



(12)发明专利申请

(10)申请公布号 CN 110951742 A

(43)申请公布日 2020.04.03

(21)申请号 201911405281.8

A01H 5/00(2018.01)

(22)申请日 2019.12.31

A01H 6/46(2018.01)

(71)申请人 北京市农林科学院

地址 100097 北京市海淀区曙光花园中路9
号北京市农林科学院玉米研究中心

(72)发明人 杨进孝 徐雯 康桂婷 李璐
赵思

(74)专利代理机构 北京纪凯知识产权代理有限公司 11245

代理人 关畅

(51)Int.Cl.

C12N 15/113(2010.01)

C12N 15/82(2006.01)

C12N 15/65(2006.01)

C12N 9/22(2006.01)

权利要求书2页 说明书10页
序列表18页 附图3页

(54)发明名称

一种不产生DNA双链断裂的实现植物基因替换的方法

(57)摘要

本发明提供了一种不产生DNA双链断裂的实现基因替换的方法。所述方法包括如下步骤:将sgRNA、Cas9切刻酶、供体DNA导入目的植物中;sgRNA靶向DNA片段甲靶点序列;供体DNA依次包括DNA片段甲靶点序列、DNA片段乙和DNA片段甲靶点序列;DNA片段乙为将DNA片段甲经过一个或几个碱基突变后得到的DNA分子;在sgRNA引导下,Cas9切刻酶在目的植物基因组中的DNA片段甲靶点序列处和所述供体DNA中的DNA片段甲靶点序列处均产生单链DNA切刻,并通过目的植物体内的修复机制将目的植物基因组中的DNA片段甲替换为DNA片段乙,实现植物基因替换。

1. 一种不产生DNA双链断裂的实现植物基因替换的方法,包括如下步骤:将sgRNA、Cas9切刻酶或其变体、供体DNA导入目的植物中;

所述sgRNA靶向DNA片段甲靶点序列;

所述供体DNA依次包括所述DNA片段甲靶点序列、DNA片段乙、所述DNA片段甲靶点序列;

所述DNA片段乙为将所述DNA片段甲经过一个或几个碱基突变后得到的DNA分子;

在所述sgRNA引导下,所述Cas9切刻酶或其变体在目的植物基因组中的DNA片段甲靶点序列处和所述供体DNA中的DNA片段甲靶点序列处均产生单链DNA切刻,并通过目的植物体内的修复机制将目的植物基因组中的所述DNA片段甲替换为所述DNA片段乙,实现植物基因替换。

2. 根据权利要求1所述的方法,其特征在于:

所述DNA片段甲和所述DNA片段乙的大小均为200-2000bp;

和/或,所述DNA片段甲为序列5第1300-1993位所示的DNA分子;

和/或,所述DNA片段乙为序列1第10695-11388位所示的DNA分子。

3. 根据权利要求1或2所述的方法,其特征在于:所述sgRNA结构如下:tRNA-所述DNA片段甲靶点序列转录的RNA-sgRNA骨架;

和/或,所述tRNA为a1) 或a2) 或a3):

a1) 将序列1第474-550位中的T替换为U得到的RNA分子;

a2) 将a1) 所示的RNA分子经过一个或几个核苷酸的取代和/或缺失和/或添加且具有相同功能的RNA分子;

a3) 与a1) 或a2) 限定的核苷酸序列具有75%或75%以上同一性且具有相同功能的RNA分子;

和/或,所述sgRNA骨架为b1) 或b2) 或b3):

b1) 将序列1第571-646位中的T替换为U得到的RNA分子;

b2) 将b1) 所示的RNA分子经过一个或几个核苷酸的取代和/或缺失和/或添加且具有相同功能的RNA分子;

b3) 与b1) 或b2) 限定的核苷酸序列具有75%或75%以上同一性且具有相同功能的RNA分子。

4. 根据权利要求1-3任一所述的方法,其特征在于:

所述Cas9切刻酶为Cas9 D10A切刻酶或Cas9 H840A切刻酶;

和/或,所述Cas9 D10A切刻酶为SpCas9n蛋白质;

和/或,所述SpCas9n蛋白质为C1) 或C2):

C1) 氨基酸序列是序列3所示的蛋白质;

C2) 将序列表中序列3所示的氨基酸序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有相同功能的蛋白质;

和/或,所述SpCas9n蛋白质的编码基因为c1) 或c2) 或c3):

c1) 序列表中序列1第2877-6977位所示的cDNA分子或DNA分子;

c2) 与c1) 限定的核苷酸序列具有75%或75%以上同一性,且编码所述SpCas9n的cDNA分子或DNA分子;

c3) 在严格条件下与c1) 或c2) 限定的核苷酸序列杂交,且编码所述SpCas9n的cDNA分子

或DNA分子。

5. 根据权利要求1-4任一所述的方法,其特征在于:

所述Cas9切刻酶或其变体带有核定位信号;

和/或,所述核定位信号为BP NLS、VirD2 NLS或SV40 NLS;

和/或,所述核定位信号为SV40 NLS;

和/或,所述SV40 NLS的氨基酸序列为序列2。

6. 根据权利要求1-5任一所述的方法,其特征在于:所述DNA片段靶点序列为序列1第551-570位;

和/或,所述供体DNA的序列为序列1第10672-11411位。

7. 根据权利要求1-6所述的方法,其特征在于:所述将sgRNA、Cas9切刻酶或其变体、供体DNA导入目的植物中的方法包括如下步骤:将转录sgRNA的DNA分子、Cas9切刻酶或其变体的编码基因和供体DNA导入目的植物中;

和/或,所述转录sgRNA的DNA分子、所述Cas9切刻酶或其变体的编码基因和所述供体DNA通过重组表达载体导入目的植物中;

和/或,所述重组表达载体包括依次由启动子、所述转录sgRNA的DNA分子和终止子组成的表达盒和依次由启动子、所述Cas9切刻酶或其变体的编码基因和终止子组成的表达盒;

和/或,所述重组表达载体的核苷酸序列如序列1所示。

8. 权利要求1-7任一所述的方法在植物基因编辑中的应用;

或,权利要求1-7任一所述的方法在制备植物突变体中的应用;

或,权利要求1-7任一所述的方法在提高植物基因替换效率中的应用;

或,权利要求1-7任一所述的方法在减少植物基因替换产生的副产物中的应用。

9. 一种植物基因编辑的方法,包括如下步骤:按照权利要求1-7任一所述的方法对植物基因组中的目的基因片段进行替换,从而实现植物基因编辑;

或,一种制备植物突变体的方法,包括如下步骤:按照权利要求1-7任一所述的方法对植物基因组中的目的基因片段进行替换,获得植物突变体;

或,一种提高植物基因替换效率的方法,包括如下步骤:按照权利要求1-7任一所述的方法对植物基因组中的目的基因片段进行替换;

或,一种减少植物基因替换产生的副产物的方法,包括如下步骤:按照权利要求1-7任一所述的方法对植物基因组中的目的基因片段进行替换。

10. 根据权利要求1-7任一所述的方法或权利要求8所述的应用或权利要求9所述的方法,其特征在于:所述植物为如下d1)-d3)中的任一种:

d1) 单子叶植物或双子叶植物;

d2) 禾本科植物;

d3) 水稻。

一种不产生DNA双链断裂的实现植物基因替换的方法

技术领域

[0001] 本发明属于生物技术领域,具体涉及一种不产生DNA双链断裂的实现植物基因替换的方法。

背景技术

[0002] 长链DNA模板介导的基因精确替换在细胞中的概率非常低,但是在待替换位点附近引入DNA双链断裂(dsDNA break,DSB)能够明显的增加替换的概率。CRISPR-Cas9技术已经成为强有力的基因组编辑手段,被广泛应用到很多组织和细胞中。CRISPR/Cas9 protein-RNA复合物通过向导RNA(guide RNA)定位于靶点上,通过切割使DNA产生DSB,从而增加长链DNA模板介导的基因精确替换效率。在产生DSB后,生物体会本能的启动DNA修复机制。修复机制一般有两种,一种是非同源末端连接(non-homologous end joining,NHEJ),此种修复所占的比例占大多数,修复DNA后一般会产生随机的indels(insertions or deletions)。另一种是同源重组(homology-directed repair,HDR),使用姐妹染色单体或外源DNA供体(donor)作为修复模板,实现基因的精确修复。在动物细胞中,修复的具体原理是:CtIP酶在DSB处起始末端切割,产生突出的3'端单链DNA(single-stranded DNA,ssDNA)尾巴,ssDNA会被重组酶Rad51识别,结合成复合物,侵入供体DNA模板,并与其同源的片段退火结合,以供体DNA作为模板,合成新的DNA链,完成修复。当供体DNA同源臂之间的序列带有外源突变时,此突变会在修复的过程中引入到DNA链中,从而实现精确定点替换。此种HDR修复起始于DSB的产生,由于NHEJ的修复概率远大于HDR,在发生了精确替换的样品中,会有较多的副产物,比如引入indels,造成DNA大片段缺失、移位等。

[0003] 为了提高产物中精确HDR:非精确HDR的比例,人们尝试使用Cas9的一种失活突变体D10A,对DNA造成单链切刻(Nick)。在动物中,相比于DSB而言,单链切刻起始的HDR的副产物较少,但同时一定程度上降低了HDR的效率。在植物中,DSB诱发的HDR在不同的基因上能够实现精确替换,但是Cas9 D10A引发的切刻能否实现HDR精确替换,其效率是否比DSB诱导的HDR低,副产物是否减少,均无报道。

发明内容

[0004] 本发明的目的是提供一种不产生DNA双链断裂的实现植物基因替换的方法。

[0005] 本发明提供的不产生DNA双链断裂的实现植物基因替换的方法包括如下步骤:将sgRNA、Cas9切刻酶(Cas9n)或其变体、供体DNA导入目的植物中;

[0006] 所述sgRNA靶向DNA片段甲靶点序列;

[0007] 所述供体DNA依次包括所述DNA片段甲靶点序列、DNA片段乙和所述DNA片段甲靶点序列;

[0008] 所述DNA片段乙为将所述DNA片段甲经过一个或几个碱基突变后得到的DNA分子;

[0009] 在所述sgRNA引导下,所述Cas9切刻酶或其变体在目的植物基因组中的DNA片段甲靶点序列处和所述供体DNA中的DNA片段甲靶点序列处均产生单链DNA切刻,并通过目的植

物体内的修复机制将目的植物基因组中的所述DNA片段甲替换为所述DNA片段乙,实现植物基因替换。

[0010] 上述方法中,所述DNA片段甲可为目的植物基因组上的任意片段,将DNA片段甲上的某一个或某几个碱基突变后获得DNA片段乙,所述DNA片段乙为供体DNA上的片段。所述碱基突变可为碱基替换和/或碱基插入和/或碱基缺失。

[0011] 在实际应用中,将sgRNA/Cas9n系统和两端添加有靶点对应的靶点序列的DNA片段乙即供体DNA导入目的植物后,可实现将目的植物基因组中的DNA片段甲替换为DNA片段乙,进而实现基因替换。此处基因替换可实现将基因突变位点引入目的植物基因组中,进而实现目的植物基因组上的基因突变(如碱基替换、碱基插入或碱基缺失),从而使目的植物中表达的相应蛋白质的氨基酸功能位点和/或种类和/或活性和/或含量等发生改变,获得具有某一功能或性状的植物突变体。在本发明的具体实施例中,所述碱基突变具体可为碱基替换。

[0012] 进一步的,所述DNA片段甲和所述DNA片段乙的大小均可为200-2000bp或200-1500bp或200-1000bp。

[0013] 更进一步的,所述DNA片段甲和所述DNA片段乙的大小均为694bp。

[0014] 所述DNA片段甲为序列5第1300-1993位所示的DNA分子。

[0015] 所述DNA片段乙为序列1第10695-11388位所示的DNA分子。

[0016] 在本发明的具体实施例中,所述DNA片段甲依次由大小为636bp的ALS基因片段和其下游大小为58bp的片段组成。所述DNA片段乙为将DNA片段甲第344位由碱基G突变为碱基T,且将第581位由碱基G突变为碱基T,且将第336位由碱基G突变为碱基C,且将第339位由碱基G突变为碱基C,且将第342位由碱基A突变为碱基G,且将第396位由碱基G突变为碱基C后得到的DNA分子。将水稻基因组中的DNA片段甲替换为DNA片段乙后,由于水稻基因组中的DNA片段甲第344位由碱基G突变为碱基T,且第581位由碱基G突变为碱基T,水稻中表达的ALS蛋白氨基酸序列第548位的氨基酸由酪氨酸(Try)突变为亮氨酸(Leu),第627位的氨基酸由丝氨酸(Ser)突变为异亮氨酸(Ile),从而产生具有除草剂抗性的精确编辑植株。

