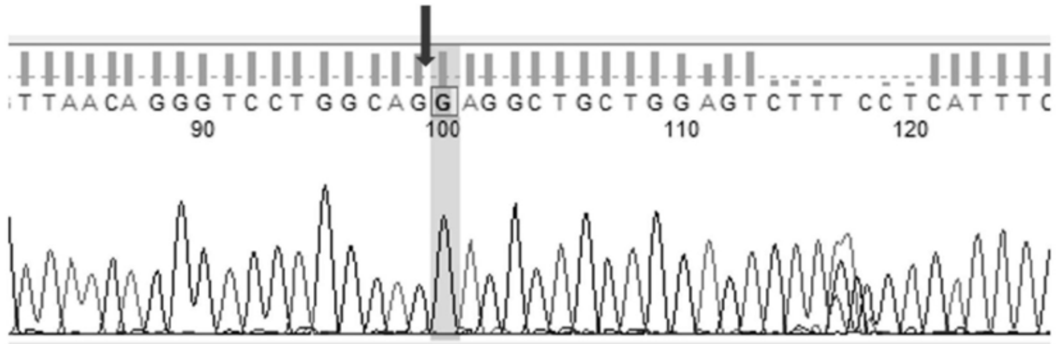


图18

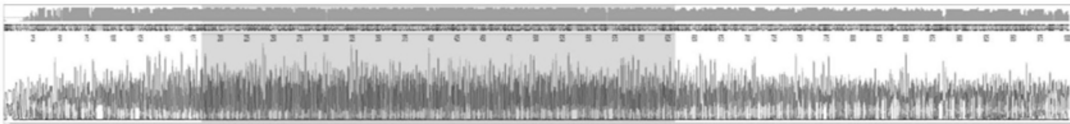


图19

```
   1 ----------------------------------------------------------------------------------------------------
   1 ggggcatgcgtagtgattagactcggtacgcgcggatcttccagagattctgcaccttcacttggacgatttgttccaaggcgtttttttgctgcctgag
   1 ggcgtcttgctgctgagtacttgccttatgctgggatttttcatgaacgtggacagcctggcttggctactcattcttctgttaacagggtcctggcag-

   1 ----------------------------------------------------------------------------------gaggctgctggagtcttc
 101 tacttgccttatgctgggatttttcatgaacgtggacagcctggcttggctactcattcttctgttaacagggtcctggcaggaggctgctggagtcttc
 100 ----------------------------------------------------------------------------------gaggctgctggagtcttt

  19 cctcatttcattatcccgttatgatggagcaggatccagagagcacccaatttacccagacccagcgagattatctcctgctgcatactacgcccagagg
 201 cctcatttcattatcccgttatgatggagcaggatccagagagcacccaatttacccagacccagcgagattatctcctgctgcatactacgcccagagg
 118 cctcatttcattatcccgttatgatggagcaggatccagagagcacccaatttacccagacccagcgagattatctcctgctgcatactacgcccagagg

 119 atgatccagtatctctcaaggagagacagtattcgccagcgctccatgcgctaccaacagaatcggcttcgttcttccacctcctcttcttcctcagaca
 301 atgatccagtatctctcaaggagagacagtattcgccagcgctccatgcgctaccaacagaatcggcttcgttcttccacctcctcttcttcctcagaca
 218 atgatccagtatctctcaaggagagacagtattcgccagcgctccatgcgctaccaacagaatcggcttcgttcttccacctcctcttcttcctcagaca

 219 accagggtccatcagtagagggaaccgacttggaatttgaggactttgaggacaatggtgacagatctaggcaccgagctccccgaaatgcccgaatgtc
 401 accagggtccatcagtagagggaaccgacttggaatttgaggactttgaggacaatggtgacagatctaggcaccgagctccccgaaatgcccgaatgtc
 318 accagggtccatcagtagagggaaccgacttggaatttgaggactttgaggacaatggtgacagatctaggcaccgagctccccgaaatgcccgaatgtc

 319 tgcaccttcacttggacgatttgttccaaggcgtttttttgctgcctgagtacttgccttatgctgggatttttcatgaacgtggacagcctggcttggct
 501 tgcaccttcacttggacgatttgttccaaggcgtttttttgctgcctgagtacttgccttatgctgggatttttcatgaacgtggacagcctggcttggct
 418 tgcaccttcacttggacgatttgttccaaggcgtttttttgctgcctgagtacttgccttatgctgggatttttcatgaacgtggacagcctggcttggct

 419 actcattcttctgttaacagggtcctggcag---------------------------------------------------------------------
 601 actcattcttctgttaacagggtcctggcaggaggctgctggagtcttaatcgtcgaacggcaggcgtgcaaacttggcgtaatcatggtcatagctgtt
 518 actcattcttctgttaacagggtcctggcaggaggctgctgggagtcttaacatccgggtttttcggggagctcggagcctacatctgtcaccattgtcc

 450 ----------------------------------------------------------------------------------------------------
 701 tcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcaca
 618 tcaaagtcctcaaattccaagtcggatccctctactggatggaccctggttgtctgagagaaaaaaaggaggtggaaaaaacgagcctattctggtggta

 450 ----------------------------------------------------------------------------------------------------
 801 ttaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgta
 718 gcgcatggagcgctggataatactgtctctccttgagagatactggatctatcatctgggcgtagtatgcagcagagagataatctcgctgagttctgag

 450 ----------------------------------------------------------------------------------------------------
 901 ttgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatc
 818 gtaaattgaggtgctctctggatcctgctcgatcataacgtgataaagaaaatgaggatagactcacgcagcctcgtgccaggaccttgataacagaaga

 450 ----------------------------------------------------------------------------------------------------
1001 c---------------------------------------------------------------------------------------------------
 918 tgagtagccagcaggctgtccacgttcatgaaacattccaagcataagcaagtactcagcagcaaaaaacggcatggaacaaatcgtcagtgaaaggtgc

 450 -------
1002 -------
1018 agactcg
```

图20