[0017] 上述方法中,所述sgRNA结构如下:tRNA-所述DNA片段甲靶点序列转录的RNA-sgRNA骨架;

[0018] 所述tRNA为a1)或a2)或a3):

[0019] a1) 将序列1第474-550位中的T替换为U得到的RNA分子;

[0020] a2) 将a1)所示的RNA分子经过一个或几个核苷酸的取代和/或缺失和/或添加且具有相同功能的RNA分子;

[0021] a3) 与a1)或a2)限定的核苷酸序列具有75%或75%以上同一性且具有相同功能的RNA分子;

[0022] 所述sgRNA骨架为b1)或b2)或b3):

[0023] b1) 将序列1第571-646位中的T替换为U得到的RNA分子;

[0024] b2) 将b1)所示的RNA分子经过一个或几个核苷酸的取代和/或缺失和/或添加且具有相同功能的RNA分子;

[0025] b3) 与b1)或b2)限定的核苷酸序列具有75%或75%以上同一性且具有相同功能的RNA分子。

[0026] 上述方法中,所述供体DNA依次由所述DNA片段甲靶点序列、PAM序列、DNA片段乙和所述DNA片段甲靶点序列和PAM序列组成。所述DNA片段甲靶点序列具体为ST215靶点序列;所述ST215靶点序列为序列1第551-570位。所述供体DNA具体为序列1第10672-11411位。

[0027] 上述方法中,所述Cas9切刻酶可为Cas9 D10A切刻酶或Cas9 H840A切刻酶;

[0028] 所述Cas9切刻酶变体包括来源于细菌的Cas9(如SaCas9、SaCas9-KKH等),识别不同PAM的Cas9变体(如xCas9、Cas9-NG、Cas9-VQR、Cas9-VRER等),Cas9高保真酶变体(如HypaCas9、eSpCas9(1.1)、Cas9-HF1等)等。

[0029] 进一步的,所述Cas9 D10A切刻酶为SpCas9n蛋白质;

[0030] 所述SpCas9n蛋白质为C1)或C2):

[0031] C1)氨基酸序列是序列3所示的蛋白质;

[0032] C2)将序列表中序列3所示的氨基酸序列经过一个或几个氨基酸残基的取代和/或缺失和/或添加且具有相同功能的蛋白质;

[0033] 更进一步的,所述SpCas9n蛋白质的编码基因为c1)或c2)或c3):

[0034] c1)序列表中序列1第2877-6977位所示的cDNA分子或DNA分子;

[0035] c2)与c1)限定的核苷酸序列具有75%或75%以上同一性,且编码所述SpCas9n的cDNA分子或DNA分子;

[0036] c3)在严格条件下与c1)或c2)限定的核苷酸序列杂交,且编码所述SpCas9n的cDNA分子或DNA分子。

[0037] 上述方法中,所述Cas9切刻酶或其变体带有核定位信号。所述核定位信号可为BP NLS、VirD2 NLS或SV40 NLS。所述核定位信号的个数可为1个或2个或多个。

[0038] 进一步的,所述核定位信号为SV40 NLS。所述SV40 NLS的氨基酸序列为序列2。所述核定位信号的个数为8个。

[0039] 更进一步的,所述SV40 NLS的编码序列为序列1第2742-2762位。所述Cas9切刻酶或其变体两端分别带有4个SV40 NLS。

[0040] 上述方法中,所述将sgRNA、Cas9切刻酶或其变体、供体DNA导入目的植物中的方法包括如下步骤:将转录sgRNA的DNA分子、Cas9切刻酶或其变体的编码基因和供体DNA导入目的植物中。

[0041] 上述方法中,所述sgRNA为tRNA-sgRNA,转录所述tRNA-sgRNA的DNA分子转录后得到的所述tRNA-sgRNA为不成熟的RNA前体,该RNA前体中的tRNA会被两种酶(RNase P和RNaseZ)切割掉后得到成熟的RNA。一个重组表达载体中有多少个靶点,就会得到多少个独立的成熟的RNA,每个成熟的RNA依次由所述靶点序列转录的RNA和所述sgRNA骨架组成,或依次由所述靶点序列转录的RNA、所述sgRNA骨架和所述tRNA残留的个别碱基组成。在本发明的具体实施例中,所述重组表达载体中含有一个靶点。

[0042] 进一步的,所述转录sgRNA的DNA分子、所述Cas9切刻酶或其变体的编码基因和所述供体DNA通过重组表达载体导入目的植物中。所述转录sgRNA的DNA分子、所述Cas9切刻酶或其变体的编码基因和所述供体DNA可通过同一个重组表达载体导入目的植物中,也可通过两个或者多个重组表达载体共同导入目的植物中。

[0043] 在本发明的具体实施例中,所述转录sgRNA的DNA分子、所述Cas9切刻酶或其变体的编码基因和所述供体DNA通过同一个重组表达载体导入目的植物中。所述重组表达载体

包括依次由启动子、所述转录sgRNA的DNA分子和终止子组成的表达盒和依次由启动子、所述Cas9切刻酶或其变体的编码基因和终止子组成的表达盒。

[0044] 更进一步的,所述重组表达载体的核苷酸序列如序列1所示。

[0045] 上述方法在植物基因编辑或制备植物突变体或提高植物基因替换效率或减少植物基因替换产生的副产物中的应用也属于本发明的保护范围。

[0046] 本发明最后还提供了方法一或方法二或方法三或方法四:

[0047] 所述方法一为一种植物基因编辑的方法;所述植物基因编辑的方法包括如下步骤:按照上述方法对植物基因组中的目的基因片段进行替换,从而实现植物基因编辑。所述编辑具体可为碱基替换。

[0048] 所述方法二为一种制备植物突变体的方法;所述制备植物突变体的方法包括如下步骤:按照上述方法对植物基因组中的目的基因片段进行替换,获得植物突变体。所述植物突变体具体可为抗除草剂突变体。

[0049] 所述方法三为一种提高植物基因替换效率的方法;所述提高植物基因替换效率的方法包括如下步骤:按照上述方法对植物基因组中的目的基因片段进行替换。所述替换效率具体可为HDR替换效率。

[0050] 所述方法四为一种减少植物基因替换产生的副产物的方法;所述减少植物基因替换产生的副产物的方法包括如下步骤:按照上述方法对植物基因组中的目的基因片段进行替换。所述减少植物基因替换产生的副产物具体体现在按照上述方法对植物基因组中的目的基因片段进行替换得到的产物没有额外的Indels产生。

[0051] 上述方法或应用中,所述植物为如下d1)-d3)中的任一种:

[0052] d1) 单子叶植物或双子叶植物;

[0053] d2) 禾本科植物;

[0054] d3) 水稻(如日本晴)。

[0055] 本发明提供的不产生DNA双链断裂的实现基因替换的方法的原理如下:Cas9 D10A切刻酶在sgRNA的引导下,在基因组ALS靶位点引发单链DNA切刻;载体上的供体DNA含有具有除草剂抗性的突变位点,同时供体DNA的5'端和3'端各含有1个靶点对应的靶点序列,在Cas9 D10A/sgrNA复合体的引导下,各产生一个切刻位点;在植物体内的修复机制下,实现供体DNA与基因组片段的精确替换,从而产生具有除草剂抗性的精确编辑植株。

[0056] 本发明具有以下优点:

[0057] 1、效率高:单链切刻(Nick)引发的HDR精确替换效率是DSB引发的HDR精确替换效率的2-3倍。

[0058] 2、副产物少:单链切刻(Nick)引发的HDR精确替换植株中不含有随机的Indels,而DSB引发的HDR精确替换含有大量的indels。

[0059] 3、成本低:本发明提供的不产生DNA双链断裂的实现基因替换的方法仅需要载体中含有Cas9n相关元件,以及对应的供体DNA,通过农杆菌侵染的方法即可实现精确替换。

[0060] 本发明提供了一种不产生DNA双链断裂的实现基因替换的方法。通过实验证明:本发明提供的不产生DNA双链断裂的实现基因替换的方法在水稻中实现了对内源的乙酰乳酸合酶ALS基因的精确替换,获得具有除草剂抗性的精确编辑植株。

附图说明

- [0061] 图1为精确替换载体构建示意图。
- [0062] 图2为Nick引发的ALS精确替换工作原理示意图。
- [0063] 图3为DSB引发的ALS精确替换工作原理示意图。
- [0064] 图4为Nick引发的ALS精确替换产生的植株,特异引物扩增检测结果。
- [0065] 图5为DSB引发的ALS精确替换产生的植株,特异引物扩增检测结果。
- [0066] 图6为Nick和DSB对应的精确替换具体形式。
- [0067] 图7为精确替换载体构建示意图。
- [0068] 图8为sgRNA/Cas9n常规载体转基因植株,特异引物扩增检测结果。
- [0069] 图9为esgRNA/Cas9n-P2A-Hpt优化载体转基因植株,特异引物扩增检测结果。

具体实施方式

- [0070] 以下的实施例便于更好地理解本发明,但并不限定本发明。
- [0071] 下述实施例中的实验方法,如无特殊说明,均为常规方法。下述实施例中所用的试验材料,如无特殊说明,均为自常规生化试剂商店购买得到的。下述实施例中,如无特殊说明,序列表中各核苷酸序列的第1位均为相应DNA/RNA的5'末端核苷酸,末位均为相应DNA/RNA的3'末端核苷酸。
- [0072] 引物对P1由引物HDR-F:5'-gcgcccgattctctatgtc-3'和引物HDR-R:5'-acctatcctccaactggacg-3'组成,用于检测植株是否发生精确替换。
- [0073] 引物对P2由引物gALS-F:5'-atcccagttacaaccactctg-3'和引物gALS-R:5'-cacttaactcagagctattgcatag-3'组成,用于扩增基因组ALS序列并测序。
- [0074] HDR苗是指在获得的T₀苗中,含有donor带入的对应突变位点的植株。
- [0075] 基于T₀苗的HDR替换效率=T₀苗中发生了精确替换的苗数/获得的T₀苗总数。
- [0076] 基于愈伤数量的HDR替换效率=T₀苗中发生了精确替换的苗数/起始侵染的愈伤总数。
- [0077] HDR中的indels效率=精确替换的T₀苗中发生了indels的苗数/精确替换的T₀苗总数。
- [0078] 日本晴水稻:参考文献:梁卫红,王高华,杜京尧,等. 硝普钠及其光解产物对日本晴水稻幼苗生长和5种激素标记基因表达的影响[J]. 河南师范大学学报(自然版), 2017(2):48-52.; 公众可以从北京市农林科学院获得。
- [0079] 恢复培养基:含有200mg/L特美汀的N6固体培养基。
- [0080] 筛选培养基1:含有50mg/L潮霉素的N6固体培养基。
- [0081] 筛选培养基2:含有0.4uM/L双草醚的N6固体培养基。
- [0082] 分化培养基:含有2mg/L KT、0.2mg/L NAA、0.5g/L谷氨酸、0.5g/L脯氨酸的N6固体培养基。
- [0083] 生根培养基:含有0.2mg/L NAA、0.5g/L谷氨酸、0.5g/L脯氨酸、0.28uM/L双草醚的N6固体培养基。
- [0084] 下述实施例中ALS蛋白的氨基酸序列如序列表中序列6所示,其编码基因序列如序列表中序列5第1-1935位所示。

[0085] 实施例1、不产生DNA双链断裂实现基因替换的载体的构建及其在水稻基因替换中的应用

[0086] 一、重组表达载体的构建及替换原理说明

[0087] 1、重组表达载体的构建

[0088] 人工合成如下重组表达载体,各载体均为环状质粒:sgRNA/Cas9重组表达载体、sgRNA/Cas9n重组表达载体。sgRNA/Cas9重组表达载体、sgRNA/Cas9n重组表达载体的结构示意图均如图1所示。具体结构描述分别如下:

[0089] sgRNA/Cas9n重组表达载体的序列为列表中的序列1。序列1的第131-467位为0sU3启动子序列,第474-550位为tRNA序列,第551-570位为ST215靶点序列,第571-646位为sgRNA骨架序列,第647-937位为0sU3终止子序列。序列1的第944-2657位为0sUbq3启动子序列,第2742-2762位、第2772-2792位、第2802-2822位、第2832-2852位、第6996-7016位、第7026-7046位、第7056-7076位、第7086-7106位均为核定位信号SV40的编码序列(编码序列2所示的核定位信号SV40),第2877-6977位为SpCas9n蛋白质的编码序列(编码序列3所示的SpCas9n蛋白质);序列1的第7122-7376位为Nos终止子序列;序列1的第7405-9397位为ZmUbi1启动子序列,第9404-10429位为潮霉素磷酸转移酶的编码序列,第10456-10671位为CaMV35S终止子序列。第10672-10694位、第11389-11411位均为ST215靶点靶序列(由ST215靶点序列和PAM序列组成);第10695-11388位为ALS供体DNA序列。

[0090] sgRNA/Cas9重组表达载体的序列为将序列1中的第2877-6977位替换为序列4所示的SpCas9蛋白质的编码序列,且保持其他序列不变后得到的序列。

[0091] 2、重组表达载体的精确替换原理

[0092] 1) sgRNA/Cas9n重组表达载体的精确替换原理

[0093] sgRNA/Cas9n重组表达载体是基于单链切刻(Nick)引导的精确替换,原理示意图如图2所示。

[0094] sgRNA/Cas9n重组表达载体包括如下元件:sgRNA、Cas9n、供体DNA。

[0095] sgRNA靶向ST215靶点靶序列。

[0096] 供体DNA(donor DNA):供体DNA依次由ST215靶点靶序列、ALS供体DNA序列和ST215靶点靶序列组成。

[0097] ALS供体DNA序列(序列1第10695-11388位)是将DNA片段甲(DNA片段甲为序列5第1300-1993位所示的大小为694bp的片段,其依次由大小为636bp的ALS基因片段和其下游大小为58bp的片段组成。DNA片段甲为水稻基因组上的片段。)进行突变后得到的DNA分子。突变包括功能位点突变和同义位点突变。

[0098] 功能位点突变:将DNA片段甲第344位由碱基G突变为碱基T(此处碱基突变可使水稻中的ALS蛋白氨基酸序列第548位的酪氨酸(Try)突变为亮氨酸(Leu)),且将第581位由碱基G突变为碱基T(此处碱基突变可使水稻中的ALS蛋白氨基酸序列第627位的丝氨酸(Ser)突变为异亮氨酸(Ile))。水稻中的ALS蛋白氨基酸序列第548位的酪氨酸(Try)突变为亮氨酸(Leu),且第627位的丝氨酸(Ser)突变为异亮氨酸(Ile)后能够抗双草醚除草剂,将水稻ALS蛋白氨基酸序列第548位记作W548L功能突变位点,将水稻ALS蛋白氨基酸序列第627位记作S627I功能突变位点。

[0099] 同义位点突变:为了方便后期设计特异检测引物检测精确替换突变体,将DNA片段

甲第336位由碱基G突变为碱基C(此处碱基突变对应水稻ALS蛋白氨基酸序列第545位),且将第339位由碱基G突变为碱基C(此处碱基突变对应水稻ALS蛋白氨基酸序列第546位),且将第342位由碱基A突变为碱基G(此处碱基突变对应水稻ALS蛋白氨基酸序列第547位),且将第396位由碱基G突变为碱基C(此处碱基突变对应水稻ALS蛋白氨基酸序列第565位)。此处同义位点突变不改变水稻ALS蛋白氨基酸序列上的相应氨基酸位点所对应的氨基酸,将水稻ALS蛋白氨基酸序列第545位记作545同义突变位点,将水稻ALS蛋白氨基酸序列第546位记作546同义突变位点,将水稻ALS蛋白氨基酸序列第547位记作547同义突变位点,将水稻ALS蛋白氨基酸序列第565位记作P565同义突变位点。

[0100] Cas9n/sgRNA复合物在ST215靶点的引导下,在载体中供体donor上的靶点序列产生两个单链切刻位点,同时在水稻基因组DNA片段甲上的靶点序列产生一个单链切刻位点,将此方案记作triNicks(三个Nick)。在水稻体内的修复机制下,供体DNA在水稻基因组上的切刻位点处发生精确替换(水稻基因组上的DNA片段甲替换为ALS供体DNA),使ALS供体DNA上的突变位点引入到水稻基因组中,获得基因替换后的植株,该基因替换后的植株具有除草剂抗性。

[0101] 2) sgRNA/Cas9重组表达载体的精确替换原理

[0102] sgRNA/Cas9重组表达载体是基于DSB引导的精确替换,原理示意图如图3所示。

[0103] sgRNA/Cas9重组表达载体包括如下元件:sgRNA、Cas9、供体DNA。

[0104] sgRNA靶向ST215靶点靶序列。

[0105] 供体DNA(donor DNA):供体DNA依次由ST215靶点靶序列、ALS供体DNA序列和ST215靶点靶序列。

[0106] Cas9/sgRNA复合物在ST215靶点的引导下,在载体中供体donor上的靶点序列产生两个DSB位点,同时在水稻基因组上的靶点序列产生一个DSB位点,将此方案记作triDSBs(三个DSB)。在水稻体内的修复机制下,供体DNA在水稻基因组上的DSB位点处发生精确替换(水稻基因组上的DNA片段甲替换为ALS供体DNA),使ALS供体DNA上的突变位点引入到水稻基因组中,获得基因替换后的植株,该基因替换后的植株具有除草剂抗性。

[0107] 二、水稻阳性抗性愈伤的获得

[0108] 将步骤一获得的sgRNA/Cas9n、sgRNA/Cas9重组表达载体分别按照如下步骤1-7进行操作:

[0109] 1、将载体导入农杆菌EHA105(上海唯地生物技术有限公司的产品,CAT#:AC1010),得到重组农杆菌。

[0110] 2、采用培养基(含50 μ g/ml卡那霉素和25 μ g/ml利福平的YEP培养基)培养重组农杆菌,28 $^{\circ}$ C,150rpm震荡培养至OD₆₀₀为1.0-2.0,室温条件下,10000rpm离心1min,用侵染液(将N6液体培养基中的糖替换为葡萄糖和蔗糖,葡萄糖和蔗糖在侵染液中的浓度分别为10g/L和20g/L)重悬菌体并稀释至OD₆₀₀为0.2,得到农杆菌侵染液。

[0111] 3、水稻品种日本晴成熟种子去壳脱粒,置于100mL三角瓶中,加入70%(v/v)乙醇水溶液浸泡30sec,再置于25%(v/v)次氯酸钠水溶液中,120rpm震荡灭菌30min,无菌水冲洗3次,用滤纸吸干水分,然后将种子胚朝下置于N6固体培养基上,28 $^{\circ}$ C暗培养4-6周,得到水稻愈伤。

[0112] 4、完成步骤3后,将水稻愈伤浸泡置于农杆菌侵染液甲(农杆菌侵染液甲为向农杆

菌侵染液中加入乙酰丁香酮得到的液体,乙酰丁香酮的添加量满足乙酰丁香酮与农杆菌侵染液的体积比为25 μ l:50ml)中浸泡10min,然后,放在铺有两层灭菌滤纸的培养皿(内含约200ml不含农杆菌的侵染液)上,21 $^{\circ}$ C暗培养1天。

[0113] 5、取步骤4得到的水稻愈伤放入恢复培养基上,25-28 $^{\circ}$ C暗培养3天。

[0114] 6、取步骤5得到的水稻愈伤,置于筛选培养基1上,28 $^{\circ}$ C暗培养2周。

[0115] 7、取步骤6得到的水稻愈伤,转移至筛选培养基2上,28 $^{\circ}$ C暗培养4周,得到水稻抗性愈伤。

[0116] 三、水稻T0苗的获得

[0117] 1、取步骤1得到的水稻抗性愈伤放入分化培养基上,25 $^{\circ}$ C光照培养1个月左右。

[0118] 2、将分化出来的小苗移至生根培养基上,25 $^{\circ}$ C光照培养2周,获取抗除草剂的水稻T0苗。

[0119] 四、精确替换植株鉴定

[0120] 1、经过含有除草剂的生根培养基筛选后,存活下来的水稻T0苗,分别提取基因组DNA并以其作为模板,采用引物HDR-F(5'-gcgcccgattctctatgtc-3')和引物HDR-R(5'-acctatctctccaactggacg-3')组成的引物对进行PCR扩增,得到PCR扩增产物;将该PCR扩增产物进行琼脂糖凝胶电泳,然后进行如下判断:如果PCR扩增产物中含有约833bp的DNA片段,则相应的水稻T0苗为发生了精确替换的阳性T0苗;如果PCR扩增产物中不含有约833bp的DNA片段,则相应的水稻T0苗为未发生精确替换的T0苗。

[0121] 2、使用引物gALS-F(5'-atcccagttacaaccactctg-3')和引物gALS-R(5'-cacttaactcagagctattgcatag-3')组成的引物对,对上述步骤1筛选出的精确替换的阳性T0苗的基因组ALS基因序列进行PCR扩增,得到PCR扩增产物;将该PCR扩增产物进行一代测序,分析对应位点是否确实发生了精确替换。

[0122] 五、结果分析

[0123] 1、水稻T0苗中引物初筛获得的精确替换植株

[0124] triNicks方案共获得32棵转基因阳性苗(独立转化事件),经过含有除草剂的生根培养基筛选后,32棵苗全部存活。经过HDR-F和HDR-R引物筛选后,其中有5棵转基因苗为精确替换植株,PCR检测结果如图4所示。

[0125] triDSBs方案共获得23棵转基因阳性苗(独立转化事件),经过含有除草剂的生根培养基筛选后,仅9棵苗存活。经过HDR-F和HDR-R引物筛选后,有2棵转基因苗为精确替换植株,PCR检测结果如图5所示。

[0126] 2、水稻T0苗中测序确认的精确替换植株及对应效率

[0127] 一代测序结果显示:PCR检测结果为阳性的植株均在对应位点上发生了精确替换。因此,triNicks和triDSBs方案在T0苗中发生精确替换的概率分别是15.6%(5/32)、8.7%(2/23)。若从整体转化效率来看,以起始侵染的840块抗性愈伤为分母计算,triNicks和triDSBs方案发生精确替换的概率分别是0.6%(5/840)、0.2%(2/840)。

[0128] 综上所述,triNicks方案在T0苗中的精确替换效率是triDSBs的1.8倍,若以起始侵染的愈伤为基础,triNicks方案的精确替换效率是triDSBs的3倍。说明Nick诱导的triNicks方案的精确替换效率明显高于DSB诱导的triDSBs方案。

[0129] 另外,将发生了精确替换的样品的PCR扩增产物连接pEASY-B(北京全式金生物),

每个样品挑选4-6个阳性克隆测序,确认替换的具体形式,具体测序结果见图6所示。triNicks方案中的5棵T0苗在基因组中均为杂合替换突变体,在基因组中未引发额外的Indels产生,而triDSBs方案中的2棵T0苗均在基因组ST215靶点处均产生了一个碱基G的插入。说明Nick诱导的triNicks方案在基因组上产生的副产物少于DSB诱导的triDSBs方案。

[0130] 对triNicks方案中发生精确替换的样品进行进一步的分析,发现不同的样品在W548、P565、S627处的替换形式不相同,有的是三个位点同时发生替换(突变形式1),有的是在W548、P565处发生替换(突变形式2),有的在W548、S627处发生替换(突变形式3),有的仅在W548处发生替换(突变形式4)。各突变形式对应的效率如表1所示。

[0131] 表1

		W548位点	P565位点	S627位点	样品数	各突变形式在总HDR样品中的相对比例 (%)
	triDSBs 突变形式4	C..C..G..T	-	-	2	100% (2/2)
[0132]	突变形式1	C..C..G..T	CCC	ATT	1	20% (1/5)
	triNicks 突变形式2	C..C..G..T	CCC	-	2	40% (2/5)
	突变形式3	C..C..G..T	-	ATT	1	20% (1/5)
	突变形式4	C..C..G..T	-	-	1	20% (1/5)

[0133] 实施例2、sgRNA/Cas9n重组表达载体的优化及其在水稻基因替换中的应用

[0134] 一、重组表达载体的构建及替换原理说明

[0135] 1、重组表达载体的构建

[0136] 人工合成如下重组表达载体,各载体均为环状质粒:sgRNA/Cas9n重组表达载体、esgRNA/Cas9n-P2A-Hpt重组表达载体。sgRNA/Cas9n重组表达载体、esgRNA/Cas9n-P2A-Hpt重组表达载体结构示意图如图7所示。具体结构描述分别如下:

[0137] sgRNA/Cas9n重组表达载体的序列为列表中的序列1。

[0138] esgRNA/Cas9n-P2A-Hpt重组表达载体的序列为将序列1中的第571-646位替换为序列7所示的esgRNA骨架序列,且将序列1的第7107-10671位替换为序列8所示的P2A-Hpt-tNos序列,且保持其他序列不变后得到的序列。序列8第22-78位为自切割寡肽P2A的编码序列(编码序列9所示的自切割寡肽P2A),第79-1104位为潮霉素磷酸转移酶的编码序列(编码序列10所示的潮霉素磷酸转移酶),第1111-1365位为Nos终止子序列。

[0139] 2、重组表达载体的精确替换原理说明

[0140] 1) sgRNA/Cas9n重组表达载体的精确替换原理

[0141] 同实施例1中的sgRNA/Cas9n重组表达载体的精确替换原理。

[0142] 2) esgRNA/Cas9n-P2A-Hpt重组表达载体的精确替换原理

[0143] esgRNA/Cas9n-P2A-Hpt重组表达载体的精确替换原理与sgRNA/Cas9n重组表达载体的精确替换原理相同,仅是将sgRNA/Cas9n重组表达载体中的sgRNA骨架替换为esgRNA骨架序列,且将sgRNA/Cas9n重组表达载体中独立的Cas9n编码基因表达盒和独立的潮霉素磷酸转移酶基因表达盒替换为由自切割寡肽P2A连接的Cas9n和潮霉素磷酸转移酶共表达的表达盒。

[0144] 二、水稻阳性抗性愈伤的获得

[0145] 将步骤一获得的sgRNA/Cas9n、esgRNA/Cas9n-P2A-Hpt载体分别按照实施例1步骤二中的方法进行的操作。

[0146] 三、水稻T0苗的获得

[0147] 同实施例1的步骤三。

[0148] 四、精确替换植株鉴定

[0149] 同实施例1的步骤四。

[0150] 五、结果分析

[0151] 1、水稻T0苗中引物初筛获得的精确替换植株

[0152] sgRNA/Cas9n普通载体,共获得32棵转基因阳性苗(独立转化事件),经过含有除草剂的生根培养基筛选后,32棵苗均存活。经过HDR-F和HDR-R引物筛选后,有5棵转基因苗为精确替换植株,PCR检测结果如图8所示。

[0153] esgRNA/Cas9n-P2A-Hpt优化载体,共获得55棵转基因阳性苗(独立转化事件),经过含有除草剂的生根培养基筛选后,24棵苗存活。经过HDR-F和HDR-R引物筛选后,有12棵转基因苗为精确替换植株,PCR检测结果如图9所示。

[0154] 2、水稻T0苗中测序确认的精确替换植株及对应效率

[0155] 一代测序结果显示:PCR检测结果为阳性的植株均在对应位点上发生了精确替换,且无额外的Indels碱基发生。因此,sgRNA/Cas9n普通载体、esgRNA/Cas9n-P2A-Hpt优化载体在T0苗中发生精确替换的概率分别是15.6% (5/32)、21.8% (12/55)。若从整体转化效率来看,以起始侵染的840块抗性愈伤为分母计算,sgRNA/Cas9n普通载体、esgRNA/Cas9n-P2A-Hpt优化载体发生精确替换的概率分别是0.6% (5/840)、1.4% (12/840)。

[0156] 综上所述,esgRNA/Cas9n-P2A-Hpt优化载体在T0苗中的精确替换效率是sgRNA/Cas9n普通载体的1.4倍,若以起始侵染的愈伤为基础,esgRNA/Cas9n-P2A-Hpt优化载体的精确替换效率是sgRNA/Cas9n普通载体的2.3倍。说明esgRNA/Cas9n-P2A-Hpt优化载体比sgRNA/Cas9n普通载体的精确替换效率高。

[0157] 表2

载体	侵染愈伤粒数	T0苗总数	BS阳性苗数	精确替换苗数	T0苗精确替换效率 (%)	基于愈伤数量的精确替换效率 (%)
sgRNA/Cas9n	840	32	32	5	15.6	0.6
esgRNA/Cas9n-T2A-Hpt	840	55	24	12	21.8	1.4

[0159] 以上所述仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明技术原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

序列表

<110>北京市农林科学院

<120>一种不产生DNA双链断裂的实现基因替换的方法

<160>10

<170>PatentIn version 3.5

<210>1

<211>11586

<212>DNA

<213>Artificial Sequence

<400>1

```

ggtggcagga tatattgtgg tgtaaacatg gactagcct caccgtcttc gcagacgagg 60
ccgctaagtc gcagctacgc tctcaacggc actgactagg tagtttaaac gtgcacttaa 120
ttaaggtacc gaagcaactt aaagttatca ggcatgcatg gatcttggag gaatcagatg 180
tgcagtcagg gaccatagca caagacaggc gtcttctact ggtgctacca gcaaatgctg 240
gaagccggga aacttgggta cgttggaaac cacgtgatgt gaagaagtaa gataaactgt 300
aggagaaaag catttcgtag tgggccatga agcctttcag gacatgtatt gcagtatggg 360
ccggccatt acgcaattgg acgacaacaa agactagtat tagtaccacc tcggctatcc 420
acatagatca aagctgattt aaaagagttg tgcagatgat ccgtggcggg tccaacaaag 480
caccagtggc ctagtggtag aatagtacc tgccacggta cagaccggg ttcgattccc 540
ggctgggtgca atttgggtat ggtgggtgcaa gtttagagc tagaaatagc aagttaaaat 600
aaggctagtc cgttatcaac ttgaaaaagt ggcaccgagt cgggtgctttt tttttcgtt 660
ttgcattgag ttttctccgt cgcattgttg cagttttatt ttccgttttg cattgaaatt 720
tctccgtctc atgtttgcag cgtgttcaaa aagtacgcag ctgtatttca cttatttacg 780
gcgccacatt tcatgccgt ttgtgccaac tatcccgagc tagtgaatac agcttggctt 840
cacacaacac tggtgaccg ctagactgct cgtacctcgt accgtcgtac ggcacagcat 900
ttggaattaa aggggtgat cgatactgct tgctgctaag cttacaaatt cgggtcaagg 960
cggaagccag cgcgccacc cagtcagca aatacggagg cgcggggttg acggcgtcac 1020
ccggtcctaa cggcgacca caaaccagcc agaagaaatt acagtaaaaa aaaagtaaat 1080
tgcactttga tccacctttt attacctaag tctcaatttg gatcaccctt aaacctatct 1140
tttcaatttg ggccgggttg tggtttgac taccatgaac aacttttcgt catgtctaac 1200
ttccctttca gcaaacatat gaaccatata tagaggagat cggccgtata ctagagctga 1260
tgtgtttaag gtcgttgatt gcacgagaaa aaaaaatcca aatcgcaaca atagcaaatt 1320
tatctggttc aaagtgaaaa gatatgttta aaggtagtcc aaagtaaac ttatagataa 1380
taaaatgtgg tccaaagcgt aattcaactc aaaaaatca acgagacgtg taccaaacgg 1440
agacaaacgg catcttctcg aaatttccca accgctcgtc cccccctc gtcttcccgg 1500
aaaccgcggt ggtttcagcg tggcggatc tccaagcaga cggagacgtc acggcacggg 1560
actcctccca ccaccaacc gccataaata ccagccccct catctctct cctcgcatac 1620
gctccacccc cgaaaaattt ctccccaatc tcgcgaggct ctcgtcgtcg aatcgaatcc 1680

```

tctcgcgtcc tcaaggtacg ctgcttctcc tctcctcget tcgtttcgat tcgatttcgg 1740
acgggtgagg ttgttttggt gctagatccg attggtggtt agggttgtcg atgtgattat 1800
cgtgagatgt ttaggggttg tagatctgat gtttgtgatt tgggcacggt tggttcgata 1860
gggtggaatcg tggttaggtt ttgggattgg atgttggttc tgatgattgg ggggaatfff 1920
tacggttaga tgaattgttg gatgattcga ttggggaaat cgggtgtagat ctgttgggga 1980
attgtggaac tagtcatgcc tgagtgattg gtgcgatttg tagcgtgttc catctttag 2040
gccttgttgc gagcatgttc agatctactg ttccgctctt gattgagtta ttgggtccat 2100
gggttgggtgc aaacacaggc ttaatatgt tatactctgt ttgtgttga ttagatctg 2160
taggtagtt cttcttagac atggttcaat tatgtagctt gtgcgtttcg atttgatttc 2220
atatgttcac agattagata atgatgaact ctttaatta attgtcaatg gtaaatagga 2280
agtcttgcg ctatatctgt cataatgatc tcatgttact atctgccagt aatttatgct 2340
aagaactata ttagaatatc atgttacaat ctgtagtaat atcatgttac aatctgtagt 2400
tcatctatat aatctattgt ggtaatttct tttactatc tgtgtgaaga ttattgccac 2460
tagttcattc tacttatttc tgaagttcag gatacgtgtg ctgttactac ctatctgaat 2520
acatgtgtga tgtgctgtt actatctttt tgaatacatg tatgttctgt tggaaatagt 2580
ttgctgtttg atccgttgtt gtgtccttaa tcttgtgcta gttcttacc tatctgtttg 2640
gtgattatft cttgcagtac gtaagcatgg actacaagga ccacgacggg gattacaaag 2700
accacgacat agactacaag gatgacgatg acaaatggc accgaagaaa aaaaggaagg 2760
tcggcggctc cccgaagaaa aaaaggaagg tcggcggctc cccgaagaaa aaaaggaagg 2820
tcggcggctc cccgaagaaa aaaaggaagg tcggaatcca tggcgttcca gctgccgaca 2880
agaagtactc catcggcctc gccatcggca ccaacagcgt cggctgggcg gtgatcaccg 2940
acgagtacaa ggtcccgtcc aagaagtca aggtcctggg caacaccgac cgccactcca 3000
tcaagaagaa cctcatcggc gccctcctct tcgactccgg cgagacggcg gaggcgacc 3060
gcctcaagcg caccgcccgc cgccgtaca cccgccgcaa gaaccgcatc tgctacctcc 3120
aggagatctt ctccaacgag atggcgaagg tcgacgactc cttcttccac cgcctcgagg 3180
agtcttctc cgtggaggag gacaagaagc acgagcgcca cccatcttc ggcaacatcg 3240
tcgacgaggt cgcctaccac gagaagtacc cactatcta ccacttctgt aagaagcttg 3300
ttgactctac tgataaggct gatcttctg tcactctacc tgctctcgct cacatgatca 3360
agttccgtgg tcacttctt atcgagggtg accttaacc tgataactcc gacgtggaca 3420
agctcttcat ccagctcgtc cagacctaca accagctctt cgaggagaac cctatcaacg 3480
cttccgggtg cgacgctaag gcgatcttt ccgctaggt ctccaagtcc aggcgtctcg 3540
agaacctcat cgcccagctc cctggtgaga agaagaacgg tcttttcggt aacctcatcg 3600
ctctctccct cggctctgacc cctaacttca agtccaactt cgacctcgtc gaggacgcta 3660
agcttcagct ctccaaggat acctacgacg atgatctcga caacctctc gctcagattg 3720
gagatcagta cgctgatctc ttcttctgt ctaagaacct ctccgatgct atctctcttt 3780
cggatatact tagggttaac actgagatca ctaaggctcc tctttctgct tccatgatca 3840
agcgtacga cgagcaccac caggacctca cctctctcaa ggctcttgtt cgtcagcagc 3900
tccccgagaa gtacaaggag atcttcttcg accagtcaa gaacggctac gccggttaca 3960
ttgacgggtg agctagccag gaggagtct acaagttcat caagccaatc cttgagaaga 4020

tggatggtac tgaggagctt ctcgttaagc ttaaccgtga ggacctcctt aggaagcaga 4080
ggactttcga taacggctct atccctcacc agatccacct tggtagctt cacgccatcc 4140
ttcgtaggca ggaggacttc taccctttcc tcaaggacaa ccgtgagaag atcgagaaga 4200
tccttacttt ccgtattcct tactacgttg gtctctttgc tcgtggtaac tcccgtttcg 4260
cttggatgac taggaagtcc gaggagacta tcacccttg gaacttcgag gaggttggtg 4320
acaaggggtc ttccgcccag tccttcatcg agcgcacgac caacttcgac aagaacctcc 4380
ccaacgagaa ggtcctcccc aagcactccc tcctctacga gtacttcacg gtctacaacg 4440
agctcaccaa ggtcaagtac gtcaccgagg gtatgacgaa gcctgccttc ctctccggcg 4500
agcagaagaa ggctatcggt gacctctct tcaagaccaa ccgcaaggtc accgtcaagc 4560
agctcaagga ggactacttc aagaagatcg agtgcttcca ctccgtcgag atcagcggcg 4620
ttgaggaccg tttcaacgct tctctcggtc cctaccacga tctctcaag atcatcaagg 4680
acaaggactt cctcgacaac gaggagaacg aggacatct cgaggacatc gtctcactc 4740
ttactctctt cgaggatagg gagatgatcg aggagaggct caagacttac gctcatctct 4800
tcgatgacaa ggttatgaag cagctcaagc gtcgccgta caccggttg gtaggctct 4860
cccgaagct catcaacggt atcagggata agcagagcgg caagactatc ctgacttcc 4920
tcaagtctga tggtttcgct aacaggaact tcatgcagct catccacgat gactctctta 4980
ccttcaagga ggatattcag aaggctcagg tgtccggtca gggcgactct ctccacgagc 5040
acattgctaa ccttgctggt tcccctgcta tcaagaagg catccttcag actgttaagg 5100
ttgtcgatga gcttgtaac gttatgggtc gtcacaagcc tgagaacatc gtcacgaga 5160
tggctcgtga gaaccagact acccagaagg gtcagaagaa ctcgaggag cgcataaga 5220
ggattgagga gggatcaag gagcttggtt ctcatctct taaggagcac cctgtcgaga 5280
acaccagct ccagaacgag aagctctacc tctactacct ccagaacggt aggatatgt 5340
acgttgacca ggagctcgac atcaacaggc tttctgacta cgacgtcgac cacattgttc 5400
ctcagtcttt ccttaaggat gactccatcg acaacaagg cctcacgagg tccgacaaga 5460
acaggggtaa gtcggacaac gtccttccg aggaggtgt caagaagatg aagaactact 5520
ggaggcagct tctcaacgct aagctcatta ccagaggaa gttcgacaac ctcacgaagg 5580
ctgagagggg tggcctttcc gagcttgaca aggtggttt catcaagagg cagcttggtg 5640
agacgaggca gattaccaag cacgttgctc agatcctcga ttctaggatg aacaccaagt 5700
acgacgagaa cgacaagctc atccgcgagg tcaagtgat caccctcaag tccaagctcg 5760
tctccgactt ccgcaaggac ttccagttct acaaggtccg cgagatcaac aactaccacc 5820
acgctcacga tgcttacctt aacgctgtcg ttggtaccgc tcttatcaag aagtacccta 5880
agcttgagtc cgagttcgtc tacgggtgact acaaggteta cgacgttctg aagatgatcg 5940
ccaagtccga gcaggagatc ggcaaggcca ccgccaagta cttcttctac tccaacatca 6000
tgaacttctt caagaccgag atcaccctcg ccaacggcga gatccgcaag cgccctctta 6060
tcgagacgaa cggtgagact ggtgagatcg tttgggacaa gggtcgcgac ttcgctactg 6120
ttcgcaaggt cctttctatg cctcaggtta acatcgtcaa gaagaccgag gtccagaccg 6180
gtggcttctc caaggagtct atccttcaa agagaaactc ggacaagctc atcgttagga 6240
agaaggattg ggacctaaag aagtacggtg gtttcgactc ccctactgtc gcctactccg 6300
tcctcgtggt cgccaagggtg gagaagggtg agtcgaagaa gctcaagtcc gtcaaggagc 6360

tcctcggcat caccatcatg gagcgtcct ccttcgagaa gaacccgatc gacttcctcg 6420
 aggccaaggg ctacaaggag gtcaagaagg acctcatcat caagctcccc aagtactctc 6480
 ttttcgagct cgagaacggt cgtaagagga tgctggcttc cgctggtgag ctccagaagg 6540
 gtaacgagct tgctcttctt tccaagtacg tgaacttctt ctacctcgcc tcccactacg 6600
 agaagctcaa gggttcccct gaggataacg agcagaagca gctcttcgtg gagcagcaca 6660
 agcactacct cgacgagatc atcgagcaga tctccgagtt ctccaagcgc gtcactctcg 6720
 ctgacgctaa cctcgacaag gtcctctccg cctacaacaa gcaccgagc aagcccatcc 6780
 gcgagcaggc cgagaacatc atccacctct tcacgctcac gaacctcggc gcccttgctg 6840
 ctttcaagta cttcgacacc accatcgaca ggaagcgta cacgtccacc aaggaggttc 6900
 tcgacgctac tctcatccac cagtccatca ccggctctta cgagactcgt atcgaccttt 6960
 cccagcttgg tggatgatgac gatgacaaaa tggcaccgaa gaaaaaaagg aaggctcggcg 7020
 gctccccgaa gaaaaaaagg aaggctcggcg gctccccgaa gaaaaaaagg aaggctcggcg 7080
 gctccccgaa gaaaaaaagg aaggctcggaa tccatggcta gtcccgatcg ttcaaacatt 7140
 tggcaataaa gtttcttaag attgaatcct gttgccggtc ttgcgatgat tatcatataa 7200
 tttctgttga attacgttaa gcatgtaata attaacatgt aatgcatgac gttatttatg 7260
 aggtggggtt ttatgattag agtcccgcaa ttatacattt aatagcgat agaaaacaaa 7320
 atatagcgcg caaactagga taaattatcg cgcgcggtgt catctatggt actagaggcg 7380
 cgctcgggtg atcgtccgcc taggctgcag tgcagcgtga cccggtcgtg cccctctcta 7440
 gagataatga gcattgcatg tctaagttat aaaaaattac cacatatttt tttgtcaca 7500
 cttgtttgaa gtgcagtta tctatcttta tacatatatt taaactttac tctacgaata 7560
 atataatcta tagtactaca ataatatcag tgttttagag aatcatataa atgaacagtt 7620
 agacatggtc taaaggacaa ttgagtattt tgacaacagg actctacagt tttatctttt 7680
 tagtgtgcat gtgttctctt tttttttg ccaatagctt acctatataa tacttcatcc 7740
 attttattag tacatccatt tagggtttag ggttaatggt ttttatagac taattttttt 7800
 agtacatcta ttttattcta ttttagcctc taaattaaga aactaaaac tctattttag 7860
 tttttttatt taataattta gatataaaat agaataaaat aaagtgacta aaaattaaac 7920
 aaataccctt taagaaatta aaaaaactaa ggaaacattt ttcttgtttc gagtagataa 7980
 tgccagcctg ttaaacgccg tcgacgagtc taacggacac caaccagcga accagcagcg 8040
 tcgctcggg ccaagcgaag cagacggcac ggcattctct tcgctgcctc tggaccctc 8100
 tcgagagttc cgctccaccg ttggacttgc tccgctgtcg gcatccagaa attgctggtc 8160
 ggagcggcag acgtgagccg gcacggcagg cggcctctc ctctctcac ggcaccggca 8220
 gctacggggg attcctttcc caccgtctct tcgcttctcc ttctcgcgc gccgtaataa 8280
 atagacacc cctccacacc ctctttccc aacctcgtgt tgttcggagc gcacacacac 8340
 acaaccagat ctcccccaaa tccaccgctc ggcacctccg cttcaaggta cgccgctcgt 8400
 cctccccccc cccccctctc taccttctct agatcggcgt tccggtccat ggttagggcc 8460
 cggtagttct acttctgttc atgtttgtgt tagatccgtg tttgtgttag atccgtgctg 8520
 ctagcgttcg tacacggatg cgacctgtac gtcagacacg ttctgattgc taacttgcca 8580
 gtgtttctct ttggggaatc ctgggatggc tctagccgtt ccgcagacgg gatcgatttc 8640
 atgatttttt ttgtttcgtt gcatagggtt tggtttgccc ttttcttta tttcaatata 8700

tgccgtgcac ttgtttgtcg ggtcatcttt tcatgctttt ttttgtcttg gttgtgatga 8760
tgtggctctgg ttgggcggtc gttctagatc ggagtagaat tctgtttcaa actacctggt 8820
ggatttatta attttggatc tgtatgtgtg tgccatacat attcatagtt acgaattgaa 8880
gatgatggat ggaaatatcg atctaggata ggtatacatg ttgatgcggg ttttactgat 8940
gcatatacag agatgctttt tgttcgcttg gttgtgatga tgtgggtgtg ttgggcggtc 9000
gttcattcgt tctagatcgg agtagaatac tgtttcaaac tacctggtgt atttattaat 9060
tttggaaactg tatgtgtgtg tcatacatct tcatagttac gagtttaaga tggatggaaa 9120
tatcgatcta ggataggtat acatgttgat gtgggtttta ctgatgcata tacatgatgg 9180
catatgcagc atctattcat atgctctaac cttgagtacc tatctattat aataaacaag 9240
tatgttttat aattattttg atcttgatat acttgatga tggcatatgc agcagctata 9300
tgtggatttt tttagccctg cttcatacag ctatttattt gcttggact gtttcttttg 9360
tcgatgctca ccctgttgtt tgggtttact tctgcaggag ctcatgaaa agcctgaact 9420
caccgcgacg tctgtcgaga agtttctgat cgaaaagttc gacagcgtct ccgacctgat 9480
gcagctctcg gagggcgaag aatctcgtgc tttcagcttc gatgtaggag ggcgtggata 9540
tgtcctcgcg gtaaatagct gcgccgatgg tttctacaaa gatcgttatg tttatcgga 9600
ctttgcatcg gccgcgctcc cgattccgga agtgcttgac attggggagt ttagcgagag 9660
cctgacctat tgcactctcc gccgttcaca ggggtgcacg ttgcaagacc tgcctgaaac 9720
cgaactgccc gctgttctac aaccggtcgc ggaggctatg gatgcgatcg ctgcggccga 9780
tcttagccag acgagcgggt tcggccatt cggaccgcaa ggaatcggtc aatacactac 9840
atggcgtgat ttcatatgcg cgattgctga tccccatgtg tactactggc aaactgtgat 9900
ggacgacacc gtcagtgcgt ccgtcgcgca ggctctcgat gagctgatgc tttgggccga 9960
ggactgcccc gaagtccggc acctcgtgca cgcggatttc ggctccaaca atgtcctgac 10020
ggacaatggc cgcataacag cggtcattga ctggagcgag gcgatgttcg gggattccca 10080
atacgaggtc gccaacatct tcttctggag gccgtggttg gcttgtatgg agcagcagac 10140
gcgctacttc gagcggaggc atccggagct tgcaggatcg ccacgactcc ggcgtatat 10200
gctccgcatt ggtcttgacc aactctatca gagcttggtt gacggcaatt tcgatgatgc 10260
agcttgggcg cagggtcgat gcgacgcaat cgtccgatcc ggagccggga ctgtcgggcg 10320
tacacaaatc gcccgcagaa gcgcgccgt ctggaccgat ggctgtgtag aagtactcgc 10380
cgatagtgga aaccgacgcc ccagcactcg tccgagggca aagaaataga gtagatgccg 10440
accgggatct gtcgatcgac aagctcgagt ttctcataa taatgtgtga gtagttccca 10500
gataagggaa ttagggttcc tatagggttt cgctcatgtg ttgagcatat aagaaacct 10560
tagtatgtat ttgtatttgt aaaatacttc tatcaataaa atttctaatt ctaaaacca 10620
aaatccagta ctaaaatcca gatccccga attaattcgg cgttaattca gatttgggta 10680
tggtggtgca acgggaagag atcccaccgc aatatgcat tcaggtgctg gatgagctga 10740
cgaaaggatga ggcaatcatc gctactggtg ttgggcagca ccagatgtgg gcggcacaat 10800
attacacctc caagcggcca cggcagtggc tgtcttcggc tggctctggc gcaatgggat 10860
ttgggctgcc tgctgcagct ggtgcttctg tggctaacc aggtgtcaca gttgttgata 10920
ttgatgggga tggtagcttc ctcatgaaca ttcaggagct ggcattgatc cgcattgaga 10980
acctccctgt gaaggatgat gtgttgaaca accaacattt gggtatggtc gtccagttgg 11040

aggatagggtt ttacaaggcg aatagggcgc atacatactt gggcaacccc gaatgtgaga 11100
 gcgagatata tccagatddd gtgactattg ctaaggggtt caatattcct gcagtcctgtg 11160
 taacaaagaa gagtgaagtc cgtgccgcca tcaagaagat gctcgagact ccagggccat 11220
 acttgttgga tatcatcgtc ccgcaccagg agcatgtgct gcctatgata ccaattgggg 11280
 gcgcattcaa ggacatgata ctggatgggtg atggcaggac tgtgtattaa tctataatct 11340
 gtatgttggtc aaagcaccag cccggcctat gtttgacctg aatgacccat ttgggtatgg 11400
 tgggtcaacg gcctgcagga cgcgtttaat taagtgcacg cggccgcta cttagtcaag 11460
 agcctcgcac gcgactgtca cgcggccagg atcgctcgtg gagcctcgca atctgtacct 11520
 agtgtttaaa ctatcagtgt ttgacaggat atattggcgg gtaaacctaa gagaaaagag 11580
 cgttta 11586

<210>2

<211>7

<212>PRT

<213>Artificial Sequence

<400>2

Pro Lys Lys Lys Arg Lys Val

1 5

<210>3

<211>1367

<212>PRT

<213>Artificial Sequence

<400>3

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly

1 5 10 15

Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys

20 25 30

Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys

50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr

65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe

85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His

100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His

115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser

130	135	140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met		
145	150	155
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp		
	165	170
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn		
	180	185
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys		
	195	200
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu		
210	215	220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu		
225	230	235
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp		
	245	250
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp		
	260	265
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu		
275	280	285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile		
290	295	300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met		
305	310	315
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala		
	325	330
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp		
	340	345
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln		
355	360	365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly		
370	375	380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys		
385	390	395
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly		
	405	410
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu		
	420	425
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro		
435	440	445

Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
 450 455 460
 Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
 465 470 475 480
 Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
 485 490 495
 Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
 500 505 510
 Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
 515 520 525
 Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
 530 535 540
 Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
 545 550 555 560
 Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
 565 570 575
 Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
 580 585 590
 Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
 595 600 605
 Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
 610 615 620
 Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
 625 630 635 640
 Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
 645 650 655
 Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
 660 665 670
 Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
 675 680 685
 Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
 690 695 700
 Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
 705 710 715 720
 Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
 725 730 735
 Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
 740 745 750
 His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755	760	765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu		
770	775	780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val		
785	790	795
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln		
805	810	815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu		
820	825	830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp		
835	840	845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly		
850	855	860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn		
865	870	875
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe		
885	890	895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys		
900	905	910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys		
915	920	925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu		
930	935	940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys		
945	950	955
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu		
965	970	975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val		
980	985	990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val		
995	1000	1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys		
1010	1015	1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr		
1025	1030	1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn		
1040	1045	1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr		
1055	1060	1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg	1070	1075	1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu	1085	1090	1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg	1100	1105	1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys	1115	1120	1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu	1130	1135	1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser	1145	1150	1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe	1160	1165	1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu	1175	1180	1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe	1190	1195	1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu	1205	1210	1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn	1220	1225	1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro	1235	1240	1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His	1250	1255	1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg	1265	1270	1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr	1280	1285	1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile	1295	1300	1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe	1310	1315	1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr	1325	1330	1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly	1340	1345	1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp			

1355	1360	1365	
<210>4			
<211>4101			
<212>DNA			
<213>Artificial Sequence			
<400>4			
gacaagaagt	actccatcgg	cctcgacatc	ggcaccaaca gcgtcggctg ggcggtgatc 60
accgacgagt	acaaggtccc	gtccaagaag	ttcaaggtcc tgggcaacac cgaccgccac 120
tccatcaaga	agaacctcat	cggcgcctc	ctcttcgact ccggcgagac ggcggaggcg 180
acccgcctca	agcgcaccgc	ccgccgccgc	tacaccgcc gcaagaaccg catctgctac 240
ctccaggaga	tcttctccaa	cgagatggcg	aaggtcgacg actccttctt ccaccgcctc 300
gaggagtctt	tctctgtgga	ggaggacaag	aagcacgagc gccaccccat ctteggcaac 360
atcgtcgacg	aggtcgctta	ccacgagaag	taccccacta tctaccacct tcgtaagaag 420
cttgttgact	ctactgataa	ggetgatctt	cgtctcatct accttgctct cgctcacatg 480
atcaagttcc	gtggtcactt	ccttatcgag	ggtgacctta acctgataa ctccgacgtg 540
gacaagctct	tcatccagct	cgtccagacc	tacaaccagc tcttcgagga gaacctatc 600
aacgcttccg	gtgtcgacgc	taaggcgatc	ctttccgcta ggctctccaa gtccaggcgt 660
ctcgagaacc	tcatcgccca	gctccctggg	gagaagaaga acggtctttt cggtaacctc 720
atcgtctctt	ccctcggctt	gaccctaacc	ttcaagtcca acttcgacct cgctgaggac 780
gctaagcttc	agctctccaa	ggatacctac	gacgatgatc tcgacaacct cctcgtctcag 840
attggagatc	agtacgctga	tctcttcctt	gctgctaaga acctctccga tgctatcctc 900
ctttcggata	tccttagggg	taaacctgag	atcactaagg ctctctttc tgcttccatg 960
atcaagcgct	acgacgagca	ccaccaggac	ctcacctcc tcaaggctct tgttcgtcag 1020
cagctccccg	agaagtacaa	ggagatcttc	ttcgaccagt ccaagaacgg ctacgccggt 1080
tacattgacg	gtggagctag	ccaggaggag	ttctacaagt tcatcaagcc aatccttgag 1140
aagatggatg	gtactgagga	gcttctcggt	aagcttaacc gtgaggacct ccttaggaag 1200
cagaggactt	tcgataacgg	ctctatccct	caccagatcc accttggtga gcttcacgcc 1260
atccttcgta	ggcaggagga	cttctaccct	ttctcaagg acaaccgtga gaagatcgag 1320
aagatcctta	ctttccgtat	tccttactac	gttggtctct ttgctcgtgg taactcccgt 1380
ttcgtcttga	tgactaggaa	gtccgaggag	actatcacc cttggaactt cgaggagggt 1440
gttgacaagg	gtgcttccgc	ccagtccttc	atcgagcgca tgaccaactt cgacaagaac 1500
ctccccaacg	agaaggtcct	ccccaaagca	tcctctctct acgagtactt cacggtctac 1560
aacgagctca	ccaaggtcaa	gtacgtcacc	gagggtatgc gcaagcctgc ctctctctcc 1620
ggcgagcaga	agaaggttat	cgttgacctc	ctcttcaaga ccaaccgcaa ggtcacctgc 1680
aagcagctca	aggaggacta	cttcaagaag	atcgagtgtc tcgactccgt cgagatcagc 1740
ggcgttgagg	accgtttcaa	cgttctctct	ggtacctacc acgatctctt caagatcacc 1800
aaggacaagg	acttctctga	caacgaggag	aacgaggaca tctcagagga catcgtctct 1860
actcttactc	tcttcgagga	tagggagatg	atcgaggaga ggctcaagac ttacgtctcat 1920
ctcttcgatg	acaaggttat	gaagcagctc	aagcgtcgcc gttacaccgg ttggggtagg 1980

ctctcccgca agctcatcaa cggatcagg gataagcaga gcggcaagac tatcctcgac 2040
ttcctcaagt ctgatggttt cgtaacagg aacttcatgc agctcatcca cgatgactct 2100
cttaccttca aggaggatat tcagaaggct caggtgtccg gtcaggcgca ctctctccac 2160
gagcacattg ctaaccttgc tggttcccct gctatcaaga agggcatcct tcagactggt 2220
aaggttgtcg atgagcttgt caaggttatg ggtcgtcaca agcctgagaa catcgtcatc 2280
gagatggctc gtgagaacca gactaccag aagggtcaga agaactcgag ggagcgcag 2340
aagaggattg aggagggtat caaggagctt gtttctcaga tccttaagga gcaccctgtc 2400
gagaacaccc agctccagaa cgagaagctc taccttact acctccagaa cggtagggat 2460
atgtacgttg accaggagct cgacatcaac aggtttctg actacgacgt cgaccacatt 2520
gttctcagt ctttcttaa ggatgactcc atcgacaaca aggtcctcac gaggtccgac 2580
aagaacagg gtaagtcgga caacgtccct tccgaggagg ttgtcaagaa gatgaagaac 2640
tactggaggc agcttctcaa cgtaagctc attaccaga ggaagttega caacctcacg 2700
aaggctgaga ggggtggcct ttccgagctt gacaaggctg gtttcatcaa gaggcagctt 2760
gttgagacga ggcagattac caagcacgtt gctcagatcc tcgattctag gatgaacacc 2820
aagtacgacg agaacgacaa gctcatccgc gaggtcaagg tgatcacct caagtccaag 2880
ctcgtctccg acttccgcaa ggacttccag ttctacaagg tccgcgagat caacaactac 2940
caccacgctc acgatgctta ccttaacgct gtcgttggtta ccgctcttat caagaagtac 3000
cctaagcttg agtccgagtt cgtctacggt gactacaagg tctacgacgt tcgtaagatg 3060
atcgccaagt ccgagcagga gatcggcaag gccaccgcca agtacttctt ctactccaac 3120
atcatgaact tcttcaagac cgagatcacc ctgccaacg gcgagatccg caagcgcctt 3180
cttatcgaga cgaacggtga gactgggtgag atcgtttggg acaagggtcg cgacttcgct 3240
actgttcgca aggtcctttc tatgcctcag gttaacatcg tcaagaagac cgaggtccag 3300
accggtggct tctccaagga gtctatcctt ccaaagagaa actcggacaa gctcatcgct 3360
aggaagaagg attgggacct taagaagtac ggtggtttcg actcccctac tgtcgcctac 3420
tccgtcctcg tggctgccaa ggtggagaag ggtaagtcca agaagctcaa gtccgtcaag 3480
gagctcctcg gcatcacat catggagcgc tctccttcg agaagaacct gatcgacttc 3540
ctcaggcca agggctacaa ggaggtcaag aaggacctc tcatcaagct cccaagtac 3600
tctcttttcg agctcgagaa cggtcgtaag aggatgctgg cttccgctgg tgagctccag 3660
aagggtaacg agcttgctct tcttccaag tacgtgaact tctctacct cgcctcccac 3720
tacgagaagc tcaagggttc cctgaggat aacgagcaga agcagctctt cgtggagcag 3780
cacaagcaact acctcgacga gatcatcgag cagatctccg agttctcaa gcgctcatc 3840
ctcgtgacg ctaacctcga caaggtctc tccgctaca acaagcaccg cgacaagccc 3900
atccgcgagc agcccgagaa catcatccac ctcttcaagc tcacgaacct cggcgcccct 3960
gctgctttca agtacttcga caccaccatc gacaggaagc gttacacgtc caccaaggag 4020
gttctcgacg ctactctcat ccaccagtc atcaccggtc tttacgagac tcgtatcgac 4080
ctttcccagc ttgggtggtga t 4101

<210>5

<211>2936

<212>DNA

<213>Artificial Sequence

<400>5

atggctacga cgcgccggc cgcggccgcc gccctgtccg ccgccgcgac ggccaagacc 60
 ggccgtaaga accaccagcg acaccacgtc cttcccgtc gaggccgggt gggggcggcg 120
 gcggtcaggt gctcggcggg gtccccggtc accccgccgt ccccgccgcc gccggccacg 180
 ccgctccggc cgtggggggc ggccgagccc cgcaagggcg cggacatcct cgtggaggcg 240
 ctggagcggg gcggcgtcag cgacgtgtt cctaccgg gcggcgcgtc catggagatc 300
 caccaggcgc tgacgcgtc cccggtcac accaaccacc tcttcccca cgagcagggc 360
 gaggcgttcg cggcgtccgg gtacgcgcgc gcgtccggc gcgtcgggg ctgcgtcgc 420
 acctccggcc ccggggcaac caacctcgtg tccgcgtcg ccgacgcgt gctcgactcc 480
 gtcccgatgg tcgccatcac gggccaggtc ccccgccga tgatcggcac cgacgccttc 540
 caggagacgc ccatagtcga ggtcaccgc tccatcaca agcacaatta cttgtcctt 600
 gatgtggagg acatccccg cgtcatacag gaagccttct tctcgcgtc ctcgggccgt 660
 cctggccccg tgctggtcga catccccaa gacatccagc agcagatggc cgtgccggtc 720
 tgggacacct cgatgaatct accagggtac atcgcacgc tgcceaagc acccgcgaca 780
 gaattgcttg agcaggctct gcgtctggtt ggcgagtcac ggcgcccgat tctctatgtc 840
 ggtggtggct gctctgcatc tggtagcga ttgcgctggt ttgttgagct gactggtatc 900
 ccagttacaa cactctgat gggcctcggc aatttccca gtgacgacc gttgtccctg 960
 cgcattgctg ggatgcatgg cacggtgtac gcaaattat ccgtggataa ggctgacctg 1020
 ttgcttgctg ttggtgtgcg gtttgatgat cgtgtgacag ggaaaattga ggcttttgca 1080
 agcagggccca agattgtgca cattgacatt gatccagcag agattgaaa gaacaagcaa 1140
 ccacatgtgt caatttgcgc agatgttaag cttgctttac agggcttgaa tgctctgcta 1200
 caacagagca caacaaagac aagtctgat ttagtgcat ggcacaatga gttggaccag 1260
 cagaagaggg agtttctct ggggtacaaa acttttggtg aagagatccc accgcaatat 1320
 gccattcagg tgctggatga gctgacgaaa ggtgaggcaa tcatcgctac tgggtttggg 1380
 cagcaccaga tgtgggcggc acaatattac acctacaagc ggccacggca gtggctgtct 1440
 tcggctggtc tgggcgcaat gggatttggg ctgcctgctg cagctggtgc ttctgtggct 1500
 aaccaggtg tcacagttgt tgatattgat ggggatggtg gcttctcat gaacattcag 1560
 gagctggcat tgatccgcat tgagaacct cctgtgaagg tgatggtgtt gaacaacca 1620
 catttgggta tgggtgtgca atgggaggat aggttttaca aggcgaatag ggcgataca 1680
 tacttgggca acccggaatg tgagagcgag atatatccag attttgtgac tattgctaag 1740
 gggttcaata ttctgcagt ccgtgtaaca aagaagagtg aagtcctgac cgccatcaag 1800
 aagatgctcg agactccagg gccatacttg ttggatata tcgtcccga ccaggagcat 1860
 gtgctgccta tgatcccaag tggggcgcca ttcaaggaca tgatcctgga tggatgagc 1920
 aggactgtgt attaatctat aatctgtatg ttggcaaagc accagcccgg cctatgtttg 1980
 acctgaatga ccataaaga gtggtatgcc tatgatgttt gtatgtgctc tatcaataac 2040
 taagggtgca actatgaacc atatgctctt ctgttttact tgtttgatgt gcttggcatg 2100
 gtaatcctaa ttagcttctt gctgtctagg tttgtagtgt gttgttttct gtaggcatat 2160
 gcatcacaag atatcatgta agtttcttgt cctacatata aataataaga gaataaagta 2220

cttctatgca atagctctga gttaagtgtt tcaacaattt ctgaacttct gaacttatgt 2280
 ttgctcaact gtcacacac gaagtactct cettgtaact acattttccc caagacttta 2340
 aatccccctca gttacagcaa aaaataaact ttgcatctac tgttttccct ctcttcggtc 2400
 gatcttattg ggtactacta tagagagagg ctgcatgaag tatttccttt ttctgtttag 2460
 ttatgccgtg taaattagca tccatgcaaa atagatgaaa aatcaagcta ttcctgactg 2520
 ctaaggatta tttttggcat aatgtattct tatatactcc ctccgtccca tattataagg 2580
 gattttgagt ttttgttat actgtttgac cactcgtctt attcaaaaaa ttttagaatt 2640
 attatttatt tttttgtga cttactttat tatctaaagt actttaagca caattttcgt 2700
 attttatatt tgcacaaatt tttgaataa gacgaatgt caacaatac aaataaaaat 2760
 tcaaaatccc ttataatatg ggacggaggt atgatagttg gtgaactgct acgtattgcc 2820
 atttgacatt ttttgatta tgcaattttg ctgtctatag tgctctaate aattcgcaat 2880
 cccgaccttg gagtattggt ctcatggaac cctcatctg agtaatctcc atattt 2936

<210>6

<211>644

<212>PRT

<213>Artificial Sequence

<400>6

Met	Ala	Thr	Thr	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Leu	Ser	Ala	Ala	Ala
1				5								10				15
Thr	Ala	Lys	Thr	Gly	Arg	Lys	Asn	His	Gln	Arg	His	His	Val	Leu	Pro	
				20												30
Ala	Arg	Gly	Arg	Val	Gly	Ala	Ala	Ala	Val	Arg	Cys	Ser	Ala	Val	Ser	
				35												45
Pro	Val	Thr	Pro	Pro	Ser	Pro	Ala	Pro	Pro	Ala	Thr	Pro	Leu	Arg	Pro	
																60
Trp	Gly	Pro	Ala	Glu	Pro	Arg	Lys	Gly	Ala	Asp	Ile	Leu	Val	Glu	Ala	
65																80
Leu	Glu	Arg	Cys	Gly	Val	Ser	Asp	Val	Phe	Ala	Tyr	Pro	Gly	Gly	Ala	
																95
Ser	Met	Glu	Ile	His	Gln	Ala	Leu	Thr	Arg	Ser	Pro	Val	Ile	Thr	Asn	
																110
His	Leu	Phe	Arg	His	Glu	Gln	Gly	Glu	Ala	Phe	Ala	Ala	Ser	Gly	Tyr	
																125
Ala	Arg	Ala	Ser	Gly	Arg	Val	Gly	Val	Cys	Val	Ala	Thr	Ser	Gly	Pro	
																140
Gly	Ala	Thr	Asn	Leu	Val	Ser	Ala	Leu	Ala	Asp	Ala	Leu	Leu	Asp	Ser	
145																160
Val	Pro	Met	Val	Ala	Ile	Thr	Gly	Gln	Val	Pro	Arg	Arg	Met	Ile	Gly	
																175

Thr Asp Ala Phe Gln Glu Thr Pro Ile Val Glu Val Thr Arg Ser Ile
 180 185 190
 Thr Lys His Asn Tyr Leu Val Leu Asp Val Glu Asp Ile Pro Arg Val
 195 200 205
 Ile Gln Glu Ala Phe Phe Leu Ala Ser Ser Gly Arg Pro Gly Pro Val
 210 215 220
 Leu Val Asp Ile Pro Lys Asp Ile Gln Gln Gln Met Ala Val Pro Val
 225 230 235 240
 Trp Asp Thr Ser Met Asn Leu Pro Gly Tyr Ile Ala Arg Leu Pro Lys
 245 250 255
 Pro Pro Ala Thr Glu Leu Leu Glu Gln Val Leu Arg Leu Val Gly Glu
 260 265 270
 Ser Arg Arg Pro Ile Leu Tyr Val Gly Gly Gly Cys Ser Ala Ser Gly
 275 280 285
 Asp Glu Leu Arg Trp Phe Val Glu Leu Thr Gly Ile Pro Val Thr Thr
 290 295 300
 Thr Leu Met Gly Leu Gly Asn Phe Pro Ser Asp Asp Pro Leu Ser Leu
 305 310 315 320
 Arg Met Leu Gly Met His Gly Thr Val Tyr Ala Asn Tyr Ala Val Asp
 325 330 335
 Lys Ala Asp Leu Leu Leu Ala Phe Gly Val Arg Phe Asp Asp Arg Val
 340 345 350
 Thr Gly Lys Ile Glu Ala Phe Ala Ser Arg Ala Lys Ile Val His Ile
 355 360 365
 Asp Ile Asp Pro Ala Glu Ile Gly Lys Asn Lys Gln Pro His Val Ser
 370 375 380
 Ile Cys Ala Asp Val Lys Leu Ala Leu Gln Gly Leu Asn Ala Leu Leu
 385 390 395 400
 Gln Gln Ser Thr Thr Lys Thr Ser Ser Asp Phe Ser Ala Trp His Asn
 405 410 415
 Glu Leu Asp Gln Gln Lys Arg Glu Phe Pro Leu Gly Tyr Lys Thr Phe
 420 425 430
 Gly Glu Glu Ile Pro Pro Gln Tyr Ala Ile Gln Val Leu Asp Glu Leu
 435 440 445
 Thr Lys Gly Glu Ala Ile Ile Ala Thr Gly Val Gly Gln His Gln Met
 450 455 460
 Trp Ala Ala Gln Tyr Tyr Thr Tyr Lys Arg Pro Arg Gln Trp Leu Ser
 465 470 475 480
 Ser Ala Gly Leu Gly Ala Met Gly Phe Gly Leu Pro Ala Ala Ala Gly

	485		490		495
Ala Ser Val	Ala Asn Pro Gly Val Thr Val Val	Asp Ile Asp Gly Asp			
	500		505		510
Gly Ser Phe	Leu Met Asn Ile Gln Glu Leu Ala Leu Ile Arg Ile Glu				
	515		520		525
Asn Leu Pro	Val Lys Val Met Val Leu Asn Asn Gln His Leu Gly Met				
	530		535		540
Val Val Gln	Trp Glu Asp Arg Phe Tyr Lys Ala Asn Arg Ala His Thr				
545		550		555	560
Tyr Leu Gly	Asn Pro Glu Cys Glu Ser Glu Ile Tyr Pro Asp Phe Val				
	565		570		575
Thr Ile Ala	Lys Gly Phe Asn Ile Pro Ala Val Arg Val Thr Lys Lys				
	580		585		590
Ser Glu Val	Arg Ala Ala Ile Lys Lys Met Leu Glu Thr Pro Gly Pro				
	595		600		605
Tyr Leu Leu	Asp Ile Ile Val Pro His Gln Glu His Val Leu Pro Met				
	610		615		620
Ile Pro Ser	Gly Gly Ala Phe Lys Asp Met Ile Leu Asp Gly Asp Gly				
625		630		635	640

Arg Thr Val Tyr

<210>7

<211>86

<212>DNA

<213>Artificial Sequence

<400>7

gtttcagagc tatgctggaa acagcatagc aagttgaaat aaggctagtc cgttatcaac 60
 ttgaaaaagt ggcaccgagt cgggtgc 86

<210>8

<211>1373

<212>DNA

<213>Artificial Sequence

<400>8

ggaatccatg gcggatcagg agccaccaac ttctcctcc tcaagcaggc cggcgacgtg 60
 gaggagaacc cgggccaat gaaaaagcct gaactcaccg cgacgtctgt cgagaagttt 120
 ctgatcgaaa agttcgacag cgtctccgac ctgatgcagc tctcggaggg cgaagaatct 180
 cgtgctttca gcttcgatgt aggaggcgt ggatatgtcc tgcgggtaaa tagctgcgcc 240
 gatggtttct acaaagatcg ttatgtttat cggcactttg catcggccgc gctcccgatt 300
 ccggaagtgc ttgacattgg ggagtttagc gagagcctga cctattgcat ctcccgcct 360
 tcacagggtg tcacgttgca agacctgcct gaaaccgaac tgcccgtgt tctacaaccg 420

gtcgaggagg ctatggatgc gatcgctgcg gccgatctta gccagacgag cgggttcggc 480
 ccattcggac cgcaaggaat cggtaatac actacatggc gtgatttcat atgcgcgatt 540
 gctgatcccc atgtgtatca ctggcaact gtgatggacg acaccgtcag tgcgtccgctc 600
 gcgcaggctc tcgatgagct gatgctttgg gccgaggact gccccgaagt ccggcacctc 660
 gtgcacgcgg atttcggctc caacaatgtc ctgacggaca atggccgcat aacagcggtc 720
 attgactgga gcgaggcgat gttcggggat tccaatacag aggtcgccaa catcttcttc 780
 tggaggccgt ggttggcttg tatggagcag cagacgcgct acttcgagcg gaggcatccg 840
 gagcttgcag gatcgccacg actccgggag tatatgctcc gcattggtct tgaccaactc 900
 tatcagagct tggttgacgg caatttcgat gatgcagctt gggcgcaggg tcgatgcgac 960
 gcaatcgctc gatccggagc cgggactgtc gggcgtaac aaatcgcccg cagaagcgcg 1020
 gccgtctgga ccgatggtg tgtagaagta ctcgccgata gtggaaaccg acgccccagc 1080
 actcgtccga gggcaaagaa atagactagt tcccgatcgt tcaaacattt ggcaataaag 1140
 tttcttaaga ttgaatcctg ttgccggtct tgcgatgatt atcatataat ttctgttgaa 1200
 ttacgttaag catgtaataa ttaacatgta atgcatgacg ttatttatga ggtgggtttt 1260
 tatgattaga gtcccgaat tatacattta atacgcgata gaaaacaaaa tatagcgcgc 1320
 aaactaggat aaattatcgc gcgcggtgtc atctatgtta ctagaggcgc gcc 1373

<210>9

<211>19

<212>PRT

<213>Artificial Sequence

<400>9

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn
 1 5 10 15

Pro Gly Pro

<210>10

<211>341

<212>PRT

<213>Artificial Sequence

<400>10

Met Lys Lys Pro Glu Leu Thr Ala Thr Ser Val Glu Lys Phe Leu Ile
 1 5 10 15

Glu Lys Phe Asp Ser Val Ser Asp Leu Met Gln Leu Ser Glu Gly Glu
 20 25 30

Glu Ser Arg Ala Phe Ser Phe Asp Val Gly Gly Arg Gly Tyr Val Leu
 35 40 45

Arg Val Asn Ser Cys Ala Asp Gly Phe Tyr Lys Asp Arg Tyr Val Tyr
 50 55 60

Arg His Phe Ala Ser Ala Ala Leu Pro Ile Pro Glu Val Leu Asp Ile
 65 70 75 80

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

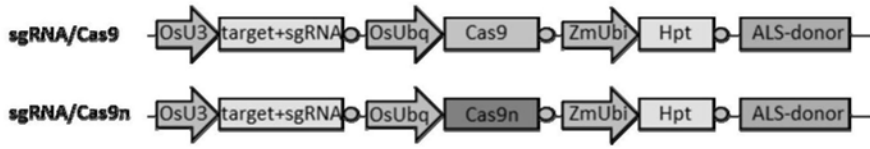


图1

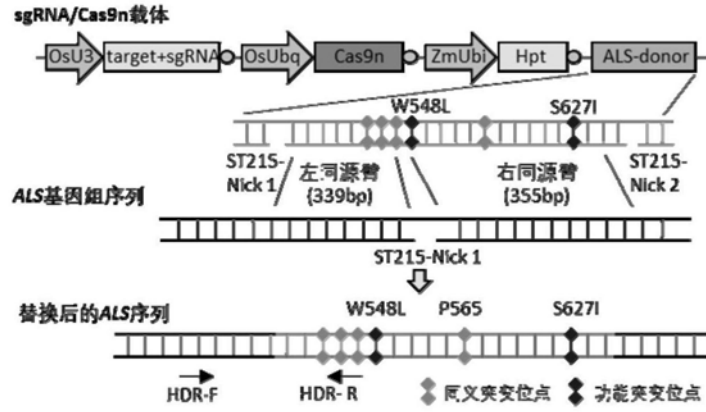


图2

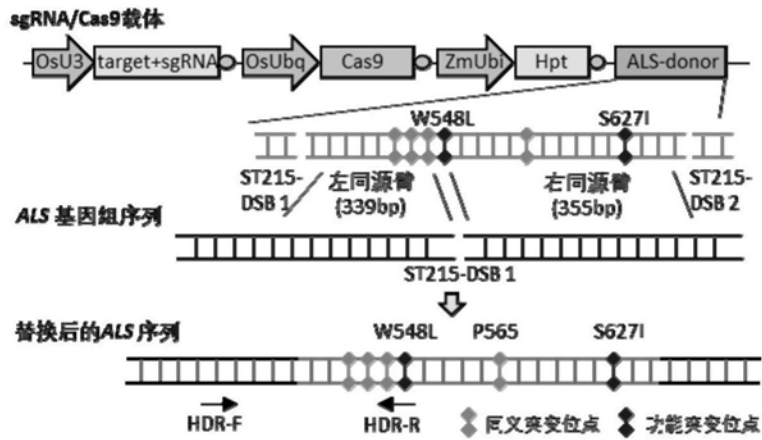


图3

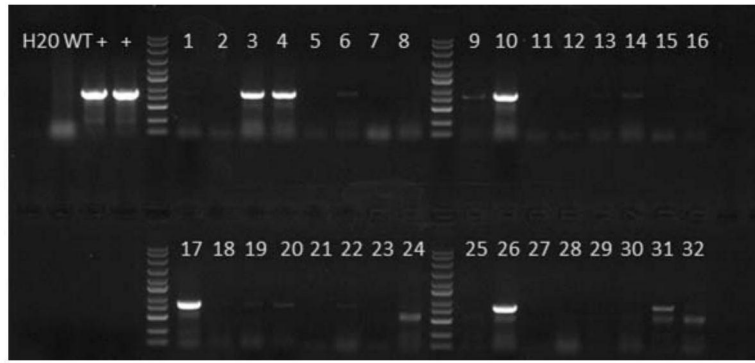


图4

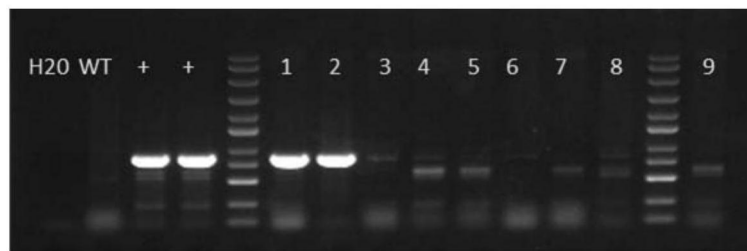


图5

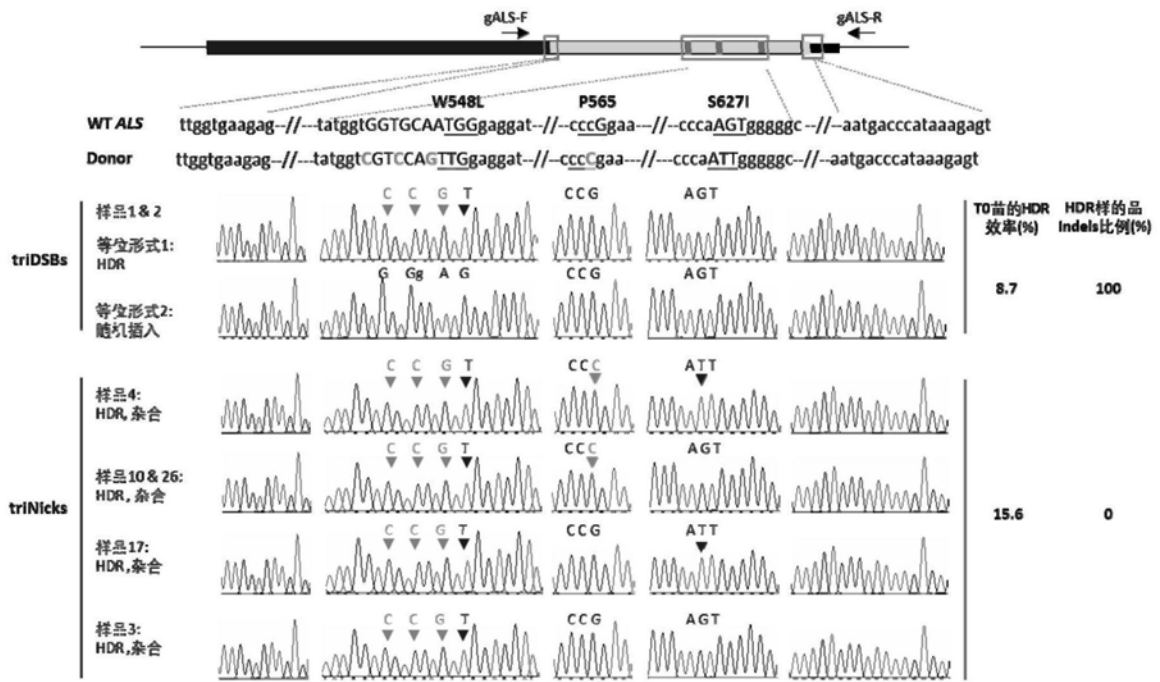


图6

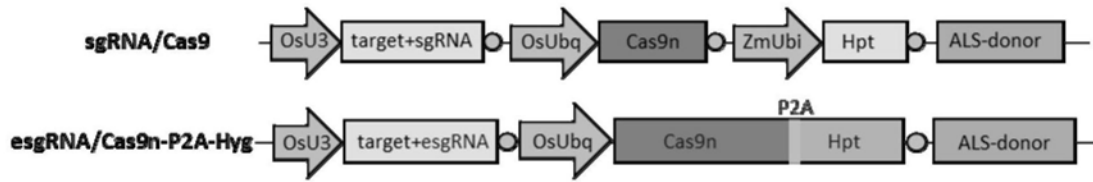


图7

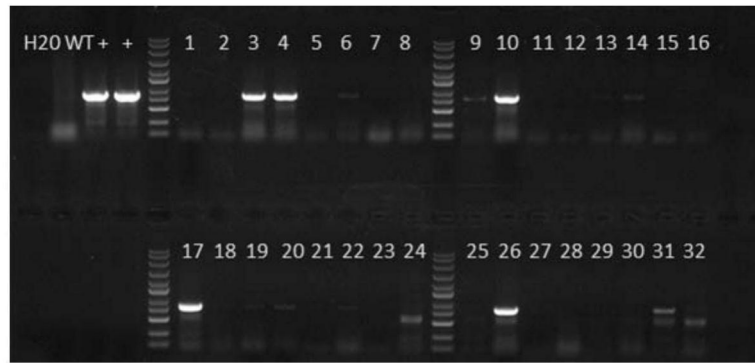


图8

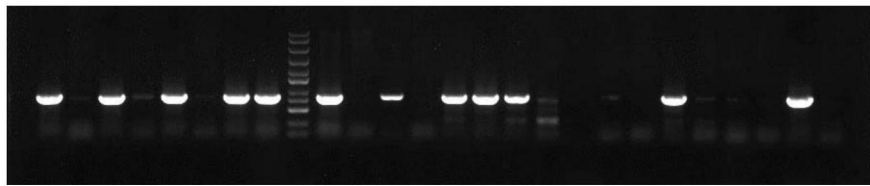


图9